frontiers | Research Topics

Artificial intelligence and the future of work: humans in control

Edited by
Ekkehard Ernst, Janine Berg and Phoebe V. Moore

Published in
Frontiers in Artificial Intelligence

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5509-5
DOI 10.3389/978-2-8325-5509-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Artificial intelligence and the future of work: humans in control

Topic editors

Ekkehard Ernst — International Labour Organization, Switzerland
Janine Berg — International Labour Organization, Switzerland
Phoebe V. Moore — University of Essex, United Kingdom

Citation

Ernst, E., Berg, J., Moore, P. V., eds. (2024). *Artificial intelligence and the future of work: humans in control*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5509-5

Table of contents

frontiers | Frontiers in Artificial Intelligence

TYPE Editorial
PUBLISHED 13 March 2024
DOI 10.3389/frai.2024.1378893

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Dursun Delen,
Oklahoma State University, United States

*CORRESPONDENCE
Ekkehard Ernst
ernste@ilo.org

RECEIVED 30 January 2024
ACCEPTED 05 March 2024
PUBLISHED 13 March 2024

CITATION
Ernst E, Berg J and Moore PV (2024) Editorial: Artificial intelligence and the future of work: humans in control.
Front. Artif. Intell. 7:1378893.
doi: 10.3389/frai.2024.1378893

COPYRIGHT
© 2024 Ernst, Berg and Moore. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Artificial intelligence and the future of work: humans in control

Ekkehard Ernst[1]*, Janine Berg[1] and Phoebe V. Moore[2]

[1]International Labour Organization, Geneva, Switzerland, [2]School of Business, Faculty of Social Sciences, University of Essex, Colchester, East of England, United Kingdom

KEYWORDS

artificial intelligence, world of work, occupational safety and health (OSH), employment, wages, recruitment, ethics, productivity

Editorial on the Research Topic
Artificial intelligence and the future of work: humans in control

Latest developments around artificial intelligence (AI) have triggered excitement about the potential to replace and complement human activities while also raising concerns about possible risks to society. Dramatic effects are specifically being felt in the world of work, including jobs, wages and working conditions but also recruitment, performance monitoring, and dismissal. So far, research in this area has focused predominantly on the potential of AI for job gains and losses. Other aspects of its transformative dynamics have received less attention, however. In particular, the impact of AI on job quality, average hours worked, mobility, or labor relations between employers and workers are often overlooked. Moreover, society-wide effects triggered by AI, including its rising environmental burden, need to be reassessed. To address these issues, this Research Topic includes nine exciting contributions that shed light on a broader range of issues that AI technologies might bring to the world of work.

To set the stage for the overall effect of AI on employment in 23 OECD countries, in our special edition, Georgieff and Hyee present research using an adapted AI occupational impact measure. The authors do not find that AI exposure affects employment growth in their sample. However, occupations where computer use is high see faster employment growth when exposed to AI. In contrast, occupations with low computer use see a decline in average hours work (yet not in employment) when exposed to AI, suggesting a distributional impact of AI rather than one on the overall number of jobs.

Whether digital technological technologies improve or worsen wages for employees remains a hotly debated topic. Fossen et al. argue that it depends on the specific application considered. Whereas, software and industrial robots seem to be associated with wage decreases, suggesting job displacement, innovations in AI are associated with wage increases, pointing toward positive productivity effects, at least as far as the labor market in the United States is concerned.

How can the income and wealth disparities that are brought about by AI be addressed? Merola looks at the various proposals that have been brought forward in recent years to address the differential effects of AI on labor markets. She discusses pros and cons of various proposals, including a robot tax, digital taxation, share price taxation, or – alternatively – wage subsidies for low-income earners and assesses their potential impact on employment growth, inequality and innovation.

Besides the impact of AI on the number of jobs or their distribution, AI will also affect working conditions for those in employment. Using a representative business survey for Germany, Warning et al. demonstrate how occupations with a high share of routine cognitive activities exposed to AI are associated with higher demand for flexibility, including employee self-organization and time management. Moreover, such worsening of working conditions predominantly affects older workers and women in the labor market.

Concerns about implications of AI for occupational health and safety (OSH) abound. Niehaus et al. report results from a large-scale study of German workers on the impact of AI on job autonomy and psychological occupational stress. The authors highlight that AI is often being used to increase autonomy of supervisory functions while lowering control for job execution. This is likely to increase work-related stress over and above possible concerns for job or earnings loss.

AI is also a tool that can be used by HR managers to improve functional mobility and ultimately employees' job satisfaction, given that internal mobility risks deteriorating job satisfaction if it increases stress to the detriment of personal life. Bossi et al. analyze various approaches using AI to help better manage internal mobility schemes with a view to improving future job satisfaction. The authors analyse alternative statistical models and compare different methods in supporting predictive Human Resources analytics.

AI has society-wide implications not only for employment and wages but also on the use of energy and its potential for increasing productivity. Ernst argues that taken together and considering the current path of technological development, AI gives rise to a trilemma, making it impossible to achieve high productivity growth, low inequality, and reduced energy consumption simultaneously. Instead, he argues, a new technological paradigm is needed to orient AI applications toward those areas where social returns are particularly high, such as in mobility and waste management, clean energy, and natural capital solutions.

Beyond automating certain tasks at individual workplaces, AI is also transforming managerial control. Woodcock analyses in detail how AI affects the work of managers, based on a case study of AI's use in call centers. He shows how the introduction of new surveillance tools based on AI are being contested by call center employees and how this shapes the extent and incidence of such tools for managerial purposes.

Society-wide implications of AI have often met with calls for "ethical Artificial Intelligence." Cole et al. conceptualize these calls and question their efficacy in sufficiently addressing the resulting societal challenges given their mostly narrow political framework focused on privacy, transparency and non-discrimination. In response, the authors identify a set of principles to facilitate fairer working conditions with AI, focusing on operationalizable processes that effectively help address the potential risks and harms resulting from AI in the workplace.

The collection of papers brought together in this Research Topic offer new insights into the multi-faceted and society-wide implications that AI is likely to bring to the world of work. Our ambition was to demonstrate how current technological changes will cause a wide-ranging transformation that goes beyond headline indicators such as the number of potential job losses. We hope that these papers can contribute to a wider discussion both among researchers and policy makers of the multi-faceted effects of AI on the world of work, and thus for the need to better understand – and find appropriate responses to – the multitude of ongoing changes in the world of work.

Author contributions

EE: Writing – original draft. JB: Writing – review & editing. PM: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/frai.2022.848015

Estimating Successful Internal Mobility: A Comparison Between Structural Equation Models and Machine Learning Algorithms

Francesco Bossi[1], Francesco Di Gruttola[1], Antonio Mastrogiorgio[2], Sonia D'Arcangelo[3], Nicola Lattanzi[2], Andrea P. Malizia[1] and Emiliano Ricciardi[1]*

[1] MoMiLab Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy, [2] Axes Research Unit, IMT School for Advanced Studies Lucca, Lucca, Italy, [3] Neuroscience Lab, Intesa Sanpaolo Innovation Center SpA, Turin, Italy

Internal mobility often depends on predicting future job satisfaction, for such employees subject to internal mobility programs. In this study, we compared the predictive power of different classes of models, i.e., (i) traditional Structural Equation Modeling (SEM), with two families of Machine Learning algorithms: (ii) regressors, specifically least absolute shrinkage and selection operator (Lasso) for feature selection and (iii) classifiers, specifically Bagging meta-model with the *k*-nearest neighbors algorithm (*k*-NN) as a base estimator. Our aim is to investigate which method better predicts job satisfaction for 348 employees (with operational duties) and 35 supervisors in the training set, and 79 employees in the test set, all subject to internal mobility programs in a large Italian banking group. Results showed average predictive power for SEM and Bagging *k*-NN (accuracy between 61 and 66%; F1 scores between 0.51 and 0.73). Both SEM and Lasso algorithms highlighted the predictive power of resistance to change and orientation to relation in all models, together with other personality and motivation variables in different models. Theoretical implications are discussed for using these variables in predicting successful job relocation in internal mobility programs. Moreover, these results showed how crucial it is to compare methods coming from different research traditions in predictive Human Resources analytics.

Keywords: internal mobility, job relocation, job satisfaction, structural equation models, machine learning, resistance to change, predictive HR analytics

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Leicester,
United Kingdom

Reviewed by:
Massimo Cannas,
University of Cagliari, Italy
Denis Helic,
Graz University of Technology, Austria

****Correspondence:***
Francesco Bossi
francesco.bossi@imtlucca.it

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *03 January 2022*
Accepted: *02 March 2022*
Published: *25 March 2022*

Citation:
Bossi F, Di Gruttola F, Mastrogiorgio A, D'Arcangelo S, Lattanzi N, Malizia AP and Ricciardi E (2022) Estimating Successful Internal Mobility: A Comparison Between Structural Equation Models and Machine Learning Algorithms. Front. Artif. Intell. 5:848015. doi: 10.3389/frai.2022.848015

INTRODUCTION

Job relocation is a traditional issue of organizational literature whose main paradigms refer to the effect of job transfer on stress and family life (Burke, 1986; Munton, 1990), where job transfer traditionally requires geographical mobility. Consolidate evidence shows that the preference for a specific location is a major predictor of post-transfer satisfaction (Pinder, 1977). In general, employees in the early career stage tend to be more willing to accept mobility opportunities as they perceive more dissonance between their current job and ideal job (Noe et al., 1988). The willingness to relocate enters the selection process in which attitudinal, biographical and social variables predict how many potential employees are prone to international mobility (Andresen and Margenfeld, 2015).

Nevertheless, post-transfer satisfaction is not simply a matter of geographical opportunity. The rise of information technologies in recent decades has made geographical relocations less problematic, as they enable a flexible and geographically independent job organization (i.e. remote working). Internal migration rates are declining across most Western countries (Haan and Cardoso, 2020), for several economic and social reasons. In contemporary economies, post-transfer satisfaction is mainly referred to the *internal mobility* where the changes—due to promotions and/or lateral transfers—occur within the same organization. Promoted workers, internal to an organization, have significantly better performance and lower exit rates than those externally hired into similar jobs (Bidwell, 2011). Indeed, upward progressions are much more likely to happen through internal than external mobility (Bidwell and Mollick, 2015). High performers are less likely to quit, and when they do quit the reasons are typically not related to work (Benson and Rissing, 2020). Furthermore, there is evidence of a negative association between performance and internal mobility for low performers as they add value to the organization by developing complex social networks through internal job transfers (Chen et al., 2020a).

Internal mobility is not just an opportunity for career development. In many cases, internal mobility is not a discretional choice but a strategic or a contingent organizational need that could involve the forced relocation of hundreds of employees. In such cases, *predicting job satisfaction* for such employees involved in mobility programs is fundamental. The literature on job satisfaction is abundant regarding the construct and its antecedents (e.g., Judge et al., 2002; Aziri, 2011), but its prediction is often problematic. In particular, job satisfaction is not always an existent construct to be simply measured in given settings. Human resources (HR) specialists are often interested in predicting post-transfer job satisfaction in such settings that include internal mobility, as organizational changes are designed precisely depending on how employees will react to the new arrangements. In short, internal mobility often depends on the prediction of future job satisfaction. In such situations, what HR practitioners have at disposal is many individual-related variables, such as individual differences in personality, motivation and emotion for workers, and leadership style and empathy for leaders. Using such variables to predict job satisfaction—where satisfaction is a general construct that also includes communication- and inclusion-related aspects—could be opportune. Machine learning comes in help as it allows predicting job satisfaction, based on the available variables.

Job Satisfaction

Job satisfaction represents a complex research domain stratified over decades, whose definition and research questions are significantly dependent on the specific historical contingencies (Latham and Budworth, 2007). Generally speaking, job satisfaction is a construct whose investigation admits different paradigms and approaches, each one with specific theoretical nuances. Such approaches include Hertzberg's motivator-hygiene theory (Herzberg, 1964), job design frameworks (Hackman and Oldham, 1976), dispositional (Staw et al., 1986) and equity approaches (Huseman et al., 1987). Traditionally, job satisfaction has been defined as "a pleasurable or positive emotional state resulting from the appraisal of one's job or job experiences" (Locke, 1976, p. 1304). Job satisfaction presents a number of facets as it can be defined with reference to specific job aspects. Spector (1997) identifies fourteen aspects that include appreciation, communication, coworkers, fringe benefits, job conditions, nature of the work, organization, personal growth, policies and procedures, promotion opportunities, recognition, security, and supervision.

The assumption that happier workers are more productive is the fundamental hypothesis of literature, showing that both cognitive and affective factors can explain, to different degrees, job satisfaction (Moorman, 1993). Managers usually look for satisfied workers, assuming that they are more engaged and performative, where job satisfaction and employee motivation, though different constructs, are fundamental for organizational performance (Vroom, 1964). The meta-analytical evidence of satisfaction-performance relationship encompasses several paradigms that flourished over the last century, whose theoretical and practical implications would deserve dedicated discussions (Schwab and Cummings, 1970; Iaffaldano and Muchinsky, 1985; Judge et al., 2001; Harter et al., 2002). Importantly, job satisfaction traditionally also extends outside of the job domains to include private life (Near et al., 1980; Rain et al., 1991). The "happy-productive worker paradigm" has been unpacked and evidence shows the role of general psychological well-being, not just job satisfaction, in explaining performance (Wright and Cropanzano, 2000). While such meta-analytical evidence emphasizes a correlation between job satisfaction and individual performance, the same cannot be maintained for organizational performance, where the less consolidated literature shows mixed evidence. Some studies show a positive relationship (e.g., Huselid, 1995; Schneider et al., 2003), others show the absence of any significant correlation (e.g., Mohr and Puck, 2007). Interestingly, the opposite relationship is also meaningful considering that organizational success affects employees' satisfaction (Ryan et al., 1996).

Predictive HR Analytics

Big data analytics represent a fundamental factor for companies to mine information to achieve competitive advantages (for a generalist literature review see Holsapple et al., 2014; Chong and Shi, 2015). Within this broad domain, HR analytics occupies a significant position as they help companies in managing human resources by exploiting data about how employees work and their individual differences. HR analytics refers to the use of statistical tools and computational methods for making decisions involving HR strategies and practices.

While HR analytics are traditionally reactive, predictive HR analytics is proactive and represents a relatively novel domain of investigation. Predictive HR analytics can be defined as "the systematic application of predictive modeling using inferential statistics to existing HR people-related data to inform judgements about possible causal factors driving key HR-related performance indicators" (Edwards and Edwards, 2019, p. 3). The increasing application of artificial intelligence (i.e., machine learning), far from being a passing fad, represents a significant trend in the last

decade (Falletta, 2014). Predictive HR analytics serve the purpose of identifying opportunities and risks in advance before they are clear to managers. Finally, predictive HR analytics is not merely devoted to improving efficiency but, more and more enables strategic human capital decisions (Kapoor and Sherif, 2012; Zang and Ye, 2015).

HR analytics is a still developing topic whose related evidence is often based on anecdotal evidence and case histories (e.g., Dow Chemical mined the employee data to predict the success of promotions and internal transfers, Davenport et al., 2010). Ben-Gal (2019), through an analytical review of the literature, highlights that empirical and conceptual studies in HR analytics are related to higher economic performances compared to technical- and case-based studies. In particular, such performances are related to the application of HR analytics to workforce planning and recruitment/selection tasks.

While the general impact of artificial intelligence on HR is a well-debated topic (Bassi, 2011; e.g., Rathi, 2018), the study of specific machine learning methods for predictive purposes in HR analytics represents a non-consolidated domain of research, characterized by a high degree of technicalities (Kakulapati et al., 2020). Such research domain includes the turnover prediction through neural networks (Quinn et al., 2002) or machine learning algorithms based on Extreme Gradient Boosting (Punnoose and Ajit, 2016), data mining for personnel selection (Chien and Chen, 2008), workforce optimization through constraint programming (Naveh et al., 2007). Nevertheless, in the past decades, the application of automated machine learning algorithms or neural networks in this field was mostly limited to the areas of intervention of a company spread in a region (Kolesar and Walker, 1974) or to the relocation of a whole company to a new geographical location (Haddad and Sanders, 2020). However, no studies have previously compared different statistical methods for predictive HR analytics or, more specifically, for automated relocation.

A crucial problem of predictive HR analytics is related to ethical issues arising from evidence-based decisions. Indeed, the use of some specific predictive variables can be problematic: what if HR specialists make human capital decisions based, e.g., on an applicant's hometown, car preference or sports habits, precisely because these variables are predictive of job performance? Such practices might be questionable and represent a matter on which the HR community will be likely called into account in the years to come (for a discussion see, Hamilton and Davison, 2021). In particular, some of the main concerns include the violation of national and international employment discrimination laws or data protection regulations, as well as employees' desires for privacy and justice.

Aim of the Study

In this study, we compared the predictive power of three classes of models. We compared (i) traditional Structural Equation Modeling (SEM), with two families of Machine Learning algorithms, i.e., (ii) regressors, specifically least absolute shrinkage and selection operator (Lasso) for feature selection and (iii) classifiers, specifically Bootstrap Aggregation meta-model using as base estimator the *k*-nearest neighbors algorithm (*k*-NN). Our aim is to validate which method better predicts successful relocation (measured in terms of job satisfaction, inclusion in the work team, and communication satisfaction) for 348 employees (with operational duties) and 35 leaders in the training set, and 79 employees in the test set, subject to internal mobility programs in a large Italian banking group. We considered an array of heterogeneous independent variables (from personality and motivation literature) which often constitute what HR directors have at disposal for making predictions about their employees, and we compared the alternative predictive methods.

MATERIALS AND METHODS

Participants

During the first part of the study (training set, February-March 2020), 503 employees and 40 supervisors in a large-scale Italian banking group volunteered for the data collection. Out of this sample, 380 employees and 37 supervisors opened the survey, but only 348 employees (147 F, mean ± sd age: 49.4 ± 6.9) and 35 supervisors (7 F, mean ± sd age: 50.7 ± 7.5) completed the survey. During the second data collection (test set, July 2020), 100 employees volunteered, 82 of them opened the survey and 79 (34 F, mean ± sd age: 48.3 ± 7.6) completed it. All employees were relocated more than 6 months before the data collection (mean training set: 15.9 months; mean test set: 14.1 months). All participants had normal or a corrected-to-normal vision, no history of auditory or psychiatric disorders.

Ethical Statement

All participants were provided with an exhaustive description of all the experimental procedures and were required to sign a written informed consent before taking part in the study. The study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and under a protocol approved by the Area Vasta Nord Ovest Ethics Committee (protocol n. 24579/2018).

Procedure

All questionnaires were administered via an online survey based on SurveyMonkey®. Survey links were sent by email to all volunteers by a collaborator bank employee. In this way, researchers could never have direct access to participants' names and they could participate anonymously. The first part of data collection (348 employees and 35 supervisors), aimed to collect the training set, was carried out between February 26th and March 16th 2020. The second part of data collection (79 employees), aimed to collect the test set, was carried out between July 6th and July 31st 2020.

Materials

Questionnaires administered to employees (in both training and test sets) included a Personality questionnaire (Jackson et al., 1996a,b; Hogan and Hogan, 1997, 2007), Motivational Orientation Test (Alessandri and Russo, 2011), Resistance to Change questionnaire (Oreg, 2003), Emotion Regulation Questionnaire (Gross and John, 2003), Rational Experiential

Inventory – Short Form (REI-S24, Pacini and Epstein, 1999), Inclusion questionnaire (Jansen et al., 2014), Job Satisfaction Index (Brayfield and Rothe, 1951), and Communication Satisfaction Scale (Madlock, 2008). Questionnaires administered to supervisors (only in the training set) included the Trait Emotional Intelligence Questionnaire – Short Form (Petrides, 2009), Interpersonal Reactivity Instrument (Davis, 1983), Prosocialness Scale for Adults (Caprara et al., 2005), Multifactor Leadership Questionnaire – 6S Form (Avolio and Bass, 2004), and REI-S24 (Pacini and Epstein, 1999).

Predictors

Questionnaires used to measure predictors (or independent variables) administered to employees are:

Personality Questionnaire

We built a 40-items questionnaire (Malizia et al., 2021) to measure four specific dimensions of personality of interest in our study. In particular, we considered three dimensions of the Six Factor Personality Questionnaire (Jackson et al., 1996a,b). Such dimensions (and their facets) are Independence (autonomy, individualism, self-reliance), Openness to Experience (change, understanding, breadth of interest), Industriousness (achievement, endurance, seriousness). A further dimension, Dutifulness, was selected from the Hogan Personality Inventory (Hogan and Hogan, 1997, 2007).

Motivational Orientation Test

The Motivational Orientation Test (Borgogni et al., 2004; Petitta et al., 2005; Test di Orientamento Motivazionale, see Alessandri and Russo, 2011) is based on 43 items and addressed to measure four drivers—Objective, Innovation, Relation, Leadership—of individual motivation.

Resistance to Change

The Resistance to Change Test (Oreg, 2003), based on 18 items, was used to measure four dimensions related to change: Routine Seeking, Emotional Reaction to Imposed Change, Cognitive Rigidity, and Short-Term Focus. We used the total index in our analyses as it showed higher reliability in the original validation paper.

Emotion Regulation

Participants' use of different emotion regulation strategies was investigated with the Emotion Regulation Questionnaire (ERQ, Balzarotti et al., 2010). This is a 10-item questionnaire, in which each item is scored on a 7-point Likert scale (from 1 = "Strongly disagree" to 7 = "Strongly agree"). Items are scored into two separate subscales investigating expressive suppression (basic emotion regulation strategy, i.e., suppressing the behavioral expression of the emotion) and cognitive reappraisal (more advanced cognitive emotion regulation strategy, aimed at modifying the internal representation of an event to change one's own emotional experience) (Gross and John, 2003). Previous literature (*ibidem*) showed that people who use cognitive reappraisal more often tend to experience and express greater positive emotion and lesser negative emotion, whereas people who use expressive suppression experience and express lesser positive emotion, yet experience greater negative emotion.

Questionnaires used to measure predictors (or independent variables) administered to supervisors are:

Trait Emotional Intelligence Questionnaire—Short Form

The Trait Emotional Intelligence Questionnaire—Short Form (Petrides, 2009), based on 30 items, was used to measure trait emotional intelligence.

Interpersonal Reactivity Instrument

The Interpersonal Reactivity Instrument (Davis, 1983) based on 28 items, is aimed at measuring dispositional empathy on four dimensions: Perspective Taking, Fantasy, Empathic Concern and Personal Distress.

Prosocialness Scale for Adults

The Prosocialness Scale for Adults (Caprara et al., 2005), based on 16 items, was used to measure individual differences in adult prosocialness.

Multifactor Leadership Questionnaire

The shortened form of the Multifactor Leadership Questionnaire (Avolio and Bass, 2004), based on 21 items, was used to measure transformational and transactional leadership.

The only questionnaire used to measure predictors (or independent variables) administered to both groups is:

Rational Experiential Inventory–Short Form (REI-S24)

We used the Rational Experiential Inventory (Pacini and Epstein, 1999), in the short version of 24 items, to measure to what degree people engage in automatic-System 1 or deliberate-System 2 modes of thinking.

Outcomes

Questionnaires used to measure outcomes (or dependent variables) administered to employees are:

Inclusion Questionnaire

The Perceived Group Inclusion Scale (Jansen et al., 2014), composed of 16 items, was used in order to measure inclusion in the workplace.

Job Satisfaction Index

We used the traditional index of job satisfaction (Brayfield and Rothe, 1951), based on 18 items.

Communication Satisfaction Scale

The 19-items Communication Satisfaction Scale (adapted from Madlock, 2008) was used to understand the influence of supervisor communication competence on employees satisfaction.

General Data Treatment

All questionnaires were scored according to official guidelines from each validation paper (which typically consisted of summing scores from all items in each factor). Since it was impossible to trace back the exact correspondence between each

colleague and individual supervisors for privacy reasons, average scores were computed for supervisors in each of the six business units involved in the project. These scores were then attributed to each colleague according to their business unit when computing models and algorithms.

In all approaches, the Inclusion questionnaire, Job Satisfaction Index, and Communication Satisfaction Scale were considered as three outcomes of success in job relocation (i.e., endogenous variables in Structural Equation Models and target measures in machine learning algorithms). Accordingly, all other variables from employees (Independence, Openness, Industriousness, Dutifulness, Orientation to Target, to Innovation, to Relation, to Leadership, Resistance to Change, Cognitive Reappraisal, Expressive Suppression, Rational Style, Experiential Style, Age, Seniority) and supervisors (Emotional Intelligence, Empathy, Prosociality, six Leadership Styles, Rational Style, Experiential Style, Age, Seniority) were used as predictors (i.e., exogenous variables in SEM and features in machine learning algorithms).

The Mean Absolute Error (MAE) was used in both SEM and machine learning algorithms to compare the prediction accuracy in the test set. The MAE is a common measure of prediction accuracy in regression models and is computed according to Formula 1:

$$MAE = \frac{\sum_{t=1}^{n} |P_t - O_t|}{n} \quad (1)$$

where O_t is the observed value and P_t is the predicted value. The absolute value in their difference is summed for every predicted point and divided by the number of fitted points n.

Data collected from participants involved in the first data collection (i.e., 348 employees and 35 supervisors) were used as a training set; data collected from participants involved in the second data collection (i.e., 79 employees) were used as an independent test set.

Structural Equation Models

When using the Structural Equation Models (SEM) approach, the analyses were aimed to find the most efficient model in predicting success in job relocation, i.e., predicting the highest variance with the lower number of parameters. Given this aim, the most suitable method to compare models is the Akaike Information Criterion (AIC, Akaike, 1974). The AIC is a goodness of fit index and therefore evaluates how well a model fits the data it was generated from. Let k be the number of estimated parameters in the model and let $\hat{L}$ be the maximum value of the likelihood function for the model: as shown in Formula 2, the AIC also takes into account the model complexity, as it is penalized for the number of parameters included in the model. This penalty is aimed at reducing overfitting. When comparing different models, the model with the lowest AIC is also the most efficient one (i.e., explaining more variance with fewer parameters).

$$AIC = 2k - ln\left(\hat{L}\right) \quad (2)$$

In the models comparison procedure, we started by testing models with the highest number of exogenous variables and then reducing the parameters estimated by the model by removing parameters that did not show a statistically significant effect on the endogenous variables. The whole models comparison procedure is detailed in the results. In all models, the estimator was Maximum Likelihood (ML) and the optimization method was Nonlinear Minimization subject to Box Constraints (NLMINB).

Fit measures (i.e., Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Akaike Information Criterion (AIC) and R^2) from all models are reported in **Table 1**. We indicate that R^2 is a goodness of fit measure, as it shows the explained variance of the outcome variable predicted by the predictors (ranging from 0 to 1), but it is not penalized for the number of parameters in the model as the AIC.

Structural equation models were analyzed in RStudio software (RStudio Inc., 2016) by using the *lavaan* package (Rosseel, 2012).

TABLE 1 | Structural Equation Models summary.

	M1	M2	M3	M4
Number of free parameters	60	57	45	17
Comparative Fit Index (CFI)	> 0.999	> 0.999	> 0.999	0.992
Tucker-Lewis Index (TLI)	> 0.999	> 0.999	> 0.999	0.962
Root Mean Square Error of Approximation (RMSEA)	< 0.001	< 0.001	< 0.001	0.049
Standardized Root Mean Square Residual (SRMR)	< 0.001	< 0.001	< 0.001	0.027
Akaike Information Criterion (AIC)	8,123	8,130	8,113	8,083
R-Square:				
INQ_Inclusion	0.173	0.169	0.159	0.138
JSI_JobSatisfaction	0.185	0.162	0.155	0.129
CSS_CommunicationSatisfaction	0.176	0.175	0.169	0.137

Acronyms for all tables: INQ, Inclusion Questionnaire; JSI, Job Satisfaction Index; CSS, Communication Satisfaction Scale; PQ, Personality Questionnaire; TOM, Motivational Orientation Test (original name: Test di Orientamento Motivazionale); ERQ, Emotion Regulation Questionnaire; REIS24, Rational-Experiential Inventory—Short form 24 items; TEIQ, Trait Emotional Intelligence Questionnaire; IRI, Interpersonal Reactivity Instrument; PSA, Prosocialness Scale for Adults.

Machine Learning Algorithms

Through Machine Learning we aimed at selecting the best features to predict the target variables with a certain degree of accuracy. We opted for using a Least Absolute Shrinkage and Selection Operator (Lasso) regression algorithm for feature selection and a Bootstrap Aggregation of K-Nearest Neighbors classifiers for final classification. Despite having continuous target variables, the latter choice was made in order to reduce data variability given the small sample size. We used Python's Pandas (Version 1.3.3), Numpy (1.21.2) and Scikit-Learn (1.0) toolboxes for this analysis setting seed of 1.

Firstly, some further pre-processing steps were carried out on the data before the model validation process. For each feature, the score of the training set was normalized to obtain a distribution with mean 0 and standard deviation 1 (Z-scores transformation). The mean and standard deviation of the training set distribution were also used as a benchmark to normalize the data points of the test set features. Regarding the target variables, to apply the regression model, no further pre-processing steps were needed. On the other hand, to use the classification algorithm, we transformed each target variable of the training set into an ordinal dichotomous output through a median split, assigning labels of 1 and 2 to the values below and above the median, respectively. We used the same median value of the training set to split and transform the test set target variables.

Then, three types of Machine Learning models were used for the analysis: Lasso Regression, K-Nearest Neighbors Classifier and Bootstrap Aggregation meta-model. The latter is an ensemble learning model created with a bootstrapping method and used to further enhance the performance of a single K-Nearest Neighbors Classifier. For each target variable (Inclusion, Job Satisfaction and Communication Satisfaction), an independent model validation process was followed, thus obtaining three Lasso, three K-Nearest Neighbors and three Bootstrap Aggregation meta-models in total. We also evaluated different classification methods to be used instead of K-Nearest Neighbors; see **Supplementary Material** for a complete description of the classification method choice.

Lasso Regression

We used Lasso regression to select the best features for each model. This step can be useful when dealing with small datasets, as in our case, in order to reduce overfitting and the curse of dimensionality (Chen et al., 2020b). Lasso is a linear regression model where the absolute value of each feature coefficient is added to the loss function (Ordinary Least Square) and multiplied to a constant parameter Alpha (Friedman et al., 2010). This type of regularization (L1) allowed us to perform feature selection, zeroing the coefficient of the less important features in predicting the target variable and using only the remaining ones in the model. This feature made the Lasso Regression one of the three reference models in our research because has a similar objective and output of SEMs and is useful in reducing the numerous features we have in our small sample size dataset.

A Randomized Search Cross-Validation fitted on the training set was used to find parameters that optimize the Lasso regression model performance (hyperparameter tuning—Bergstra and Bengio, 2012). Considering the trade-off between search quality and computational efficiency, we set n= 100 randomized search iterations. For each loop, the algorithm assesses a certain number of random combinations by picking up from a starting grid an entry for each validation parameter. In our case, we made the Lasso hyperparameter tuning only on the Alpha parameter. Accordingly, to select the best Alpha for each model we used a grid of values ranging from 0.1 to 20 in steps of 0.1 (n = 200 maximum Alpha values to select). Then, for each iteration, the model performance is measured using a K-Fold Cross-Validation score (Géron, 2019). This technique avoids a fixed split of the data into a training and a validation set. Accordingly, the algorithm divides the training set into K parts, in our case K = 5 (as a trade-off between quality and computation timing), computing 5 iterations. For each K, one-fifth of the training set was in turn used as validation and the other part as the training set computing a performance score each time. The mean of the MAE for the 5-Fold iterations was the final cross-validation score of each randomized search iteration. Then, the Alpha value which led the model to the lowest MAE was picked and the best model was used for feature selection. For each feature, we obtained a regression coefficient computed on the training set that reflects feature importance in predicting the target variable. This coefficient could be positive or negative and can be interpreted as the increase or decrease, respectively, in the target variable score for one standard deviation change of the feature. For each target variable, features with a coefficient equal to zero have been discarded and not included in the validation process as starting predictors of the K-Nearest Neighbors Classifier.

K-Nearest Neighbors Classifier

The K-Nearest Neighbors was used as the base model for the Bootstrap Aggregation meta-model in order to solve the classification problem. This approach was chosen because it is one of the simplest machine learning classification models. In fact, for each data point, this model predicts the target label by looking at the K closest data points (Géron, 2019). For the Randomized Cross-Validation of the model, we tuned three parameters: *number of neighbors*, *weight* and *metric*. For the *number of neighbors*, thus the K closest point to consider for the classification, a grid was used with odd values ranging from 1 to 85 for both the Inclusion and the Communication Satisfaction target variables, while from 1 to 81 for the Job Satisfaction. This range was chosen in order to avoid overfitting because the maximum possible number of K was equal to half the data points belonging to the least represented class of each target variable. The *weight* parameter, that controls the importance assigned to the K neighbors, had two entries: *uniform,* which assigns the same weight to the neighbors, leading to choose the predicted label according to the most frequent and closest K and *distance* that gives proportionally high weight to the nearest points. *Metric* is the formula used to calculate the distance between data points. The entries are *Manhattan* and *Euclidean* that use an L1 and L2 norm formula, respectively. The model performance was assessed by the accuracy score–i.e., the number of correct predictions out of the total, the higher, the better. Finally, the optimized model was used on the test set. This operation was carried out with two purposes: to assess the goodness of the single K-Nearest

Neighbors classifier, thus lower or better model performance for the training set compared to the test set, respectively; and as a control measure to understand whether the ensemble learning technique would actually lead to improvements to the single base model in terms of performance. This possibility was evaluated using again the accuracy score. Moreover, given the small sample size and the imbalance between classes, we also computed, for each target class, the *precision*, that is the number of correctly predicted samples of the class with respect to all the samples predicted of that class by the classifier, *recall,* which is the number of correctly classified samples of that class compared to the total samples of the same class and *F1 score*, that is a weighted average of precision and recall scores (Saito and Rehmsmeier, 2015).

Bagging Meta-Model

We used an ensemble meta-estimator called Bootstrap Aggregation (Bagging–Breiman, 1996) with each single validated K-Nearest Neighbors classification model. We chose this approach in order to reduce the variance of the single model, that is the error deriving from the noise present in the training sample, whose consequence could be overfitting. The Bagging allows us to train each validated model n times picking at random and with replacement for each iteration 348 data points, corresponding to the sample size of the training set. Thus, some data points may be picked more than once in the same iteration, while others may never be drawn (out of bagging). For each iteration, the trained model makes a prediction. For the classifier (Bagging Classifier) the final prediction of the target variable is the most frequent predicted class.

In the validation process of each bagging meta-model, we chose the best number of iterations with a for loop testing. We set a range of n going from 10 to 100 with a step of 1 as a trade-off between accuracy and computation power. For each iteration, a 5-Fold Cross-Validation on the training set was done by computing the mean and standard deviation of the accuracy score as the performance metric. The iteration parameter that reflected the model with the highest mean accuracy score was considered to be the best one. Then, an optimized meta-model was fitted on the training set and the model performance was evaluated on the test set with accuracy, precision, recall and F1 scores. Specifically, the accuracy was used to assess the possible presence of underfitting or overfitting. The latter conditions were operationalized as a variation of more than one standard deviation between the test set and the training set accuracy scores.

RESULTS

Structural Equation Models

All model summaries can be found in **Table 1**, while statistically significant parameters are reported in **Table 2**. A report of all parameters in all models can be found in the **Supplementary Material**, in which statistically significant effects are highlighted in bold.

The first model included all variables collected from both employees and supervisors (excluding age and seniority). Nevertheless, in this model, the sample covariance matrix was not positive-definite. This result typically implies multicollinearity in the model (i.e., means that at least one of the exogenous variables can be expressed as a linear combination of the others) or the number of observations is less than the number of variables. Best practice, in this case, is to remove highly correlated variables from the model (Field et al., 2012); in our case, the questionnaire showing the highest number of highly correlated variables (i.e., $|r| > 0.5$) was the Multifactor Leadership Questionnaire (MLQ-6S). For this reason, this model was re-run without the 6 Leadership Styles variables.

The following model (M1) included variables collected from both employees and supervisors (excluding Leadership Styles, age and seniority). Information and fit indices from this and the following models are summarized in **Table 1**. Since no variables collected from supervisors showed statistically significant effects on any of the three endogenous variables, we removed these exogenous variables and added age and seniority (from both employees and supervisors) to the next model (M2). Also in this case, age and seniority (from both employees and supervisors) showed no statistically significant effects on any of the three endogenous variables. Therefore, in M3 only variables collected from employees were included. This model was further reduced by including only parameters showing significant effects in M3 (in a feature selection fashion), leading to an optimized model (M4). As shown in **Table 1**, the AIC was lower in M4 (8083) than in M3 (8113), displaying thus increased efficiency in explaining data in the optimized model (i.e., M4) compared to M3. Nevertheless, a Likelihood Ratio Test (LRT) was performed between the two best models (i.e., M3 and M4) to compare the likelihood of the two models. This LRT showed that the models' likelihood was not significantly different ($\Delta\chi^2$ (4) = 7.35, p = 0.119), despite the relevant change in the number of estimated parameters (45 in M3 vs. 17 in M4).

Statistically significant effects are reported in **Table 2**. Significant effects showed noteworthy consistency across different models (only two effects were not significant in M4) and are summarized in **Figure 1**. Orientation to relation showed a significant positive effect toward all three outcome measures, while resistance to change presented a significant negative effect toward all outcome variables. Therefore, employees higher in orientation to relation and lower in resistance to change showed better success in relocation. Industriousness showed significant negative effects toward inclusion and communication satisfaction, while dutifulness displayed significant positive effects toward the two same outcome variables. Finally, orientation to objective showed a significant positive effect toward communication satisfaction. This latter effect and the effect of industriousness toward communication satisfaction did not show to be statistically significant in model M4.

Testing Sample

Predicted scores for the three outcome measures were computed in the testing sample (n = 79) according to the parameters found in models M3 and M4 (i.e., the two models with the lowest AIC, representing the largest explained variance with the lowest number of parameters). Predicted scores were then compared to observed scores to test the predictive accuracy of the models by

TABLE 2 | Summary of statistically significant parameters.

	Estimate	Std. Err	*z* value	P (>|z|)	Std. lv	Std. all
M1						
Regressions						
INQ_Inclusion ~						
PQ_Industriousness	−0.422	0.182	−2.316	0.021	−0.422	−0.137
PQ_Dutifulness	0.439	0.173	2.534	0.011	0.439	0.149
TOM_Relation	0.577	0.176	3.271	0.001	0.577	0.220
Resistance to change	−0.170	0.077	−2.192	0.028	−0.170	−0.147
JSI_JobSatisfaction ~						
TOM_Relation	0.560	0.191	2.929	0.003	0.560	0.196
Resistance to change	−0.287	0.084	−3.420	0.001	−0.287	−0.227
CSS_CommunicationSatisfaction~						
PQ_Industriousness	−0.410	0.194	−2.115	0.034	−0.410	−0.125
PQ_Dutifulness	0.429	0.184	2.327	0.020	0.429	0.137
TOM_Target	0.384	0.174	2.207	0.027	0.384	0.207
TOM_Relation	0.647	0.188	3.446	0.001	0.647	0.231
Resistance to change	−0.214	0.082	−2.597	0.009	−0.214	−0.173
M2						
Regressions						
INQ_Inclusion ~						
PQ_Industriousness	−0.418	0.182	−2.300	0.021	−0.418	−0.136
PQ_Dutifulness	0.444	0.174	2.555	0.011	0.444	0.151
TOM_Relation	0.615	0.179	3.438	0.001	0.615	0.234
Resistance to change	−0.188	0.077	−2.448	0.014	−0.188	−0.162
JSI_JobSatisfaction ~						
TOM_Relation	0.593	0.196	3.028	0.002	0.593	0.207
Resistance to change	−0.323	0.084	−3.846	<0.001	−0.323	−0.255
CSS_CommunicationSatisfaction~						
PQ_Industriousness	−0.413	0.193	−2.137	0.033	−0.413	−0.126
PQ_Dutifulness	0.439	0.185	2.378	0.017	0.439	0.140
TOM_Target	0.379	0.175	2.169	0.030	0.379	0.204
TOM_Relation	0.664	0.190	3.493	<0.001	0.664	0.237
Resistance to change	−0.212	0.081	−2.605	0.009	−0.212	−0.172
M3						
Regressions						
INQ_Inclusion ~						
PQ_Industriousness	−0.368	0.180	−2.044	0.041	−0.368	−0.119
PQ_Dutifulness	0.421	0.174	2.422	0.015	0.421	0.143
TOM_Relation	0.598	0.177	3.369	0.001	0.598	0.228
Resistance to change	−0.187	0.077	−2.433	0.015	−0.187	−0.161
JSI_JobSatisfaction ~						
TOM_Relation	0.606	0.194	3.120	0.002	0.606	0.211
Resistance to change	−0.320	0.084	−3.804	<0.001	−0.320	−0.253
CSS_CommunicationSatisfaction~						
PQ_Industriousness	−0.401	0.191	−2.099	0.036	−0.401	−0.122
PQ_Dutifulness	0.429	0.184	2.330	0.020	0.429	0.137
TOM_Target	0.384	0.174	2.202	0.028	0.384	0.206
TOM_Relation	0.652	0.188	3.466	0.001	0.652	0.233
Resistance to change	−0.208	0.081	−2.558	0.011	−0.208	−0.169

(Continued)

TABLE 2 | Continued

	Estimate	Std. Err	*z* value	P (>|z|)	Std. lv	Std. all
			M4			
			Regressions			
INQ_Inclusion ~						
PQ_Industriousness	−0.299	0.141	−2.117	0.034	−0.299	−0.097
PQ_Dutifulness	0.292	0.133	2.193	0.028	0.292	0.099
TOM_Relation	0.793	0.136	5.827	<0.001	0.793	0.303
Resistance to change	−0.215	0.058	−3.708	<0.001	−0.215	−0.186
JSI_JobSatisfaction ~						
TOM_Relation	0.755	0.143	5.269	<0.001	0.755	0.264
Resistance to change	−0.298	0.063	−4.712	<0.001	−0.298	−0.236
CSS_CommunicationSatisfaction~						
PQ_Dutifulness	0.348	0.148	2.351	0.019	0.348	0.112
TOM_Relation	0.755	0.155	4.864	<0.001	0.755	0.271
Resistance to change	−0.242	0.069	−3.493	<0.001	−0.242	−0.196

Std. lv, effects estimate standardized on the first manifest variable; in our models, this corresponds to the default estimate. Std. all, effects estimate standardized on all manifest variables.

using the Mean Absolute Error (MAE). MAE values for the two models M3 and M4 are reported in **Table 3**. Higher MAE values in M4 compared to M3 showed that prediction accuracy in the testing phase was higher in M3 than M4. This result indicates that the higher number of parameters in M3 contributed to predicting more accurately the outcome scores in the testing sample, thus generalizing better the results to an external dataset.

Machine Learning Algorithms

Like SEMs, supervisor-related feature variables have been dropped from Machine Learning models due to their multicollinearity. The rest of the features have all been included in the validation process of each model.

Feature Selection

Inclusion

For the inclusion target variable, the best Lasso Regression model (Alpha = 0.6, MAE = 9.71) reported a major influence of TOM Relation (2.9), Resistance to change (−1.7), PQ Dutifulness (1.3), TOM Target (0.42), PQ Industriousness (−0.14), ERQ Suppression (−0.06) and REI Rational (0.03—see **Figure 2A**).

Job Satisfaction

The best Lasso Regression model (Alpha = 0.6, MAE = 10.91) for the Job Satisfaction target variable showed a notable influence of TOM Relation (2.74), Resistance to change (−2.74), PQ Dutifulness (1.08), TOM Target (0.46), PQ Industriousness (0.14), and Seniority (0.05—see **Figure 2B**).

Communication Satisfaction

Considering as target variable the Communication Satisfaction, the best Lasso Regression model (Alpha = 0.6, MAE = 10.45) reported a major influence of TOM Relation (2.74), Resistance to change (−1.99), PQ Dutifulness (1.46), ERQ Suppression (−0.93), TOM Target (0.58), TOM Leadership (−0.24), PQ Industriousness (−0.18—see **Figure 2C**).

Classification Models

Inclusion

The best K-Nearest Neighbor Classifier (weights = distance, n neighbors = 31 and metric = euclidean) reported a cross-validation accuracy score of the training set higher (0.66) compared to the test set (0.61). The prediction of the low inclusion class (1, N = 32) on the test set reached a precision of 0.52 and a recall of 0.50 (F1 score = 0.51). This was lower compared to the high inclusion class (2, N = 47) that scored 0.67 on precision and 0.68 on the recall (F1 score = 0.67).

The best Bagging Classifier (n estimators = 43—see **Figure 3A**) did not show overfitting or underfitting. Indeed, the cross-validation accuracy score of the training set was 0.65 ± 0.05 compared to 0.63 of the test set. Moreover, the Bagging Classifier slightly enhanced the prediction performance of both the low inclusion (precision = 0.55, recall = 0.53 and F1 score = 0.54) and the high inclusion class (precision = 0.60, recall = 0.70 and F1 score = 0.69) compared to the single K-Nearest Neighbor Classifier.

Job Satisfaction

The best K-Nearest Neighbor Classifier (weights = distance, n neighbors = 47 and metric = manhattan) reported a training set cross-validation accuracy score of 0.65 compared to 0.63 of the test set . Moreover, on the test set, the prediction performance of the low job satisfaction class (N = 28) scored lower (precision = 0.48, recall = 0.54 and F1 score = 0.51) compared to the high job satisfaction class (precision = 0.73, recall = 0.69 and F1 score = 0.71, N = 51).

The best Bagging Classifier (n estimators = 56—see **Figure 3B**) did not show overfitting or underfitting and displayed a slight increase in model performance. Accordingly, the cross-validation accuracy score of the training set was 0.64 ± 0.03, while of the test set was 0.66. Moreover, the Bagging Classifier slightly increased the prediction performance of both the low

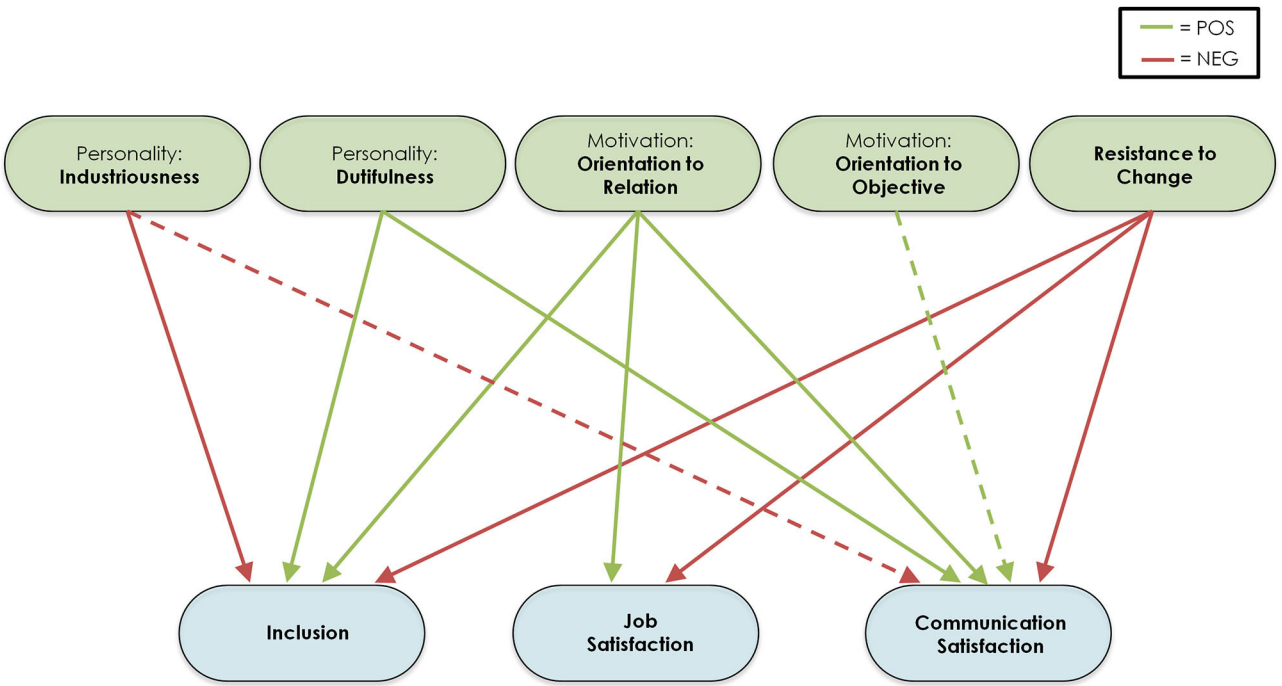


FIGURE 1 | Summary of results from Structural Equation Models. This figure represents the statistically significant effects found across all Structural Equation Models tested. Green boxes represent exogenous (independent) variables measured in employees, while blue boxes represent endogenous (dependent) variables from employees. Green arrows represent statistically significant positive effects, red arrows represent statistically significant negative effects. Dashed lines represent two effects found as statistically significant in all tested models except for the model with reduced parameters (M4).

TABLE 3 | Mean Absolute Error (MAE) in the testing sample in Structural Equation Models.

MAE	M3	M4
INQ_Inclusion	11.04	13.33
JSI_JobSatisfaction	10.54	16.97
CSS_CommunicationSatisfaction	10.00	13.46

job satisfaction (precision = 0.52, recall = 0.54 and F1 score = 0.53) and the high job satisfaction class (precision = 0.74, recall = 0.73 and F1 score = 0.73) compared to the single K-Nearest Neighbor Classifier.

Communication Satisfaction

We observed that the best K-Nearest Neighbor Classifier (weights = uniform, n neighbors = 59 and metric = euclidean) reported a higher cross-validation accuracy score of the training set (0.66) compared to the test set (0.65). Again, on the test set, the prediction performance of the low communication satisfaction class (N = 32) scored lower (precision = 0.56, recall = 0.56 and F1 score = 0.56) compared to the high communication satisfaction class (precision = 0.70, recall = 0.70 and F1 score = 0.70, N = 47).

The best Bagging Classifier (n estimators = 38—see **Figure 3C**) did not show overfitting or underfitting, but a small increment in model performance. Indeed, the cross-validation accuracy score of the training set reached 0.66 ± 0.04, while of the test set was 0.66. Moreover, the Bagging Classifier increased the prediction performance of both the low communication satisfaction class (precision = 0.58, recall = 0.56 and F1 score = 0.57) and the high communication satisfaction class (precision = 0.71, recall = 0.72 and F1 score = 0.72) compared to the single K-Nearest Neighbor Classifier.

DISCUSSION

Internal mobility has been previously investigated as a specific form of job relocation. Nevertheless, no studies have identified what characteristics (in both employees and supervisors) can predict successful mobility in terms of job satisfaction. In this study, we compared different classes of models to identify the most efficient technique to predict successful mobility, i.e., (i) traditional Structural Equation Modeling (SEM), with two families of Machine Learning algorithms: (ii) regressors, specifically least absolute shrinkage and selection operator (Lasso) aimed at feature selection and (iii) classifiers, specifically *k*-nearest neighbors algorithm (*k*-NN). Results showed different performances for the three classes of models, ranging from low to medium accuracy.

A crucial aspect in results is the consistency among statistically significant effects. All SEMs replicated the statistical significance of the effects involving five predictors: (i) orientation to relation

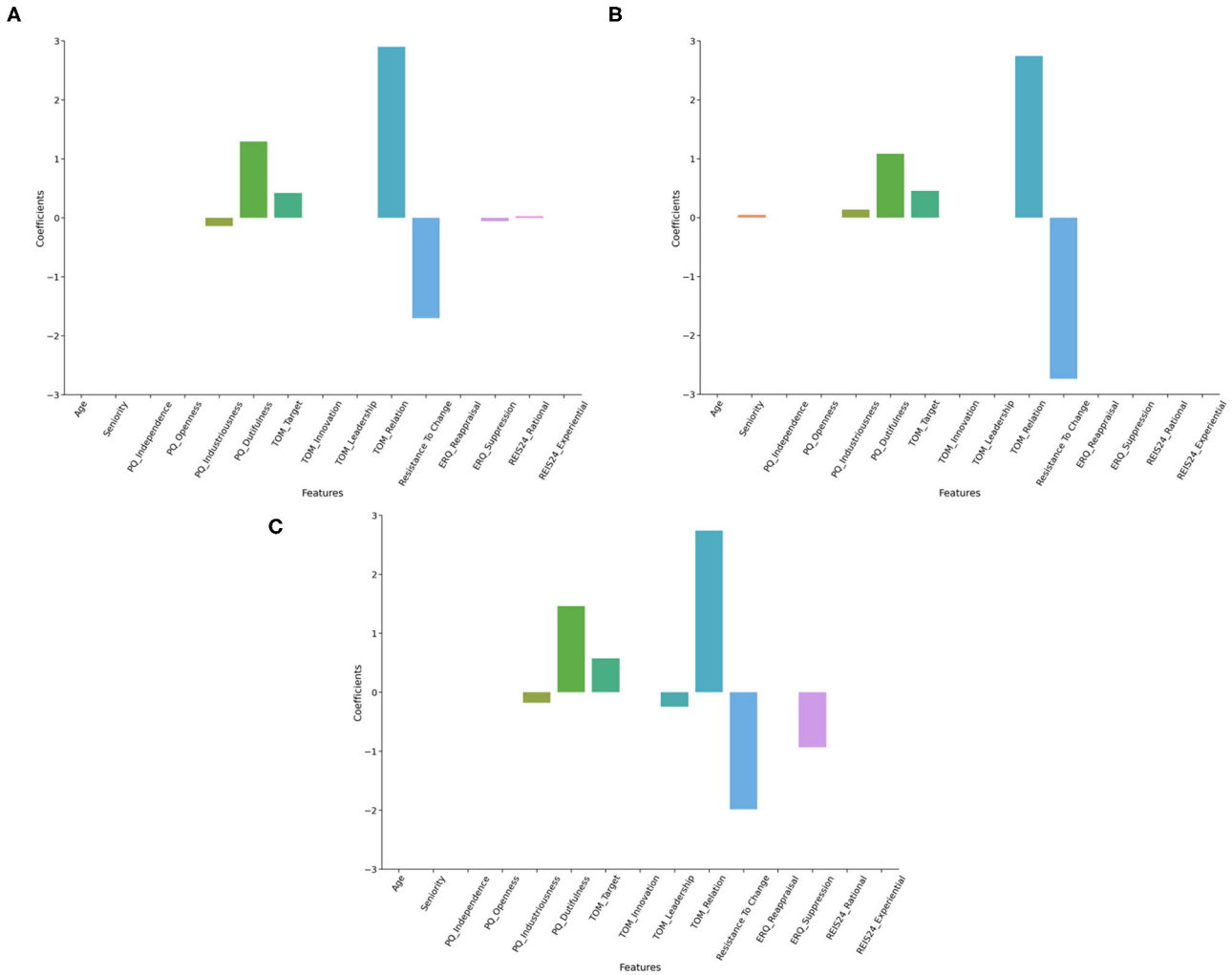


FIGURE 2 | Feature coefficients in Lasso regressors. This figure represents the coefficients of features surviving the regularization in Lasso regressors. Models predicting inclusion **(A)**, job satisfaction **(B)**, and communication satisfaction **(C)** are represented.

and resistance to change showed to influence all endogenous variables with high effect sizes, (ii) dutifulness, industriousness and orientation to objective displayed a relevant but less consistent contribution, as they did not influence all endogenous variables, and in the reduced model (M4) two of the parameters appeared not to be statistically significant. A decisive result is a consistency with the Lasso results, in particular for what concerns orientation to relation and resistance to change. Minor results also identified employees' seniority, rational style, expressive suppression and orientation to leadership as relevant in Lasso models. Feature selection in Lasso models allows predicting the target variable by zeroing the coefficient of the less important features and using only the remaining ones in the model. The fact that two models stemming from different approaches to data analysis replicated comparable results is acceptable evidence for results consistency.

Previous literature showed that personality dispositions, resistance to change and social orientations are crucial for mobility relocation (Otto and Dalbert, 2012), and managing resistance to change in the workplace appears to be fundamental to stimulate job satisfaction (Laframboise et al., 2003). Resistance to change is defined as an individual's dispositional inclination to resist changes (declined in routine seeking, emotional reaction to imposed change, cognitive rigidity, and short-term focus, Oreg, 2003) and therefore the influence of this construct on job relocation appears to be self-evident. Nevertheless, having measured the influence of this variable on all three indices of successful relocation with such consistency is meaningful evidence for the robustness of this relation. At the same time, the orientation to social relation was previously shown to be fundamental in some careers with a great number of employee-customer exchanges (Alessandri and Russo, 2011). According to our data, orientation to relation appears to be crucial in successful relocation for any employee, and not only for a specific personality phenotype as previously found (Otto and Dalbert, 2012). This result shows that motivational orientation toward

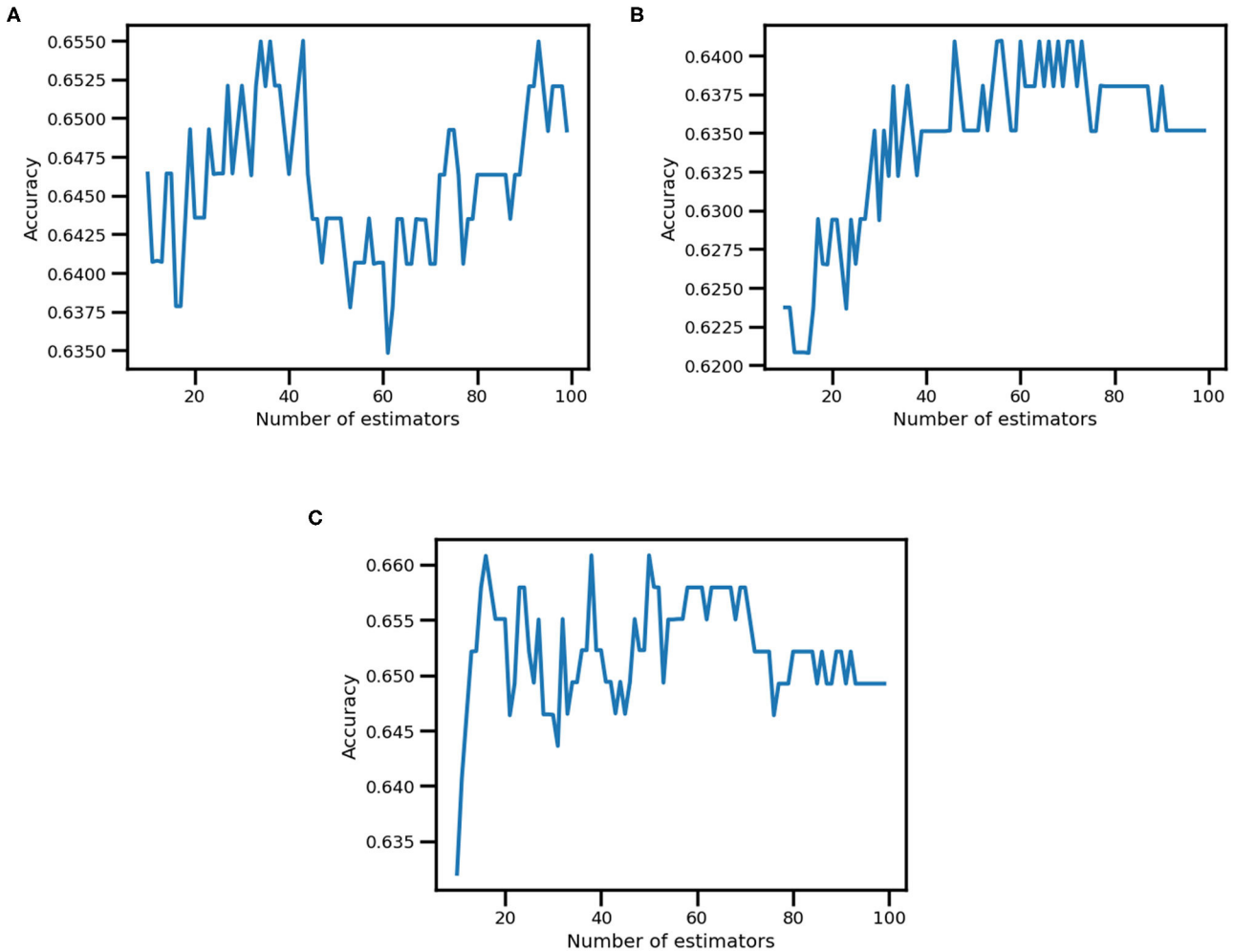


FIGURE 3 | Accuracy in Ensemble learning *k*-NN. This figure represents the accuracy variation (on the y-axis) based on the number of estimators (i.e., number of classifiers included in the Bagging Classifier, on the x-axis) for inclusion **(A)**, job satisfaction **(B)**, and communication satisfaction **(C)** output variables.

social bonding is crucial when a change in the social group (i.e., job relocation) is experienced. This motivational orientation can stimulate a person to be accepted by the new social group and, consequently, experience higher job satisfaction in the workplace.

In some models we reported, industriousness predicted negatively successful relocation. This result is related to the fact that the industriousness facet captured several aspects related to workaholism (e.g., being under constant pressure, putting work before pleasure) (Jackson et al., 1996a), that can be considered as the extreme opposite of job satisfaction (Burke, 2001). Dutifulness appears to be relevant for a successful relocation, as it explains aspects related to prudence and compliance with rules, which are fundamental for inclusion in a new social group. Finally, the suppression emotion regulation strategy negatively predicts success in job relocation in some Lasso models, as it represents a basic regulation strategy, which is often not effective in intrapersonal and interpersonal functioning (Gross and John, 2003).

In SEMs, M3 and M4 are the most efficient models, considering only exogenous variables from employees. When choosing the best model, on the one hand, the likelihood ratio test (LRT) in the training set did not find a significant difference in explained variance. This result favored M4, as it explained a comparable level of variance with an extremely lower number of parameters (as displayed by lower AIC). On the other hand, validation on the test set appeared to favor M3, as it showed lower MAE values than M4 across all three endogenous variables. Therefore, the higher number of parameters in M3 occurred to describe better data in the test set. To sum up, we cannot univocally prefer one of the two models, as M3 generalized better results on the test set, while M4 performed more efficiently on the training set. On the contrary, models M1 and M2 (the least efficient ones, in terms of AIC) showed that

supervisors' data, as well as age and seniority, are not relevant in predicting success in relocation. The interpretation of these (null) results would imply low relevance of the supervisors' role in relocation success; nevertheless, we cannot interpret this result because multicollinearity generated by using aggregate data for supervisors could invalidate them. Moreover, there is a substantial overlap between the features surviving in Lasso regressors and statistically significant exogenous variables in SEMs. Since in Lasso models only the most relevant features survive, the meaningful agreement between these two classes of models shows the statistical relevance of these effects and the reliability of the results.

When chunking the information with feature selection and classifying high vs. low relocation success (i.e., job satisfaction, inclusion and communication satisfaction) by means of *k*-NN algorithms, it was possible to predict success with 64–66% accuracy in the training set and 61–65% in the test set, which represent medium performance. Overall, bagging meta-models displayed slightly higher performances compared to the single *k*-NNs and, despite the small dataset, they did not show underfitting or overfitting. The ensemble meta-model reached a 64–66% accuracy in the training set and a 63–66% on the test set, raising also the F1 score of each predicted class. Indeed, bagging is based on bootstrapping techniques, which is frequently used with small datasets. This represents further evidence that results could be improved by using larger datasets in both training and testing phases (see next paragraph). These results also show that the algorithms often generalized well on the test set. However, the algorithms showed to predict high success in relocation (values above the median) with generally higher precision and recall than low success (below the median). An explanation of this result can be related to the relatively small sample size, which was generally biased toward high scores (i.e., more participants with high values in the three outcomes than with low values). Because of this imbalance in classes, algorithms were most probably better trained on identifying participants with high relocation success than low success.

In summary, our results suggest that SEMs (more broadly used in HR literature, Borgogni et al., 2010) can estimate successful relocation with average accuracy. Resistance to change and orientation to relation were found to be the most relevant predictors, as confirmed by Lasso regressors. Bagging Classifier with *k*-NN as base estimator displayed good performance in classifying data, showing potentiality in using machine learning techniques in predictive HR analytics. The performance of these algorithms could be increased in future research by increasing sample size and including further predictors, as specified hereafter.

Limitations

The main methodological limitation of this study is represented by having used aggregate data for supervisors. Unfortunately, for privacy reasons, this was our better option, since we could trace supervisors back to Business Units, but not to individual direct reports. These aggregate data generated multicollinearity in both SEMs and machine learning algorithms using data from supervisors, thus making null effects in these predictors hard to interpret. The lack of influence from supervisors' features on the relocation success in employees would be an outstanding result in terms of implications, but, unfortunately, this interpretation is impractical for methodological limitations.

Despite a discretely large sample size in both training and testing sets, this study could have benefitted from larger samples. The main reason is the large number of predictors involved in models, which would need an adequate number of participants in order to fit the data (Sawyer, 1982; Fursov et al., 2018).

Another theoretical limitation of our study is the fact that we considered only post-relocation data. Dependent and independent variables have been measured only after the transfer had occurred (more than 6 months before data collection). Hence, we do not have information about such variables before the transfer, that is, we do not know the level of job satisfaction, inclusion, and communication satisfaction related to the old job position. Hence, we cannot exclude that there are differences in job satisfaction between old and new positions inherently related to the specificities of the new job position. Actually, in our study, we adopted the point of view of such practitioners, interested in estimating future relocation success for normative purposes.

Different classification methods could have been chosen instead of K-Nearest Neighbors. In our specific case, these alternative methods would have yielded similar results (**Supplementary Material**). This aspect may represent a future development of this study, consisting in comparing different algorithms in several fields of predictive HR analytics according to different research or market questions.

Future directions for this study may also consider adding predictors which are known to be predictive of satisfaction in relocation. These predictors would include both stable psychological traits (e.g., personality factors) and social-environmental features (e.g., economic conditions, family characteristics, differences among industries and occupations).

Conclusions

To sum up, we found that traditional SEMs predicted with average accuracy successful relocation, thus identifying the relevance of resistance to change and orientation to relation. Lasso regressors confirmed the influence of these variables; while *k*-NN classifiers displayed good performance in classifying data.

The practical application of these results is prominent in the field of HR, as we have empirical evidence for pushing for training employees who are going to be relocated in reducing their resistance to change, thus promoting resilience, and improving their social skills, aside from training in hard skills. Moreover, we show that artificial intelligence algorithms could help in selecting employees who are more prone to be relocated to a new job position, with all due ethical reservations and in conjunction with further methods such as interviews, validated questionnaires, et cetera.

In the field of predictive HR analytics, this is a seminal result comparing methods stemming from different research traditions. There is considerable room for improvement since the models' efficiency was typically not high. Future research will have to consider different variables and different approaches, but we believe it is crucial to start comparing the performance from

divergent methods. The aim of this comparison is not to find that one method is better than others in the entire field of HR analytics, but to make them all available and comparable according to different research (and market) questions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Area Vasta Nord Ovest Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FB, AMal, and ER contributed to the design, the conception of the research, and contributed to the manuscript revision. FB contributed to the implementation of the study and data collection. FB and FD contributed to analyzing data. FB, FD, and AMas contributed to writing the manuscript. SD'A as a member of Intesa Sanpaolo Innovation Center S.p.A., assisted with the project management between IMT School for Advanced Studies Lucca and Intesa Sanpaolo Group. All authors contributed to the article, read, and approved the submitted version.

FUNDING

The authors declare that this study received funding from Intesa Sanpaolo Innovation Center S.p.A. The funder was not involved in the study design, collection, analysis, interpretation of data, and the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The research was conducted under a cooperative agreement between the IMT School for Advanced Studies Lucca and Intesa Sanpaolo Innovation Center S.p.A. and Intesa Sanpaolo banking group. The authors would like to thank Francesca Maggi from Intesa Sanpaolo Innovation Center—Neuroscience Lab for her precious contribution during data collection, Alessia Patuelli for her suggestions in formalizing the study, Luigi Ruggerone (Head of Trend Analysis and Applied Research—Intesa Sanpaolo Innovation Center S.p.A.) for his insightful comments, and Linda Fiorini for her precious observations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2022.848015/full#supplementary-material

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19, 716–723. doi: 10.1109/TAC.1974.1100705

Alessandri, G., and Russo, S. D. (2011). Concurrent and predictive validity of the Motivational Orientation Test General Version (TOM-VG). *Giornale Italiano di Psicologia* 3, 691–700.

Andresen, M., and Margenfeld, J. (2015). International relocation mobility readiness and its antecedents. *J. Manag. Psychol.* 30. doi: 10.1108/JMP-11-2012-0362

Avolio, B., and Bass, B. (2004). *Multifactor leadership questionnaire (TM)*. Menlo Park, CA: Mind Garden, Inc.

Aziri, B. (2011). Job satisfaction: a literature review. *Manag. Res. Pract.* 3, 77–86.

Balzarotti, S., John, O. P., and Gross, J. J. (2010). An Italian adaptation of the emotion regulation questionnaire. *Eur J Psychol Assess. 26*, 61–67. doi: 10.1027/1015-5759/a000009

Bassi, L. (2011). Raging debates in HR analytics. *People Strat.* 34, 14.

Ben-Gal, H. C. (2019). An ROI-based review of HR analytics: practical implementation tools. *Personnel Rev.* 48. doi: 10.1108/PR-11-2017-0362

Benson, A., and Rissing, B. A. (2020). Strength from within: internal mobility and the retention of high performers. *Organization Sci.* 31, 1475–1496. doi: 10.1287/orsc.2020.1362

Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *J. Mach. Learn Res.* 13, 281–305.

Bidwell, M. (2011). Paying more to get less: the effects of external hiring versus internal mobility. *Adm. Sci. Q.* 56, 369–407. doi: 10.1177/0001839211433562

Bidwell, M., and Mollick, E. (2015). Shifts and ladders: comparing the role of internal and external mobility in managerial careers. *Organization Sci.* 26, 1629–1645. doi: 10.1287/orsc.2015.1003

Borgogni, L., Dello Russo, S., Petitta, L., and Vecchione, M. (2010). Predicting job satisfaction and job performance in a privatized organization. *Int. J. Public Adm. 13*, 275–296. doi: 10.1080/10967494.2010.504114

Borgogni, L., Petitta, L., and Barbaranelli, C. (2004). Il test di orientamento motivazionale (tom) come strumento per la misura della motivazione al lavoro. *Bollettino Di Psicologia Applicata.* 43–52.

Brayfield, A. H., and Rothe, H. F. (1951). An index of job satisfaction. *Am. J. Appl. Psychol.* 35, 307. doi: 10.1037/h0055617

Breiman, L. (1996). Bagging predictors. *Mach. Learn.* 24, 123–140. doi: 10.1007/BF00058655

Burke, R. J. (1986). Occupational and life stress and the family: Conceptual frameworks and research findings. *Appl Psychol.* 35, 347–368. doi: 10.1111/j.1464-0597.1986.tb00934.x

Burke, R. J. (2001). Workaholism Components, Job Satisfaction, and Career Progress1. *J. Appl. Soc. Psychol.* 31, 2339–2356. doi: 10.1111/j.1559-1816.2001.tb00179.x

Caprara, G. V., Steca, P., Zelli, A., and Capanna, C. (2005). A new scale for measuring adults' prosocialness. *Eur. J. Psychol. Assess.* 21, 77–89. doi: 10.1027/1015-5759.21.2.77

Chen, H., Dunford, B. B., and Boss, W. (2020a). (Non-) star struck: internal mobility and the network evolution of B-performers. *Acad. Manag. Ann.* 1, 18169. doi: 10.5465/AMBPP.2020.18169abstract

Chen, R.-C., Dewi, C., Huang, S.-W., and Caraka, R. E. (2020b). Selecting critical features for data classification based on machine learning methods. *J. Big Data. 7*, 52. doi: 10.1186/s40537-020-00327-4

Chien, C. F., and Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Syst. Appl.* 34, 280–290. doi: 10.1016/j.eswa.2006.09.003

Chong, D., and Shi, H. (2015). Big data analytics: a literature review. *J. Manag. Anal.* 2, 175–201. doi: 10.1080/23270012.2015.1082449

Davenport, T. H., Harris, J., and Shapiro, J. (2010). Competing on talent analytics. *Harv. Bus. Rev.* 88, 52–58.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

Edwards, M. R., and Edwards, K. (2019). *Predictive HR analytics*: *Mastering the HR metric*. London, UK: Kogan Page Publishers.

Falletta, S. (2014). In search of HR intelligence: evidence-based HR analytics practices in high performing companies. *People Strat.* 36, 28.

Field, A., Miles, J., and Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: SAGE Publications.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33. doi: 10.18637/jss.v033.i01

Fursov, V. A., Gavrilov, A. V., and Kotov, A. P. (2018). Prediction of estimates' accuracy for linear regression with a small sample size. *2018 41st Int. Conf. Telecommun. Signal Process.* 1–7. doi: 10.1109/TSP.2018.8441385

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.)*. O'Reilly Media.

Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* 85, 348–362. doi: 10.1037/0022-3514.85.2.348

Haan, M., and Cardoso, M. (2020). Job changing and internal mobility: Insights into the "declining duo" from Canadian administrative data. *Popul. Space Place.* 26, e2324. doi: 10.1002/psp.2324

Hackman, J. R., and Oldham, G. R. (1976). Motivation through the design of work: test of a theory. *Organ. Behav. Hum. Decis. Process.* 16, 250–279. doi: 10.1016/0030-5073(76)90016-7

Haddad, M. J., and Sanders, D. A. (2020). Artificial neural network approach for business decision making applied to a corporate relocation problem. *J. Bus. Res.* 8. 180–195. doi: 10.14738/abr.86.8202

Hamilton, R. H., and Davison, H. K. (2021). Legal and ethical challenges for HR in machine learning. *Employee Responsibilities Rights J.* 1–21. doi: 10.1007/s10672-021-09377-z

Harter, J. K., Schmidt, F. L., and Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *Am. J. Appl. Psychol.* 87, 268. doi: 10.1037/0021-9010.87.2.268

Herzberg, F. (1964). The motivation-hygiene concept and problems of manpower. *Pers. Adm.* 27, 3–7.

Hogan, R., and Hogan, J. (1997). *Hogan development survey* manual. OK: Hogan Assessment Systems.

Hogan, R., and Hogan, J. (2007). *Hogan personality inventory* manual. OK: Hogan Assessment Systems.

Holsapple, C., Lee-Post, A., and Pakath, R. (2014). A unified foundation for business analytics. *Decis. Support Syst.* 64, 130–141. doi: 10.1016/j.dss.2014.05.013

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Acad. Manag. Ann.* 38, 635–672. doi: 10.5465/256741

Huseman, R. C., Hatfield, J. D., and Miles, E. W. (1987). A new perspective on equity theory: The equity sensitivity construct. *Acad Manage Rev.* 12, 222–234. doi: 10.2307/258531

Iaffaldano, M. T., and Muchinsky, P. M. (1985). Job satisfaction and job performance: a meta-analysis. *Psychol. Bull.* 97, 251–273. doi: 10.1037/0033-2909.97.2.251

Jackson, D. N., Ashton, M. C., and Tomes, J. L. (1996a). The six-factor model of personality: facets from the big five. *Pers. Individ. Differ.* 21, 391–402. doi: 10.1016/0191-8869(96)00046-3

Jackson, D. N., Paunonen, S. V., Fraboni, M., and Goffin, R. D. (1996b). A five-factor versus six-factor model of personality structure. *Pers. Individ. Differ.* 20, 33–45. doi: 10.1016/0191-8869(95)00143-T

Jansen, W. S., Otten, S., van der Zee, K. I., and Jans, L. (2014). Inclusion: conceptualization and measurement: inclusion: conceptualization and measurement. *Eur. J. Soc. Psychol.* 44, 370–385. doi: 10.1002/ejsp.2011

Judge, T. A., Parker, S. K., Colbert, A. E., Heller, D., and Ilies, R. (2002). "Job satisfaction: A cross-cultural review," in *Handbook of Industrial, Work and Organizational Psychology, Vol. 2, Organizational Psychology*, eds N. Anderson, D. S. Ones, H. K. Sinangil and C. Viswesvaran (Sage Publications, Inc.), 25–52.

Judge, T. A., Thoresen, C. J., Bono, J. E., and Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychol. Bull.* 127, 376–407. doi: 10.1037/0033-2909.127.3.376

Kakulapati, V., Chaitanya, K. K., Chaitanya, K. V. G., and Akshay, P. (2020). Predictive analytics of HR-A machine learning approach. *J. Stat. Softw.* 23, 959–969. doi: 10.1080/09720510.2020.1799497

Kapoor, B., and Sherif, J. (2012). Human resources in an enriched environment of business intelligence. *Kybernetes.* 41, 1625–1637. doi: 10.1108/03684921211276792

Kolesar, P., and Walker, W. E. (1974). An algorithm for the dynamic relocation of fire companies. *Operat. Res.* 22, 249–274. doi: 10.1287/opre.22.2.249

Laframboise, D., Nelson, R. L., and Schmaltz, J. (2003). Managing resistance to change in workplace accommodation projects. *J. Facil. Manag.* 1, 306–321. doi: 10.1108/14725960310808024

Latham, G. P., and Budworth, M. H. (2007). "The study of work motivation in the 20th century", in *Historical perspectives in industrial and organizational psychology*. Koppes, L. L. (Ed.), Mahwah, NJ: Lawrence Erlbaum. p. 353–381

Locke, E. A. (1976). "The nature and causes of job satisfaction", in, *Handbook of industrial and organizational psychology*, Dunnette, M. D. (Ed.). Chicago: Rand McNally. p.1297–1349.

Madlock, P. E. (2008). The link between leadership style, communicator competence, and employee satisfaction. *J Busi Communicat.* 45, 61–78. doi: 10.1177/0021943607309351

Malizia, A. P., Bassetti, T., Menicagli, D., Patuelli, A., D'Arcangelo, S., Lattanzi, N., et al. (2021). Not all sales performance is created equal: personality and interpersonal traits in inbound and outbound marketing activities. *Arch. Ital. Biol.* 159(3-4), 107–122.

Mohr, A. T., and Puck, J. F. (2007). Role conflict, general manager job satisfaction and stress and the performance of international joint ventures. *Eur. Manag. J.* 25, 25–35. doi: 10.1016/j.emj.2006.11.003

Moorman, R. H. (1993). The influence of cognitive and affective based job satisfaction measures on the relationship between satisfaction and organizational citizenship behavior. *Human Relat.* 46, 759–776. doi: 10.1177/001872679304600604

Munton, A. G. (1990). Job relocation, stress and the family. *J. Organ. Behav.* 11, 401–406. doi: 10.1002/job.4030110507

Naveh, Y., Richter, Y., Altshuler, Y., Gresh, D. L., and Connors, D. P. (2007). Workforce optimization: Identification and assignment of professional workers using constraint programming. *IBM J Res Dev.* 51, 263–279. doi: 10.1147/rd.513.0263

Near, J. P., Rice, R. W., and Hunt, R. G. (1980). The relationship between work and nonwork domains: A review of empirical research. *Acad Manage Rev.* 5, 415–429. doi: 10.5465/amr.1980.4288868

Noe, R. A., Steffy, B. D., and Barber, A. E. (1988). An investigation of the factors influencing employees' willingness to accept mobility opportunities. *Pers. Psychol.* 41, 559–580. doi: 10.1111/j.1744-6570.1988.tb00644.x

Oreg, S. (2003). Resistance to change: developing an individual differences measure. *Am. J. Appl. Psychol.* 88, 680–693. doi: 10.1037/0021-9010.88.4.680

Otto, K., and Dalbert, C. (2012). Individual differences in job-related relocation readiness: The impact of personality dispositions and social orientations. *Career Dev. Int.* 17, 168–186. doi: 10.1108/13620431211225340

Pacini, R., and Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *J. Pers. Soc. Psychol.* 76, 972. doi: 10.1037/0022-3514.76.6.972

Petitta, L., Borgogni, L., and Mastrorilli, A. (2005). Il Test di Orientamento Motivazionale-Versione Generale (TOM-VG) come strumento per la misura delle inclinazioni motivazionali. *Giornale Italiano Di Psicologia.* 32, 653–650.

Petrides, K. V. (2009). "Psychometric properties of the trait emotional intelligence questionnaire (TEIQue)", in *Assessing Emotional Intelligence: Theory, Research, and Applications* (pp. 85–101), Parker, J. D. A., Saklofske, D. H., and Stough, C. (Eds.). US: Springer. doi: 10.1007/978-0-387-88370-0_5

Pinder, C. (1977). Multiple predictors of post-transfer satisfaction: the role of urban factors. *Pers. Psychol.* 30, 543–556. doi: 10.1111/j.1744-6570.1977.tb02326.x

Punnoose, R., and Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *Int. J. Adv. Res. Artif. Intell.* 5, 22–26. doi: 10.14569/IJARAI.2016.050904

Quinn, A., Rycraft, J. R., and Schoech, D. (2002). Building a model to predict caseworker and supervisor turnover using a neural network and logistic regression. *J. Technol. Hum. Serv.* 19, 65–85. doi: 10.1300/J017v19v04_05

Rain, J. S., Lane, I. M., and Steiner, D. D. (1991). A current look at the job satisfaction/life satisfaction relationship: review and future considerations. *Human Relat.* 44, 287–307. doi: 10.1177/001872679104400305

Rathi, D. R. (2018). Artificial intelligence and the future of hr practices. *Int. J. Appl. Res.* 4, 113–116.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *J Statistical Software*. 48, 1–36. doi: 10.18637/jss.v048.i02

RStudio Inc. (2016). RStudio, integrated development environment for R. *Version:* 1.0.44. Boston, MA: R Studio Inc.

Ryan, A. M., Schmitt, M. J., and Johnson, R. (1996). Attitudes and effectiveness: examining relations at an organizational level. *Pers. Psychol.* 49, 853–882. doi: 10.1111/j.1744-6570.1996.tb02452.x

Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 10, e0118432. doi: 10.1371/journal.pone.0118432

Sawyer, R. (1982). Sample size and the accuracy of predictions made from multiple regression equations. *J Educat. Statist.* 7, 91. doi: 10.3102/10769986007002091

Schneider, B., Hanges, P. J., Smith, D. B., and Salvaggio, A. N. (2003). Which comes first: Employee attitudes or organizational financial and market performance? *J Appl. Psychol.* 88, 836–851. doi: 10.1037/0021-9010.88.5.836

Schwab, D. P., and Cummings, L. L. (1970). Theories of performance and satisfaction: A review. *Indust. Relat.* 9, 408–430. doi: 10.1111/j.1468-232X.1970.tb00524.x

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes and consequences*. Thousand Oaks, CA: SAGE. doi: 10.4135/9781452231549

Staw, B. M., Bell, N. E., and Clausen, J. A. (1986). The dispositional approach to job attitudes: A lifetime longitudinal test. *Adm. Sci. Q.* 56–77. doi: 10.2307/2392766

Vroom, V. H. (1964). *Work and Motivation*. New York, NY: Wiley.

Wright, T. A., and Cropanzano, R. (2000). Psychological well-being and job satisfaction as predictors of job performance. *J. Occup. Health Psychol.* 5, 84–94. doi: 10.1037/1076-8998.5.1.84

Zang, S., and Ye, M. (2015). Human resources management in the Era of big data. *J. Hum. Environ. Stud.* 3, 41–45. doi: 10.4236/jhrss.2015.31006

Conflict of Interest: SD'A was employed by Intesa Sanpaolo Innovation Center S.p.A.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bossi, Di Gruttola, Mastrogiorgio, D'Arcangelo, Lattanzi, Malizia and Ricciardi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/frai.2022.868789

On the Impact of Digitalization and Artificial Intelligence on Employers' Flexibility Requirements in Occupations—Empirical Evidence for Germany

Anja Warning[1]*, *Enzo Weber*[1,2] *and Anouk Püffel*[3]

[1] Department Forecasts and Macroeconomic Analyses, Institute for Employment Research (IAB), Nuremberg, Germany, [2] Institute of Economics and Econometrics, Universität Regensburg, Regensburg, Germany, [3] Universität Regensburg, Regensburg, Germany

Artificial intelligence (AI) has a high application potential in many areas of the economy, and its use is expected to accelerate strongly in the coming years. This is linked with changes in working conditions that may be substantial and entail serious health risks for employees. With our paper we are the first to conduct an empirical analysis of employers' increasing flexibility requirements in the course of advancing digitalization, based on a representative business survey, the IAB Job Vacancy Survey. We combine establishment-level data from the survey and occupation-specific characteristics from other sources and apply non-linear random effects estimations. According to employers' assessments, office and secretarial occupations are undergoing the largest changes in terms of flexibility requirements, followed by other occupations that are highly relevant in the context of AI: occupations in company organization and strategy, vehicle/aerospace/shipbuilding technicians and occupations in insurance and financial services. The increasing requirements we observe most frequently are those concerning demands on employees' self-organization, although short-term working-time flexibility and workplace flexibility also play an important role. The estimation results show that the occupational characteristics, independently of the individual employer, play a major role for increasing flexibility requirements. For example, occupations with a larger share of routine cognitive activities (which in the literature are usually more closely associated with artificial intelligence than others) reveal a significantly higher probability of increasing flexibility demands, specifically with regard to the employees' self-organization. This supports the argument that AI changes above all work content and work processes. For the average age of the workforce and the unemployment rate in an occupation we find significantly negative effects. At the establishment level the share of female employees plays a significant negative role. Our findings provide clear indications for targeted action in labor market and education policy in order to minimize the risks and to strengthen the chances of an increasing application of AI technologies.

Keywords: digitalization, artificial intelligence, flexible working conditions, occupations, employer survey

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Leicester,
United Kingdom

Reviewed by:
Abdul Naser Ibrahim Nour,
An-Najah National University, Palestine
Idiano D'Adamo,
Sapienza University of Rome, Italy

***Correspondence:**
Anja Warning
anja.warning@iab.de

Specialty section:
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

Received: *03 February 2022*
Accepted: *16 March 2022*
Published: *03 May 2022*

Citation:
Warning A, Weber E and Püffel A (2022) On the Impact of Digitalization and Artificial Intelligence on Employers' Flexibility Requirements in Occupations—Empirical Evidence for Germany. Front. Artif. Intell. 5:868789. doi: 10.3389/frai.2022.868789

INTRODUCTION

Increasing digitalization, including the development and use of artificial intelligence (AI), has substantially changed working conditions in establishments and administrations. This is one of the main results obtained in the empirical analyses conducted by Warning and Weber (2018) and Warning et al. (2020) on the basis of data from a representative German employer survey. The analyses show, among other things, that employers with digitalization activities—including the application of artificial intelligence—specify higher flexibility requirements with respect to place of work, working time, and self-organization for their newly hired employees significantly more frequently compared to employers without digitalization activities.

As far as we know, that study was one of the first to deal with changes in qualitative working conditions in the course of digitalization. To date, most analyses from labor market research focus on the quantitative effects, and the debate surrounding whether digitalization and its components creates or suppresses employment remains in the foreground (DeCanio, 2016; Arntz et al., 2017, 2020; Acemoglu and Restrepo, 2020a).

Yet, serious research from both labor and health economics and sociology point to the possible negative effects of precisely that type of qualitative changes reported by Warning and Weber (2018) and Warning et al. (2020). According to that research, changing requirements of employers with regard to working place, working time and work organization are not regarded as positive by all employees, and digitalization causes a significant proportion of individual psychological stress (Diebig et al., 2020; Hartwig et al., 2020). In Germany almost half of all employees (46%) associate digitalization with an increasing workload, while only 9% experience a reduction of their workload (Institut DGB Index Gute Arbeit, 2016).

Health insurance providers, in turn, report an increase in illnesses related to such increasing workloads, deadlines and time pressures, as well as changing working hours, and warn of the negative health effects of digitalization, see for Germany Marschall et al. (2017). The increase in stress-related illnesses is not only associated with lost hours of work and a strain on health and social security funds, employers must also expect significant reductions in the performance of those who continue to work despite illness (Diebig et al., 2020).

Sociological research intensively discusses the possible effects of increasing flexibility in working-time. It can entail considerable negative aspects for workers if they face the challenge of reconciling changing working times with other areas of their life, which is not always possible without conflict and is not always cost neutral (Allen et al., 2000; Ford et al., 2007; Dettmers et al., 2013; Brough et al., 2020). Of course, other individuals benefit from more time flexibility in their jobs in terms of work-life balance, particularly when increasing flexibility goes hand in hand with a high level of individual freedom, rather than increasing control over what employees do minute by minute.

Potential negative effects have been documented in a large number of studies and are likely to be relevant in most areas of digitalization. Not least due to the challenges in the wake of the COVID-19 pandemic, the dynamics of digitalization processes have accelerated enormously and AI is gaining importance in modern economies (Brynjolfsson et al., 2018; Al Momani et al., 2021; Amankwah-Amoah et al., 2021). As is discussed by Warning and Weber (2018), establishments and administrations first develop their internal and external digitalization technologies and networks, whereas artificial intelligence is integrated at a later date, so far in only a minority of establishments. However, its speed of dissemination is strongly increasing and a broader discussion of the effects on employees—besides the question of whether jobs are being created or destroyed—is needed to counteract at an early stage any negative developments that might burden not only individuals, but also businesses and society. In doing so, we consider it highly important to take account of the specificities of occupations, since, as has already been discussed in the literature, the applications of AI may differ considerably between occupations and fields of activity (see section Available Research on AI and the Labor Market), which in turn may have an impact on the respective working conditions.

With our analyses we make a substantive empirical contribution to the discussion surrounding qualitative changes in working conditions in the course of digitalization and the use of AI, with a special focus on the role of occupation-specific characteristics. On the basis of data from a large, representative German employer survey we shed light on employers' changing flexibility demands regarding their employees' place of work, short-term changes in their working time and requirements regarding their self-organization. As far as we know, there is no other representative study available in this context, based on concrete assessments by a large number of employers in all industries and establishment sizes. Germany is a country with a strong digital development and high investments in the development and application of AI (OECD, 2020). Therefore, the results presented here are also highly relevant for other advanced economies and contribute to discussions at the European level dealing with changing working conditions.

Our article is structured as follows: Section Available Research on AI and the Labor Market provides an overview of the research conducted to date on labor market changes related to artificial intelligence, which so far mainly comprises research on potential quantitative effects. Section Method presents the data that we use for our study, explains the transformation of the data into a panel data set and justifies the selection of a non-linear random effects estimator. This part is followed by a description of some of the digital developments in Germany and of the occupations that are relevant in the context of AI applications in section Some Descriptive Results. Section Estimation Results discusses the results of the random effects estimations and the factors that emerge as relevant for employers' increasing requirements regarding their employees' flexibility in terms of their place of work, their working time and their self-organization. We summarize our results in section Discussion and Outlook and provide an outlook for future empirical research on the qualitative labor market effects of AI.

AVAILABLE RESEARCH ON AI AND THE LABOR MARKET

As is the case for digitalization in general, there is no unique definition of AI that expresses the diversity and breadth of both the technology and its potential applications, although we do not yet know all of the potential AI applications. Therefore, labor market researchers currently address above all the possible labor market effects of AI, while the actual labor market effects remain largely unknown, with little empirical work conducted on the topic so far.

Current research deals partly with conceptual boundaries and the ways that AI can be operationalized for empirical research (Ernst et al., 2019; Acemoglu and Restrepo, 2020b; Tolan et al., 2021). Building on or parallel to this, empirical work has also been conducted on the quantitative effects of AI on employment, wages, hires, and fluctuation (Felten et al., 2019; Webb, 2020; Georgieff and Hyee, 2021; Fossen and Sorgner, 2022). These quantitative studies have to make assumptions about how certain capabilities and tasks are changed by the application of AI technologies, which have to be defined initially, for example on the basis of interviews with experts from the AI field. The aim is to assess how the characteristics of occupations change with regard to the tasks to be performed and the skills required and to estimate the quantitative effects resulting from these changes. Research on changing tasks and the shifting importance of specific task types (types of manual and cognitive tasks) is usually a crucial element of these approaches.

For instance, in German labor market research, occupations are distinguished according to five task types (Spitz-Oener, 2006), see **Table 1** for a description and examples. Using this concept Genz et al. (2021) discuss the idea of different stages of digital development that include AI in the youngest stage. They find that establishments that are active in this youngest stage ("4.0 adopters") have a comparatively larger share of employees performing routine cognitive tasks in their job activities (36%), followed by non-routine analytical tasks and non-routine manual tasks. The degree of complexity involved in the job increases with ongoing digitalization, as does demand for IT staff (AI specialists, IT security consultants, cloud engineers) and staff in business services.

From the available studies, it can be deduced that AI is mainly used in occupational fields involving a high proportion of cognitive and analytical tasks. In these fields, based on a large amount of data, AI can strengthen the basis for decision-making by making it possible to systematically monitor and evaluate processes, thereby supporting people in their decision-making. In some areas AI can also take over the control of processes entirely. On the other hand, AI is used less in areas in which people interact strongly, as not all elements of human behavior can be replaced by technological systems.

The OECD recently published an article reviewing what is known about the labor market effects of AI, showing the potential of AI on the one hand and our very limited knowledge about the real labor market effects on the other hand (Lane and Saint-Martin, 2021). This applies in particular to knowledge about changing working conditions and employers' changing demands regarding flexibility, what might be even more important than in previous stages of digitalization. The authors provide an example of this for the case of AI-supported robots: Such robots might take over activities that are dangerous or physically very strenuous for humans, which has clear positive effects on the tasks to be performed, as they become less dangerous and less strenuous. However, if the humans have to adapt their work intensity and rhythm to the robot in a close human-machine-interaction, the work pressure might simultaneously increase and the freedom of action may decrease, leading to increasing stress and growing dissatisfaction, in turn causing (new) psychological stress for the employee. Another open issue in the context of AI is the availability of big data, which enables employers to closely monitor employees' activities and to steer these activities automatically in the short term. This not only raises questions concerning data protection and personal rights, but in practice pressurizes employees to respond at short notice to adaptations intended by the AI system and to avoid any mistakes and misconduct while carrying out work.

METHOD

Establishment Data From the IAB Job Vacancy Survey

In the study presented here, we examined the role of occupation- and establishment-specific characteristics for increasing flexibility requirements expressed by employers.

We took up some of the findings obtained by Warning and Weber (2018) on significant changes in working conditions and again use the IAB Job Vacancy Survey (JVS) for our new approach. The JVS is a representative employer survey conducted at regular intervals among employers in Germany. Its overall aim is to determine the current demand for labor and to observe staff-search and hiring processes in detail (Davis et al., 2014; Bossler et al., 2020). Every year some 12,000 establishments and administrations of all sizes and from all sectors of the economy complete the written questionnaire in the fourth quarter of the year. (According to the sampling method, the term "establishment" always refers to establishments and public administrations with at least one employee covered by social security contributions.) The information they report on vacancies, employment, and the development of search and hiring processes are extrapolated to all establishments and all new hires in Germany, thereby providing a unique, representative picture of the current labor market development in Germany (on the extrapolation, see Brenzel et al., 2016). The JVS is quality assured in accordance with the regulations laid down by the European Commission concerning the collection, measurement and calculation of job vacancy and employment data that are gathered in this survey and are officially published by Eurostat in the context of labor demand data for the European countries (Eurostat, n. d).

In 2016 we integrated new detailed questions into a special questionnaire section of the JVS. It focused on changing flexibility requirements in occupations by those employers who expect increasing digitalization in the subsequent 5 years, see

TABLE 1 | Task types of occupations and examples.

Task-type	Description	Occupations with highest shares in the task-type
Non-routine analytical activities	Doing research, analyze, evaluate, plan, construct, design, develop rules/regulations, apply and interpret rules	Members of Parliament, Ministers, Architects, Civil Engineers, Veterinarians, Publicists
Interactive non-routine activities	Negotiate, represent interests, coordinate, organize, teach or train, sell, buy, advertise, entertain, present, employ or manage clients	Interpreters, translators, sales representatives, employment and professional advisers, consumer advisers
Cognitive routine activities	Calculate, make bookkeeping, correct text/data, measure length/height/temperature	Chemical laboratory technician, radio operator, data typist, telecommunication assembler
Non-routine manual activities	Repair or renovation of houses/flats/machinery/vehicles, restoration of art/monuments, service or accommodation of guests	Paving, earthmoving machine drivers, machine cleaners, railway drivers
Routine manual activities	Operating or controlling machines, equipping machines	Rubber converters, metal pullers, leather manufacturers, sheet metal presses

Sources: *Spitz-Oener (2006) p. 243, Dengler et al. (2014) p. 38.*

Figure 1. In the first question (question 36 in the JVS) the participating establishments, or their managers or personnel managers, are asked whether their particular establishment is expecting an increase in digital development over the following 5 years. As in the previous analysis of Warning and Weber, digital development is defined as internal digital networking, networking with customers/suppliers and the use of learning systems. Learning systems as part of artificial intelligence systems are thus included in our study.

All establishments that answer the first question with YES (a total of 4,262 establishments) are then asked to report the occupations for which they expect particularly strong changes in employees' qualitative working conditions as a result of increasing digitalization. The questionnaire gives the possibility to state a maximum of three occupations. The changes in working conditions refer to flexibility in terms of workplace, flexibility regarding working time on short notice and demands regarding employees' self-organization. The wording in the special questionnaire section deliberately refers only to (great and small) increasing or unchanging flexibility requirements, because our research focuses only on increases, not decreases.

Restricting the number of occupations that establishments could mention here to a maximum of three was a compromise: On the one hand, we wanted to investigate positive changes in flexibility requirements by individual occupations. On the other hand, an already extensive written survey like the JSV cannot be extended by too many additional questions, as this may lead to a drop in establishments' willingness to participate, thereby endangering the success of the entire survey. However, the restriction to three occupations proved in retrospect to be very meaningful and does not lead to a distortion of the results: The vast majority of those establishments expecting an increase in digitalization provided detailed information on flexibility requirements for one or two occupations. Only rarely did an establishment report three occupations in the questionnaire. Therefore the answers reflect employers' assessments of the occupations that they consider to be most strongly affected, this has to be taken into account when interpreting the survey results.

For the subsequent estimations we calculated three new binary variables from the JVS data. They are independent of each other and are the dependent variables in our models:

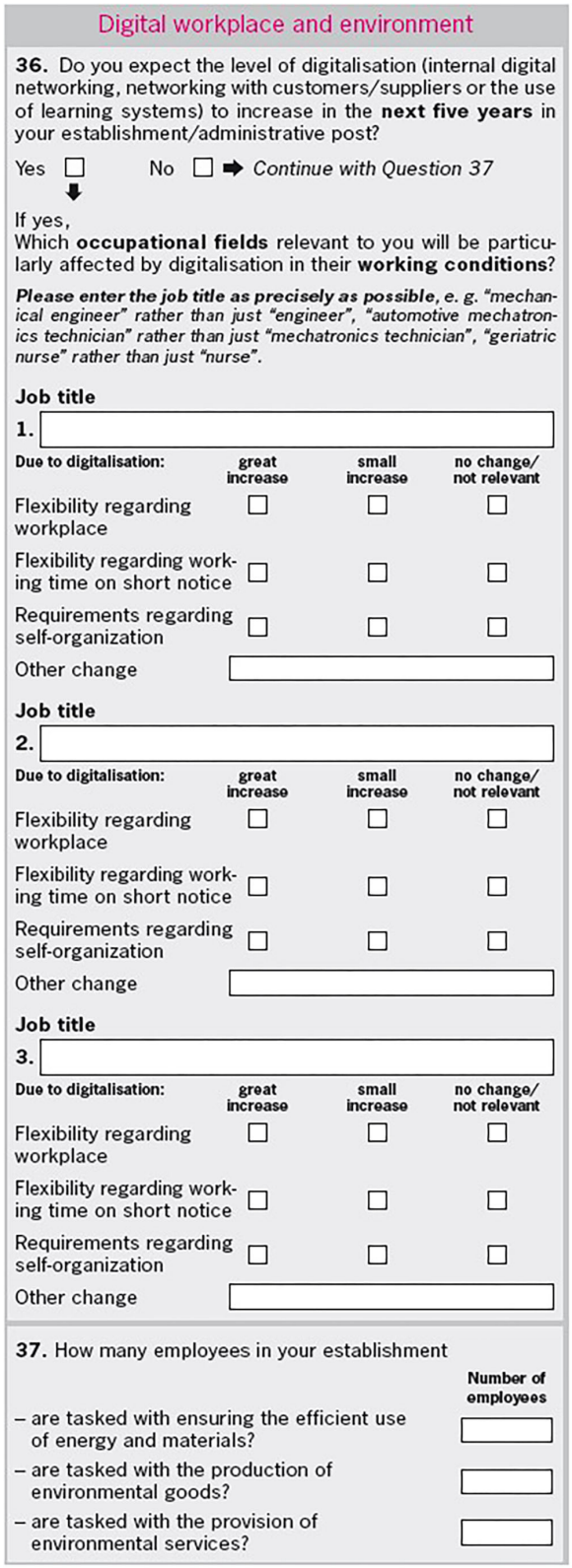
1) increasing requirements regarding flexibility in terms of place of work,
2) increasing requirements regarding short-term flexibility in working time and
3) increasing demands regarding self-organization.

Each binary variable took the value 1 if the establishment reported a small or large increase in the flexibility required in the specific occupation. It took the value 0 if the establishment indicated no change or no relevance of this requirement.

In addition to the data on changing requirements by occupation we utilized standard establishment-specific structural data from the JVS. They describe the establishment's individual employment and labor demand situation that might affect the employer's individual decisions regarding the flexibility required of their employees. Specifically, we used information on region and workforce size, the share of academics, the share of employees with vocational qualifications and the share of women. We included data on the establishment's overall labor demand, such as the expected employment development, the number of new hires, job vacancies as a proportion of employment and the fluctuation in the particular economic sector. We also included data on the existence of a works council and collective agreements, as this might hinder or delay the implementation of new technologies and the associated changes in working conditions (Warning and Weber, 2018). **Table 2** provides a descriptive overview of all establishment-specific variables used in our models.

Data on Occupation-Specific Characteristics

In order to be able to depict occupation-specific characteristics in the best possible way, we added various occupation-specific variables that are independent of the individual establishments.



Digital workplace and environment

36. Do you expect the level of digitalisation (internal digital networking, networking with customers/suppliers or the use of learning systems) to increase in the **next five years** in your establishment/administrative post?

Yes ☐ No ☐ ➡ *Continue with Question 37*

If yes,
Which **occupational fields** relevant to you will be particularly affected by digitalisation in their **working conditions**?

Please enter the job title as precisely as possible, e. g. "mechanical engineer" rather than just "engineer", "automotive mechatronics technician" rather than just "mechatronics technician", "geriatric nurse" rather than just "nurse".

Job title
1. ☐

Due to digitalisation:	**great increase**	**small increase**	**no change/ not relevant**
Flexibility regarding workplace	☐	☐	☐
Flexibility regarding working time on short notice	☐	☐	☐
Requirements regarding self-organization	☐	☐	☐
Other change	☐		

Job title
2. ☐

Due to digitalisation:	**great increase**	**small increase**	**no change/ not relevant**
Flexibility regarding workplace	☐	☐	☐
Flexibility regarding working time on short notice	☐	☐	☐
Requirements regarding self-organization	☐	☐	☐
Other change	☐		

Job title
3. ☐

Due to digitalisation:	**great increase**	**small increase**	**no change/ not relevant**
Flexibility regarding workplace	☐	☐	☐
Flexibility regarding working time on short notice	☐	☐	☐
Requirements regarding self-organization	☐	☐	☐
Other change	☐		

37. How many employees in your establishment

	Number of employees
– are tasked with ensuring the efficient use of energy and materials?	☐
– are tasked with the production of environmental goods?	☐
– are tasked with the provision of environmental services?	☐

FIGURE 1 | IAB Job Vacancy Survey 2016, written questionnaire, p. 5.

TABLE 2 | Descriptives of the variables used in the estimation models.

Variables	**Mean**	**Std. dev.**	**Min**	**Max**
Dependent variables				
Work place flexibility	0.0428	0.2023	0	1
Short-term flexibility in working time	0.0632	0.2433	0	1
Demands on self-organization	0.0775	0.2674	0	1
Independent variables				
Occupation-specific:				
Share of interactive non-routine activities	10.8945	12.0753	0.214	39.199
Share of non-routine analytical activities	19.6029	12.2761	4.098	51.101
Share of non-routine manual activities	24.6106	20.4172	0.619	57.080
Share of routine cognitive activities	28.7483	16.5478	8.978	59.502
Average age of employees	41.1220	1.6945	38.613	45.523
Employment growth rate 2013–2016	8.2091	3.8719	1.642	15.325
Fluctuation rate in 2016	2.0818	1.1458	0.403	3.927
Unemployment rate in 2016	7.2371	3.8274	2.447	15.985
Establishment-specific:				
Region	0.5622	0.4961	0	1
Size class	1.9315	0.5377	1	3
Labor-turnover rate by sector	65.3976	39.1471	27.3	152.1
Expected employment trend	1.7320	0.6170	1	3
New employees hired in the previous year	0.7842	0.4114	0	1
Vacancies as a proportion of total employment	5.2215	14.1550	0	200
Collective agreement in place	0.4940	0.5000	0	1
Existence of works council	0.3071	0.4613	0	1
Share of skilled workers	65.0450	30.0866	0	100
Share of academics	17.9145	24.8894	0	100
Share of women	41.7667	27.3834	0	100

Source: IAB Job Vacancy Survey 2016, Data Warehouse of the Federal Employment Agency 2019, own calculations.

First, we used information on the shares of five task types in each occupational group (Spitz-Oener, 2006). Data for the year 2016 come from IAB task research, providing the shares of non-routine analytical, non-routine interactive, routine cognitive, non-routine manual, and routine manual activities in each occupation (Dengler et al., 2014). **Table 1** provides a description of these types, as well as examples of occupations that have a relatively large share of the respective task type.

Second, we used structural information from the Federal Employment Agency related to the occupational group: the average age of the workforce, the employment growth rate between 2013 and 2016, the labor turnover rate in 2016 and the unemployment rate in 2016. These data allow us to describe

TABLE 3 | Sectors of the economy with the respective shares of companies that expect increasing digitalization over the next 5 years.

Financial services, Insurance	63%
Liberal professions, scientific and technical services	50%
Machines, electronics, vehicles	41%
Information and communication	41%
Public administration	39%
Health and social services	36%
Education, child care	34%
Trade, retail, repairs	33%
Other services	31%
Chemistry, plastics, glass, construction materials	31%
Energy utilities	30%
Transport, warehouses	30%
Metals, metal production	29%
Nutrition, textiles, clothing, furniture, etc.	27%
Water, waste management	26%
Real estate	26%
Other commercial services	25%
Agriculture, forestry	24%
Wood, paper, printing	24%
Construction	18%
Hospitality	18%
Art, entertainment, recreation	15%
Mining, ores and earths	13%

Source: IAB Job Vacancy Survey 2016, own calculations, weighted results.

general differences between the occupational groups as precisely as possible, thereby minimizing the risk of omitted variables in our estimation models. **Table 2** contains a descriptive overview of the occupation-specific variables.

Creation of a Panel Dataset for Random Effects Estimations

The reported occupations were originally coded according to the German Classification of Occupations 2010 at the 4-digit level (Statistical Offices of the Federation and the Länder, n. d). To ensure that the number of cases per occupational unit was sufficiently high for the analyses, we aggregated the original data at the level of 14 occupational groups and finally obtained a dataset containing information on changing requirements in 14 occupational groups from about 4,200 establishments.

In order to take into account heterogeneity effects and to analyze increasing flexibility requirements in the context of occupations, we transformed this original cross-sectional dataset into a panel data structure. This allows the use of a panel data model, we specifically chose the non-linear random effects model (Cameron and Trivedi, 2010; Wooldridge, 2010). A fixed effects model would not yield estimates for the occupation-specific variables which are the focus of our interest (see next paragraph on these variables). Besides that argument, fixed effects models do not function in the specific case of our data structure. This is characterized by the peculiarity that the three binary dependent variables have a relatively high number of zeros and a relatively low number of ones, meaning that there is relatively little variation in the dependent variables by 14 occupational groups and about 4,200 establishments. As a result, the estimation coefficients (see section Estimation Results) are small, but as is shown with the parameter rho in the estimations in **Tables 5**–**7**, a standard pooled estimation would lead to inconsistent parameter estimates and a panel data estimation is the preferred approach here.

SOME DESCRIPTIVE RESULTS

Digital Development in German Establishments

The following results are weighted with the standard weighting factors calculated for the data of the IAB Job Vacancy Survey. The figures in **Tables 3**, **4** thus represent the total numbers of the respective establishments in the economy.

A total of 4,262 establishments in the survey expected increasing digitalization in the following 5 years. Altogether, they represent 700,000 establishments in the German economy, which is equivalent to a share of about 32%. The highest shares by economic sector are found in financial and 256 insurance services, at 63%, followed by liberal professions, scientific and technical services at 257 50%, see **Table 3**. The sectors with the lowest shares of establishments expecting an increase in digitalization include for instance art, entertainment and recreation, and hospitality.

Establishments with more than 250 employees are more likely to expect increasing digitalization than medium-sized and small ones. On the whole our results are similar to those obtained in other studies on the spread of digitalization in Germany (Reimann et al., 2020).

Occupations and Increasing Flexibility Requirements

Table 4 shows a list of the most frequently mentioned occupations and the number of establishments with positive digitalization expectations and positive expectations regarding increasing flexibility requirements in these occupations. Office and secretarial occupations were mentioned most frequently, by about 58,000 establishments and administrations, followed by three occupations that are highly relevant in the context of artificial intelligence: occupations in company organization and strategy (34,000), vehicle/aerospace/shipbuilding technicians (32,000) and occupations in insurance and financial services (32,000).

The table reveals the high relevance of changes in employees' self-organization during the course of digitalization: In all the occupations listed there, this kind of flexibility requirement was mentioned most often by the employers, followed by increasing temporal flexibility and increasing workplace flexibility. As we know, digitalization and in specific the introduction of artificial intelligence systems are closely linked to changes in working structures (Quelle). Our results on the special relevance

TABLE 4 | Number of establishments with positive expectations of increasing flexibility requirements in the respective occupation.

Occupational field	Number of establishments expecting a change in working conditions in the occupational field	Number of establishments in which changing working conditions are accompanied by increasing demands of the following types:		
		Increasing workplace flexibility	Increasing temporal flexibility	Increasing self-organization
Office and secretarial	57,738	19,401	41,721	53,081
Company organization and strategy	34,467	21,207	27,654	32,328
Vehicle manufacture, aerospace, shipbuilding technicians	32,220	14,534	15,420	22,641
Insurance and financial services	31,589	19,419	24,773	28,445
Tax consultancy	27,475	15,101	15,791	24,794
Purchasing and sales	27,232	19,839	23,203	24,051
Construction planning and supervision, architecture	23,965	10,245	15,676	19,753
Accounting, controlling and auditing	18,629	11,289	14,720	16,843
Public administration	17,499	8,745	11,276	15,247
Mechanical engineering and operating technology	14,734	8,167	12,662	13,129

Source: IAB Job Vacancy Survey 2016, own calculations, weighted results.

of increasing demands regarding self-organization underline this statement.

ESTIMATION RESULTS

Occupational Characteristics

Tables 5–**10** show the coefficients and marginal effects calculated from our three random effects estimations. In the following we use the marginal effects as the basis for the discussion of our findings, see **Table 11** for a comparison between the models. The effects are small in quantitative terms, which is due to the characteristics of the data structure (see section Method). Nevertheless, the effects are highly meaningful, as is confirmed by both the error probabilities and the quality criteria of our estimations.

For all three kinds of flexibility requirements the share of routine cognitive activities is highly significant, with the highest value for increasing demands regarding self-organization. A one-percent increase in the share of routine cognitive activities raises the probability of increasing demands on self-organization by 0.16% points, the probability of increasing short-term working-time flexibility by 0.14% points and of increasing workplace flexibility by about 0.09% points. According to the literature occupations affected strongly by AI applications are often defined by relatively high shares of routine cognitive tasks or non-routine analytical tasks (Genz et al., 2021; Lane and Saint-Martin, 2021). Looking at the shares of routine cognitive activities in the occupational groups in **Table 12**, our estimates suggest this discussion with regard to occupations with a high share of routine cognitive activities: For instance, in business services and in business management and organization more than half of all tasks are routine cognitive tasks (59 and 56%, respectively). Here increasing digitalization, including the increasing use of AI, is more likely to be associated with employers demanding more flexibility, in particular with regard to self-organization and short-term flexibility in working time.

As the marginal effects show, the share of non-routine analytical tasks is negatively significant regarding increasing short-term flexibility in working time, it is not relevant regarding the other two types of flexibility. Looking at the examples of occupations with large shares of such non-routine analytical tasks in **Table 12**, this result is not surprising in the AI context. If AI is usable at all, it is used more as a supplementary technology. Human beings still have to make decisions and need to understand the AI technology and its applications. Specifically, the work involved in developing and implementing new AI technologies in the establishments may initially be very time-consuming and require a lot of attention from the people involved. It is necessary to understand in detail the interplay between technologies and humans, for which increasing requirements on short-term flexibility in working time, which workers often associate with increasing time pressure, is not a good basis.

Non-routine manual activities show no significant effects on the probability of increasing flexibility requirements. In the context of AI, as a special form of digital development, this result substantiates the discussions about the potential relevance of AI for certain occupations, but not for others.

In all three models, the average age of the employees in the occupational group is negatively and highly significantly related

TABLE 5 | Estimation results: increasing requirements regarding workplace flexibility.

	Coefficient		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.01017		0.00856	−0.00662	0.02696
Share of non-routine analytical activities	−0.01822		0.01046	−0.03871	0.00227
Share of non-routine manual activities	0.00886		0.00779	−0.02413	0.00641
Share of routine cognitive activities	0.03691	***	0.00687	0.02344	0.05037
Average age of employees	0.11902	**	0.05593	−0.22864	−0.00940
Employment growth rate 2013–2016	0.16675	***	0.04629	−0.25747	−0.07603
Fluctuation rate in 2016	0.49464	***	0.16873	0.16394	0.82533
Unemployment rate in 2016	0.07555	**	0.03171	−0.13771	−0.01340
Establishment-specific:					
Region (east)	0.08300		0.05018	−0.01536	0.18136
Establishment size class (<10)					
10–249	0.14026		0.07369	−0.28469	0.00418
>250	0.01945		0.11307	−0.20217	0.24106
Labor-turnover rate by sector	0.00014		0.00066	−0.00115	0.00143
Expected employment trend (constant)					
Increasing	0.08414		0.05562	−0.02488	0.19315
Decreasing	0.25326	***	0.08202	0.09251	0.41402
New employees hired in the previous year	0.10874		0.06711	−0.02280	0.24028
Vacancies as a proportion of total employment	0.00281		0.00164	−0.00041	0.00603
Collective agreement in place	0.06956		0.05580	−0.17893	0.03980
Existence of works council	0.00414		0.06666	−0.12652	0.13480
Share of skilled workers	0.00181		0.00112	−0.00038	0.00401
Share of academics	0.00372	***	0.00129	0.00119	0.00625
Share of women	0.00416	***	0.00093	−0.00598	−0.00234
Constant	1.48933		2.22000	−2.86178	5.84044
Rho	0.01507	***	0.00862	0.00488	0.04558

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

TABLE 6 | Estimation results: increasing requirements regarding short-term flexibility in working time.

	Coefficient		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.00593		0.00794	−0.00963	0.02149
Share of non-routine analytical activities	0.02591	***	0.00937	−0.04428	−0.00755
Share of non-routine manual activities	0.01332		0.00716	−0.02735	0.00071
Share of routine cognitive activities	0.03651	***	0.00628	0.02421	0.04881
Average age of employees	0.15857	***	0.05051	−0.25757	−0.05956
Employment growth rate 2013–2016	0.15568	***	0.04296	−0.23988	−0.07149
Fluctuation rate in 2016	0.47546	***	0.15443	0.17278	0.77814
Unemployment rate in 2016	0.07443	***	0.02871	−0.13070	−0.01817
Establishment-specific:					
Region (east)	0.04496		0.04179	−0.03696	0.12688
Establishment size class (<10)					
10–249	0.04996		0.06386	−0.07521	0.17512
>250	0.17310		0.09624	−0.01553	0.36173
Labor-turnover rate by sector	0.00060		0.00056	−0.00170	0.00049
Expected employment trend (constant)					
Increasing	0.11208	**	0.04614	0.02164	0.20251
Decreasing	0.14334	*	0.07034	0.00547	0.28120
New employees hired in the previous year	0.06931		0.05608	−0.04060	0.17923
Vacancies as a proportion of total employment	−0.00101		0.00163	−0.00420	0.00217
Collective agreement in place	−0.07340		0.04648	−0.16450	0.01769
Existence of works council	0.03594		0.05527	−0.14427	0.07240
Share of skilled workers	0.00070		0.00091	−0.00107	0.00248
Share of academics	0.00077		0.00108	−0.00134	0.00289
Share of women	0.00304	***	0.00077	−0.00455	−0.00153
Constant	3.86104		2.01311	−0.08459	7.80667
Rho	0.01333	***	0.00686	0.00483	0.03619

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, * $p<0.1$; ** $p<0.05$; *** $p<0.01$.*

to increasing requirements, with the highest value regarding the demands for self-organization. This result is expectable and reflects the relatively high level of regulation of the German labor market, which protects older employees in many ways. The question also arises of whether older employees who are unwilling or unable to adapt to their employers' changing flexibility requirements are more likely to take up occupations with a lower (or slower) level of digital development or whether they are more frequently forced by their employers to change to other occupational fields or even to change the employer.

The occupation-related employment growth rate between 2013 and 2016, the period directly before the field period of the survey, shows a negative and highly significant value in all three models. An increase in the employment growth rate by 1% reduces the probability of increasing demands on self-organization by 0.7% points. Negative effects are also

TABLE 7 | Estimation results: increasing requirements regarding self-organization.

	Coefficient		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.01087		0.00968	−0.00811	0.02985
Share of non-routine analytical activities	0.02112		0.01106	−0.04279	0.00055
Share of non-routine manual activities	−0.00865		0.00861	−0.02553	0.00823
Share of routine cognitive activities	0.03422	***	0.00757	0.01938	0.04906
Average age of employees	0.14493	**	0.05673	−0.25613	−0.03374
Employment growth rate 2013–2016	0.15120	***	0.05115	−0.25145	−0.05096
Fluctuation rate in 2016	0.41074	**	0.18199	0.05405	0.76742
Unemployment rate in 2016	0.08282	**	0.03416	−0.14978	−0.01587
Establishment-specific:					
Region (east)	0.03579		0.03818	−0.03905	0.11062
Establishment size class (<10)					
10–249	0.09868		0.05886	−0.01668	0.21405
>250	0.16452		0.08867	−0.00927	0.33832
Labor-turnover rate by sector	0.00145	**	0.00052	−0.00247	−0.00044
Expected employment trend (constant)					
Increasing	0.05175		0.04221	−0.03097	0.13447
Decreasing	0.02201		0.06596	−0.10728	0.15129
New employees hired in the previous year	0.09557		0.05141	−0.00520	0.19634
Vacancies as a proportion of total employment	0.00064		0.00149	−0.00356	0.00228
Collective agreement in place	0.04365		0.04249	−0.12693	0.03964
Existence of works council	0.01638		0.05011	−0.11460	0.08183
Share of skilled workers	0.00089		0.00083	−0.00074	0.00252
Share of academics	0.00062		0.00099	−0.00132	0.00256
Share of women	0.00155	**	0.00070	−0.00292	−0.00019
Constant	3.44720		2.25166	−0.96598	7.86038
Rho	0.02101	***	0.00923	0.00883	0.04918

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

TABLE 8 | Marginal effects: increasing requirements regarding workplace flexibility.

	Marginal effect		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.00026		0.00022	−0.00017	0.00069
Share of non-routine analytical activities	−0.00047		0.00027	−0.00099	0.00006
Share of non-routine manual activities	−0.00023		0.00020	−0.00062	0.00016
Share of routine cognitive activities	0.00095	***	0.00018	0.00059	0.00130
Average age of employees	−0.00305	**	0.00142	−0.00583	−0.00027
Employment growth rate 2013–2016	−0.00427	***	0.00118	−0.00659	−0.00195
Fluctuation rate in 2016	0.01267	***	0.00430	0.00424	0.02109
Unemployment rate in 2016	−0.00193	**	0.00082	−0.00354	−0.00033
Establishment-specific:					
Region (east)	0.00213		0.00129	−0.00041	0.00466
Establishment size class (<10)					
10–249	−0.00370		0.00205	−0.00771	0.00031
>250	0.00055		0.00322	−0.00575	0.00686
Labor-turnover rate by sector	0.00000		0.00002	−0.00003	0.00004
Expected employment trend (constant)					
Increasing	0.00213		0.00143	−0.00067	0.00492
Decreasing	0.00696	***	0.00249	0.00208	0.01183
New employees hired in the previous year	0.00278		0.00173	−0.00060	0.00617
Vacancies as a proportion of total employment	0.00007		0.00004	−0.00001	0.00015
Collective agreement in place	−0.00178		0.00143	−0.00459	0.00103
Existence of works council	0.00011		0.00171	−0.00324	0.00345
Share of skilled workers	0.00005		0.00003	−0.00001	0.00010
Share of academics	0.00010	***	0.00003	0.00003	0.00016
Share of women	−0.00011	***	0.00002	−0.00015	−0.00006

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, margins at means, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

estimated for the unemployment rate. The fields of the labor market in which digital developments are particularly dynamic and where working conditions may change as a result are more likely to be those in which employers complain of worker and skills shortages. The unemployment rate is correspondingly low and workers' demands for a good work-life balance are likely to be correspondingly high. This is likely to limit employers' possibilities to further increase their flexibility requirements and may even force them to reduce their demands.

The fluctuation rate, i.e., the dynamics of entry and exit from employment in the respective occupational group, exhibits a significant positive effect in all models. High fluctuation means that a relatively large proportion of new employees are recruited relative to the existing workforce. Whereas in the case of the existing workforce, employers are dependent on employees' willingness to change and are not always able to implement changes with the scope and speed desired, in the case of new hires the employers can formulate the precise requirements and conditions that they consider to be in line with the new challenges and opportunities of digitalization. Effects on working conditions

TABLE 9 | Marginal effects: increasing requirements regarding short term flexibility in working time.

	Marginal effect		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.00023		0.00030	−0.00037	0.00082
Share of non-routine analytical activities	−0.00099	***	0.00036	−0.00169	−0.00029
Share of non-routine manual activities	−0.00051		0.00027	−0.00105	0.00003
Share of routine cognitive activities	0.00140	***	0.00025	0.00091	0.00189
Average age of employees	−0.00607	***	0.00191	−0.00981	−0.00232
Employment growth rate 2013–2016	−0.00596	***	0.00164	−0.00917	−0.00274
Fluctuation rate in 2016	0.01819	***	0.00588	0.00666	0.02971
Unemployment rate in 2016	−0.00285	***	0.00110	−0.00501	−0.00068
Establishment-specific:					
Region (east)	0.00172		0.00160	−0.00142	0.00486
Establishment size class (<10)					
10–249	0.00186		0.00234	−0.00273	0.00645
>250	0.00682		0.00386	−0.00075	0.01439
Labor-turnover rate by sector	−0.00002		0.00002	−0.00007	0.00002
Expected employment trend (constant)					
Increasing	0.00430	**	0.00180	0.00076	0.00783
Decreasing	0.00558		0.00288	−0.00007	0.01122
New employees hired in the previous year	0.00265		0.00215	−0.00156	0.00687
Vacancies as a proportion of total employment	−0.00004		0.00006	−0.00016	0.00008
Collective agreement in place	−0.00281		0.00178	−0.00631	0.00069
Existence of works council	−0.00137		0.00212	−0.00552	0.00277
Share of skilled workers	0.00003		0.00003	−0.00004	0.00009
Share of academics	0.00003		0.00004	−0.00005	0.00011
Share of women	−0.00012	***	0.00003	−0.00018	−0.00006

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, margins at means, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

TABLE 10 | Marginal effects: increasing requirements regarding self-organization.

	Marginal effect		Std. err.	95% Confidence interval	
Occupation-specific:					
Share of interactive non-routine activities	0.00052		0.00046	−0.00039	0.00143
Share of non-routine analytical activities	−0.00101		0.00053	−0.00204	0.00003
Share of non-routine manual activities	−0.00041		0.00041	−0.00122	0.00039
Share of routine cognitive activities	0.00163	***	0.00038	0.00090	0.00237
Average age of employees	−0.00691	**	0.00271	−0.01222	−0.00161
Employment growth rate 2013–2016	−0.00721	***	0.00246	−0.01204	−0.00239
Fluctuation rate in 2016	0.01960	**	0.00870	0.00254	0.03665
Unemployment rate in 2016	−0.00395	**	0.00165	−0.00718	−0.00073
Establishment-specific:					
Region (east)	0.00171		0.00182	−0.00187	0.00528
Establishment size class (<10)					
10–249	0.00454		0.00264	−0.00064	0.00972
>250	0.00780		0.00429	−0.00061	0.01621
Labor-turnover rate by sector	−0.00007	**	0.00003	−0.00012	−0.00002
Expected employment trend (constant)					
Increasing	0.00248		0.00204	−0.00152	0.00648
Decreasing	0.00104		0.00314	−0.00512	0.00720
New employees hired in the previous year	0.00456		0.00247	−0.00028	0.00940
Vacancies as a proportion of total employment	−0.00003		0.00007	−0.00017	0.00011
Collective agreement in place	−0.00208		0.00203	−0.00606	0.00190
Existence of works council	−0.00078		0.00239	−0.00547	0.00391
Share of skilled workers	0.00004		0.00004	−0.00004	0.00012
Share of academics	0.00003		0.00005	−0.00006	0.00012
Share of women	−0.00007	**	0.00003	−0.00014	−0.00001

*Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, margins at means, * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

and flexibility requirements will therefore be more visible in the more dynamic occupational fields.

Establishment Characteristics

In contrast to the occupational effects, the characteristics of the individual establishments play a minor role in explaining increasing flexibility requirements. The size of an establishment and the region in which it is located are not explanatory. Those operating in an industry with a high labor-turnover rate, and thus having to recruit and train new staff more often, are more likely to define increasing demands on employees' self-organization. This is not the case for the other two types of flexibility.

Positive employment expectations increase the probability of increasing demands for short-term flexible working hours. This is not true of the number of new hires in the previous year or of current vacancies as a proportion of the total workforce, (It should be taken into account that all the establishments in our estimates assume increasing digitalization over the next 5 years, see section Establishment Data From the IAB Job Vacancy Survey of this article).

The skill structure in the establishment shows no significance, except for the proportion of academics in model 1. Differences

TABLE 11 | Comparison of the marginal effects of the three estimations.

	Work place flexibility		Short term flexibility in working time		Demands on self-organization	
Occupation-specific:						
Share of interactive non-routine activities	0.00026		0.00023		0.00052	
Share of non-routine analytical activities	−0.00047		−0.00099	***	−0.00101	
Share of non-routine manual activities	−0.00023		−0.00051		−0.00041	
Share of routine cognitive activities	0.00095	***	0.00140	***	0.00163	***
Average age of employees	−0.00305	**	−0.00607	***	−0.00691	**
Employment growth rate 2013–2016	−0.00427	***	−0.00596	***	−0.00721	***
Fluctuation rate in 2016	0.01267	***	0.01819	***	0.01960	**
Unemployment rate in 2016	−0.00193	**	−0.00285	***	−0.00395	**
Establishment-specific:						
Region (east)	0.00213		0.00172		0.00171	
Establishment size class (<10)						
10–249	−0.00370		0.00186		0.00454	
>250	0.00055		0.00682		0.00780	
Labor-turnover rate by sector	0.00000		−0.00002		−0.00007	**
Expected employment trend (constant)						
Increasing	0.00213		0.00430	**	0.00248	
Decreasing	0.00696	***	0.00558		0.00104	
New employees hired in the previous year	0.00278		0.00265		0.00456	
Vacancies as a proportion of total employment	0.00007		−0.00004		−0.00003	
Collective agreement in place	−0.00178		−0.00281		−0.00208	
Existence of works council	0.00011		−0.00137		−0.00078	
Share of skilled workers	0.00005		0.00003		0.00004	
Share of academics	0.00010	****	0.00003		0.00003	
Share of women	−0.00011	***	−0.00012	***	−0.00007	**

Source: IAB Job Vacancy Survey 2016, own calculations with a random effects estimation, margins at means, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

in skill levels are at least partly captured by the differences in the occupations. In our analyses differences at the occupational level are more relevant than differences at the establishment level.

The proportion of women in the workforce exhibits a significant negative marginal effect in all models. For instance, a one-percent increase in the share of female employees reduces the probability of increasing demands on short term flexibility in working time by 0.012% points. The possibilities for negotiation

TABLE 12 | Shares of task by occupational group 2016, as percentages.

	Occupations in agriculture, forestry and horticulture	Manufacturing occupations	Occupations in manufacturing technology	Occupations in construction and building completion	Occupations in food and hospitality sector	Medical and non-medical health occupations	Social and cultural service occupations	Commercial occupations	Occupations in business management and organization	Occupations in business services	Occupations in IT and scientific services	Security occupations	Occupations in transport and logistics	Cleaning occupations
Non-routine analytical tasks	15.4	7.4	18.1	14.6	8.7	19.5	36.2	12.6	31.4	22.2	51.1	22.6	10.4	4.1
Non-routine interactive tasks	2.6	0.7	1.0	0.4	12.3	17.1	39.2	34.3	10.8	17.5	8.8	5.8	1.8	0.2
Routine cognitive tasks	10.0	18.1	47.6	23.2	22.7	17.2	11.4	46.3	56.1	59.5	34.2	19.3	27.9	9.0
Routine manual tasks	34.3	64.7	22.0	13.2	17.2	4.1	2.0	3.2	1.1	0.0	5.2	0.9	28.4	29.6
Non-routine manual tasks	37.7	9.1	11.3	48.6	39.1	42.0	11.1	3.6	0.6	0.8	0.7	51.4	31.4	57.1

Source: Calculation of the shares for the year 2016 by Dengler, K., Matthes, B. and Paulus, W. on basis of Dengler et al. (2014).

with female employees regarding increased workplace and short-term working-time flexibility are likely to be fewer than is the case with male employees, at least as far as employees with children or other caring responsibilities are concerned. In many families it is still the mothers who perform the majority of the care work and who have to reconcile this with their employment in terms of space and time. This means that they are tied to existing and stable agreements with their employers to a greater extent, which tends to oppose greater flexibility. The existence of a works council or collective agreements shows no effects in the three estimations.

DISCUSSION AND OUTLOOK

Our analyses contribute to the largely unexplored area of research on the qualitative effects of digitalization and the use of AI on working conditions, especially with regard to the demand for increasing flexibility in work assignments. We pay particular attention to the role played by differences between occupations, because, as is discussed in the literature, AI is affecting different occupational fields in different ways. To our knowledge, our study is the first one to present estimation results based on data from a large representative employer survey.

First of all, our study confirms some findings from previous literature on digitalization and AI: occupations for which employers expect the most substantial changes in working conditions as a result of digitalization include office and secretarial occupations as well as occupations in business organization. Occupations in vehicle, aerospace, space, and shipping technology and occupations in tax consulting are also frequently mentioned by the employers in the survey. According to the descriptive results, increasing requirements regarding workplace flexibility play a less significant role than short-term working-time flexibility and specifically the demands on employees' self-organization. These findings indirectly support the discussion surrounding the potential labor market effects of AI, according to which AI primarily changes work content and work processes, which is directly related to aspects of employees' self-organization. According to our results, the flexibility requirements are changing especially in those occupational fields that are undergoing particular strong changes in the context of AI, as discussed for instance by Lane and Saint-Martin (2021).

Using random effects estimations and including numerous establishment- and occupation-specific control variables, we show that it is above all the occupational and less the establishment-specific characteristics that determine the probability of employers demanding increasing flexibility. Increasing demands in terms of flexibility are particularly prevalent in occupational groups that involve a large proportion of routine cognitive activities. These are the fields that are likely to change more strongly with increasing use of AI.

The largest effect of the share of routine cognitive activities in quantitative terms is measured for the probability of increasing demands for employees' self-organization, again supporting arguments, that AI mainly changes work content and work processes. This is particularly important for public employment services: people seeking jobs in occupations with a large proportion of routine cognitive activities can be supported in a targeted manner with regard to their individual abilities and opportunities for a more flexible work engagement than they might be familiar with from previous jobs. This may concern skills in self-organization at work or advice about the advantages and disadvantages of more flexible working time. In fact, policy can focus on very specific areas of the labor market, because possible risks do not affect all occupational fields in which AI is used or might be relevant in future. In our estimations the proportion of manual tasks does not show any significant effect on the flexibility requirements. And occupations involving a large amount of interaction between employees are also less at risk of negative effects. Here, AI is likely to be used somewhat less, since interactions between people are more difficult to replace by machines.

Besides labor market policy also education policy plays a crucial role for the question of whether AI mainly has a negative impact on working conditions or not. Decisive possibilities for policy action are, for instance, the strategic development of the education and vocational training systems and the provision of a child care infrastructure that supports the reconciliation of a more flexible working and private life. For female employees in particular, the increasing use of AI and the associated demand for greater working-time flexibility is likely to be a major challenge and might even become an employment risk if adequate and flexible childcare facilities are not available.

Apart from the share of women, the establishment-specific characteristics play a subordinate role compared to the occupational characteristics. Employers see the challenge of compensating for additional individual burdens on employees in order to maintain the employees' productivity and job satisfaction, especially if the employers are to be increasingly threatened by labor shortages.

Future empirical research on the qualitative labor market effects of digitalization and AI should deal in depth with the role of certain occupations, which requires a larger number of cases in survey-based studies. How does AI change productivity on the one hand and individual stress on the other hand for different employee groups (female/male, young/old, employees with families/without families, etc.) in different occupational fields? Here the gender-related effects should be paid special attention in order to be able to counteract possible replacement effects at an early stage. What options exist for employers to compensate their employees for additional burdens, for example attractive holiday arrangements, further training opportunities, setting up long-term working time accounts with attractive conditions for the employee, through to financial compensation for increasing flexibility in work assignments? What are sustainable good and healthy working conditions that keep the workforce productive and satisfied in times of accelerating digitalization? The employer's perspective is important here for negotiating joint solutions, which makes a combination of both

employer surveys and employee surveys highly attractive in this research field. Finally, international comparative analyses could take into account the specifics of different national labor market policies in the context of ongoing digitalization, which in general has been further accelerated by the current COVID-19 pandemic.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Research Data Centre (FDZ) of the Federal Employment Agency at the Institute for Employment Research. https://fdz.iab.de/en.aspx.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We thank all participants of the ILO Workshop: Artificial Intelligence and the Future of Work: Humans in Control on October 25/26, 2021 for an inspiring discussion and very helpful comments. We also thank two referees for their questions and constructive tips.

REFERENCES

Acemoglu, D., and Restrepo, P. (2020a). Robots and jobs: evidence from US labor markets. *J. Politic. Econ*. 128, 2188–2244. doi: 10.1086/705716

Acemoglu, D., and Restrepo, P. (2020b). The wrong kind of AI? artificial intelligence and the future of labour demand. *Cambridge J. Reg. Econ. Soc*. 13, 25–35. doi: 10.1093/cjres/rsz022

Al Momani, K., Nour, A. N., Jamaludin, N., and Zanani Wan Abdullah, W. Z. W. (2021). "Fourth Industrial Revolution, Artificial Intelligence, Intellectual Capital, and COVID-19 Pandemic" in *Applications of Artificial Intelligence in Business, Education and Healthcare. Studies in Computational Intelligence*, eds. A. Hamdan, A.E. Hassanien, R. Khamis, B. Alareeni, A. Razzaque and B. Awwad, (Cham: Springer).

Allen, T., Herst, D., Bruck, C., and Sutton, M. (2000). Consequences associated with work-to-family conflict: a review and agenda for future research. *J. Occupation. Health Psychol*. 5, 278–308. doi: 10.1037/1076-8998.5.2.278

Amankwah-Amoah, J., Khan, Z., Wood, G., and Knight, G. (2021). COVID-19 and digitalization: the great acceleration. *J. Bus. Res*. 136, 602–611. doi: 10.1016/j.jbusres.2021.08.011

Arntz, M., Gregory, T., and Zierahn, U. (2017). Revisiting the risk of automation. *Econ. Lett*. 159, 157–160. http://dx.doi.org/10.1016/j.econlet.2017.07.001 doi: 10.1016/j.econlet.2017.07.001

Arntz, M., Gregory, T., and Zierahn, U. (2020). "Digitization and the Future of Work: Macroeconomic Consequences" in *Handbook of Labor, Human Resources and Population Economics*, ed. K. Zimmermann, (Cham: Springer).

Bossler, M., Gürtzgen, N., Kubis, A., Küfner, B., and Lochner, B. (2020). The IAB job vacancy survey: design and research potential. *J. Labour Mark. Res*. 54, 13. doi: 10.1186/s12651-020-00278-6

Brenzel, H., Czepek, J., Kiesl, H., Kriechel, B., Kubis, A., and Moczall, A. (2016). *Revision of the IAB Job Vacancy Survey. Backgrounds, methods and results. IAB-Forschungsbericht. 04/2016*. Available online at: https://doku.iab.de/forschungsbericht/2016/fb0416_en.pdf (accessed January 27, 2021).

Brough, P., Timms, C., Chan, X. W., Hawkes, A., and Rasmussen, L. (2020). "Work–Life Balance: Definitions, Causes, and Consequences." in *Handbook of Socioeconomic Determinants of Occupational Health. Handbook Series in Occupational Health Sciences*., ed. T. Theorell, (Cham: Springer), p. 1–15.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers Proceed*. 108, 43–47. doi: 10.1257/pandp.20181019

Cameron, A. C., and Trivedi, P. K. (2010). *Microeconometrics Using Stata*. College Station: Stata Press, p. 706.

Davis, S., Röttger, C.;, Warning, A., and Weber, E. (2014). *Job Recruitment and Vacancy Durations in Germany. University of Regensburg Working Papers in Business, Economics and Management Information Systems*. Available online at: https://epub.uni-regensburg.de/29914/. (accessed January 27, 2021).

DeCanio, S. J. (2016). Robots and humans—complements or substitutes? *J. Macroecon*. 49, 280–291. doi: 10.1016/j.jmacro.2016.08.003

Dengler, K., Matthes, B., and Paulus, W. (2014). *Occupational Tasks in the German Labour Market—An alternative measurement on the basis of an expert database. FDZ-Methodenreport*. Available online at: https://doku.iab.de/fdz/reporte/2014/MR_12-14_EN.pdf (accessed January 27, 2021).

Dettmers, J., Kaiser, S., and Fietze, S. (2013). Theory and practice of flexible work: organizational and individual perspectives. introduction to the special issue. *Manage. Revue*. 24, 155–161. https://www.jstor.org/stable/23610676 doi: 10.5771/0935-9915-2013-3-155

Diebig, M., Müller, A., and Angerer, P. (2020). "Impact of the Digitization in the Industry Sector on Work, Employment, and Health" in *Handbook of Socioeconomic Determinants of Occupational Health. Handbook Series in Occupational Health Sciences*, ed. T. Theorell, (Cham: Springer), p. 1–15.

Ernst, E., Merola, R., and Samaan, D. (2019). Economics of artificial intelligence: implications for the future of work. *IZA J. Labor Policy*. 9, 1. doi: 10.2478/izajolp-2019-0004

Eurostat (n. d.). *Job Vacancies*. Available online at: https://ec.europa.eu/eurostat/web/labour-market/job-vacancies (accessed April 20, 2022).

Felten, E., Raj, M., and Seamans, R. (2019). The occupational impact of artificial intelligence on labor: the role of complementary skills and technologies". *NYU Stern School Bus*. 19, 605. doi: 10.2139/ssrn.3368605

Ford, M., Heinen, B., and Langkamer, K. (2007). Work and family satisfaction and conflict: A meta-analysis of cross-domain relations. *J. Appl. Psychol*. 92, 57–80. doi: 10.1037/0021-9010.92.1.57

Fossen, F. M., and Sorgner, A. (2022). New digital technologies and heterogeneous wage and employment dynamics in the United States: evidence from individual-level data. *Technol. Forecast. Soc. Change*. 22, 175. doi: 10.1016/j.techfore.2021.121381

Genz, S., Gregory, T., Janser, M., Lehmer, F., and Matthes, B. (2021). How Do Workers Adjust When Firms Adopt New Technologies? *ZEW—Centre Euro. Econ. Res*. 21, 21–073. doi: 10.2139/ssrn.3949800

Georgieff, A., and Hyee, R. (2021). *Artificial intelligence and employment: New cross-country evidence. OECD Social, Employment and Migration Working Papers*. 265. Paris: OECD Publishing.

Hartwig, M., Wirth, M., and Bonin, D. (2020). Insights about mental health aspects at intralogistics workplaces—a field study. *Int. J. Industr. Ergon*. 2, 76. doi: 10.1016/j.ergon.2020.102944

Institut DGB and Index Gute Arbeit (2016). *Arbeitshetze und Arbeitsintensivierung bei digitaler Arbeit. So beurteilen die Beschäftigten ihre Arbeitsbedingungen, Ergebnisse einer Sonderauswertung der Repräsentativumfrage DGB-Index Gute Arbeit (Work rush and work intensification in digital work. How the employees evaluate their working conditions. Results of a special evaluation of the representative survey DGB Index Good Work)*. Available online at: https://index-gute-arbeit.dgb.de/veroeffentlichungen/sonderauswertungen/++co++70aa62ec-2b31-11e7-83c1-525400e5a74a (accessed January 27, 2021).

Lane, M., and Saint-Martin, A. (2021). *The impact of Artificial Intelligence on the labour market: What do we know so far? OECD Social, Employment and Migration Working Papers. 256*. Paris: OECD Publishing.

Marschall, J., Hildebrandt, S., Sydow, H., and Nolting, H.-D. (2017). *Gesundheitsreport 2017 (Health report 2017). Beiträge zur Gesundheitsökonomie und Versorgungsforschung.16*. Available online at: https://www.dak.de/dak/download/gesundheitsreport-2017-2108948.pdf (accessed January 27, 2021).

OECD (2020). *OECD Digital Economy Outlook 2020*. Paris: OECD Publishing.

Reimann, M., Abendroth, A.-K., and Diewald, M. (2020). *How Digitalized is Work in Large German Workplaces, and How is Digitalized Work Perceived by Workers? A New Employer-Employee Survey Instrument. IAB-Forschungsbericht*. Available online at: https://doku.iab.de/forschungsbericht/2020/fb0820.pdf (accessed January 27, 2021).

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: looking outside the wage structure. *J. Labor Econ*. 24, 235–270. doi: 10.1086/499972

Statistical Offices of the Federation and the Länder (n. d.). *Klassifikationsserver*. Available online at: https://www.klassifikationsserver.de/ (accessed April 20, 2022).

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., and Gómez, E. (2021). Measuring the occupational impact of ai: tasks, cognitive abilities and ai benchmarks. *J. Artific. Intell. Res*. 71, 191–236. doi: 10.1613/jair.1.12647

Warning, A., Sellhorn, T., and Kummer, J.-P. (2020). Digitalisierung und Beschäftigung: Empirische Befunde für die Rechts- und Steuerberatung sowie Wirtschaftsprüfung (Digitalization and employment: Empirical findings for legal, tax consulting, and audit firms). *Betriebswirtschaftliche Forschung und Praxis*. 72, 391–412. doi: 10.1007/s41471-020-00086-1

Warning, A., and Weber, E. (2018). *Digitalisation, hiring and personnel policy: evidence from a representative business survey. IAB-Discussion Paper*. Available online at: https://doku.iab.de/discussionpapers/2018/dp1018.pdf (accessed January 19, 2022).

Webb, M. (2020). The Impact of Artificial Intelligence on the Labor Market. *SSRN Electron. J*. 20, 2150. doi: 10.2139/ssrn.3482150

Wooldridge, M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, p. 1064.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Warning, Weber and Püffel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 10 May 2022
doi: 10.3389/frai.2022.832736

Artificial Intelligence and Employment: New Cross-Country Evidence

Alexandre Georgieff and Raphaela Hyee*

Organisation for Economic Co-operation and Development, Paris, France

Recent years have seen impressive advances in artificial intelligence (AI) and this has stoked renewed concern about the impact of technological progress on the labor market, including on worker displacement. This paper looks at the possible links between AI and employment in a cross-country context. It adapts the *AI occupational impact measure* developed by Felten, Raj and Seamans—an indicator measuring the degree to which occupations rely on abilities in which AI has made the most progress—and extends it to 23 OECD countries. Overall, there appears to be no clear relationship between AI exposure and employment growth. However, in occupations where computer use is high, greater exposure to AI is linked to higher employment growth. The paper also finds suggestive evidence of a negative relationship between AI exposure and growth in average hours worked among occupations where computer use is low. One possible explanation is that partial automation by AI increases productivity directly as well as by shifting the task composition of occupations toward higher value-added tasks. This increase in labor productivity and output counteracts the direct displacement effect of automation through AI for workers with good digital skills, who may find it easier to use AI effectively and shift to non-automatable, higher-value added tasks within their occupations. The opposite could be true for workers with poor digital skills, who may not be able to interact efficiently with AI and thus reap all potential benefits of the technology[1].

Keywords: J21, J23, J24, O33, artificial intelligence

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Leicester,
United Kingdom

Reviewed by:
Jorge Davalos,
University of the Pacific, Peru
Stefan Kühn,
International Labour Organization, Switzerland

****Correspondence:***
Alexandre Georgieff
alexandre.georgieff@oecd.org

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *10 December 2021*
Accepted: *05 April 2022*
Published: *10 May 2022*

Citation:
Georgieff A and Hyee R (2022) Artificial Intelligence and Employment: New Cross-Country Evidence. Front. Artif. Intell. 5:832736. doi: 10.3389/frai.2022.832736

[1]This publication contributes to the OECD's Artificial Intelligence in Work, Innovation, Productivity and Skills (AI-WIPS) programme, which provides policymakers with new evidence and analysis to keep abreast of the fast-evolving changes in AI capabilities and diffusion and their implications for the world of work. The programme aims to help ensure that adoption of AI in the world of work is effective, beneficial to all, people-centred and accepted by the population at large. AI-WIPS is supported by the German Federal Ministry of Labour and Social Affairs (BMAS) and will complement the work of the German AI Observatory in the Ministry's Policy Lab Digital, Work & Society. For more information, visit https://oecd.ai/work-innovation-productivity-skills and https://denkfabrik-bmas.de/.

INTRODUCTION

Recent years have seen impressive advances in Artificial Intelligence (AI), particularly in the areas of image and speech recognition, natural language processing, translation, reading comprehension, computer programming, and predictive analytics.

This rapid progress has been accompanied by concern about the possible effects of AI deployment on the labor market, including on worker displacement. There are reasons to believe that its impact on employment may be different from previous waves of technological progress. Autor et al. (2003) postulated that jobs consist of routine (and thus in principle programmable) and non-routine tasks. Previous waves of technological progress were primarily associated with the automation of routine tasks. Computers, for example, are capable of performing routine cognitive tasks including record-keeping, calculation, and searching for information. Similarly, industrial robots are programmable manipulators of physical objects and therefore associated with the automation of routine manual tasks such as welding, painting or packaging (Raj and Seamans, 2019)[2]. These technologies therefore mainly substitute for workers in low- and middle-skill occupations.

Tasks typically associated with high-skilled occupations, such as non-routine manual tasks (requiring dexterity) and non-routine cognitive tasks (requiring abstract reasoning, creativity, and social intelligence) were previously thought to be outside the scope of automation (Autor et al., 2003; Acemoglu and Restrepo, 2020).

However, recent advances in AI mean that non-routine cognitive tasks can also increasingly be automated (Lane and Saint-Martin, 2021). In most of its current applications, AI refers to computer software that relies on highly sophisticated algorithmic techniques to find patterns in data and make predictions about the future. Analysis of patent texts suggests AI is capable of formulating medical prognosis and suggesting treatment, detecting cancer and identifying fraud (Webb, 2020). Thus, in contrast to previous waves of automation, AI might disproportionally affect high-skilled workers.

Even if AI automates non-routine, cognitive tasks, this does not necessarily mean that AI will displace workers. In general, technological progress improves labor efficiency by (partially) taking over/speeding up tasks performed by workers. This leads to an increase in output per effective labor input and a reduction in production costs. The employment effects of this process are ex-ante ambiguous: employment may fall as tasks are automated (substitution effect). On the other hand, lower production costs may increase output if there is sufficient demand for the good/service (productivity effect)[3].

To harness this productivity effect, workers need to both learn to work effectively with the new technology and to adapt to a changing task composition that puts more emphasis on tasks that AI cannot yet perform. Such adaptation is costly and the cost will depend on worker characteristics.

The areas where AI is currently making the most progress are associated with non-routine, cognitive tasks often performed by medium- to high-skilled, white collar workers. However, these workers also rely more than other workers on abilities AI does not currently possess, such as inductive reasoning or social intelligence. Moreover, highly educated workers often find it easier to adapt to new technologies because they are more likely to already work with digital technologies and participate more in training, which puts them in a better position than lower-skilled workers to reap the potential benefits of AI. That being said, more educated workers also tend to have more task-specific human capital[4], which might make adaption more costly for them (Fossen and Sorgner, 2019).

As AI is a relatively new technology, there is little empirical evidence on its effect on the labor market to date. The literature that does exist is mostly limited to the US and finds little evidence for AI-driven worker displacement (Lane and Saint-Martin, 2021). Felten et al. (2019) look at the effect of exposure to AI[5] on employment and wages in the US at the occupational level. They do not find any link between AI exposure and (aggregate) employment, but they do find a positive effect of AI exposure on wage growth, suggesting that the productivity effect of AI may outweigh the substitution effect. This effect on wage growth is concentrated in occupations that require software skills and in high-wage occupations.

[2]AI may however be used in robotics ("smart robots"), which blurs the line between the two technologies (Raj and Seamans, 2019). For example, AI has improved the vision of robots, enabling them to identify and sort unorganised objects such as harvested fruit. AI can also be used to transfer knowledge between robots, such as the layout of hospital rooms between cleaning robots (Nolan, 2021).

[3]This can only be the case if an occupation is only partially automated, but depending on the price elasticity of demand for a given product or service, the productivity effect can be strong. For example, during the nineteenth century, 98% of the tasks required to weave fabric were automated, decreasing the price of fabric. Because of highly price elastic demand for fabric, the demand for fabric increased as did the number of weavers (Bessen, 2016).

[4]Education directly increases task-specific human capital as well as the rate of learning-by-doing on the job, at least some of which is task-specific (Gibbons and Waldman, 2004, 2006). This can be seen by looking at the likelihood of lateral moves within the same firm: lateral moves have a direct productivity cost to the firm as workers cannot utilise their entire task-specific human capital stock in another area (e.g., when moving from marketing to logistics). However, accumulating at least some task-specific human capital in a lateral position makes sense if a worker is scheduled to be promoted to a position that oversees both areas. If a worker's task-specific human capital is sufficiently high, however, the immediate productivity loss associated with a lateral move is higher than any expected productivity gain from the lateral move following a promotion. For example, in academic settings, Ph.D., economists are not typically moved to the HR department prior to becoming the dean of a department. Using a large employer-employee linked dataset on executives at US corporations, Jin and Waldman (2019) show that workers with 17 years of education were twice as likely to be laterally moved before promotion than workers with 19 years of education.

[5]An occupation is "exposed" to AI if it has a high intensity in skills that AI can perform, see section What Do These Indicators Measure? for details.

Again for the US, Fossen and Sorgner (2019) look at the effect of exposure to AI[6] on job stability and wage growth at the individual level. They find that exposure to AI leads to higher employment stability and higher wages, and that this effect is stronger for higher educated and more experienced workers, again indicating that the productivity effect dominates and that it is stronger for high-skilled workers.

Finally, Acemoglu et al. (2020) look at hiring in US firms with task structures compatible with AI capabilities[7]. They find that firms' exposure to AI is linked to changes in the structure of skills that firms demand. They find no evidence of employment effects at the occupational level, but they do find that firms that are exposed to AI restrict their hiring in non-AI positions compared to other firms. They conclude that the employment effect of AI might still be too small to be detected in aggregate data (given also how recent a phenomenon AI is), but that it might emerge in the future as AI adoption spreads.

This paper adds to the literature by looking at the links between AI and employment growth in a cross-country context. It adapts the *AI occupational impact measure* proposed by Felten et al. (2018, 2019)—an indicator measuring the degree to which occupations rely on abilities in which AI has made the most progress in recent years—and extends it to 23 OECD countries by linking it to the Survey of Adult Skills, PIAAC. This indicator, which allows for variations in AI exposure across occupations, as well as within occupations and across countries, is matched to Labor Force Surveys to analyse the relationship with employment growth.

The paper finds that, over the period 2012–2019, there is no clear relationship between AI exposure and employment growth across all occupations. Moreover, in occupations where computer use is high, AI appears to be positively associated with employment growth. There is also some evidence of a negative relationship between AI exposure and growth in average hours worked among occupations where computer use is low. While further research is needed to identify the exact mechanisms driving these results, one possible explanation is that partial automation by AI increases productivity directly as well as by shifting the task composition of occupations toward higher value-added tasks. This increase in labor productivity and output counteracts the direct displacement effect of automation through AI for workers with good digital skills, who may find it easier to use AI effectively and shift to non-automatable, higher-value tasks within their occupations. The opposite could be true for workers with poor digital skills, who may be unable to interact efficiently with AI and thus reap all potential benefits of the technology.

[6]Fossen and Sorgner (2019) use the occupational impact measure developed by Felten et al. (2018, 2019) and the Suitability for Machine Learning indicator developed by Brynjolfsson and Mitchell (2017) and Brynjolfsson et al. (2018) discussed in Section What Do These Indicators Measure?

[7]Acemoglu et al. (2020) use data from Brynjolfsson and Mitchell, 2017; Brynjolfsson et al., 2018, Felten et al. (2018, 2019), and (Webb, 2020) to identify tasks compatible with AI capabilities; and data from online job postings to identify firms that use AI, see Section Indicators of Occupational Exposure to AI for details.

The paper starts out by presenting indicators of AI deployment that have been proposed in the literature and discussing their relative merits (Section Indicators of Occupational Exposure to AI). It then goes on to present the indicator developed in this paper and builds some intuition on the channels through which occupations are potentially affected by AI (Section Data). Section Results presents the main results.

INDICATORS OF OCCUPATIONAL EXPOSURE TO AI

To analyse the links between AI and employment, it is necessary to determine where in the economy AI is currently deployed. In the absence of comprehensive data on the adoption of AI by firms, several proxies for (potential) AI deployment have been proposed in the literature. They can be grouped into two broad categories. The first group of indicators uses information on labor demand to infer AI activity across occupations, sectors and locations. In practice, these indicators use online job postings that provide information on skills requirements and they therefore will only capture AI deployment if it requires workers to have AI skills. The second group of indicators uses information on AI capabilities—that is, information on what AI can currently do—and links it to occupations. These indicators measure potential exposure to AI and not actual AI adoption. This section presents some of these indicators and discusses their advantages and drawbacks.

Indicators Based on AI-Related Job Posting Frequencies

The first set of indicators use data on AI-related skill requirements in job postings as a proxy for AI deployment in firms. The main data source for these indicators is Burning Glass Technologies (BGT), which collects detailed information—including job title, sector, required skills etc. —on online job postings (see **Box 1** for details). Because of the rich and up-to-date information BGT data offers, these indicators allow for a timely tracking of the demand for AI skills across the labor market.

Squicciarini and Nachtigall (2021) identify AI-related job postings by using keywords extracted from scientific publications, augmented by text mining techniques and expert validation [see Baruffaldi et al. (2020) for details]. These keywords belong to four broad groups: (i) generic AI keywords, e.g., "artificial intelligence," "machine learning;" (ii) AI approaches or methods: e.g., "decision trees," "deep learning;" (iii) AI applications: e.g., "computer vision," "image recognition;" (iv) AI software and libraries: e.g., Python or TensorFlow. Since some of these keywords may be used in job postings for non AI-related jobs (e.g., "Python" or "Bayesian"), the authors only tag a job as AI-related if the posting contains at least two AI keywords from at least two distinct concepts. This indicator is available on an annual basis for Canada, Singapore, the United Kingdom and the United States, for 2012–2018[8].

[8]Sectors are available according to the North American Industry classification system (NAICS) for the US and Canada and according to the UK Standard

Acemoglu et al. (2020) take a simpler approach by defining vacancies as AI-related if they contain any keyword belonging to a simple list of skills related to AI[9]. As this indicator will tag any job posting that contains one of the keywords, it is less precise than the indicator proposed by Squicciarini and Nachtigall (2021), but also easier to reproduce.

Dawson et al. (2021) develop the *skills-space* or *skills-similarity* indicator. This approach defines two skills as similar if they often occur together in BGT job postings and are both simultaneously important for the job posting. A skill is assumed to be less "important" for a particular job posting if it is common across job postings. For example, "communication" and "team work" occur in about a quarter of all job adds, and would therefore be less important than "machine learning" in a job posting requiring both "communication" and "team work[10]." The idea behind this approach is that, if two skills are often simultaneously required for jobs, (i) they are complementary and (ii) mastery of one skill means it is easier to acquire the other. In that way, similar skills may act as "bridges" for workers wanting to change occupations. It also means that workers who possess skills that are similar to AI skills may find it easier to work with AI, even if they are not capable of developing the technology themselves. For example, the skill "copy writing" is similar to "journalism," meaning that a copy writer might transition to journalism at a lower cost than, say, a social worker, and that a copy writer might find it comparatively easier to use databases and other digital tools created for journalists.

Skill similarity allows the identification and tracking of emerging skills: using a short list of "seed skills[11]," the indicator can track similar skills as they appear in job ads over time, keeping the indicator up to date. For example, TensorFlow is a deep learning framework introduced in 2016. Many job postings now list it as a requirement without additionally specifying "deep learning" (Dawson et al., 2021).

The skill similarity approach is preferable to the simple job posting frequency indicators mentioned above (Acemoglu et al., 2020; Squicciarini and Nachtigall, 2021) as it does not only pick up specific AI job postings, but also job postings with skills that are similar (but not identical) to AI skills, and may thus enable workers to work with AI technologies. Another advantage of this indicator is its dynamic nature: as technologies develop and skill requirements evolve, skill similarity can identify new skills that appear in job postings together with familiar skills, and keep the relative skill indicators up-to-date. This indicator is available at the annual level from 2012 to 2019 for Australia and New Zealand[12].

Task-Based Indicators

Task-based indicators for AI adoption are based on measures of AI capabilities linked to tasks workers perform, often at the occupational level. They identify occupations as exposed to AI if they perform tasks that AI is increasingly capable of performing.

The *AI occupational exposure measure* developed by Felten et al. (2018, 2019) is based on progress scores in nine AI applications[13] (such as reading comprehension or image recognition) from the AI progress measurement dataset provided by the Electronic Frontier Foundation (EFF). The EFF monitors progress in AI applications using a mixture of academic literature, blog posts and websites focused on AI. Each application may have several progress scores. One example of a progress score would be a recognition error rate for image recognition. The authors rescale these scores to arrive at a composite score that measures progress in each application between 2010 and 2015.

Felten et al. (2018, 2019) then link these AI applications to abilities in the US Department of Labor's O*NET database. Abilities are defined as "enduring attributes of the individual that influence performance," e.g., "peripheral vision" or "oral comprehension." They enable workers to perform tasks in their jobs (such as driving a car or answering a call), but are distinct from skills in that they cannot typically be acquired or learned. Thus, linking O*NET abilities to AI applications means linking human to AI abilities.

The link between O*NET abilities and AI applications (a correlation matrix) is made via an Amazon Mechanical Turk survey of 200 gig workers per AI application, who are asked whether a given AI application—e.g., image recognition—can be used for a certain ability—e.g., peripheral vision[14]. The correlation matrix between applications and abilities is then calculated as the share of respondents who thought that a given AI application could be used for a given ability. These abilities are subsequently linked to occupations using the O*NET database. This indicator is available for the US for 2010–2015[15].

Similarly, the *Suitability for Machine Learning* indicator developed by Brynjolfsson and Mitchell (2017), Brynjolfsson et al. (2018) assigns a suitability for machine learning score to each of the 2,069 narrowly defined work activities from the O*NET

Industrial Classification (SIC) and Singapore Industrial Classification (SSIC) for the UK and Singapore. Occupational codes are available according to the O*NET classification for Canada, SOC for the UK, and the US and SSOC for Singapore. These codes can be converted to ISCO at the one-digit level.

[9]This paper uses the same list of skills to look at AI job-postings, see Footnote 44 for the complete list of skills.

[10]To measure importance of skills in job ads, the authors use the Revealed Comparative Advantage (RCA) measure, loaned from trade economics, that weighs the importance of a skill in a job posting up if the number of skills for this specific posting is low, and weighs it down if the skill is ubiquitous in all job adds. That is, the skill "team work" will be generally less important given its ubiquity in all job ads, but its importance in an individual job posting would increase if only few other skills were required for that job.

[11]"Artificial Intelligence," "Machine Learning," "Data Science," "Data Mining," and "Big Data".

[12]The indicator is calculated at the division level (19 industries) according to the Australian and New Zealand Standard Industrial Classification Level (ANZSIC).

[13]Abstract strategy games, real-time video games, image recognition, visual question answering, image generation, reading comprehension, language modelling, translation, and speech recognition. Abstract strategy games, for example are defined as "the ability to play abstract games involving sometimes complex strategy and reasoning ability, such as chess, go, or checkers, at a high level." While the EFF tracks progress on 16 applications, AI has not made any progress on 7 of these over the relevant time period (Felten et al., 2021).

[14]The background of the gig workers is not known and so they may not necessarily be AI experts. This could be a potential weakness of this indicator. In contrast (Tolan et al., 2021) rely on expert assessments for the link between AI applications and worker abilities (Tolan et al., 2021).

[15]At the six digit SOC 2010 occupational level, this can be aggregated across sectors and geographical regions, see Felten et al. (2021).

BOX 1 | Burning Glass Technologies (BGT) online job postings data

Burning Glass Technologies (BGT) collects data on online job postings by web-scraping 40 000 online job boards and company websites. It claims to cover the near-universe of online job postings. Data are currently available for Australia, Canada, New Zealand, Singapore, the United Kingdom, and the United States for the time period 2012–2020 (2014–2020 for Germany and 2018–2020 for other European Union countries). BGT extracts information such as location, sector, occupation, required skills, education, and experience levels from the text of job postings (deleting duplicates) and organizes it into up to 70 variables that can be linked to labor force surveys, providing detailed, and timely information on labor demand.

Despite its strengths, BGT data has a number of limitations:

- It misses vacancies that are not posted online. Carnevale et al. (2014) compare vacancies from survey data according to the Job Openings and Labor Turnover Survey (JOLTS) from the US Bureau of Labor Statistics, a representative survey of 16,000 US businesses, with BGT data for 2013. They find that roughly 70% of vacancies were posted online, with vacancies requiring a college degree significantly more likely to be posted online compared to jobs with lower education requirements.
  - There is not necessarily a direct, one-to-one correspondence between an online job ad and an actual vacancy: firms might post one job ad for several vacancies, or post job ads without firm plans to hire, e.g., because they want to learn about available talent for future hiring needs.
- BGT data might over-represent growing firms that cannot draw on internal labor markets to the same extent as the average firm.
- Higher turnover in some occupations and industries can produce a skewed image of actual labor demand since vacancies reflect a mixture of replacement demand as well as expansion.

In addition, since BGT data draws on published job advertisements, it is a proxy of current vacancies, and not of hiring or actual employment. As a proxy for vacancies, BGT data performs reasonably well, although some occupations and sectors are over-represented. Hershbein and Kahn (2018) show for the US that, compared to vacancy data from the U.S. Bureau of Labor Statistics' Job Openings and Labor Turnover Survey (JOLTS), BGT over-represents health care and social assistance, finance and insurance, and education, while under-representing accommodation, food services and construction (where informal hiring is more prevalent) as well as public administration/government. These differences are stable across time, however, such that *changes* in labor demand in BGT track well with JOLTS data. Regarding hiring, they also compare BGT data with new jobs according to the Current Population Survey (CPS). BGT data strongly over-represents computer and mathematical occupations (by a factor of over four, which is a concern when looking at growth in demand for AI skills as compared to other skills), as well as occupations in management, healthcare, and business and financial operations. It under-represents all remaining occupations, including transportation, food preparation and serving, production, or construction.

Cammeraat and Squicciarini (2020) argue that, because of differences in turnover across occupations, countries and time, as well as differences in the collection of national vacancy statistics, the representativeness of BGT data as an indicator for labor and skills demand should be measured against employment growth. They compare growth rates in employment with growth rates in BGT job postings on the occupational level in the six countries for which a BGT timeline exists. They find that, across countries, the deviation between BGT and employment growth rates by occupation is lower than 10 percentage points for 65% of the employed population. They observe the biggest deviations for agricultural, forestry and fishery workers, as well as community and personal service workers, again occupations where informal hiring may be more prevalent.

database that are shared across occupations (e.g., "assisting and caring for others," "coaching others," "coordinating the work of others"). For these scores, they use a Machine Learning suitability rubric consisting of 23 distinct statements describing a work activity. For example, for the statement "Task is describable by rules," the highest score would be "Task can be fully described by a detailed set of rules (e.g., following a recipe)," whereas the lowest score would be "The task has no clear, well-known set of rules on what is and is not effective (e.g., writing a book)." They use the human intelligence task crowdsourcing platform CrowdFlower to score each direct work activity by seven to ten respondents. The direct work activities are then aggregated to tasks (e.g., "assisting and caring for others," "coaching others," "coordinating the work of others" aggregate to "interacting with others"), and the tasks to occupations. This indicator is available for the US for the year 2016/2017.

Tolan et al. (2021) introduce a layer of cognitive abilities to connect AI applications (that they call benchmarks) to tasks. The authors define 14 cognitive abilities (e.g., visual processing, planning and sequential decision-making and acting, communication, etc.) from the psychometrics, comparative psychology, cognitive science, and AI literature[16]. They link these abilities to 328 different AI benchmarks (or applications) stemming from the authors' own previous analysis and annotation of AI papers as well as from open resources such as Papers with Code[17]. These sources in turn draw on data from multiple verified sources, including academic literature, review articles etc. on machine learning and AI. They use the research intensity in a specific benchmark (number of publications, news stories, blog entries etc.) obtained from AI topics[18]. Tasks are measured at the worker level using the European Working Conditions Survey (EWCS), PIAAC and the O*NET database. Task intensity is derived as a measure of how much time an individual worker spends on a task and how often the task is performed.

The mapping between cognitive abilities and AI benchmarks, as well as between cognitive abilities and tasks, relies on a correspondence matrix that assigns a value of 1 if the ability is absolutely required to solve a benchmark or complete a task, and 0 if it is not necessary at all. This correspondence matrix was populated by a group of multidisciplinary researchers for the mapping between tasks and cognitive abilities, and by a group of AI-specialized researchers for the mapping between AI benchmarks and cognitive abilities. This indicator is available from 2008 to 2018, at the ISCO-3 level, and

[16]The abilities are chosen from Hernández-Orallo (2017) to be at an intermediate level of detail, excluding very general abilities that would influence all others, such as general intelligence, and too specific abilities and skills, such as being able to drive a car or music skills. They also exclude any personality traits that do not apply to machines. The abilities are: Memory processing, Sensorimotor interaction, Visual processing, Auditory processing, Attention and search, Planning, sequential decision-making and acting, Comprehension and expression, Communication, Emotion and self-control, Navigation, Conceptualisation, learning and abstraction, Quantitative and logical reasoning, Mind modelling and social interaction, and Metacognition and confidence assessment.

[17]Free and open repository of machine learning code and results, which includes data from several repositories (including EFF, NLPD progress etc.).

[18]An archive kept by the by the Association for the Advancement of Artificial Intelligence (AAI).

constructed to be country-invariant (as it combines data covering different countries).

Webb (2020) constructs his *exposure of occupations to any technology* indicator by directly comparing the text of patents from Google patents public data to the texts of job descriptions from the O*NET database to quantify the overlap between patent descriptions and job task descriptions. By limiting the patents to AI patents (using a list of key-words), this indicator can be narrowed to only apply to AI. Each particular task is then assigned a score according to the prevalence of such patents that mention this task; tasks are then aggregated to occupations.

What Do These Indicators Measure?

To gauge the link between AI and employment, the chosen indicator for this study should proxy actual AI deployment in the economy as closely as possible. Furthermore, it should proxy AI deployment at the occupation level because switching occupations is more costly for workers than switching firms or sectors, making the occupation the relevant level for the automation risk of individual workers.

Task-based approaches measure *potential automatability* of tasks (and occupations), so they are measures of AI exposure, not deployment. Because task-based measures look at potential automatability, they cannot capture uneven adoption of AI across occupations, sectors or countries. Thus, in a cross-country analysis, the only source of variation in a task-based indicator are differences in the occupational task composition across countries, as well as cross-country differences in the occupational distribution.

Indicators based on job posting data measure *demand for AI skills* (albeit with some noise, see **Box 1**), as opposed to *AI use*. Thus, they rely on the assumption that AI use in a firm, sector or occupation will lead to employer demand for AI skills *in that particular firm, sector, or occupation*. This is not necessarily the case, however:

- Some firms will decide to train workers in AI rather than recruit workers with AI skills; their propensity to do so may vary across occupations.
- Many AI applications will not require AI skills to work with them.
- Even where AI skills are needed, many firms, especially smaller ones, are likely to outsource AI development and support with its adoption to specialized AI development firms. In this case, vacancies associated with AI adoption would emerge in a different firm or sector to where the technology was actually being deployed.
- The assumption that AI deployment requires hiring of staff with AI skills is even more problematic when the indicator is applied at the occupation level. Firms that adopt AI may seek workers with AI skills in completely different occupations than the workers whose tasks are being automated by AI. For instance, an insurance company wanting to substitute or enhance some of the tasks of insurance clerks with AI would not necessarily hire insurance clerks with AI skills, but AI professionals to develop or deploy the technology. Insurance clerks may only have to interact with this technology, which might not require AI development skills (but may well-require other specialized skills). Thus, even with broad-based deployment of AI in the financial industry, this indicator may not show an increasing number of job postings for insurance clerks with AI skills. This effect could also be heterogeneous across countries and time. For example, Qian et al. (2020) show that law firms in the UK tend to hire AI professionals without legal knowledge, while law firms in Singapore and the US do advertise jobs with hybrid legal-AI skillsets.

Thus, indicators based on labor demand data are a good proxy for AI deployment at the firm and sector level as long as there is no significant outsourcing of AI development and maintenance, and the production process is such that using the technology requires specialized AI skills. If these assumptions do not hold, these indicators will be incomplete. Whether or not this is the case is an empirical question that requires further research. To date the only empirical reference on this question is Acemoglu et al. (2020) who show for the US that the share of job postings that require AI skills increases faster in firms that are heavily exposed to AI (according to task-based indicators). For example, a one standard deviation increase in the measure of AI exposure according to Felten et al. (2018, 2019) leads to a 15% increase in the number of published AI vacancies.

To shed further light on the relationship between the two types of indicators, **Figure 1** plots the 2012–2019 percentage point change in the share of BGT job postings that require AI skills[19] across 36 sectors against a sector-level task-based AI exposure score, similar to the occupational AI exposure score developed in this paper (see Section Construction of the AI Occupational Exposure Measure)[20]. This analysis only covers the United Kingdom and the United States[21] because of data availability. For both countries, a positive relationship is apparent, suggesting that, overall, (i) the two measures are consistent and (ii) AI deployment does require some AI talent *at the sector level*. Specifically, a one standard deviation increase in AI exposure (approximately the difference in exposure between finance and public administration) is associated with a 0.33 higher percentage point change in the share of job postings that require AI skills in the United-Kingdom; a similar relationship emerges in the United-States[22].

While it is reassuring that, at the sector level, the two measures appear consistent, it is also clear that job postings that require AI skills fail to identify certain sectors that are, from a task

[19] AI-related technical skills are identified based on the list provided in Acemoglu et al. (2020), and detailed in Footnote 44.

[20] As with occupations, the industry-level scores are derived using the average frequency with which workers in each industry perform a set of 33 tasks, separately for each country.

[21] The United Kingdom and the United States are the only countries in the sample analysed (see Section Construction of the AI Occupational Exposure Measure) with 2012 Burning Glass Technologies data available, thereby allowing for the examination of trends over the past decade.

[22] The standard deviation of exposure to AI is 0.083 in the United-Kingdom and 0.075 in the United-States. These values are multiplied by the slopes of the linear relationships displayed in **Figure 1**: 3.90 and 4.95, respectively. The average share of job postings that require AI skills was 0.14% in the United-Kingdom and 0.26% in the United-States in 2012, and this has increased to 0.67 and 0.94%, respectively, in 2019.

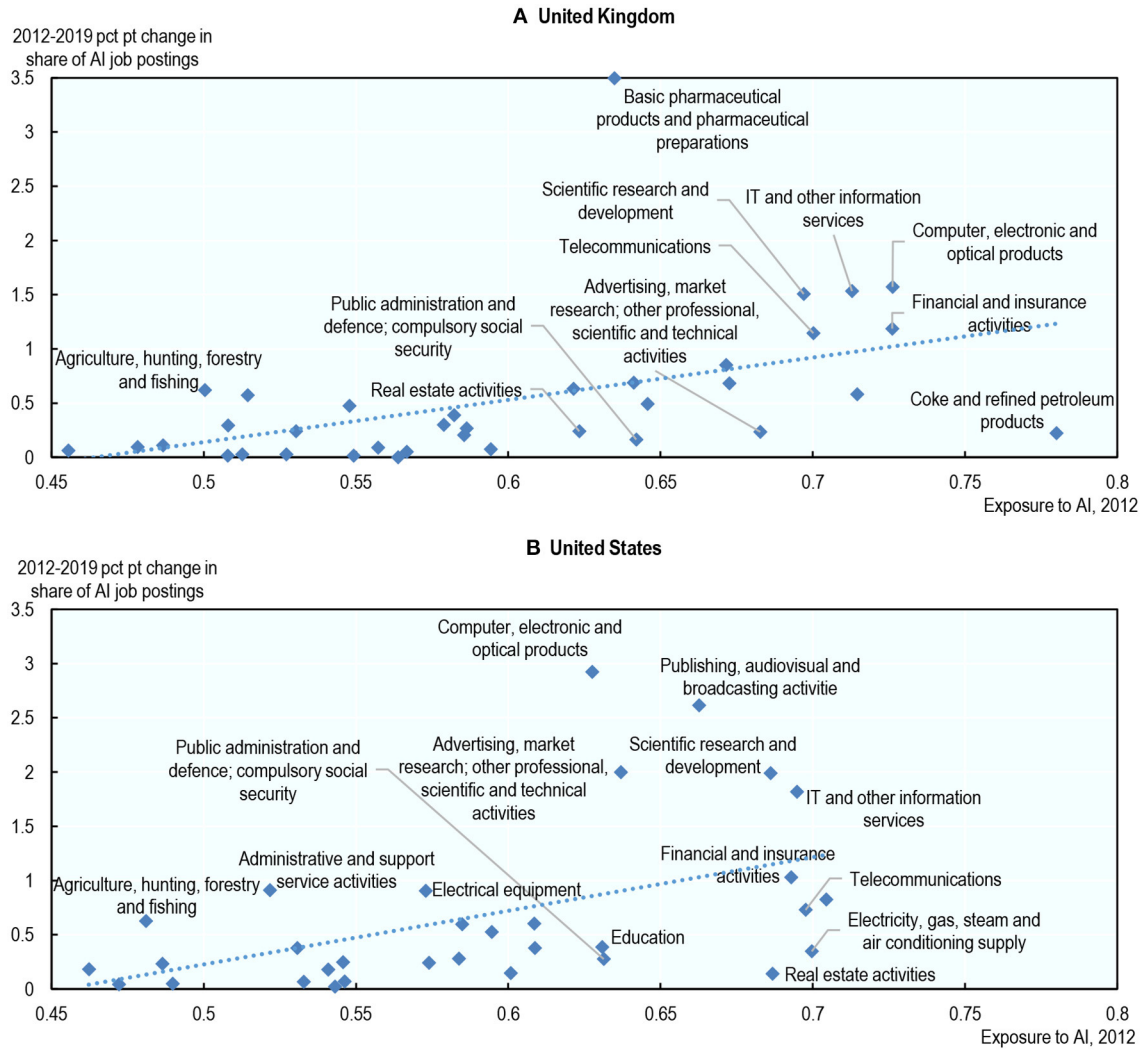


FIGURE 1 | Sectors with higher exposure to AI saw a higher increase in their share of job postings that require AI skills. Percentage point* change in the share of job postings that require AI skills (2012–2019) vs. average exposure to AI (2012), by sector. The share of job postings that require AI skills in a sector is the number of job postings requiring such skills in that sector divided by the total number of job postings in that same sector. Not all sectors have marker labels due to space constraints. *Percentage point changes are preferred over percentage changes because the share of job postings that require AI skills is equal to zero in some sectors in 2012. Source: Author' calculations using data from Burning Glass Technologies, PIAAC and Felten et al. (2019). **(A)** United Kingdom and **(B)** United States.

perspective, highly exposed to AI, such as education, the energy sector, the oil industry, public administration and real estate activities. This suggests that AI development and support may be outsourced and/or that the use of AI does not require AI skills in these sectors.

In addition, and as stated above, there is a priori no reason that demand-based indicators would pick up AI deployment at the occupational level, as firms that adopt AI may seek workers with AI skills in completely different occupations than the workers whose tasks are being automated by AI. This is also borne out in the analysis in this paper (see Section Exposure to AI and Demand for AI-Related Technical Skills: A Weak but Positive Relationship Among Occupations Where Computer Use is High). Thus, labor demand-based indicators are unlikely to be good proxies for AI deployment at the occupational level and, in the analysis described in this paper, preference will be given to task-based measures even though they, too, are only an imperfect proxy for AI adoption.

Which Employment Effects Can These Indicators Capture?

This paper analyses the relationship between AI adoption and employment at the occupational level, since it is automation risk at the occupational level that is most relevant for individual workers. The analysis will therefore require a measure of AI adoption at the occupational level and this section assesses which type of indicator might be best suited to that purpose.

It is useful to think of AI-driven automation as having two possible, but opposed, employment effects. On the one hand, AI may depress employment via automation/substitution. On the other, it may increase it by raising worker productivity.

Focusing on the substitution effect first, task-based indicators will pick up such effects since they measure what tasks could potentially be automated by AI. By contrast, labor-demand based indicators identify occupational AI exposure only if AI skills are mentioned in online job postings for a particular occupation. Thus, they will only pick up substitution effects (that is, a subsequent decline in employment for a particular occupation) if the production process is such that workers whose tasks are being automated need AI skills to interact with the technology.

Regarding the productivity effect, there are several ways in which AI might increase employment. The most straightforward way is that AI increases productivity in a given task, and thus lowers production costs, which can lead to increased employment if demand for a product or service is sufficiently price elastic. This was the case, for example, for weavers in the industrial revolution [see Footnote 4, Bessen (2016)].

In addition, technological progress may allow workers to focus on higher value-added tasks within their occupation that the technology cannot (yet) perform. For example, AI is increasingly deployed in the financial services industry to forecast stock performance. Grennan and Michaely (2017) show that stock analysts have shifted their attention away from stocks for which an abundance of data is available (which lends itself to analysis by AI) toward stocks for which data is scarce. To predict the performance of "low-AI" stocks, analysts gather "soft" information directly from companies' management, suppliers and clients, thus concentrating on tasks requiring a capacity for complex human interaction, of which AI is not (yet) capable.

Task-based indicators will pick up these productivity effects (as they identify exposed occupations directly via their task structure), while labor-demand based indicators will only do so if workers whose tasks are being automated need to interact with the technology, and interacting with the technology requires specialized AI skills.

AI can also be used to augment other technologies, that subsequently automate certain tasks. For example, in robotics, AI supports the efficient automation of physical tasks by improving the vision of robots, or by enabling robots to "learn" from the experience of other robots, e.g., by facilitating the exchange of information on the layout of rooms between cleaning robots (Nolan, 2021). While these improvements to robotics are connected to AI applications (in this example: image recognition and sensory perception of room layouts), the tasks that are being automated (cleaning of rooms) mostly consist of the physical manipulation of objects and thus pertain to the field of robotics. Thus, AI improves the effectiveness of robots to perform tasks associated with cleaners, without performing physical cleaning tasks. As task-based indicators only identify tasks that AI itself can perform (and not tasks that it merely facilitates), they would not capture this effect. In robotics, this would mostly affect physical tasks often performed by low and medium-skilled workers. Indicators based on online vacancies would also be unlikely to capture AI augmenting other technologies at the occupation level—unless cleaners require AI skills to work with cleaning robots.

TABLE 1 | Which potential employment effects of AI can task-based and labor-demand based indicators capture?

	Task-Based indicators	Labor demand-based indicators
Substitution effect (–)	Yes	Only if the production process is such that workers in the partially automated occupation require AI skills to interact with the technology
Productivity effect (+)	Yes	Only if the production process is such that workers in the partially automated occupation require AI skills to interact with the technology
Augmentation of other technologies (e.g., robotics) (–)	No	Only if the production process is such that workers in the partially automated occupation require AI skills to interact with the technology
Job creation through new products and services enabled by AI (+)	No	Only if these new jobs require AI skills

The table only refers to employment effects identified at the occupational level. +/– denote the sign of the employment effect.

Finally, AI could enable the launch of completely new products or services, that lead to job creation, e.g., in marketing or sales of AI-based products and services (Acemoglu et al., 2020). Both task- and labor-demand-based indicators cannot generally measure this effect (unless marketing/selling of AI products requires AI-skills).

To conclude, both types of indicators are likely to understate actual AI deployment at the occupational level (see **Table 1**). Labor-demand based indicators in particular will miss a significant part of AI deployment if workers whose tasks are being automated do not need to interact with AI or if the use of AI does not require any AI skills. Task-based indicators, on the other hand, are not capable of picking up differences in actual AI deployment across time and space (this is because they only measure exposure, not actual adoption). Finally, neither indicator will capture AI augmenting other automating technologies, such as robotics, which is likely to disproportionally affect low-skilled, blue collar occupations.

On the whole, for assessing the links between AI and employment at the occupational level, indicators based on labor demand data are likely to be incomplete. Task-based indicators are therefore more appropriate for the analysis carried out in this paper. Keeping their limitations in mind, however, is crucial.

DATA

This paper extends the *occupational exposure measure*, proposed by Felten et al. (2018, 2019) to 23 OECD countries[23] to look

[23]The 23 countries are Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Mexico,

at the links between AI and labor market outcomes for 36 occupations[24,25] in recent years (2012–2019). The *measure of occupational exposure to AI* proxies the degree to which tasks in those occupations can be automated by AI. Thus, the analysis compares occupations with a high degree of automatability by AI to those with a low degree.

This section presents the data used for the analysis. It begins by describing the construction of the measure of occupational exposure to AI developed and used in this paper, and builds some intuition as to why some occupations are exposed to a higher degree of potential automation by AI than others. It then shows some descriptive statistics for AI exposure and labor market outcomes: employment, working hours, and job postings that require AI skills. Finally, it describes different measures of the task composition of occupations, which will help shed light on the relationship between AI exposure and labor market outcomes.

Occupational Exposure to AI

Several indicators for (potential) AI deployment have been proposed in the literature (see Section Indicators of Occupational Exposure to AI), most of them geared to the US. Since this paper looks at the links between AI and employment across several countries, country coverage is a key criterion for the choice of indicator. This excludes indicators based on AI-related job-posting frequencies, as pre-2018 BGT data is only available for English-speaking countries)[26]. In addition to data availability issues, indicators based on labor demand data are also likely to be less complete than task-based indicators (see Section What Do These Indicators Measure?). Among the task-based measures, the *suitability for machine learning* indicator (Brynjolfsson and Mitchell, 2017; Brynjolfsson et al., 2018) was not publicly accessible at the time of publication. Webb's (2020) indicator captures the stock of patents until 2020, and is therefore too recent to look at the links between AI and the labor market during the observation period (2012–2019), particularly given that major advancements in AI occurred between 2015 and 2020, and the slow pace of diffusion of technology in the economy. The paper therefore uses the occupational exposure measure (Felten et al., 2018, 2019), which has the advantage of capturing AI developments until 2015, leaving some time for the technology to be deployed in the economy. It is also based on actual scientific progress in AI, as opposed to research activity as the indicator proposed by Tolan et al. (2021).

While the preferred measure for this analysis is the *AI occupational exposure measure* proposed by Felten et al. (2018, 2019), the paper also presents additional results using Agrawal's, Gans and Goldfarb (2019) job-posting indicator (an indicator based on job postings), as well as robustness checks using task-based indicators by Webb (2020) and Tolan et al. (2021)[27]. This section describes the construction of the main indicator, and some descriptive statistics.

Construction of the AI Occupational Exposure Measure

The *AI occupational exposure measure* links progress in nine AI applications to 52 abilities in the US Department of Labor's O*NET database (see Section What Do These Indicators Measure? for more details). This paper extends it to 23 OECD countries by mapping the O*NET abilities to tasks from the OECD's Survey of Adult Skills (PIAAC), and then back to occupations (see **Figure 2** for an illustration of the link). Specifically, instead of using the O*NET US-specific measures of an ability's "prevalence" and "importance" in an occupation, country-specific measures have been developed based on data from PIAAC, which reports the frequency with which a number of tasks are performed on the job by each surveyed individual. This information was used to measure the average frequency with which workers in each occupation (classified using two-digit ISCO-08) perform 33 tasks, and this was done separately for each country. Each O*NET ability was then linked to each of these 33 tasks, based on the authors' binary assessments of whether the ability is needed to perform the task or not[28].

This allows for task-content variations in AI exposure across occupations, as well as within occupations and across countries that may arise because of institutional or socio-economic differences across countries. Thus, the indicator proposed in this paper differs from that of Felten et al. (2019) only in that it relies on PIAAC data to take into account occupational task-content heterogeneity across countries. That is, the indicator adopted in this paper is defined at the occupation-country cell level rather than at the occupation level [as in Felten et al. (2019)]. It is scaled such that the minimum is zero and the maximum is one over the full sample of occupation-country cells. It indicates *relative*

the Netherlands, Norway, Poland, Slovenia, the Slovak Republic, Spain, Sweden, United Kingdom, and the United States.

[24]This paper aims to explore the links between employment and AI deployment in the economy, rather than the direct employment increase due to AI development. Two occupations are particularly likely to be involved in AI development: IT technology professionals and IT technicians. These two occupations both have high levels of exposure to AI and some of the highest employment growth over this paper's observation period, which may be partly related to increased activity in AI development. These occupations may bias the analysis and they are therefore excluded from the sample. Nevertheless, the results are not sensitive to the inclusion of IT technology professionals and IT technicians in the analysis.

[25]A few occupation/country cells are missing due to data unavailability for the construction of the indicator of occupational exposure to AI: Skilled forestry, fishery, hunting workers in Belgium and Germany; Assemblers in Greece; Agricultural, forestry, fishery labourers in Austria and France, and Food preparation assistants in the United Kingdom.

[26]This paper uses BGT data for additional results for the countries for which they are available.

[27]While the three task-based indicators point to the same relationships between exposure to AI and employment, the results are less clearcut for the relationship between exposure to AI and average working hours.

[28]The 33 tasks were then grouped into 12 broad categories to address differences in data availability between types of task. For example, "read letters," "read bills," and "write letters" were grouped into one category ("literacy–business"), so that this type of task does not weight more in the final score than tasks types associated with a single PIAAC task (e.g., "dexterity" or "management"). For each ability and each occupation, 12 measures were constructed to reflect the frequency with which workers use the ability in the occupation to perform tasks under the 12 broad task categories. This was done by taking, within each category of tasks, the sum of the frequencies of the tasks assigned to the ability divided by the total number of tasks in the category. Finally, the frequency with which workers use the ability at the two-digit ISCO-08 level and by country was obtained by taking the sum of these 12 measures. The methodology, including the definition of the broad categories of tasks, is adapted from Fernández-Macías and Bisello (2020) and Tolan et al. (2021).

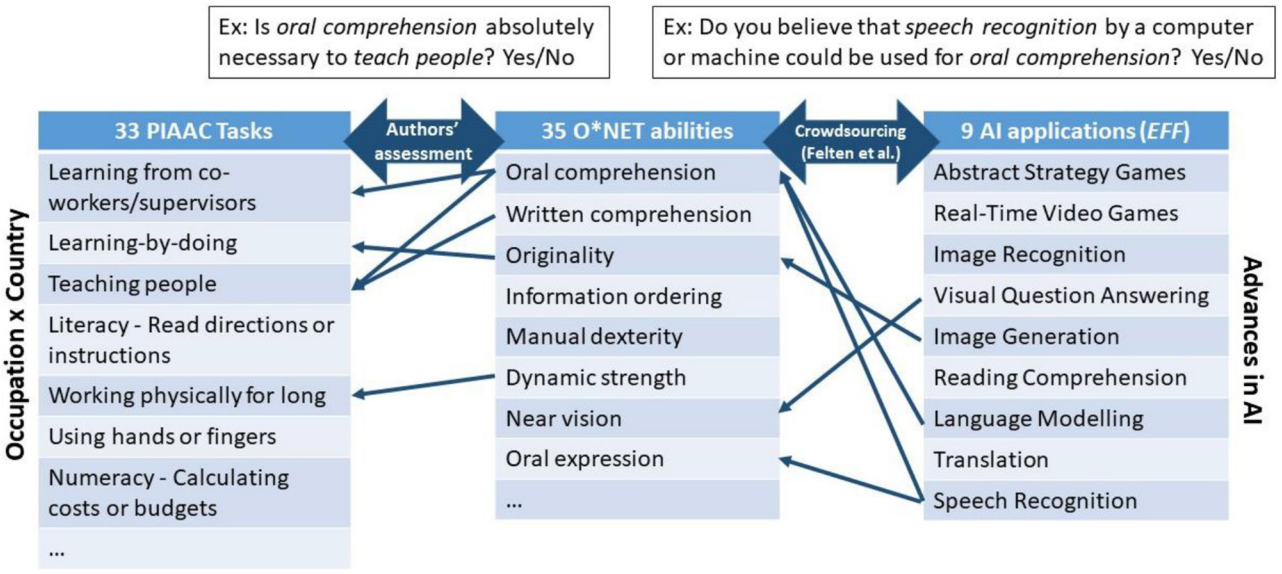


FIGURE 2 | Construction of the measure of occupational exposure to AI. Adaptation from Felten et al. (2018) to 23 OECD countries. The authors link O*NET abilities and PIAAC tasks manually by asking whether a given ability is indispensable for performing a given task. The link between O*NET abilities and AI applications (a correlation matrix) is taken from Felten et al. (2019). The matrix was built by an Amazon Mechanical Turk survey of 200 gig workers per AI application, who were asked whether a given AI application can be used for a certain ability. The correlation matrix between applications and abilities is then calculated as the share of respondents who thought that a given AI application could be used for a given ability. This chart is for illustrative purposes and is not an exhaustive representation of the links between the tasks, abilities and AI applications displayed.

exposure to AI, and no other meaningful interpretation can be given to its actual values.

In this paper, the link between O*NET abilities and PIAAC tasks is performed manually by asking whether a given ability is indispensable for performing a given task, e.g., *is oral comprehension absolutely necessary to teach people*? A given O*NET ability can therefore be linked to several PIAAC tasks, and conversely, a given PIAAC task can be linked to several O*NET abilities. This link was made by the authors of the paper and, in case of diverging answers, agreement was reached through an iterative discussion and consensus method, similar to the Delphi method described in Tolan et al. (2021). Of the 52 O*NET abilities, 35 are related to at least one task in PIAAC. Thus, the indicator loses 17 abilities compared to Felten's et al. (2018, 2019) measure. All the measures that are lost in this way are physical, psychomotor or sensory, as there are no tasks requiring these abilities in PIAAC[29]. As a result, the occupational intensity of physical, psychomotor, or sensory abilities is poorly estimated using PIAAC data. Therefore, whenever possible, robustness checks use O*NET scores of "prevalence" and "importance" of abilities within occupations for the United States (as in Felten et al., 2018) instead of PIAAC-based measures. These robustness tests necessarily assume that the importance and prevalence of abilities are the same in other countries as in the United States. Another approach would have been to assign the EFF applications directly to the PIAAC tasks. However, we preferred to preserve the robustly established mapping of Felten et al. (2018).

The level of exposure to AI in a particular occupation reflects: (i) the progress made by AI in specific applications and (ii) the extent to which those applications are related to abilities required in that occupation. Like all task-based measures, it is at its core a measure of potential automation of occupations by AI, as it indicates which occupations rely most on abilities in which AI has made progress in recent years. It should capture potential positive productivity effects of AI, as well as negative substitution effects caused by (partial) automation of tasks by AI. However, it cannot capture any effects of AI progress on occupations when these effects do not rely on worker abilities that are directly related to the capabilities of AI, such as might be the case when AI augments other technologies, which consequently make progress in the abilities that a person needs in his/her job (see also Section What Do These Indicators Measure?). Section Occupational Exposure to AI shows AI exposure across occupations and builds some intuition on why the indicator identifies some occupations as more exposed to AI than others.

[29]The 17 lost abilities are: control prevision, multilimb coordination, response orientation, reaction time, speed of limb movement, explosive strength, extent flexibility, dynamic flexibility, gross body coordination, gross body equilibrium, far vision, night vision, peripheral vision, glare sensitivity, hearing sensitivity, auditory attention, and sound localization.

AI Progress and Abilities

Over the period 2010–2015, AI has made the most progress in applications that affect abilities required to perform non-routine cognitive tasks, in particular: information ordering,

memorisation, perceptual speed, speed of closure, and flexibility of closure (**Figure 3**)[30]. By contrast, AI has made the least progress in applications that affect physical and psychomotor abilities[31]. This is consistent with emerging evidence that AI is capable of performing cognitive, non-routine tasks (Lane and Saint-Martin, 2021).

Occupational Exposure to AI

The kind of abilities AI has made the most progress in are disproportionately used in highly-educated, white-collar occupations. As a result, white-collar occupations requiring high levels of formal education are among the occupations with the highest exposure to AI: Science and Engineering Professionals, but also Business and Administration Professionals, Managers; Chiefs Executives; and Legal, Social, and Cultural Professionals (**Figure 4**). By contrast, occupations with the lowest exposure include occupations with an emphasis on physical tasks: Cleaners and Helpers; Agricultural Forestry, Fishery Laborers; Food Preparation Assistants and Laborers[32].

The occupational intensity of some abilities is poorly estimated due to PIAAC data limitations. In particular, the 33 PIAAC tasks used in the analysis include only two non-cognitive tasks, and some of the O*NET abilities are not related to any of these tasks. Therefore, as a robustness exercise, **Figure A A.1** displays the level of exposure to AI obtained when using O*NET scores of "prevalence" and "importance" of abilities within occupations for the United States (as in Felten et al., 2018) instead of the PIAAC-based measures. That is, the robustness test assumes that the importance and prevalence of abilities is the same in other countries as in the United States. The robustness test shows the same patterns in terms of AI exposure by occupation, suggesting that it is fine to use the measure linked to PIAAC abilities.

Cleaners and Helpers, the least exposed occupation according to this measure, have a low score of occupational exposure to AI because they rely less than other workers on cognitive abilities (including those in which AI has made the most progress), whereas they rely more on physical and psychomotor abilities (in which AI has made little progress). **Figure 5A** illustrates this by plotting the extent to which Cleaners and Helpers use any of the 35 abilities (relative to the average use of that ability across all occupations) against AI progress in that ability. Compared to the average worker, Cleaners and Helpers rely heavily on physical abilities such as dynamic / static/trunk strength and dexterity, areas in which AI has made the least progress in recent years. They rely less than other occupations on abilities with the fastest AI progress, such as information ordering and memorisation. Business Professionals, in contrast, are heavily exposed to AI because they rely more than other workers on cognitive abilities, and less on physical and psychomotor abilities (**Figure 5B**).

As a robustness check, **Figure A A.2** replicates this analysis using O*NET scores of "prevalence" and "importance" of abilities within occupations instead of PIAAC-based measures, and it shows the same patterns.

As abilities are the only link between occupations and progress in AI, the occupational exposure measure cannot detect any effects of AI that do not work directly through AI capabilities, for example if AI is employed to make other technologies more efficient. Consider the example of drivers, an occupation often discussed as at-risk of being substituted by AI. Drivers receive a below-average score in the AI occupational exposure measure (see **Figure 4**). This is because the driving component of autonomous vehicle technologies relies on the physical manipulation of objects, which is in the realm of robotics, not on AI. AI does touch upon some abilities needed to drive a car—such as the ability to plan a route or perceive and distinguish objects at a distance—but the majority of tasks performed when driving a car are physical. AI might well be essential for driverless cars, but mainly by enabling robotic technology, which possesses the physical abilities necessary to drive a vehicle. Thus, this indicator can be seen as isolating the "pure" effects of AI (Felten et al., 2019).

Cross-Country Differences in Occupational Exposure to AI

On average, an occupation's exposure to AI varies little across countries—differences across occupations tend to be greater. The average score of AI exposure across occupations ranges from 0.52 (Lithuania) to 0.72 (Finland, **Figure 6**) among the 23 countries analyzed[33]. By contrast, the average score across countries for the 36 occupations ranges from 0.26 (cleaners and helpers) to 0.87 (business professionals). Even the most exposed cleaners and helpers (in Finland) are only about half as exposed to AI as the least exposed business professionals (in Lithuania) (**Figure A A.3**). That being said, occupations tend to be slightly more exposed to AI in Northern European countries than in Eastern European ones (**Figure 6**).

A different way of showing that AI exposure varies more across occupations than across countries for a given occupation is by contrasting the distribution of exposure to AI across occupations in the most exposed country in the sample (Finland) with that in the least exposed country (Lithuania, **Figure 7**). The distributions are very similar. In both countries, highly educated

[30]Perceptual speed is the ability to quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns. Speed of closure is the ability to quickly make sense of, combine, and organize information into meaningful patterns. Flexibility of closure is the ability to identify or detect a known pattern (a figure, object, word, or sound) that is hidden in other distracting material.

[31]Only one psychomotor ability has an intermediate score: rate control, which is the ability to time one's movements or the movement of a piece of equipment in anticipation of changes in the speed and/or direction of a moving object or scene.

[32]To get results at the ISCO-08 2-digit level, scores were mapped from the SOC 2010 6-digits classification to the ISCO-08 4-digit classification, and aggregated at the 2-digit level by using average scores weighted by the number of full-time equivalent employees in each occupation in the United States, as provided by Webb (2020) and based on American Community Survey 2010 data.

[33]Averages are unweighted averages across occupations, so that cross-country differences only reflect differences in the ability requirements of occupations between countries, not differences in the occupational composition across countries.

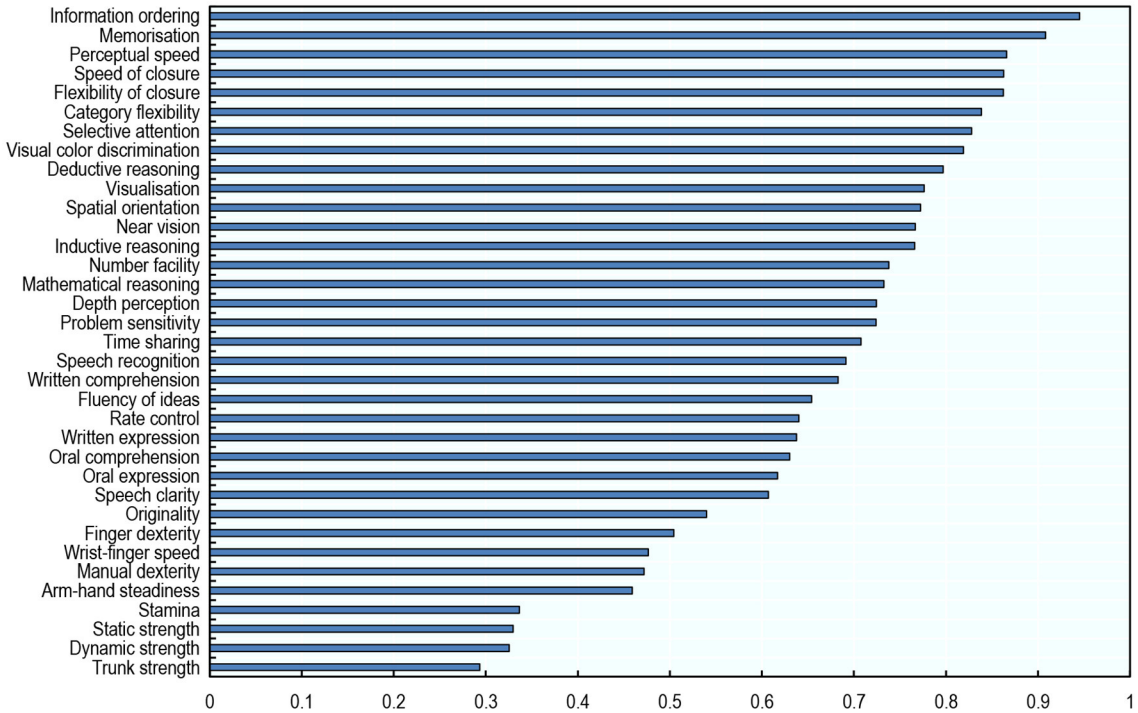


FIGURE 3 | AI has made the most progress in abilities that are required to perform non-routine, cognitive tasks. Progress made by AI in relation to each ability, 2010–2015. The link between O*NET abilities and AI applications (a correlation matrix) is taken from Felten et al. (2019). The matrix was built by an Amazon Mechanical Turk survey of 200 gig workers per AI application, who were asked whether a given AI application—e.g., image recognition—can be used for a certain ability—e.g., near vision. The correlation matrix between applications and abilities is then calculated as the share of respondents who thought that a given AI application could be used for a given ability. To obtain the score of progress made by AI in relation to a given ability, the shares corresponding to that ability are first multiplied by the Electronic Frontier Foundation (EFF) progress scores in the AI applications; these products are then summed over all nine AI applications. Authors' calculations using data from Felten et al. (2019).

white-collar occupations have the highest exposure to AI and non-office-based, physical occupations have the lowest exposure.

Differences in exposure to AI between Finland and Lithuania are greater for occupations in the lower half of the distribution of exposure to AI (**Figure 7**). For example, Food Preparation Assistants in Finland are more than twice as exposed to AI than food preparation assistants in Lithuania, while the score for Business and Administration Professionals is only 12% higher in Finland than in Lithuania.

This is because, while occupations across the entire spectrum of exposure to AI rely more on physical than on cognitive abilities in Lithuania than in Finland, this reliance is more pronounced at the low end of the exposure spectrum. **Figure 8** illustrates this for the least (Cleaners and Helpers) and the most exposed occupations (Business and Administration Professionals). The top panel displays: (i) the difference in the intensity of use of each ability by Cleaners and Helpers between Finland and Lithuania; and (ii) the progress made by AI in relation to that ability. The bottom panel shows the same for Business and Administration Professionals.

For both occupations, workers in Lithuania tend to rely more on physical and psychomotor abilities (which are little exposed to AI), and less on cognitive abilities, including cognitive abilities in which AI has made the most progress. The differences in the intensity of use of cognitive, physical, and psychomotor abilities between Finland and Lithuania are however greater for Cleaners and Helpers than they are for Business and Administration Professionals (**Figure 8**). As an example of how cleaners may be more exposed to AI in Finland than in Lithuania, AI navigation tools may help cleaning robots map out their route. They could therefore substitute for cleaners in supervising cleaning robots, especially in countries where cleaning robots are more prevalent (e.g., probably in Finland[34]). More generally, it is likely that cleaners in Finland use more sophisticated equipment and protocols, resulting in a greater reliance on more exposed cognitive abilities. That being said, even in Finland, the least exposed occupation remains Cleaners and Helpers (**Figure 7**).

Workers in Lithuania may rely more on physical abilities than in Finland because, in 2012, when these ability requirements were measured, technology adoption was more advanced in Finland than in Lithuania. That is, in 2012, technology may have already automated some physical tasks (e.g., cleaning) and created more cognitive tasks (e.g., reading instructions,

[34]Although specific data on cleaning robots are not available, data from the International Federation of Robotics show that, in 2012, industrial robots were more prevalent in Finland than in Lithuania in all areas for which data are available.

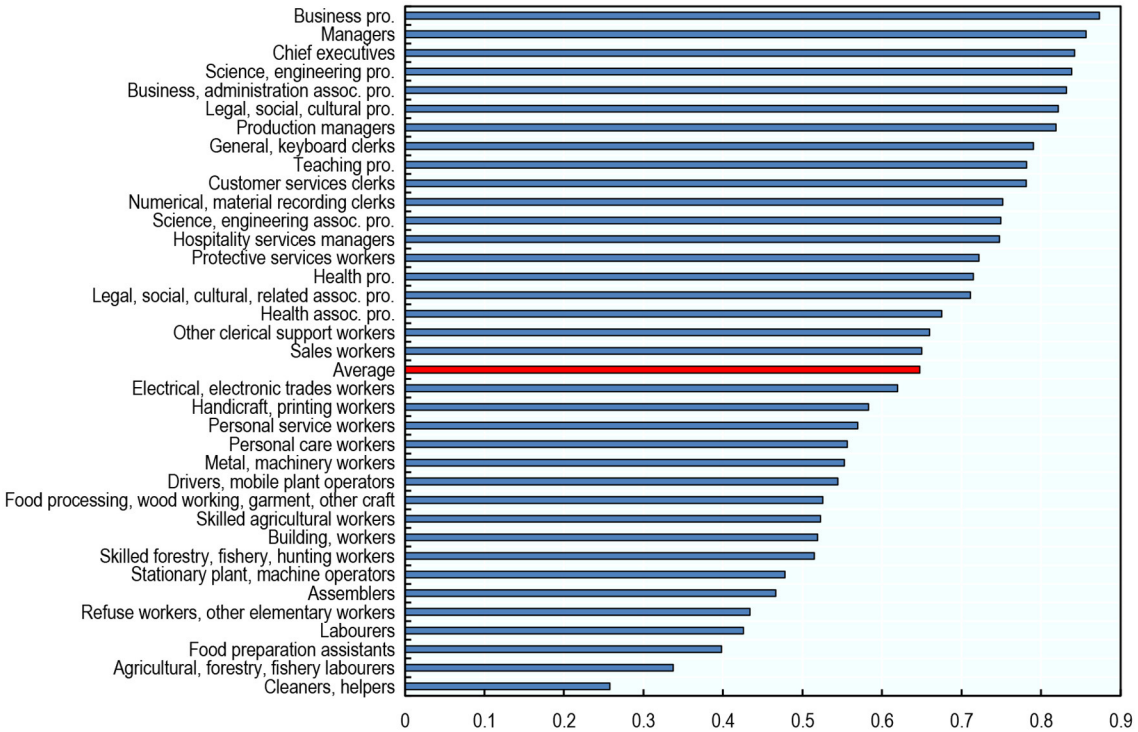


FIGURE 4 | Highly educated white-collar occupations are among the occupations with the highest exposure to AI. Average exposure to AI across countries by occupation, 2012. The averages presented are unweighted. Cross-country averages are taken over the 23 countries included in the analysis. Authors' calculations using data from the Programme for the International Assessment of Adult Competencies (PIAAC) and Felten et al. (2019).

filling out documentation, supervising cleaning robots) in Finland than in Lithuania, and this might have had a bigger effect on occupations that rely more on physical tasks (like cleaning).

Occupational Exposure to AI and Education

Section Occupational Exposure to AI showed that white-collar occupations requiring high levels of formal education are the most exposed to AI, while low-educated physical occupations are the least exposed[35]. **Figure 9** confirms this pattern. It shows a clear positive relationship between the share of highly educated workers within an occupation in 2012 and the AI exposure score in that occupation in that year (red line). By contrast, low-educated workers were less likely to work in occupations with high exposure to AI (blue line). The relationship is almost flat for middle-educated workers. In 2012, 82% of highly educated workers were in the most exposed half of occupations, compared to 37% of middle-educated and only 16% of low-educated[36].

[35]Again, as in the rest of the paper, *exposure to AI* specifically refers to potential automation of tasks, as this is primarily what task-based measures of exposure capture.

[36]On average across countries, there is no clear relationship between AI exposure and gender and age, see **Figures A A.4**, **A A.5** in the Annex.

Labor Market Outcomes

The analysis links occupational exposure to AI to a number of labor market outcomes: employment[37], average hours worked[38], the share of part-time workers, and the share of job postings that require AI-related technical skills. This section presents some descriptive statistics on labor market outcomes for the period 2012 and 2019. Two thousand twelve is chosen as the first year for the period of analysis because it ensures consistency with the measure of occupational exposure to AI, for two reasons. First, the measure of exposure to AI is based on the task composition of occupations in 2012 for most countries[39]. Second,

[37]Employment includes all people engaged in productive activities, whether as employees or self-employed. Employment data is taken from the Mexican National Survey of Occupation and Employment (ENOE), the European Union Labour Force Survey (EU-LFS), and the US Current Population Survey (US-CPS). The occupation classification was mapped to ISCO-08 where necessary. More specifically, the ENOE SINCO occupation code was directly mapped to the ISCO-08 classification. The US-CPS occupation census code variable was first mapped to the SOC 2010 classification. Next, it was mapped to the ISCO-08 classification.

[38]Hours worked refer to the average of individuals' usual weekly hours, which include the number of hours worked during a normal week without any extra-ordinary events (such as leave, public holidays, strikes, sickness, or extra-ordinary overtime).

[39]2012 is available in PIAAC for most countries except Hungary (2017), Lithuania (2014), and Mexico (2017).

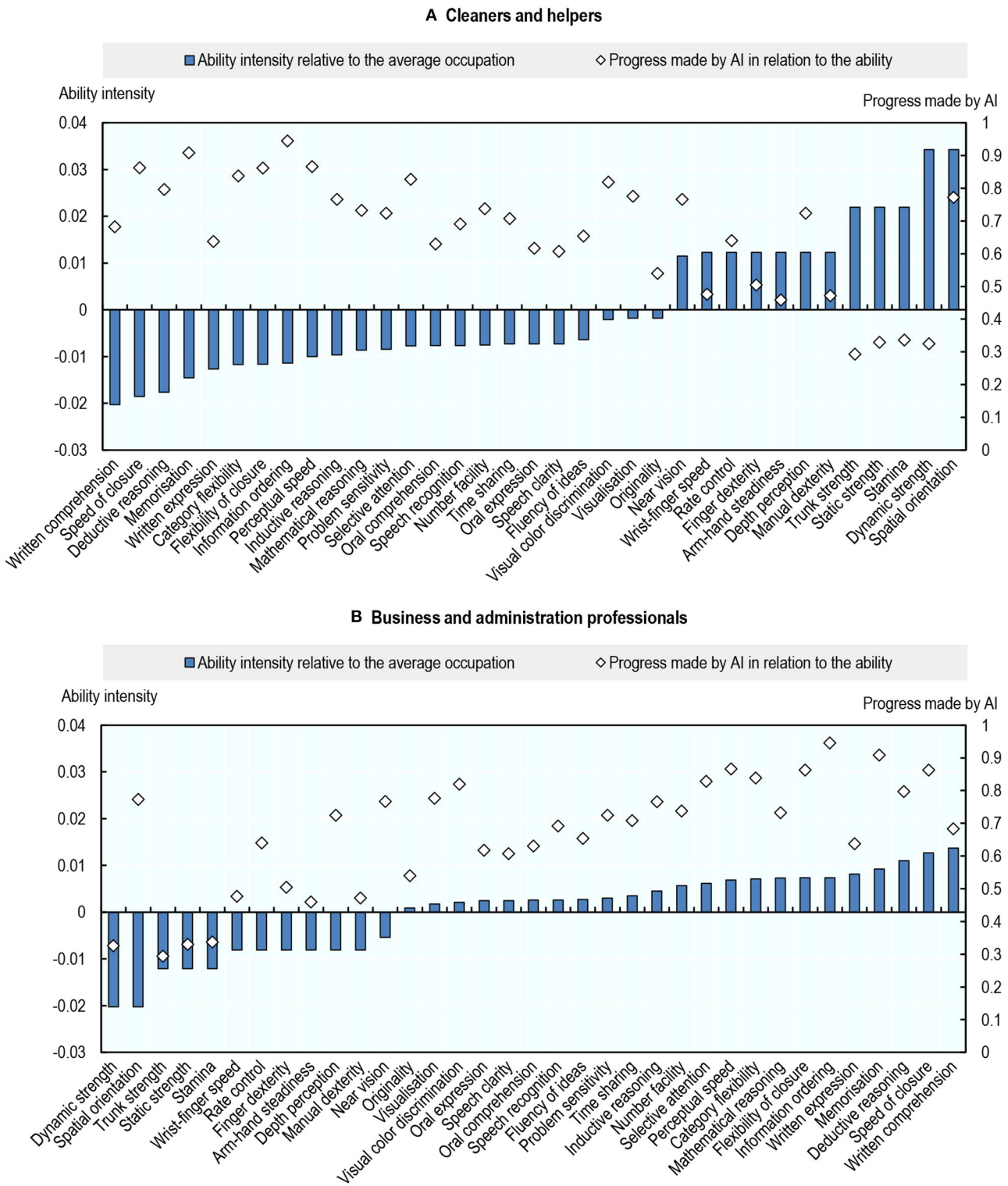


FIGURE 5 | Cross-occupation differences in AI exposure are caused by differences in the intensity of use of abilities. Intensity of use of an ability relative to the average across occupations, and progress made by AI in relation to that ability, 2012. Ability intensity represents the cross-country average frequency of the use of an ability among Cleaner and helpers (top) or Business professionals (bottom) minus the cross-country average frequency of the use of that ability, averaged across the 36 occupations in the sample. Authors' calculations using data from the Programme for the International Assessment of Adult Competencies (PIAAC) and Felten et al. (2019). **(A)** Cleaners and helpers and **(B)** Business and administration professionals.

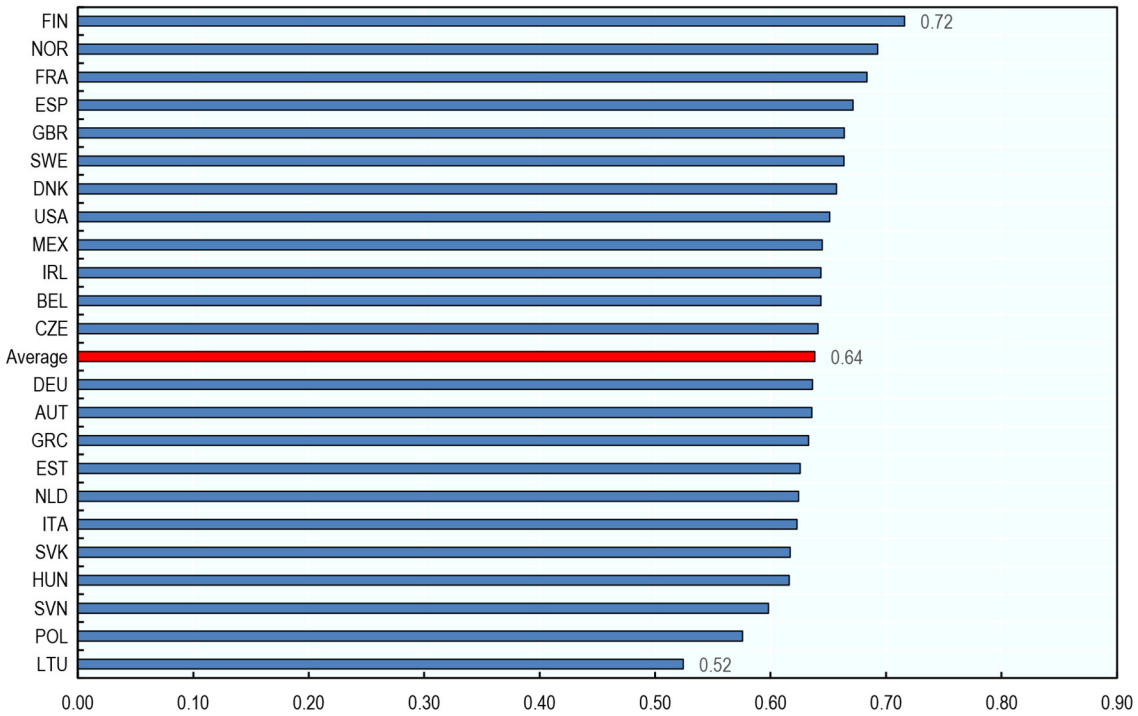


FIGURE 6 | Cross-country differences in exposure to AI for a given occupation are small compared to cross-occupation differences. Average exposure to AI across occupations by country, 2012. The averages presented are unweighted averages across the 36 occupations in the sample. Authors' calculations using data from the Programme for the International Assessment of Adult Competencies (PIAAC) and Felten et al. (2019).

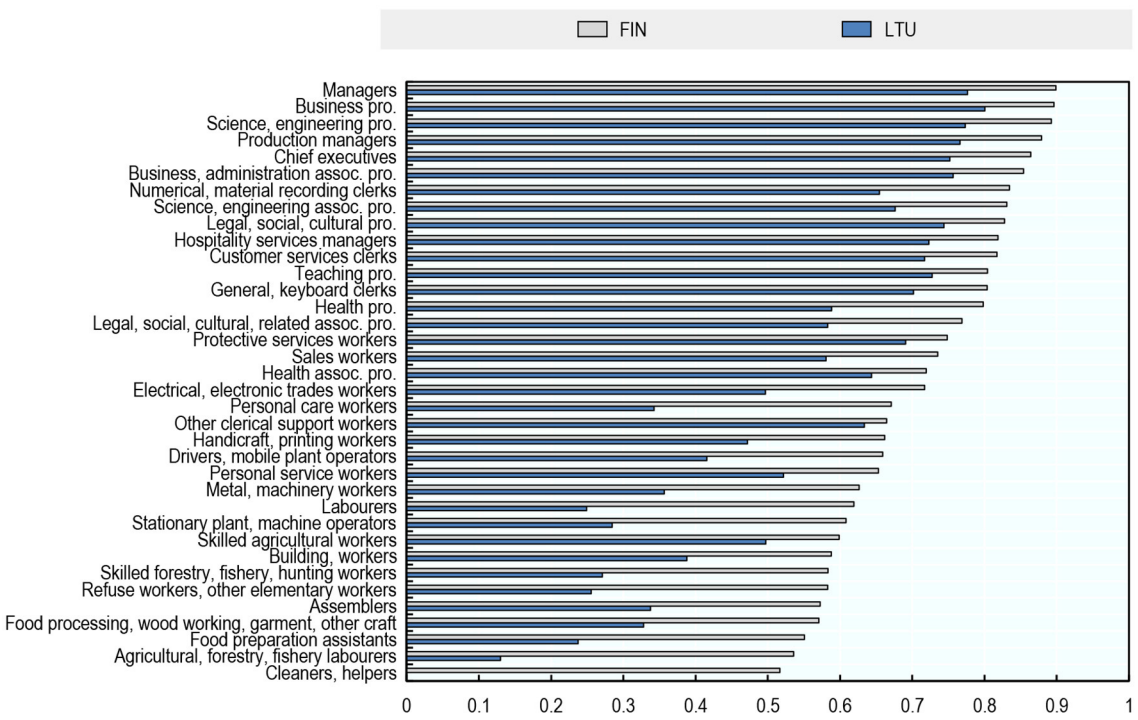


FIGURE 7 | The distribution of AI exposure across occupations is similar in Finland and Lithuania. Exposure to AI, 2012. Authors' calculations using data from the Programme for the International Assessment of Adult Competencies (PIAAC) and Felten et al. (2019).

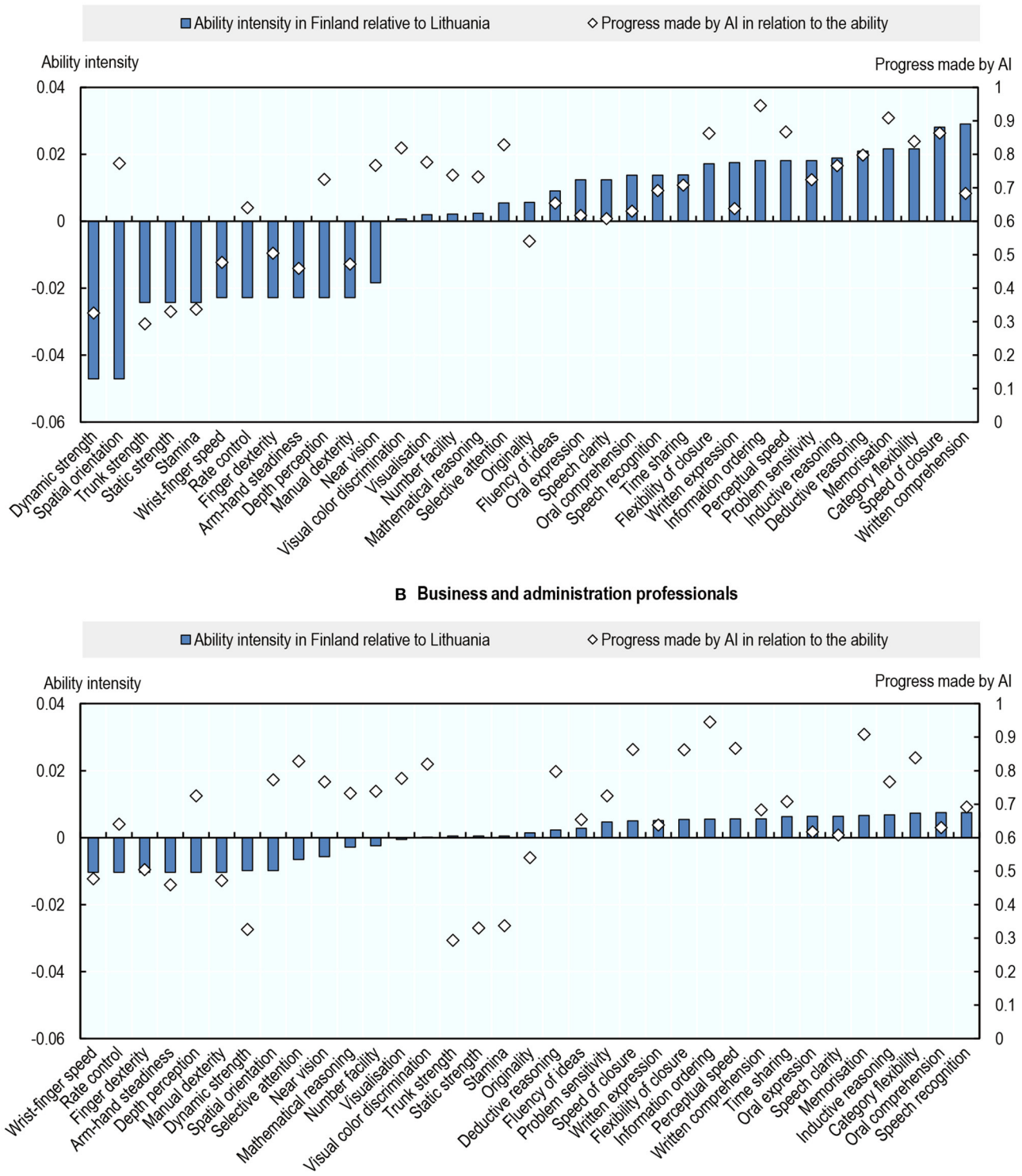


FIGURE 8 | Cross-country differences in occupational AI exposure are caused by differences in the intensity of use of abilities. Intensity of use of an ability in Finland relative to Lithuania and progress made by AI in relation to that ability, 2012. Ability intensity represents the difference in the frequency of the use of an ability among Cleaner and helpers (top) or Business professionals (bottom) between Finland and Lithuania. Authors' calculations using data from the Programme for the International Assessment of Adult Competencies (PIAAC) and Felten et al. (2019). **(A)** Cleaners and helpers and **(B)** Business and administration professionals.

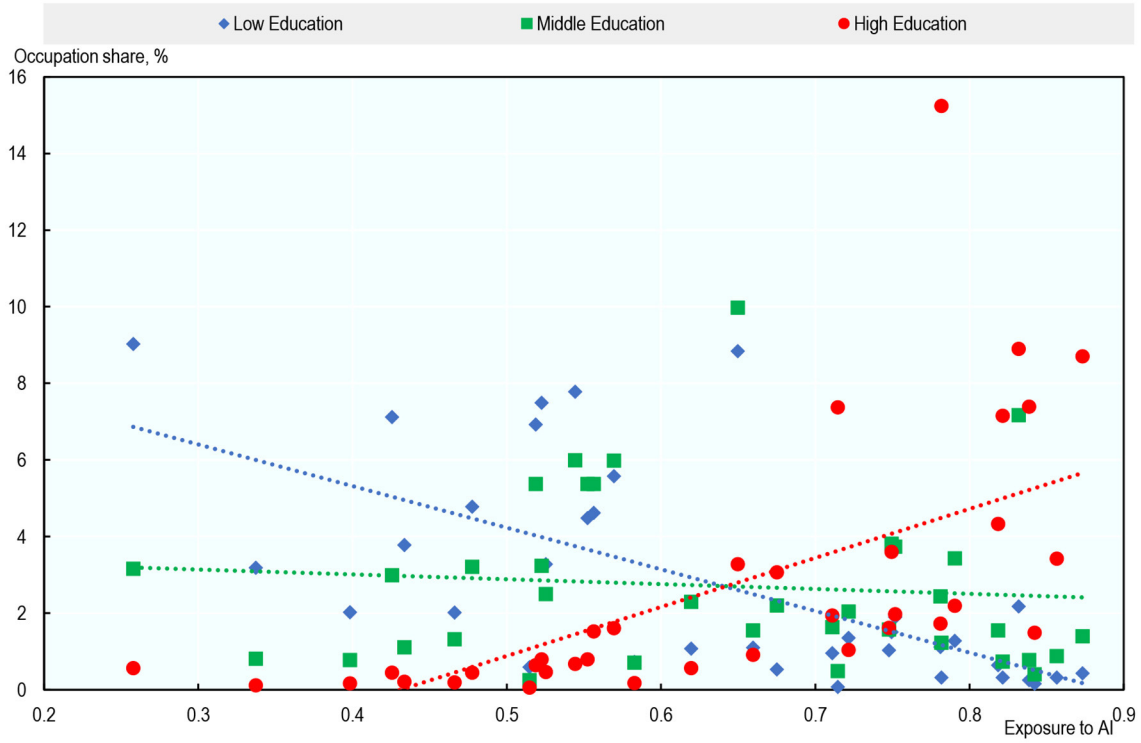


FIGURE 9 | Highly educated workers are disproportionately exposed to AI. Average share of workers with low, medium or high education within occupations vs. average exposure to AI, across countries (2012). For each education group, occupation shares represent the share of workers of that group in a particular occupation. Each dot reports the unweighted average across the 23 countries analyzed of the share of workers with a particular education in an occupation. Authors' calculations using data from the European Union Labor Force Survey (EU-LFS), the Mexican National Survey of Occupation and Employment (ENOE), the US Current Population Survey (US-CPS) PIAAC, and Felten et al. (2019).

progress in AI applications is measured over the period 2010–2015. As a result, AI, as proxied by the occupational AI exposure indicator, could affect the labor market starting from 2010 and fully from 2015 onwards. Starting in 2012 provides a long enough observation period, while closely tracking the measure of recent developments in AI.

Employment and Working Hours

Overall, in most occupations and on average across the 23 countries, employment grew between 2012 and 2019, a period that coincides with the economic recovery from the global financial crisis. Employment grew by 10.8% on average across all occupations and countries in the sample (**Figure 10**). Average employment growth was negative for only four occupations: Other Clerical Support Workers (−9.2%), Skilled Agricultural Workers (−8.2%), Handicraft and Printing Workers (−7.9%), and Metal and Machinery Workers (-1.7%).

By contrast, average usual weekly hours declined by 0.40% (equivalent to 9 min per week[40] average over the same period (**Figure 11**)[41]. On average across countries, working hours declined in most occupations. Occupations with the largest drops in working hours include (but are not limited to) occupations that most often use part-time employment, such as Sales Workers (−2.0%); Legal, Social, Cultural Related Associate Professionals (−1.8%); and Agricultural, Forestry, Fishery Laborers (−1.8%).

Job Postings That Require AI Skills

Beyond its effects on job quantity, AI may transform occupations by changing their task composition, as certain tasks are automated and workers are increasingly expected to focus on other tasks. This may result in a higher demand for AI-related technical skills as workers interact with these new technologies. However, it is not necessarily the case that working with AI requires technical AI skills. For example, a translator using an AI translation tool does not necessarily need any AI technical skills.

This section looks at the share of job postings that require AI-related technical skills (*AI skills*) by occupation using job postings data from Burning Glass Technologies[42] for the United Kingdom

[40] Estimated at the average over the sample (37.7 average usual weekly hours).
[41] Mexico is excluded from the analysis of working time due to lack of data.
[42] See **Box 1** for more details on Burning Glass Technologies data. The Burning Glass Occupation job classification (derived from SOC 2010) was directly mapped to the ISCO-08 classification.

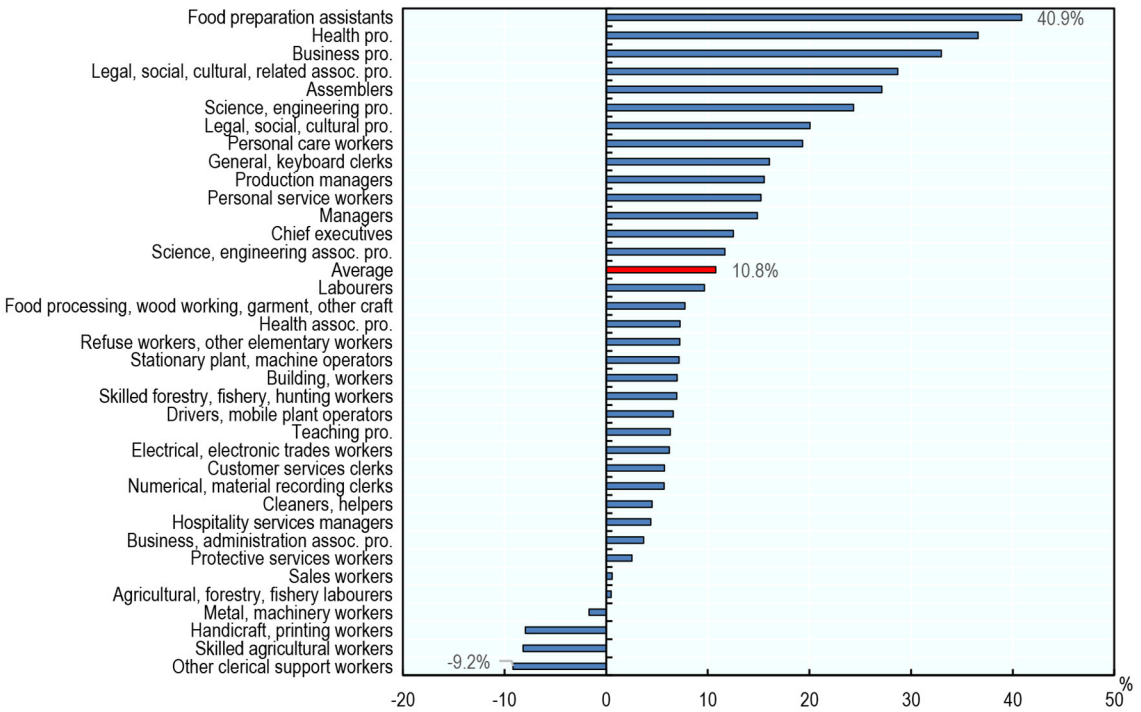


FIGURE 10 | Employment has grown in most occupations between 2012 and 2019. Average percentage change in employment level across countries by occupation, 2012–2019. Occupations are classified using two-digit ISCO-08. The averages presented are unweighted averages across the 23 countries analyzed. Source: ENOE, EU-LFS, and US-CPS.

and the United States[43]. AI-related technical skills are identified based on the list provided in Acemoglu et al. (2020)[44].

In the United States, the share of job postings requiring AI skills has increased in almost all occupations between 2012 and 2019 (**Figure 12**). Science and Engineering Professionals experienced the largest increase, but growth was also substantial for Managers, Chief Executives, Business and Administration Professionals, and Legal, Social, Cultural Professionals. That being said, the share of job postings that require AI skills remains very low overall, with an average across occupations of 0.24% in 2019 (against 0.10% in 2012). These orders of magnitude are in line with Acemoglu et al. (2020) and Squicciarini and Nachtigall (2021).

[43]United Kingdom and the United States are the only countries in the sample with 2012 Burning Glass Technologies data available, thereby allowing for the examination of trends over the past decade.

[44]Job postings that require AI-related technical skills are defined as those that include at least one keyword from the following list: Machine Learning, Computer Vision, Machine Vision, Deep Learning, Virtual Agents, Image Recognition, Natural Language Processing, Speech Recognition, Pattern Recognition, Object Recognition, Neural Networks, AI ChatBot, Supervised Learning, Text Mining, Support Vector Machines, Unsupervised Learning, Image Processing, Mahout, Recommender Systems, Support Vector Machines (SVM), Random Forests, Latent Semantic Analysis, Sentiment Analysis/Opinion Mining, Latent Dirichlet Allocation, Predictive Models, Kernel Methods, Keras, Gradient boosting, OpenCV, Xgboost, Libsvm, Word2Vec, Chatbot, Machine Translation, and Sentiment Classification.

RESULTS

This section looks at the link between an occupation's exposure to AI in 2012 and changes in employment, working hours, and the demand for AI-related technical skills between 2012 and 2019. Exposure to AI appears to be associated with greater employment growth in occupations where computer use is high, and larger reductions in hours worked in occupations where computer use is low. So, even though AI may substitute for workers in certain tasks, it also appears to create job opportunities in occupations that require digital skills. In addition, there is some evidence that greater exposure to AI is associated with greater increase in demand for AI-related technical skills (such as natural language processing, machine translation, or image recognition) in occupations where computer use is high. However, as the share of jobs requiring AI skills remains very small, this increase in jobs requiring AI skills cannot account for the additional employment growth observed in computer-intensive occupations that are exposed to AI.

Empirical Strategy

The analysis links changes in employment levels within occupations and across countries to AI exposure[45]. The

[45]The analysis is performed at the 2-digit level of the International Standard Classification of Occupations 2008 (ISCO-08).

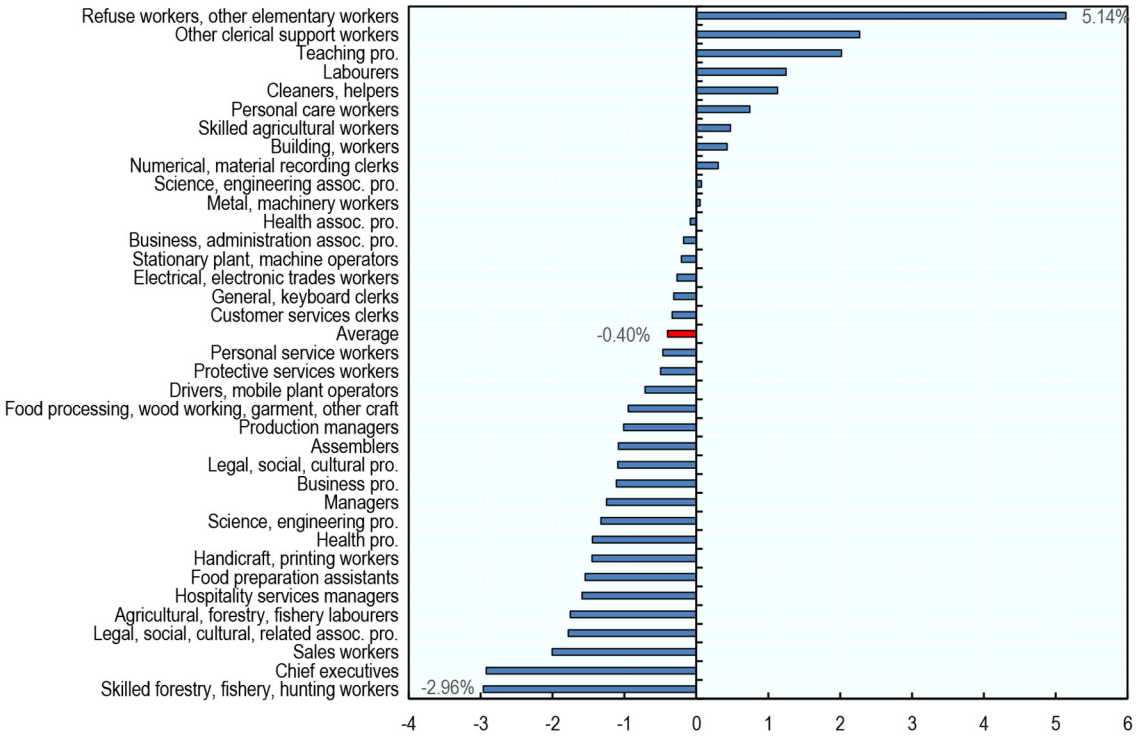


FIGURE 11 | Average usual working hours have decreased in most occupations between 2012 and 2019. Average percentage change in average usual weekly hours across countries by occupation, 2012–2019. Occupations are classified using two-digit ISCO-08. The averages presented are unweighted averages across the 22 countries analyzed (Mexico is excluded from the analysis of working time due to data availability). Usual weekly working hours by country-occupation cell are calculated by taking the average across individuals within that cell. Source: ENOE, EU-LFS, and US-CPS.

regression equation is the following:

$$Y_{ij} = \alpha_j + \beta\, AI_{ij} + \gamma\, X_{ij} + u_{ij} \quad (1)$$

where Y_{ij} is the percentage change in the number of workers (both dependent employees and self-employed) in occupation i in country j over the period 2012–2019[46]; AI_{ij} is the index of exposure to AI for occupation i in country j as measured in 2012; X_{ij} is a vector of controls including exposure to other technological advances (software and industrial robots), offshorability, exposure to international trade, and 1-digit occupational ISCO dummies; α_j are country fixed effects; and u_{ij} is the error term. The coefficient of interest β captures the link between exposure to AI and changes in employment. The inclusion of country fixed effects means that the analysis only exploits within-country variation in AI exposure to estimate the parameter of interest. The specifications that include 1-digit occupational dummies only exploit variation within broad occupational groups, thereby controlling for any factors that are constant across these groups.

To control for the effect of non-AI technologies, the analysis includes measures of exposure to software and industrial robots developed by Webb (2020) based on the overlap between the text of job descriptions provided in the O*NET database and the text of patents in the fields corresponding to each of these technologies[47]. Offshoring is proxied by an index of offshorability developed by Firpo et al. (2011) and made available by Autor and Dorn (2013), which measures the potential offshoring of job tasks using the average between the two variables "Face-to-Face Contact" and "On-Site Job" that Firpo et al. (2011) derive from the O*NET database[48]. This measure captures the extent to which an occupation requires direct interpersonal interaction or proximity to a specific work location[49].

[46]In a second step, Y_{ij} will stand for the percentage change in average weekly working hours and the percentage change in the share of part-time workers.

[47]To select software patents, Webb uses an algorithm developed by Bessen and Hunt (2007) which requires one of the keywords "software," "computer," or "programme" to be present, but none of the keywords "chip," "semiconductor," "bus," "circuity," or "circuitry." To select patents in the field of industrial robots, Webb develops an algorithm that results in the following search criteria: the title and abstract should include "robot" or "manipulate," and the patent should not fall within the categories: "medical or veterinary science; hygiene" or "physical or chemical processes or apparatus in general".

[48]They reverse the sign to measure offshorability instead of non-offshorability.

[49]Firpo et al. (2011) define "face-to-face contact" as the average value between the O*NET variables "face-to-face discussions," "establishing and maintaining interpersonal relationships," "assisting and caring for others," "performing for or working directly with the public", and "coaching and developing others." They define "on-site job" as the average between the O*NET variables "inspecting equipment, structures, or material," "handling and moving objects," "operating vehicles, mechanized devices, or equipment," and the mean of "repairing

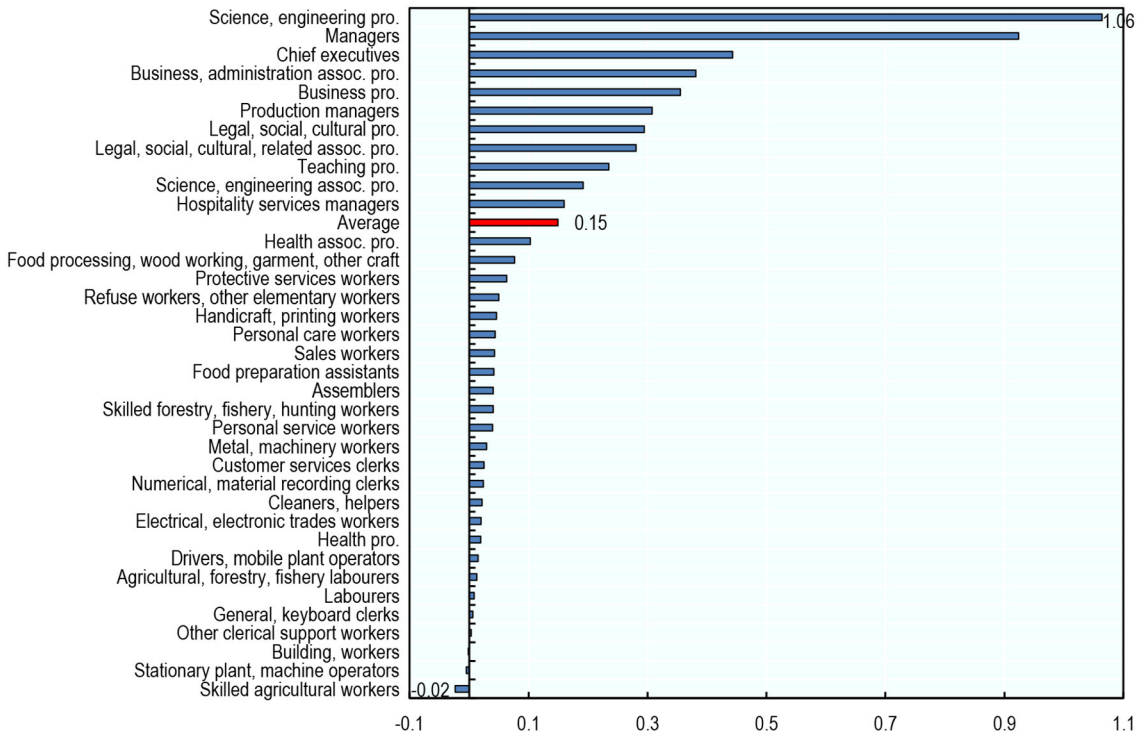


FIGURE 12 | Nearly all occupations have increasingly demanded AI skills between 2012 and 2019 in the United States. Percentage point* change in the share of job postings that require AI skills, 2012–2019, USA. The share of job postings that require AI skills in an occupation is the number of job postings requiring such skills in that occupation divided by the total number of job postings in that same occupation. *Percentage point changes are preferred over percentage changes because the share of job postings that require AI skills is equal to zero in some occupations in 2012. Source: Burning Glass Technologies.

The three above indices are occupation-level task-based measures derived from the O*NET database for the United States; this analysis uses those measures for all 23 countries, assuming that the cross-occupation distribution of these indicators is similar across countries[50]. Exposure to international trade is proxied by the share of employment within occupations that is in tradable sectors[51]. These shares are derived from the European Union Labor Force Survey (EU-LFS), the Mexican National Survey of Occupation and Employment (ENOE), the US Current Population Survey (US-CPS).

Exposure to AI and Employment: A Positive Relationship in Occupations Where Computer Use Is High

As discussed in Section Introduction, the effect of exposure to AI on employment is theoretically ambiguous. On the one hand, employment may fall as tasks are automated (substitution effect). On the other hand, productivity gains may increase labor demand (productivity effect) (Acemoglu and Restrepo, 2019a,b; Bessen, 2019; Lane and Saint-Martin, 2021)[52]. The labor market impact of AI on a given occupation is likely to depend on the task composition of that occupation—the prevalence of high-value added tasks that AI cannot automate (e.g., tasks that require creativity or social intelligence) or the extent to which the occupation already uses other digital technologies [since AI applications are often similar to software in their use, workers with digital skills may find it easier to use AI effectively (Felten et al., 2019)]. Therefore, the following analysis will not only look at the entire sample of occupation-country cells, but will also split the sample according to what people do in these occupations and countries.

In particular, the level of computer use within an occupation is proxied by the share of workers reporting the use of a

and maintaining mechanical equipment" and "repairing and maintaining electronic equipment".

[50]All three indices are available by occupation based on U.S. Census occupation codes. They were first mapped to the SOC 2010 6-digits classification and then to the ISCO-08 4-digit classification. They were finally aggregated at the 2-digit level using average scores weighted by the number of full-time equivalent employees in each occupation in the United-States, as provided by Webb (2020) and based on American Community Survey 2010 data.

[51]The tradable sectors considered are agriculture, industry, and financial and insurance activities.

[52]Partial worker substitution in an occupation may increase worker productivity and employment in the same occupation, but also in other occupations and sectors (Autor and Salomons, 2018). These AI-induced productivity effects are relevant to the present cross-occupation analysis to the extent that they predominantly affect the same occupation where AI substitutes for workers. For example, although AI translation algorithms may substitute for part of the work of translators, they may increase the demand for translators by significantly reducing translation costs.

computer at work in that occupation, calculated for each of the 23 countries in the sample. It is based on individuals' answers to the question "Do you use a computer in your job?," taken from the Survey of Adult Skills (PIAAC). Occupation-country cells are then classified into three categories of computer use (low, medium, and high), where the terciles are calculated based on the full sample of occupation-country cells[53]. Another classification used is the country-invariant classification developed by Goos et al. (2014), which classifies occupations based on their average wage relying on European Community Household Panel (ECHP) data. For example, occupations with an average wage in the middle of the occupation-wage distribution would be classified in the middle with respect to this classification[54]. Finally, the prevalence of creative and social tasks is derived from PIAAC data. PIAAC data include the frequency with which a number of tasks are performed at the individual level. Respondents' self-assessment are based on a 5-point scale ranging from "Never" to "Every day." This information is used to measure the average frequency with which workers in each occupation perform creative or social tasks, and this is done separately for each country[55].

While employment grew faster in occupations more exposed to AI, this relationship is not robust. There is stronger evidence that AI exposure is positively related to employment growth in occupations where computer use is high. **Table 2** displays the results of regression equation (1) without controls. When looking at the entire sample, the coefficient on AI exposure is both positive and statistically significant (Column 1), but the coefficient is no longer statistically significant as soon as any of the controls described in Section Empirical Strategy are included (with the exception of offshorability)[56]. When the sample is split by level of computer use (low, medium, high), the coefficient on AI exposure remains positive and statistically significant only for the subsample where computer use is high (Columns 2–4). It remains so after successive inclusion of controls for international trade (i.e., shares of workers in tradable sectors), offshorability, exposure to other technological advances (software and industrial robots) and 1-digit occupational dummies (**Table 3**)[57]. In occupations where computer use is high, a one standard deviation increase in AI exposure is associated with 5.7 percentage points higher employment growth (**Table 2**, Column 4)[58].

TABLE 2 | Exposure to AI is positively associated with employment growth in occupations where computer use is high.

	(1) All occupations	(2) Low computer use	(3) Medium computer use	(4) High computer use
Exposure to AI	13.3**	−3.7	8.3	85.7**
	(6.4)	(13.2)	(18.4)	(36.5)
Country FEs	Yes	Yes	Yes	Yes
Observations	822	274	274	274
R-squared	0.058	0.127	0.172	0.098

Dependent variable: 2012–2019% change in employment level.
Robust standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Each observation is a country-occupation cell. Each column shows the results of regression equation (1) applied to one of the subsamples obtained by splitting the overall sample by level of computer use. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. Source: Authors' calculations using data from ENOE, EU-LFS, US-CPS, PIAAC, and Felten et al. (2019).

By contrast, the average wage level of the occupation or the prevalence of creative or social tasks matter little in the link between exposure to AI and employment growth. **Table A A.1** in Appendix shows the results obtained when replicating the analysis on the subsamples obtained by splitting the overall sample by average wage level, prevalence of creative tasks, or prevalence of social tasks. All coefficients on exposure to AI remain positive, but are weakly statistically significant and of lower magnitude than those obtained on the subsample of occupations where computer use is high (**Table 3**).

As a robustness check, **Table A A.2** in the Appendix replicates the analysis in **Table 2** using the score of exposure to AI obtained when using O*NET scores of "prevalence" and "importance" of abilities within occupations instead of PIAAC-based measures. The results remain unchanged. **Table A A.3** replicates the analysis using the alternative indicators of exposure to AI constructed by Webb (2020) and Tolan et al. (2021), described in Section What Do These Indicators Measure?[59] While the Webb (2020) indicator confirms the positive relationship between employment growth and exposure to AI in occupations where computer use is high, the coefficient obtained with the

[53] Data are from 2012, with the exception of Hungary (2017), Lithuania (2014), and Mexico (2017).

[54] Low-skill occupations include the ISCO-08 1-digit occupation groups: Services and Sales Workers; and Elementary Occupations. Middle-skill occupations include the groups: Clerical Support Workers; Skilled Agricultural, Forestry, and Fishery Workers; Craft and Related Trades Workers; and Plant and Machine Operators and Assemblers. High-skill occupations include: Managers; Professionals, and Technicians; and Associate Professionals.

[55] In line with Nedelkoska and Quintini (2018), creative tasks include: problem solving—simple problems, and problem solving—complex problems; and social tasks include: teaching, advising, planning for others, communicating, negotiating, influencing, and selling. For each measure, occupation-country cells are then classified into three categories depending on the average frequency with which these tasks are performed (low, medium, and high). These three categories are calculated by applying terciles across the full sample of occupation-country cells. Data are from 2012, with the exception of Hungary (2017), Lithuania (2014), and Mexico (2017).

[56] These results are not displayed but are available on request.

[57] **Tables 2**, **3** correspond to unweighted regressions, but the results hold when each observation is weighted by the inverse of the number of country observations in the subsample considered, so that each country has the same weight. These results are not displayed but are available on request.

[58] The standard deviation of exposure to AI is 0.067 among high computer use occupations. Multiplying this by the coefficient in Column 4 gives 0.067*85.73 = 5.74.

[59] The Webb (2020) indicator is available by occupation based on U.S. Census occupation codes. It was first mapped to the SOC 2010 6-digits classification and then to the ISCO-08 4-digit classification. It was finally aggregated at the 2-digit level by using average scores weighted by the number of full-time equivalent employees in each occupation in the United States, as provided by Webb (2020) and based on American Community Survey 2010 data. The Tolan et al. (2021) indicator is available at the ISCO-08 3-digit level and was aggregated at the 2-digit level by taking average scores.

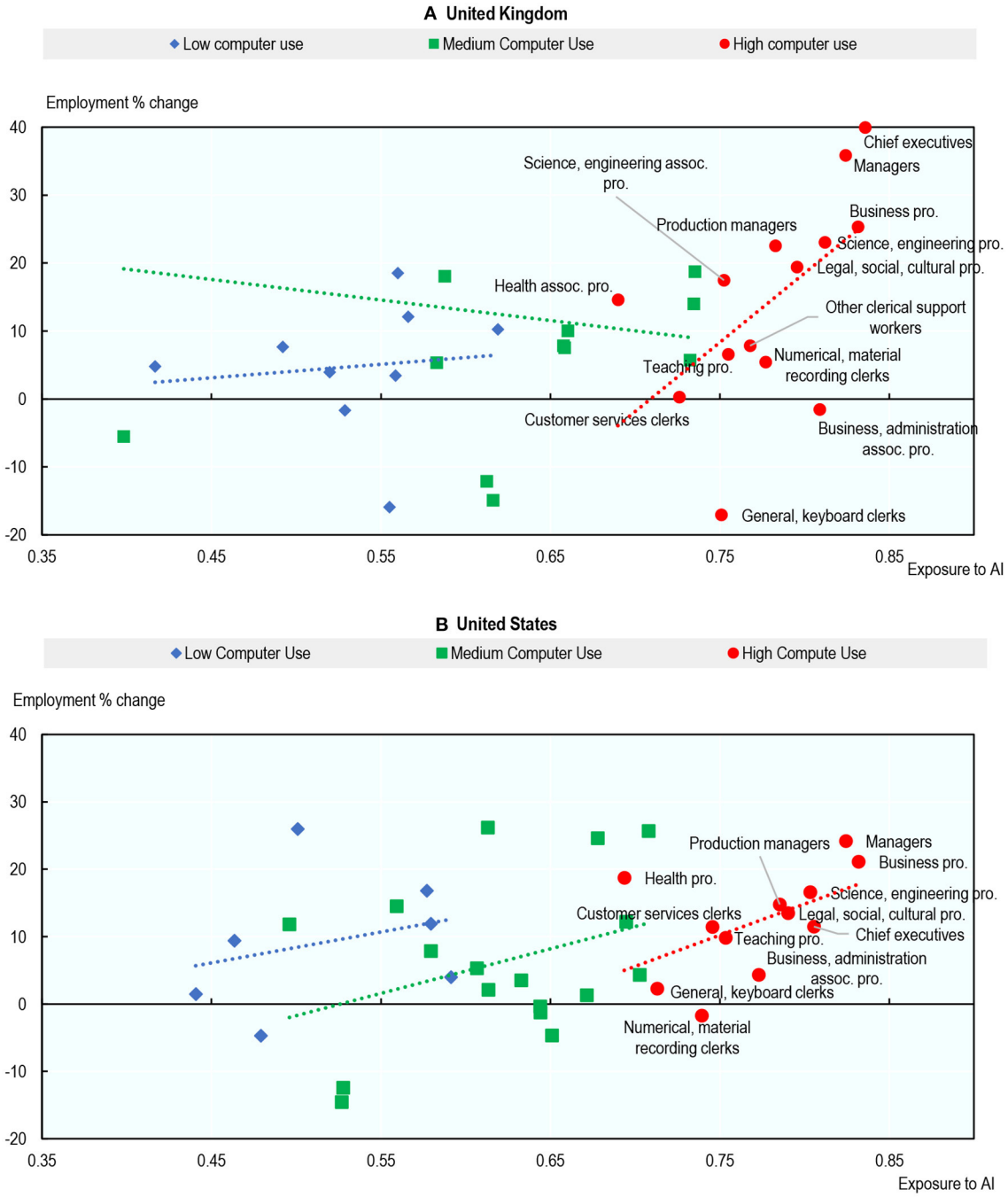


FIGURE 13 | Exposure to AI is associated with higher employment growth in occupations where computer use is high. Percentage change in employment level (2012–2019) and exposure to AI (2012). Occupations are classified using two-digit ISCO-08. Not all occupations have marker labels due to space constraints. Skilled forestry, fishery, hunting workers excluded from **(A)** for readability reasons. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. Source: Authors' calculations using data from EU-LFS, US-CPS, PIAAC, and Felten et al. (2019). **(A)** United Kingdom and **(B)** United States.

Tolan et al. (2021) indicator is positive but not statistically significant. This could be due to the fact that the Tolan et al. (2021) indicator reflects different aspects of AI advances, as it focuses more on cognitive abilities and is based on research intensity rather than on measures of progress in AI applications.

TABLE 3 | The relationship between exposure to AI and employment growth is robust to the inclusion of a number of controls.

	(1)	(2)	(3)	(4)	(5)
High computer use occupations					
Exposure to AI	85.7**	94.4***	137.7***	135.4***	144.6**
	(36.5)	(34.7)	(36.5)	(40.6)	(62.6)
Share of tradable sectors		−0.143	−0.0120	−0.00931	0.157
		(0.151)	(0.145)	(0.166)	(0.256)
Offshorability			−7.4**	−7.4***	−9.7**
			(2.9)	(2.8)	(4.6)
Exposure to softwares				0.0103	0.00429
				(0.190)	(0.253)
Exposure to robots				−0.0241	0.258
				(0.280)	(0.341)
1-digit occupation FEs	No	No	No	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes
Observations	274	274	274	274	274
R-squared	0.098	0.101	0.127	0.127	0.173

*Dependent variable: 2012–2019 % change in employment level. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each observation is a country-occupation cell. The sample is restricted to occupations with high computer use. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. Offshorability is an occupation-level measure from Autor and Dorn (2013) based on data from the United States. Exposure to software and exposure to robots are occupation-level measures developed by Webb (2020) based on data from the United States. The share of tradable sector represents the 2012 share of workers in the country-occupation cell working in agriculture, industry, and financial and insurance activities. Source: Authors' calculations using data from ENOE, EU-LFS, US-CPS, PIAAC, Autor and Dorn (2013), Felten et al. (2019), and Webb (2020).*

The examples of the United Kingdom and the United States illustrate these findings clearly[60]. **Figure 13** shows the percentage change in employment from 2012 to 2019 for each occupation against that occupation's exposure to AI in 2012, both in the United Kingdom (**Figure 13A**) and the United States (**Figure 13B**). Occupations are classified according to their level of computer use. The relationship between exposure to AI and employment growth within computer use groups is generally positive, but the correlation is stronger in occupations where computer use is high. For occupations with high computer use, the most exposed occupations tend to have experienced higher employment growth between 2012 and 2019: Business Professionals; Legal, Social and Cultural Professionals; Managers; and Science & Engineering Professionals. AI applications relevant to these occupations include: identifying investment opportunities, optimizing production in manufacturing plants, identifying problems on assembly lines, analyzing and filtering recorded job interviews, and translation. In contrast, high computer-use occupations with low or negative employment growth were occupations with relatively low exposure to AI, such as clerical workers and teaching professionals.

While further research is needed to test the causal nature of these patterns and to identify the exact mechanism behind them, it is possible that a high level of digital skills (as proxied by computer use) indicates a greater ability of workers to adapt to and use new technologies at work and, hence, to reap the benefits that these technologies bring. If AI allows these workers to interact with AI and to substantially increase their productivity and/or the quality of their output, this may, under certain conditions, lead to an increase in demand for their labor[61].

Exposure to AI and Working Time: A Negative Relationship Among Occupations Where Computer Use Is Low

This subsection extends the analysis by shifting the focus from the number of working individuals (extensive margin of employment) to how much these individuals work (intensive margin).

In general, the higher the level of exposure to AI in an occupation, the greater the drop in average hours worked over the period 2012–2019; and this relationship is particularly marked in occupations where computer use is low. Column (1) of **Table 4** presents the results of regression equation (1) using the percentage change in average usual weekly working hours as the variable of interest. The statistically significant and negative coefficient on exposure to AI highlights a negative relationship across the entire sample. Splitting the sample by computer use category shows that this relationship is stronger among occupations with lower computer use (Column 2–4). The size of the coefficients in Column 2 indicates that, within countries and across occupations with low computer use, a one standard deviation increase in exposure to AI is associated with a 0.60 percentage point greater drop in usual weekly working hours[62] (equivalent to 13 min per week)[63]. Columns 1–4 of **Table 5** show that the result is robust to the successive inclusion of controls for international trade, offshorability, and exposure to other technologies. However, the coefficient on exposure to AI loses statistical significance when controlling for 1 digit occupational dummies (**Table 5**, Column 5), which could stem from attenuation bias, as measurement errors may be significant relative to the variation in actual exposure within the 1 digit occupation groups[64].

The relationship between exposure to AI and the drop in average hours worked was driven by part-time employment[65].

[60] Although statistically significant on aggregate, the relationships between employment growth and exposure to AI suggested by **Table 2** are not visible for some countries.

[61] For productivity-enhancing technologies to have a positive effect on product and labour demand, product demand needs to be price elastic (Bessen, 2019).

[62] The standard deviation of exposure to AI is 0.125 among low computer use occupations. Multiplying this by the coefficient in Column 2 gives $0.125*(-4.823) = -0.60$.

[63] Estimated at the average working hours among low computer use occupations (37.2 h).

[64] **Tables 4**, **5** correspond to unweighted regressions, but most of the results hold when each observation is weighted by the inverse of the number of country observations in the subsample considered, so that each country has the same weight. These results are not displayed but are available on request.

[65] Part-time workers are defined as workers usually working 30 hours or less per week in their main job.

TABLE 4 | Exposure to AI is negatively associated with the growth in average working hours in occupations where computer use is low.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: 2012–2019% change in working hours				Dependent variable: 2012–2019% change in part-time employment			
	All occupations	Low computer use	Medium computer use	High computer use	All occupations	Low computer use	Medium computer use	High computer use
Exposure to AI	−2.7***	−4.8**	−4.1	−3.2	14.9	56.6**	−37.6	2.4
	(0.9)	(2.3)	(3.1)	(3.1)	(10.0)	(24.7)	(94.1)	(53.7)
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	781	252	261	268	781	252	261	268
R-squared	0.143	0.133	0.209	0.304	0.143	0.206	0.193	0.211

*Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each observation is a country-occupation cell. Each column shows the results of regression equation (1) applied to one of the subsamples obtained by splitting the overall sample by level of computer use. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. In columns 1–4, the dependent variable is the percentage change in average usual weekly working hours. In columns 5–8, the dependent variable is the percentage change in the share of part-time workers. Mexico is excluded from the analysis of working time due to data availability. Source: Authors' calculations using data from EU-LFS, US-CPS, PIAAC, and Felten et al. (2019).*

TABLE 5 | The relationship between exposure to AI and growth in average working hours is robust to the inclusion of a number of controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: 2012–2019% change in working hours					Dependent variable: 2012–2019% change in part-time employment				
	Low computer use occupations					Low computer use occupations				
Exposure to AI	−4.8**	−4.9**	−9.2***	−9.2**	−7.2	56.6**	56.6**	49.4**	53.0	23.5
	(2.3)	(2.3)	(3.2)	(4.0)	(4.6)	(24.7)	(24.7)	(24.5)	(35.3)	(41.7)
Share of tradable sectors		−0.0148	−0.0194*	−0.0267**	−0.0222		0.0135	0.00582	−0.00142	−0.0721
		(0.0111)	(0.0116)	(0.0133)	(0.0176)		(0.113)	(0.124)	(0.139)	(0.167)
Offshorability			−1.4**	−0.970	−0.887			−2.4	−1.6	−2.9
			(0.7)	(0.8)	(0.9)			(8.4)	(11.9)	(12.8)
Exposure to softwares				0.0289	0.0350				0.0358	−0.0567
				(0.0263)	(0.0300)				(0.314)	(0.376)
Exposure to robots				−0.0270	−0.0364				0.0151	0.00943
				(0.0385)	(0.0619)				(0.447)	(0.744)
1-digit occupation FEs	No	No	No	No	Yes	No	No	No	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	252	252	252	252	252	252	252	252	252	252
R-squared	0.133	0.141	0.157	0.161	0.166	0.206	0.206	0.207	0.207	0.214

*Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each observation is a country-occupation cell. The sample is restricted to occupations with low computer use. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. In columns 1–4, the dependent variable is the percentage change in average usual weekly working hours. In columns 5–8, the dependent variable is the percentage change in the share of part-time workers. Offshorability is an occupation-level measure from Autor and Dorn (2013) based on data from the United States. Exposure to software and exposure to robots are occupation-level measures developed by Webb (2020) based on data from the United States. The share of tradable sector represents the 2012 share of workers in the country-occupation cell working in: agriculture, industry, and financial and insurance activities. Mexico is excluded from the analysis of working time due to data availability. Source: Authors' calculations using data from EU-LFS, US-CPS, PIAAC, Autor and Dorn (2013), Felten et al. (2019), and Webb (2020).*

Columns 5–8 of **Table 4** replicate the analysis in Columns 1–4 using the change in the occupation-level share of part-time workers as the variable of interest. The results are consistent with those in columns 2–4: the coefficient on exposure to AI is positive and statistically significant only for the subsample of occupations where computer use is low (Columns 6–8). The coefficient remains statistically significant and positive when controlling for international trade and offshorability, but loses statistical significance when controlling for exposure to other technological advances and 1-digit occupational dummies (**Table 5**, columns 6–10)[66]. The results hold when replacing the share of part-time workers with the share of involuntary part-time workers[67] (**Table A A.7**), suggesting that the additional decline in working hours among low computer use occupations that are exposed to AI is not a voluntary choice by workers.

The examples of Germany and Spain provide a good illustration of these results[68]. **Figure 14** shows the percentage change in average usual weekly working hours from 2012 to 2019 for each occupation against that occupation's exposure to AI, both in Germany (**Figure 14A**) and in Spain (**Figure 14B**). As before, occupations are classified according to their degree of computer use (low, medium, high). In both countries, there is a clear negative relationship between exposure to AI and the change in working hours among occupations where computer use is low. In particular, within the low computer use category, most occupations with negative growth in working hours are relatively exposed to AI. These occupations include: Drivers and Mobile Plant Operators, Personal Service Workers, and Skilled Agricultural Workers. AI applications relevant to these occupations include route optimisation for drivers, personalized chatbots and demand forecasting in the tourism industry[69], or the use of computer vision in the agricultural sector to identify plants that need special attention. By contrast, low computer use occupations with the strongest growth in working hours are generally less exposed to AI. This is for example the case for Laborers (which includes laborers in transport and storage, manufacturing, or mining and construction).

Again, while further research is required, a lack of digital skills may mean that workers are not able to interact efficiently with AI and thus cannot reap all potential benefits of the technology. The substitution effect of AI in those occupations therefore appears to outweigh the productivity effect, resulting in reduced working hours, possibly as a result of more involuntary part-time employment. However, these results remain suggestive, as they are not robust to the inclusion of the full set of controls and the use of alternative indicators of exposure to AI.

Exposure to AI and Demand for AI-Related Technical Skills: A Weak but Positive Relationship Among Occupations Where Computer Use Is High

Beyond its effects on employment, AI may also transform occupations as workers are increasingly expected to interact with the technology. This may result in a higher demand for AI-related technical skills in affected occupations, although it is not necessarily the case that working with AI requires technical AI skills.

Indeed, exposure to AI is positively associated with the growth in the demand for AI technical skills, especially in occupations where computer use is high. **Figure 15** shows the correlation between the growth in the share of job postings that require AI skills from 2012 to 2019 within occupations and occupation-level exposure to AI for the United Kingdom (**Figure 15A**) and the United States (**Figure 15B**), the only countries in the sample with BGT time series available. Occupations are again classified according to their computer use. There is a positive correlation between the growth in the share of job postings requiring AI skills and the AI exposure measure, particularly among occupations where computer use is high. The most exposed of these occupations (Science and Engineering Professionals; Managers; Chief Executives; Business and Administration Professionals; Legal, Social, Cultural professionals) are also experiencing the largest increases in job postings requiring AI skills.

However, the increase in jobs requiring AI skills cannot account for the additional employment growth observed in computer-intensive occupations that are exposed to AI (despite the similarities between the patterns displayed in **Figures 13**, **15**). As highlighted by the different scales in those two charts, the order of magnitude of the correlation between exposure to AI and the percentage change in employment (**Figure 13**) is more than ten times that of the correlation between exposure to AI and the percentage point change in the share of job postings requiring AI skills (**Figure 15**)[70]. This is because job

[66] As an additional robustness exercise, **Table A A.4** in the Appendix replicates the analysis using the score of exposure to AI obtained when using O*NET scores of "prevalence" and "importance" of abilities within occupations instead of PIAAC-based measures. The results remain qualitatively unchanged, but the coefficients on exposure to AI are no longer statistically significant on the subsample of occupations where computer use is low, when using working hours as the variable of interest. **Tables A A.5, A.6** replicate the analysis using the alternative indicators of exposure to AI constructed by Webb (2020) and Tolan et al. (2021). When using the Webb (2020) indicator, the results hold on the entire sample but are not robust on the subsample of occupations where computer use is low. Using the Tolan et al. (2021) indicator, the results by subgroups hold qualitatively but the coefficients are not statistically significant.

[67] Involuntary part-time workers are defined as part-time workers (i.e., workers working 30 h or less per week) who report either that they could not find a full-time job or that they would like to work more hours.

[68] Although statistically significant on aggregate, the relationships between the percentage change in average usual weekly working hours and exposure to AI suggested by **Table 4** are not visible for some countries.

[69] For example, personalised chatbots can partially substitute for travel attendants. Demand forecasting algorithms may facilitate the operation of hotels, including the work of housekeeping supervisors. Travel Attendants and Housekeeping Supervisors both fall into the Personal Service Workers category.

[70] The results of the regression equation (1) on the subsample (of only 26 observations) of high computer use occupations in the United Kingdom and the United States give a coefficient on exposure to AI equal to 151.4 when using percentage employment growth as the variable of interest, which is about forty times greater than the 4.1 obtained when using percentage point change in the share of job postings that require AI skills as the variable of interest.

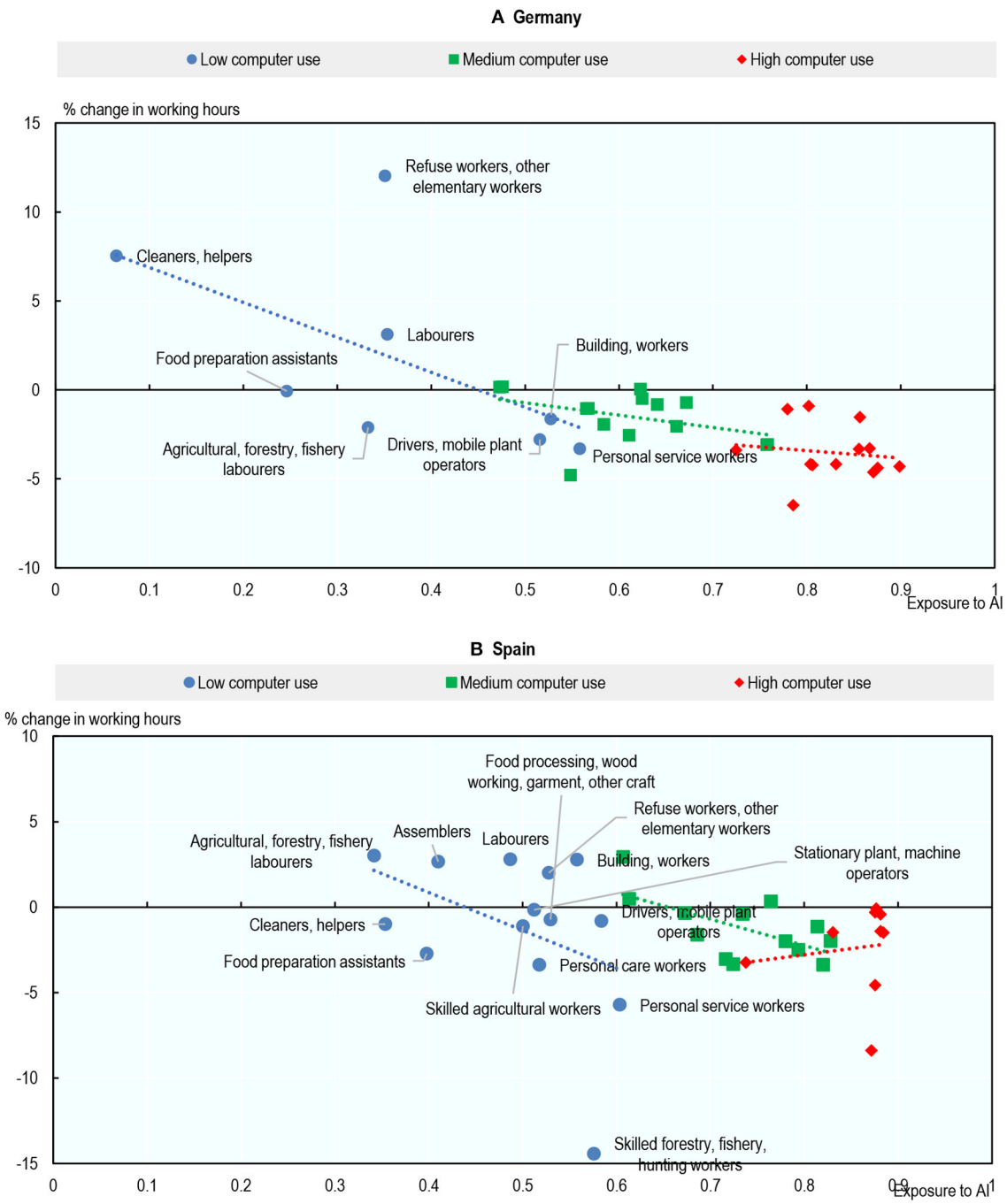


FIGURE 14 | In occupations where computer use is low, exposure to AI is negatively associated with the growth in average working hours. Percentage change in average usual working hour (2012–2019) and exposure to AI (2012). Occupations are classified using two-digit ISCO-08. Not all occupations have marker labels due to space constraints. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. Source: Author' calculations using data from EU-LFS, PIAAC, and Felten et al. (2019). **(A)** Germany and **(B)** Spain.

postings requiring AI skills remain a very small share of overall job postings. In 2019, on average across the 36 occupations analyzed, job postings that require AI skills accounted for only 0.14% of overall postings in the United Kingdom and 0.24% in the United States. By contrast, across the same 36 occupations, employment grew by 8.82% on average in the United States and 11.15% in the United Kingdom between 2012 and 2019.

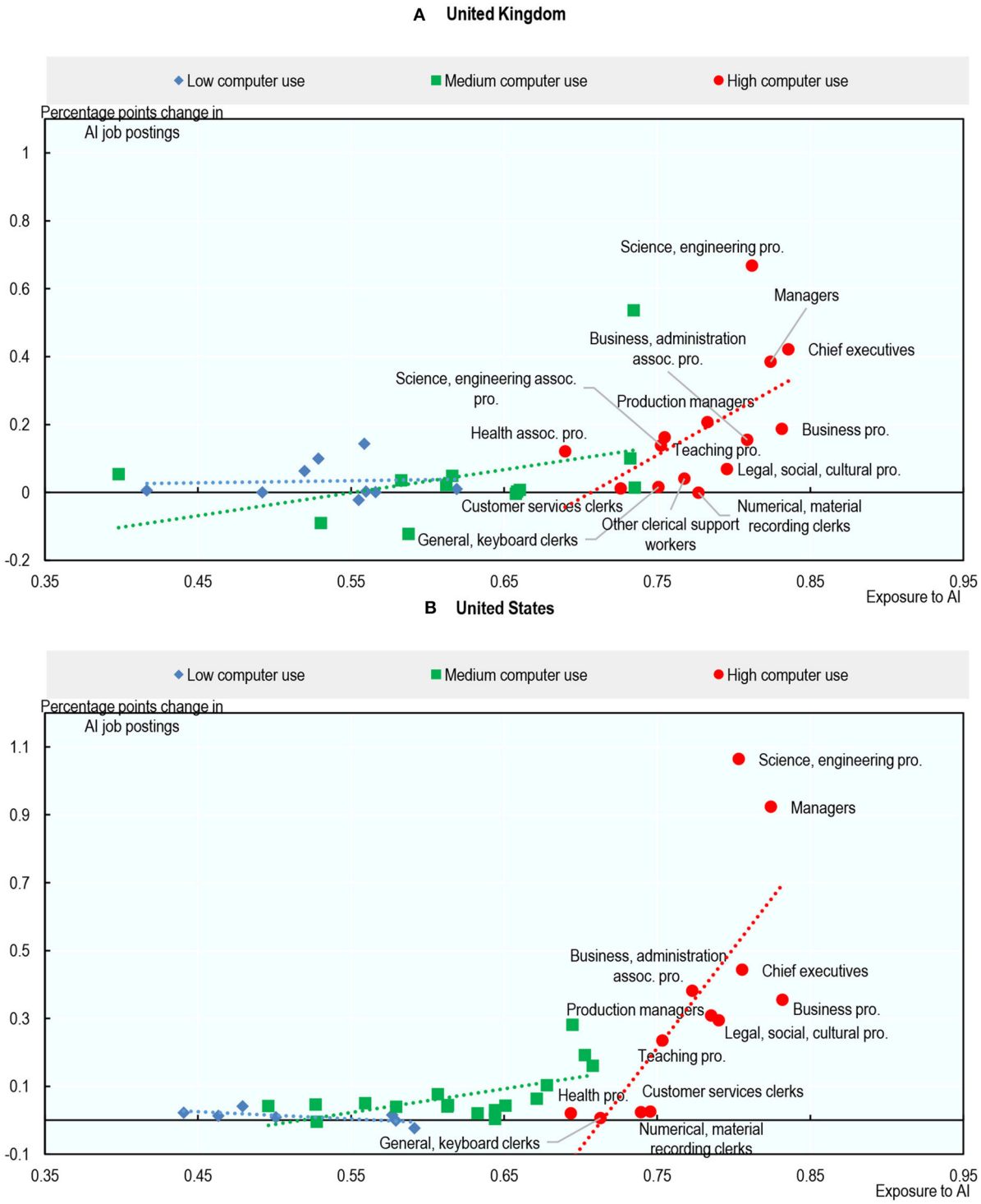


FIGURE 15 | High computer use occupations with higher exposure to AI saw a higher increase in their share of job postings that require AI skills. Percentage point change in the share of job postings that require AI skills (2012–2019) and exposure to AI (2012). The share of job postings that require AI skills in an occupation is taken as a share of the total number of job postings in that occupation. Occupation-country cells are classified into low, medium or high computer use by tercile of computer use applied across the full sample of occupation-country cells. Source: Author' calculations using data from Burning Glass Technologies, PIAAC, and Felten et al. (2019). **(A)** United Kingdom and **(B)** United States.

CONCLUSION

Recent years have seen impressive advances in artificial intelligence (AI) and this has stoked renewed concern about the impact of technological progress on the labor market, including on worker displacement.

This paper looks at the possible links between AI and employment in a cross-country context. It adapts the *AI occupational impact measure* developed by Felten et al. (2018, 2019)—an indicator measuring the degree to which occupations rely on abilities in which AI has made the most progress—and extends it to 23 OECD countries. The indicator, which allows for variations in AI exposure across occupations, as well as within occupations and across countries, is then matched to Labor Force Surveys, to analyse the relationship with employment.

Over the period 2012–2019, employment grew in nearly all occupations analyzed. Overall, there appears to be no clear relationship between AI exposure and employment growth. However, in occupations where computer use is high, greater exposure to AI is linked to higher employment growth. The paper also finds suggestive evidence of a negative relationship between AI exposure and growth in average hours worked among occupations where computer use is low.

While further research is needed to identify the exact mechanisms driving these results, one possible explanation is that partial automation by AI increases productivity directly as well as by shifting the task composition of occupations toward higher value-added tasks. This increase in labor productivity and output counteracts the direct displacement effect of automation through AI for workers with good digital skills, who may find it easier to use AI effectively and shift to non-automatable, higher-value added tasks within their occupations. The opposite could be true for workers with poor digital skills, who may not be able to interact efficiently with AI and thus reap all potential benefits of the technology.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://www.oecd.org/skills/piaac/data/.

AUTHOR CONTRIBUTIONS

Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Special thanks must go to Stijn Broecke for his supervision of the project and to Mark Keese for his guidance and support throughout the project. The report also benefitted from helpful comments provided by colleagues from the Directorate for Employment, Labour and Social Affairs (Andrew Green, Marguerita Lane, Luca Marcolin, and Stefan Thewissen) and from the Directorate for Science, Technology and Innovation (Lea Samek). Thanks to Katerina Kodlova for providing publication support. The comments and feedback received from participants in the February 2021 OECD Expert Meeting on AI indicators (Nik Dawson, Joe Hazell, Manav Raj, Robert Seamans, Alina Sorgner, and Songul Tolan) and the March 2021 OECD Future of Work Seminar are also gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2022.832736/full#supplementary-material

REFERENCES

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2020). *AI and Jobs: Evidence from Online Vacancies*. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w28257

Acemoglu, D., and Restrepo, P. (2019a). The wrong kind of AI? Artificial intelligence and the future of labour demand. *Cambridge J. Reg. Econ. Soc.* 13, 25–35. doi: 10.1093/cjres/rsz022

Acemoglu, D., and Restrepo, P. (2019b). Automation and new tasks: How technology displaces and reinstates labor. *J. Econom. Pers.* 33, 3–30. Available online at: https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.33.2.3

Acemoglu, D., and Restrepo, P. (2020). Robots and jobs: evidence from us labor markets. *J. Polit. Econ.* 128, 2188–2244. doi: 10.1086/705716

Agrawal, A., Gans, J., and Goldfarb, A (eds.). (2019). *8. Artificial Intelligence, Automation, and Work*. University of Chicago Press.

Autor, D., and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *Am. Econ. Rev.* 103, 1553–1597. doi: 10.1257/aer.103.5.1553

Autor, D., Levy, F., and Murnane, R. (2003). The skill content of recent technological change: an empirical exploration. *Q. J. Econ.* 118, 1279–1333. doi: 10.1162/003355303322552801

Autor, D., and Salomons, A. (2018). *Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share*. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w24871

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., et al. (2020). *Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible*. Available online at: https://www.oecd-ilibrary.org/science-and-technology/identifying-and-measuring-developments-in-artificial-intelligence_5f65ff7e-en

Bessen, J. (2016). *How Computer Automation Affects Occupations: Technology, Jobs, and Skills. Boston University School of Law, Law and Economics Research Paper 15-49.* Available online at: https://scholarship.law.bu.edu/faculty_scholarship/813

Bessen, J. (2019). Automation and jobs: when technology boosts employment. *Econ. Policy* 34, 589–626. doi: 10.1093/epolic/eiaa001

Bessen, J., and Hunt, R. (2007). An empirical look at software patents. *J. Econ. Manage. Strat.* 16, 157–189. doi: 10.1111/j.1530-9134.2007.00136.x

Brynjolfsson, E., and Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science* 358, 1530–1534. doi: 10.1126/science.aap8062

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). *Replication Data For*: What can machines learn and what does it mean for occupations and the economy? *AEA Pap*. Proc. 108, 43–47. doi: 10.1257/pandp.20181019

Cammeraat, E., and Squicciarini, M. (2020). *Assessing the Properties of Burning Glass Technologies' Data to Inform Use in Policy Relevant Analysis*. OECD.

Carnevale, A. P., Jayasundera, T., and Repnikov, D. (2014). *Understanding online job ads data*. Center on Education and the Workforce, Georgetown University, Washington, DC, United States. Available online at: https://cew.georgetown.edu/wp-content/uploads/2014/11/OCLM.Tech_.Web_.pdf

Dawson, N., Rizoiu, M., and Williams, M. (2021). Skill-driven recommendations for job transition pathways. *PLoS ONE* 16, e0254722. doi: 10.1371/journal.pone.0254722

Felten, E., Raj, M., and Seamans, R. (2018). A method to link advances in artificial intelligence to occupational abilities. *AEA Pap. Proc.* 108, 54–57. doi: 10.1257/pandp.20181021

Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: a novel dataset and its potential uses. *Strat. Manage. J.* 42, 2195–2217. doi: 10.1002/smj.3286

Felten, E. W., Raj, M., and Seamans, R. (2019). *The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization*. NYU Stern School of Business.

Fernández-Macías, E., and Bisello, M. (2020). *A Taxonomy of Tasks for Assessing the Impact of New technologies on Work (No. 2020/04)*. JRC Working Papers Series on Labour, Education and Technology.

Firpo, S., Fortin, N. M., and Lemieux, T. (2011). *Occupational tasks and changes in the wage structure*. Available online at: https://ftp.iza.org/dp5542.pdf

Fossen, F., and Sorgner, A. (2019). *New Digital Technologies and Heterogeneous Employment and Wage Dynamics in the United States: Evidence From Individual-Level Data*. IZA Discussion Paper 12242. Available online at: https://www.iza.org/publications/dp/12242/new-digital-technologies-and-heterogeneous-employment-and-wage-dynamics-in-the-united-states-evidence-from-individual-level-data

Gibbons, R., and Waldman, M. (2004). Task-specific human capital. *Am. Econ. Rev.* 94, 203–207. doi: 10.1257/0002828041301579

Gibbons, R., and Waldman, M. (2006). Enriching a theory of wage and promotion dynamics inside firms. *J. Lab. Econ.* 24, 59–107. doi: 10.1086/497819

Goos, M., Manning, A., and Salomons,A. (2014). Explaining job polarization: routine-biased technological change and offshoring. *Am. Econ. Rev.* 104, 2509–2526. doi: 10.1257/aer.104.8.2509

Grennan, J., and Michaely, R. (2017). *Artificial Intelligence and the Future of Work: Evidence From Analysts*. Available online at: https://conference.nber.org/conf_papers/f130049.pdf

Hernández-Orallo, J. (2017). *The Measure of all Minds: Evaluating Natural and Artificial Intelligence*. Cambridge University Press. Available online at: https://www.cambridge.org/core/books/measure-of-all-minds/DC3DFD0C1D5B3A3AD6F56CD6A397ABCA

Hershbein, B., and Kahn, L. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *Am. Econ. Rev.* 108, 1737–1772, doi: 10.1257/aer.20161570

Jin, X., and Waldman, M. (2019). Lateral moves, promotions, and task-specific human capital: theory and evidence. *J. Law Econ. Organ.* 36, 1–46. doi: 10.1093/jleo/ewz017

Lane, M., and Saint-Martin, A. (2021). *The Impact of Artificial Intelligence on the Labour Market: What Do We Know So Far?* OECD Social, Employment and Migration Working Papers, No. 256. Paris: OECD Publishing.

Nedelkoska, L., and Quintini, G. (2018). *Automation, Skills Use and Training. OECD Social, Employment and Migration Working Papers, No. 202*. Paris: OECD Publishing.

Nolan, A. (2021). *Making life easier, richer and healthier:* Robots, their future and the roles of public policy.

Qian, M., Saunders, A., and Ahrens, M. (2020). "Mapping legaltech adoption and skill demand," in *The Legal Tech Book: The Legal Technology Handbook for Investors, Entrepreneurs and FinTech Visionaries*, eds S. Chishti, S. A. Bhatti, A. Datoo, and D. Indjic (John Wiley and Sons), 211–214.

Raj, M., and Seamans, R. (2019). Primer on artificial intelligence and robotics. *J. Organ. Des.* 8, 11. doi: 10.1186/s41469-019-0050-0

Squicciarini, M., and Nachtigall, H. (2021). *Demand for AI skills in jobs: Evidence from online job postings, OECD Science, Technology and Industry Working Papers, No. 2021/03*. Paris: OECD Publishing. doi: 10.1787/3ed32d94-en

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., and Gómez, E. (2021). Measuring the occupational impact of AI: tasks, cognitive abilities and AI benchmarks. *J. Artif. Intell. Res.* 71, 191–236. doi: 10.1613/jair.1.12647

Webb, M. (2020). *The Impact of Artificial Intelligence on the Labor Market. Working Paper*. Standford University. Available online at: https://web.stanford.edu/

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Georgieff and Hyee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/frai.2022.867832

Inclusive Growth in the Era of Automation and AI: How Can Taxation Help?

*Rossana Merola**

International Labour Organization (ILO), Research Department, Geneva, Switzerland

In the last decades, the world economy is facing a massive rise in automation, robotics and Artificial Intelligence (AI) which, according to some analysts, could lead to significant job losses or job polarization and hence widen income and wealth disparities. This scenario may impede the achievement of the Sustainable Development Goal 8 (SDG 8). In this context, the role of government and regulation becomes crucial in order to prevent an undesirable scenario, where technological change, namely automation and AI, comes at the cost of mass unemployment and growing inequality. This paper focuses on the role of taxation as a possible tool for sharing the gains from automation and AI. Nowadays, advances in technology may have a direct impact on tax systems, which should be re-adapted to take into account new forms of jobs and new business models. The paper discusses pros and cons of several possible solutions and then compares progresses achieved in different countries. Concerning robot tax and digital taxes there are already some concrete steps undertaken both at national and international level, while other proposals remain still nebulous. Of course, taxation *per se*, and any single policy in general, is not sufficient to achieve a more inclusive and equal growth. It is instead crucial to create synergies across policies and a strong link between employment creation strategies, redistributive policies, skill development and social protection systems.

Keywords: robot tax, digital tax, automation, artificial intelligence, tax policy, inequality, technological unemployment

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Essex, United Kingdom

Reviewed by:
Idiano D'Adamo,
Sapienza University of Rome, Italy
Luigi Aldieri,
University of Salerno, Italy

***Correspondence:**
Rossana Merola
merola@ilo.org

Specialty section:
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

Received: *01 February 2022*
Accepted: *03 May 2022*
Published: *31 May 2022*

Citation:
Merola R (2022) Inclusive Growth in the Era of Automation and AI: How Can Taxation Help? Front. Artif. Intell. 5:867832. doi: 10.3389/frai.2022.867832

INTRODUCTION

In the last decades, the world economy has witnessed a massive process of automation, robotization and artificial intelligence (AI), which can already replace humans in a range of activities. Advanced robotics, machine learning and AI already find diverse applications, including digital assistants such as the Google Assistant or Siri, speech and image recognition, text translation and automatic text generation. More sophisticated applications include medical systems for diagnosis of pathologies (medtech), automated review of legal contracts (lawtech), self-driving cars, the detection of patterns in stock markets for successful trading (algorithmic trading) and the estimation of building's interior temperature (Villa and Sassanelli, 2020).

Many analysts are rising concerns about the risk that advances in robotization and AI may lead to significant job losses or job polarization and ultimately result in widening income and wealth disparities (Méda, 2016; Korinek and Stiglitz, 2017). Among these, Frey and Osborne (2017) find

that over the next 20 years technology may displace a large share of human workers, precisely 35% in the United Kingdom and 47% in the United States. In a report published in 2018,[1] the World Economic Forum warned that by 2025 more than 50% of current jobs will be automated. Jobs in Eastern and Southern Europe, Germany, Chili and Japan are more automatable than those in Anglo-Saxon and Nordic countries (Nedelkoska and Quintini, 2018). While some studies cast doubts on the job loss effect of technology in advanced economies, there is consensus on the effects in emerging economies which rely more on manufactory and are facing robot-driven reshoring (see Carbonero et al., 2018; De Backer et al., 2018). According to the World Bank (2016), the risk of job loss in developing countries is even higher than in advanced economies: 69% in India, 72% in Thailand, 77% in China, and a massive 85% in Ethiopia.[2]

Job losses due to automation are likely to widen inequality. According to the common view, automation is likely to penalize medium-skilled workers more than low- and high-skilled workers, as their tasks can be more easily replaced by AI and robots. Many commenters have hence argued that technological progress should not come at the expense of more vulnerable people and that solving inequity should be a priority for governments. However, decreasing labor income could create limitations for governments in the use of labor taxation as a tool for redistributing wealth, which further exacerbates inequality.

This scenario threatens the achievement of Sustainable Development Goal 8 (SDG 8) in the United Nations 2030 Agenda for Sustainable Development. SDG 8 exhorts the international community to "promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all".

It is evident that while technological progress certainly improves life quality, it may nevertheless produce serious social, economic and political harms if it remains unregulated (Acemoglu, 2021). In light of this context, the role of governments and regulation becomes crucial in order to prevent an undesirable scenario, where technological change comes at the cost of mass unemployment and growing inequality. Therefore, governments and enterprises should take steps to preserve competition and avoid monopolistic power, updating skills and redistribute profits.

This paper focuses only on the role of taxation as a possible tool for sharing the gains from automation and AI. The aim is to shed light on possible solutions, being aware that each of them presents strengths and weaknesses. Nowadays, technological progress is radically changing the society and may have a direct impact on tax systems, which should be re-adapted to take into account new forms of jobs and new business models. Section Challenges Arising From Robots and Artificial Intelligence summarizes the discussion in the policy debate on the possible effects of robots and AI on employment and inequality. The lack of agreement makes policy interventions even more relevant in order to minimize possible negative effects of technological change and to make sure that gains from robotization and AI are equally shared. Section Tax Solutions presents several tax policy solutions, discusses pros and cons of each of them and then compares progresses achieved in different countries and at international level. Concerning the robot tax and the digital tax, there are already some concrete steps undertaken by some governments in advanced economies, while other proposals remain more nebulous. Finally, Section Conclusions concludes.

Of course, taxation *per se*, and any single policy in general, is not sufficient to achieve a more inclusive and equal growth. It is instead crucial to create synergies across policies and a strong link between employment creation strategies, redistributive policies, skill development and social protection systems.

CHALLENGES ARISING FROM ROBOTS AND ARTIFICIAL INTELLIGENCE

The widespread adoption of AI poses several challenges, related to modalities of consumers' data collection which are often intrusive and not transparent, privacy protection and cyber-security in e-commerce (D'Adamo et al., 2021; Puntoni et al., 2021).

This section focuses on challenges for labor market and equity. Despite some afore-mentioned studies are warning that technological progress may cause job losses and widening inequality, so far there is no agreement in the literature on the effects of robotization and AI on employment and inequality.

According to some studies, employment effects specifically from adopting robots remain rather limited or are even positive at aggregate level. Among these, Dixon et al. (2021) compare employment and performance outcome between robot-adopting and non-adopting firms in Canada. They find that employment increases in robot-adopting firms, especially for low-skilled workers. Similar results are found by Acemoglu et al. (2020) for French manufacturing firms and by Koch et al. (2019) for Spanish firms. These studies also find an increase in performance (i.e., TFP or total revenues) in those firms adopting robots. Using data on US textile, steel and auto industries, Bessen (2017) argues that technological progress may at the same time be beneficial for some industries and hit some others. Ryan Avent, an editor and columnist for *The Economist*, points out that employment remains very high in many advanced countries, such as Germany and Japan, although they make an intense use of robots.

Looking more specifically at AI, the final effect on employment will be determined by the coexistence of three effects: task-substitution, task-complementarity and creation of new jobs. In the case of matching applications (e.g., Linkedin, Amazon), algorithms are already used to match supply and demand and hence easily replace human workers. In the case of classification/screening tasks, AI can assist workers but without substituting them. An example might be computer-assisted surgery which allows surgeons to perform surgical intervention remotely. In this case, there is no substitution, but a kind of "cobotisation", that is a co-working between humans and artificial intelligence, which can ultimately increase overall productivity. Finally, concerning process-management tasks,

[1]The Future of Jobs Report 2018 (World Economic Forum, 2018).

[2]For a wider discussion of the literature on job implications of AI, we refer to Ernst et al. (2019).

AI can perform tasks that human workers are not capable to perform. Moreover, the digital economy has created new types of jobs (e.g., AI-programmers, e-commerce specialists, apps and software developers, crowd-workers, influencers and those working on social media).

Keeping this in mind and considering the scarcity of data and difficulties in measuring the exposure to AI, it is difficult nowadays to predict the overall effect of AI and automation on jobs. The final effect will depend on which effect will dominate. Georgieff and Hyee (2021) find that task substitution dominates only for workers with low digital skills, while productivity effects dominate for workers with good digital skills. In addition, the final effect also depends on the adaptability of jobs in the digital transformation (Arntz et al., 2017). In this light, some studies share an optimistic view. Brynjolfsson et al. (2018) state that digitalization could also lead to reorganization of occupations rather than replacement. In a similar vein, Bessen (2017) argue that "automation might not cause mass unemployment, but it may well require workers to make disruptive transitions to new industries, requiring new skills and occupations".

Concerning the effects on inequality, new technologies in the last years have been associated to greater inequality and job polarization. Automation due to AI, robots and computers is likely to affect mostly middle-class jobs. Humans are already being replaced, partially or fully, in some tasks as legal services, accounting, logistics and retail. Displaced workers are likely to compete downwards, rather than falling into unemployment. This scenario suggests further job polarization in the next years. However, according to a recent study by Michael Webb, while robots and software may take over middle-skilled tasks, AI may perform high-skilled tasks and hence is expected to have the reverse effect on inequality, since better-educated and better-paid workers will be the most affected by the new AI-based technologies. Still, this study warned that while AI will reduce 90:10 wage inequality, it will not have an impact on the top 1% earners.

Inequality has increased not only across workers undertaking different tasks in the same firm, but also across firms. According to recent research conducted at the World Bank by Kelly et al. (2017), at least in Europe, the main driver of wage inequality is the wage gap across firms, which is determined by differences in the rate of adoption of digital technologies. As pointed out by Ernst (2019), in this era of AI, we are witnessing the emergence of a new business model, called "surveillance capitalism", which is based on collecting data without barriers to access and exploited with proprietary algorithms. While the data come free—and users are often all too willing to give up their privacy—data collection is not since it is protected by intellectual property rights. While on the one side the rise of new "big data" platforms, able to collect huge information on consumer behaviors and preferences, can certainly improve the efficiency in the economy, on the other side "big data" have encouraged the emergence of "superstar" firms which are outperforming compared to the other firms in the economy. These "superstars", mostly digital companies such as Facebook, Google, Amazon and Netflix, collect huge amounts of data which allow them to individualize prices and product offers and cumulate profits and wealth. "Superstar" firms are then able to gain market power and not surprisingly, concentrated winner-take-all markets are associated with the fall in the labor share (see Autor et al., 2017; Barkai, 2020).

These different forms of inequalities require different forms of tax interventions. We will discuss different alternative tax policies in the following sections. One of the main arguments in favor of a tax on robots is that it preserves low-skill jobs which are more likely to be automated. In this regard, a robot tax can address inequality caused by skill-biased technological change. Another option could be wage subsidies for low-skilled workers. However, inequality may also arise because the emergence of a new business model, called "surveillance capitalism", concentrates profits and wealth in the hands of few "superstars" firms, mostly digital companies. In this case, other types of taxes would be preferable, such as digital taxation, new tax on corporations' stock shares or the creation of sovereign funds. In particular, the latter two solution are deemed to be progressive since stock ownership is highly concentrated among the richest.

To achieve more inclusive and equal growth, taxation should go hand in hand with other type of policies. Digital businesses can easily collect a huge amount of data from their users. Governments and private businesses should acknowledge that users' data represent an incredibly valuable source of profits and take steps to ensure that markets remain contestable and competitive. On the one side, there are proposals to tax the income or rents generated by the exploitation of users' data.[3] On the other side, there are proposals to share data for free in order to guarantee market competition. In a recent article, Ekkehard Ernst discussed several solutions to address potential rise in inequality in the era of "surveillance capitalism".[4] Considering data as a common good which allows the extraction of rents would help restore the balance between individual data suppliers and corporate platform providers. Treating data ownership as a collective-action problem can limit the increases in concentration and market power and will ultimately help to address the continuous rise in inequality. Moreover, it is crucial that both governments and enterprises support the existing workforces through reskilling and upskilling. Governments should implement effective policies to facilitate the transition to the new world of work where humans will co-work with artificial intelligence, without leaving anybody behind. In this light, a necessary step is readapting the current education system to support the transit to new tasks required by AI-based technologies.

TAX SOLUTIONS

This section discusses different tax solutions which can ensure that gains from AI and technological changes are equally shared. Some proposals are already on track, while others remain more nebulous and/or limited to a few countries. Each proposal presents both strengths and weakness.

[3]For a discussion on alternative options for taxing profits and rents generated by the collection and the process of users' data, see Aslam and Shah (2020).

[4]I refer the reader to "Big Data and its enclosure of the commons", published in *Social Europe* on June 12, 2019.

Robot Tax

The most immediate solution, which has been strongly supported among others by Bill Gates, Elon Musk and Nobel Laureate in Economics Robert Shiller, is taxing robots. A robot tax stems from the idea that robot-adopting firms should pay a tax since they replace human workers with robots. There are several arguments in favor of robot tax. The first one is preserving human employment by introducing disincentives for firms from replacing humans with robots. Second, even though firms prefer replacing humans with robots, a robot tax would generate revenues for the government to cover the loss of revenues from payroll taxes and income tax.[5] A third argument in favor of the robot tax is allocation efficiency: robots do not pay neither payroll taxes, nor income taxes. Taxing robots improves the efficiency in the economy, because governments already tax labor, so they should also tax robots at the same rate to avoid distortion in the resource allocation. In most of advanced economies, and in particular in the United States, taxation favors AI and automation over human employment.[6] This may distort investment toward automation simply because companies benefit from tax windfalls and not because automation may increase profitability.[7] Finally, not taxing robots will increase income inequality, because of the decreasing share of national income going to labor.

Revenues from the robot tax can be redistributed as universal basic income or as transfers to workers displaced in their jobs by robotic systems and AI and not able to be relocated in new jobs. New York Mayor, Bill de Blasio, proposed to use revenues from the robot tax to create new jobs in green energy, health care and education.

There are also arguments against the robot tax. First of all, as discussed in Section Challenges Arising From Robots and Artificial Intelligence, according to some studies employment effects from adopting robots remain rather limited or even positive at aggregate level (Acemoglu and Restrepo, 2017; Bessen, 2017; Graetz and Michaels, 2018; Koch et al., 2019; Acemoglu et al., 2020; Dixon et al., 2021).

The main argument against taxing robots, however, is that it might impede innovation in an era of productivity slump. Over the last decades, advanced economies have experienced stagnating productivity. Taxing new technologies could make that slowdown worse, while according to some studies investing in robots enhances growth and productivity. A CEBR (2017) study finds that investment in robots contributed to 10 percent of GDP growth per capita in OECD countries from 1993 to 2016. Graetz and Michaels (2018) find that a unit increase in robotics density (defined as the number of robots per million of hours worked) is associated with a 0.04 percent increase in labor productivity. An analysis carried out by the Institute for Employment Research and the Düsseldorf Institute for Competition Economics finds that from 2004 to 2014 GDP has increased by 0.5% per person per robot as result of robotization (CEBR, 2017).

Finally, another argument against the robot tax is that it would reduce the incentive for companies to invest in innovation and will make low wage traps more persistent, as argued by Robert D. Atkinson, president of the Information Technology and Innovation Foundation (ITIF). According to Atkinson, the main reason behind wage and GDP growth stagnation in advanced economies is the productivity slow-down. As mentioned above, there is empirical evidence that robots are driving labor productivity and GDP growth (CEBR, 2017; Graetz and Michaels, 2018). Therefore, creating disincentives to robotization may further impede labor productivity and perpetuate wage stagnation.

Provided that automation increases overall productivity and efficiency and hence is beneficial to the society, hence the robot tax should be designed so to avoid discouraging the use of robots and automation. Some research shows that it is optimal to tax robots only for a limited time span. In this view, Guerreiro et al. (2020) propose to tax robots for three decades.

Beside the opinion in favor or against the robot tax, there is however still discussion on how companies should pay the robot tax. A first proposal could be to tax robots themselves, in the amount of the salary paid to the hypothetical displaced human worker. This solution is however extremely complicated to be put in practice, since robots are unlikely to replace human workers in the entire set of their tasks. It is more common that robots take over only some tasks previously performed by humans and hence it is quite difficult to find a one-to-one link between the robot and the displaced worker difficult.

Alternatively, another option could be to levy a tax on the use of robots, that is imposing a higher rate of corporate tax for using robots, since companies make higher profits due to the powerful efficiency of robots. This proposal is also complicated to be implemented, because what we see nowadays is a form of "cobotization", which is a collaboration between robots and human workers to complete a task and jointly contribute to make profits. Therefore, it is not so straightforward to disentangle the profits or value created by the robot from that one created by the human worker.

Another proposal is subjecting robots to VAT, since robots can replace humans in the supply of goods or services which are subject to the VAT. To avoid obstacles to the adoption of new technologies and innovation, a simpler approach could be levying a lump-sum tax, payable at the same level by everyone, which would not create distortions in the economy. However, lump-sum taxes present trade-off in terms of equity and distributional effects to be considered. A lump-sum tax would be regressive and bear more on small businesses. Since every business will pay the same amount of robot tax no matter the profits it runs, absorbing the fixed cost of a robot tax would be more arduous for small family businesses than for large companies.

[5]On this argument, Acemoglu and Restrepo (2018) wrote, "The vast majority of tax revenues are now derived from labour income, so firms avoid taxes by eliminating employees." *New York Times* journalist Eduardo Porter wrote, "Machines don't incur payroll taxes, which are used to fund Social Security and Medicare. For every worker replaced by a robot, the employer saves on payroll taxes."

[6]In OECD countries, in 2015 individual income taxes and social insurance taxes represented approximately 50% of all tax revenues. In the United States, the reliance on labor taxation is even more pronounced, with more than 60% of all tax revenue coming from individual income taxes or payroll taxes (see https://taxfoundation.org/publications/sources-of-government-revenue-in-the-oecd/).

[7]See Eduardo Porter, "Don't Fight the Robots, Tax Them", N.Y. TIMES (Februaty 23, 2019).

Overall, these proposals require international coordination to avoid that income could be taxed twice, at the robot level, in the amount of the imputed salary or higher profits associated to the use of robots, and at the corporation level (Oberson, 2017).

Another problem with the robot tax is the definition of robot itself. Some institutions (e.g., the EU Parliament, the International Federation of Robotics) have proposed criteria to define robots.[8] All definitions include two main criteria: the level of autonomy and the capacity to learn. However, there is still a lack of consensus on the definition of robots. The distinction between a machine and a robot or between a computer program and AI is still not clear. For example, a ticket-vending machine replaces a human but could not be considered a robot.

Moving to more "philosophical" issues, some thorny questions deserve more consideration. First, governments may choose whether to tax robots themselves as they were persons or whether to levy a tax on the use of robots. If they opt for the first solution, then governments should give legal-person status to robots in order to make them taxable, as Professor Xavier Oberson points out. The status of legal person implies that robots would have rights and obligations, so they could collect social security and retire or go to jail if they do not pay taxes.

At this stage, proposals on how to implement a robot tax in practice remain very nebulous. South Korea is the only country to have introduced a kind of robot tax. An extensive talk about the need of a robot tax is starting to emerge in the United Kingdom, the United States, Japan and Canada. We discuss below some country cases in this respect.

South Korea has been the first country to have levied a robot tax on August 6, 2017. Korea is one of the countries with the highest share of robots in the workplace, particularly in the manufacturing industry. However, South Korea has not exactly introduced a tax on individual robots or on the use of robots, rather a reduction in the deductions for increasing automation. Under previous governments, Article 24 of the Restriction of Special Taxation Act established that companies could have between 3 and 7% of their corporate tax deducted, depending on the size of the business. Since August 2017, the new administration of President Moon Jae-in has lowered the tax deduction rate by up to 2% points.

In the United States, New York's Mayor and 2020 presidential candidate, Bill de Blasio, has pointed out the need of to adopt a kind of robot tax to protect those jobs at risk of obsolescence. Revenues from the robot tax might be used to create new jobs in green energy, health care and education. Another example of possible proposal of a robot tax has been put forward by a political candidate in Chicago, Ameya Pawar, who has suggested a two-fold approach: on the one side redeeming subsidies given to companies who do not create the promised number of jobs, and on the other side taxing companies who adopt robots to displace human workers. While calls for a robot tax have emerged in the political debate in the United States, the only concrete example attempting to deal with automation, although a very specific type of automation, is the Autonomous Vehicles Tax Legislation. However, there is not agreement on the definition of "fully autonomous vehicle". In 2017, the Nevada legislature imposed an excise tax on transportation network companies using fully autonomous vehicles. Similarly, in 2018 the California legislature authorized San Francisco to impose a local tax on transportation network companies using autonomous vehicles. Calls for a similar legislation have emerged in two other states, Massachusetts and Tennessee, but not concrete steps have been taken so far.

In Italy a law proposal in August 2017 suggested to increase the corporate income tax rate by 1% for companies "if the production activity of the company is implemented and managed predominantly from artificial intelligence systems and robotics". However, no further action has been taken. The proposal presented some pitfalls, in particular the legislation provided neither a definition of "artificial intelligence systems" or "robotics", nor clear criteria to determine whether a company's activity may be considered "predominantly" implemented and managed by AI or robotics.

In 2017, Ms Mady Delvaux, a member of the European Parliament, tried to introduce a recommendation of a robot tax in a Committee on Legal Affairs Report. However, ultimately the resolution adopted by the European Parliament did not include a robot tax. Although the majority of European leaders agreed on the urgence to control the possible side effects of automation on human employment, the EU was concerned about the risk that a robot tax may impede innovation. In particular Andrus Ansip, the former European Commissioner for Digital Single Market, opposed the robot tax.

There is no large empirical evidence on the effects of the robot tax. In South Korea the introduction of the robot tax is associated with a slow-down in investment in robotics. Koracev (2020) reports that in 2017 the new industrial robot installations in South Korea decreased for the first time since 2012. However, it is difficult to establish with certainty the causality between the reduction in the automation tax credit and the slowdown in robotization.

Conversely, Bogenschneider (2021) reports empirical evidence suggesting that higher taxation does not seem to discourage robotization. The empirical evidence shows that "robot density is positively associated with high corporate tax rates, such as in Germany, Japan, South Korea and the Nordic countries, with little or no automation occurring in tax havens where the value of tax deductions for capital investment is zero".

Digital Taxes

Another solution is digital taxation. The debate on digital taxation focuses on two main aspects. First, how to ensure that tax policy remains neutral in targeting traditional and digital businesses? Digital businesses have benefitted from preferential tax regimes, e.g., tax advantage for income earned from intellectual property, shorter amortization for intangibles, R&D tax relief. The risk is that preferences for digitalized businesses may create tax windfalls that can be used in ways that distort investment, rather than focusing on innovation.

Second, digital companies may operate without having physical presence in countries where digital enterprises have customers, since they can reach customers through remote

[8]I refer to Oberson (2017) for more details.

sales and service platforms. The ability of digitalized firms to make profits though cross-border sales without a physical presence poses challenges on the traditional corporate income tax rule. Up to now, digital businesses have paid corporate taxes on profits only in those countries where they had a permanent establishment, so either the headquarter or factory or storefront. This means that the countries where sales are made or where online users are located have no taxing rights over the firm's income.

To tax digital profits, several tools have been considered. A first option consists in extending existing rules. For instance, a country may extend its Value-Added Tax (VAT) and Goods and Services Tax (GST) to include digital services or extend the tax base so to include revenues generated from the provision of digital goods and services. A second option is to levy a Digital Service Tax (DST).

Over the past years, many countries have introduced DST and VAT on digital goods and services at unilateral level, which has highlighted how lack of coordination and alignment of standards may be harmful for the global economy and can potentially lead to economically harmful trade wars. The lack of international coordination over the last years has shed lights on some crucial steps which need to be urgently taken. First of all, the VAT and GST rules need to be revised to ensure that foreign suppliers are accountable for the collection and remittance of these taxes in countries where they sell their goods and services, even without having a physical presence. Lack of coordination may also lead to confusion and impede economic activity, since digital business who sells in different countries where they do not have a permanent establishment need to conform to a large diversity of requirements in each of the countries where they have customers. Moreover, lack of coordination can also facilitate tax avoidance, since multinational enterprises can exploit differences in corporate tax rates. Finally, the risk of double taxation can easily arise, since digital businesses may be taxed twice in the hosting country under the national CIT regime and in the countries where they have customers under the DST.

Countries and international organizations are undertaking various initiatives at national level and more recently also at international level.

Regarding VAT and GST, in most of the OECD countries VAT or GST are levied on a large set of goods and services.[9]

Regarding DST, the situation is more complex. Up to now, digital enterprises have paid corporate income tax in the country where they had a permanent establishment, rather than where consumers or users are located. In practice, a digital enterprise may provide services abroad through digital means without having physical presence abroad and make profits without being subject to corporate income tax in foreign countries. Several countries over the past years have decided to tax digital goods and services and they have unilaterally introduced a DST, which rate was varying across countries.

As of May 2020, Austria, France, Hungary, Italy, Turkey and the United Kingdom have introduced a DST, while a proposal for a DST has been put forward in Spain, the Czech Republic, Slovakia and Poland. Some more timid steps in this direction have been taken in Latvia, Norway and Slovenia. Some cases are discussed more in detail below.[10]

In France in July 2019 a 3% DST has been levied on revenues from digital interface services and sale of data for advertising purposes. The United States Trade Representative considered this policy to be discriminatory against US companies and proposed retaliatory tariffs. Following the US reaction, France postponed the collection of the DST.

In the United Kingdom in April 2020, a 2% DST has been levied on revenues from social media platforms, internet search engines and online marketplaces.

In Austria in January 2020 a 5% DST has been levied on revenues from online advertising. This measure applied only to companies whose revenues exceed €750 millions worldwide and exceeding €25 millions in Austria.

Outside Europe, other countries have also adopted DST (e.g., India, Indonesia and Tunisia) or announced or show intention to adopt DST (e.g., Brazil, Kenya, Canada, Israel and New Zealand). On the contrary, Chili has rejected the proposal of a DST.

This experience has created potential rooms for retaliation, trade wars, tax avoidance and hence has highlighted the need of international coordination.

Over the last years the OECD and the European Commission have put forth proposals and started negotiations. An agreement was reached only in the second half of 2021.

Over the last years, the OECD has hosted negotiations with 139 countries to revise the international tax system and require that profits run by multinational enterprises are subject to taxation also in those countries where enterprises sell their products and services even without having a physical presence.

On 1 July 2021, the OECD Inclusive Framework issued the key principles defining the new taxation system for multinational companies.[11] The agreement has been signed on 8 October 2021. The new agreement establishes two pillars. Pillar 1 states that business with an annual turnover exceeding EUR 20 billions and a margin of profit above 10% will be subject to taxation in those countries where customers are located. Pillar 2 establishes a minimum tax rate of 15% for multinational companies with an annual turnover exceeding EUR 750 millions.

New taxing rights for market countries at the expense of residence countries, along the lines of proposals discussed under Pillar 1 of the OECD-Inclusive Framework (IF) will change the geographic distribution of tax revenues paid by digital enterprises. Countries imposing low corporate tax and with investment hubs are likely to lose revenues as less profits will be shifted toward them. Conversely, those countries

[9]In some countries, some categories of goods or services are not subject to VAT (e.g. e-books, online courses), See Bunn et al. (2020), in particular Table 4.

[10]For further information, I refer the reader to Bunn et al. (2020).

[11]For more details we refer the reader to OECD/G20 Base Erosion Profit Shifting Project (2021). https://www.oecd.org/tax/beps/statement-on-a-two-pillar-solution-to-address-the-tax-challenges-arising-from-the-digitalisation-of-the-economy-july-2021.pdf and https://www.oecd.org/tax/beps/brochure-two-pillar-solution-to-address-the-tax-challenges-arising-from-the-digitalisation-of-the-economy-october-2021.pdf.

where multinational enterprises are not headquartered but have customers are likely to gain revenues from the reallocation.[12]

Other Proposals

Some other alternatives to the robot tax are imposing a higher VAT tax on buying robot systems, or government's purchase of shares in companies and participation in dividends that can be redistributed to the population.

Recently, Saez and Zucman (2021) have proposed to introduce a new tax on corporations' stock shares for all companies with headquarter in G20 countries. This proposal stems from the idea that in the globalized world some companies may establish market power and raise enormous profits and wealth. Since stock ownership is highly concentrated in the hands of the richest, this tax on corporation stock shares would be progressive. To avoid liquidity issues, the tax could be paid by issuing new stock.

In a similar vein, Miles Kimball and Bloomberg writer Noah Smith suggest the creation a sovereign-wealth fund, split into many smaller funds, to avoid ownership concentration. Government could buy stocks and real estate using tax revenues and then distribute the profits to the society. In this way, governments would redistribute some of the profits arising from robotization.

Finally, another solution could be a wage subsidy for low-income workers. The most direct way is to cut payroll taxes, which overly burden low-paid workers. To fund social security, governments can use other sources, for instance increasing income taxes on the richest or a value-added tax. This is basically a shortcut to make human workers cheaper. However, while this solution reduces inequality in the short run, it may slow down productivity in the long run since it preserves unskilled labor employment which is less productive than robots. Therefore, in adopting this policy governments should balance trade-off effects in the short and long run (Berg et al., 2021).

CONCLUSIONS

While there are several proposals on the table, the only concrete steps, although very timid, undertaken so far concern the robot tax and the digital tax. There are a few ideas defined as "robot tax", but they vary significantly in design and magnitude. For example, the so-called robot tax in South Korea is a measure to reduce tax incentives for investment in automation rather than a tax on robots as proposed by Bill Gates. The idea of a robot tax as a way to levy companies directly on their use of robots and to apply those revenues toward a universal basic income is indeed philosophically appealing. However, it is overly unrealistic to expect that companies will pay for it through an income tax on their robots and AI networks.

Finally, possible widening inequalities caused by technological change may require different tax policies, depending on whether inequality is arising from skill-biased technological change or from the emergence of "superstar" firms in the digital economy. In the first case, to preserve employment and especially low-skilled workers, the robot tax could be a valid solution. However, it is also true that, as side effect, the robot tax can impede innovation. To avoid this side effect, the "design" of the robot tax is crucial. A solution could be to levy the robot tax just for a limited time, to preserve employment and have the necessary time to re-skill workers and provide them with the new skills and competencies requested on the market. The alternative, levying the robot tax as a lump-sum tax, may be not distorsive, but it will bear more on small businesses with high costs in terms of inequality. To preserve low-skill and low-paid employment, an alternative to the robot tax could be to provide wage subsidies for low-income workers. However, in choosing tax instruments governments should find the right balance between reducing inequality and preserving long-term productivity and growth. Wage subsidies for low-paid workers may be successful in preserving low-skill employment and reduce inequality in the short-run, but at cost of lower productivity in the long-run.

In the second case, when the main driver of inequality is the dichotomy between digital "superstar" firms and traditional business, the digital tax is a valid tool. There are no side effects arising from a digital tax *per se*. However, due the cross-border nature of digital businesses, digital taxation requires international coordination and multilateral action to avoid harmful retaliation and trade wars. Other solutions consist in redistributing profits from "superstars" to the society though the creation of a sovereign wealth fund or the introduction of a tax on corporation stock shares.

Not necessarily a tax option is preferable compared to the others and the discussed proposals are not mutually exclusive. Of course, policy-makers always have to keep in mind synergies between policy instruments.

To reduce inequality and achieve a more inclusive growth, tax policies should go hand in hand with other types of policies, such as education and training to guarantee that workers gain competencies demanded by the new digital economies, as well as competition policies to avoid concentration of market power in the hands of a limited number of "superstar" firms.

Within this debate, new points of discussion are emerging. Data are necessary for machine learning projects and predictive models which allow companies to provide better customer service, refine and personalize marketing and ultimately increase their profits. Users often disclose their personal data without being aware of how much information they are providing and how much digital firms monetize it. An interesting point of discussion which is recently arising is the opportunity to tax digital companies for profiting from users' personal data. In 2018 the European Commission has proposed to adopt tax measures on revenues created from activities where users play a major role in value creation. However, no further measures have been adopted at European level. Alternatively, if data can be treated as labor, users should be compensated for providing data. Since consumers have no bargaining power

[12]For a discussion on Asia, see IMF (2021).

vis-à-vis digital firms, it is quite unrealistic that consumers can be compensated if they sell data individually. A solution could be the creation of "mediators of individual data" that would collect users' data and negotiate agreements with firms according to a transparent setting price mechanism (Lanier and Weyl, 2018). This field certainly deserves more analysis and research.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The work greatly benefitted from helpful comments received by participants in the workshop on Artificial Intelligence and the Future of Work: Humans in Control organized by the ILO on 25-26 October 2021. The author thanks two referees for carefully reading the paper and for providing constructive comments.

REFERENCES

Acemoglu, D. (2021). "Harms of AI", in *NBER Working Paper No. 29247*.

Acemoglu, D., Lelarge, C., and Restrepo, P. (2020). Competing with robots: firm-level evidence from France. *AEA Papers Proc.* 110, 383–388. doi: 10.1257/pandp.20201003

Acemoglu, D., and Restrepo, P. (2017). "Robots and Jobs: Evidence from US Labor Markets", in *NBER Working Paper No. 22252*. Cambridge, MA: National Bureau of Economic Research.

Acemoglu, D., and Restrepo, P. (2018). "Artificial intelligence, automation and work," in *Economics of Artificial Intelligence*. Cambridge, MA: NBER Working Paper 24196, National Bureau of Economic Research.

Arntz, M., Terry, G., and Zierahn, U. (2017). Revisiting the risk of automation. *Econ. Lett.* 159, 157–160. doi: 10.1016/j.econlet.2017.07.001

Aslam, A., and Shah, A. (2020). "Tec(h)tonic shifts. Taxing the 'digital economy", in *IMF Working Paper No. 20/76*.

Autor, D. H., Dorn, D., Katz, L., Patterson, C., and Van Reenen, J. (2017). "The fall of the labor share and the rise of star firms," in *NBER Working Paper No. 23396*.

Barkai, S. (2020). Declining labor and capital shares. *J. Finance* 75, 2421–2463. doi: 10.1111/jofi.12909

Berg, A., Bounader, L., Gueorguiev, N., Miyamoto, H., Moriyama, K., Nakatani, R., et al. (2021). "For the benefit of all: fiscal policies and equity-efficiency trade-offs in the age of automation", in *IMF Working Paper No. 21/187*.

Bessen, J. (2017). "Automation and jobs: when technology boosts employment", in *Law and Economics Research Paper No. 17-09*. Boston: Boston University School of Law.

Bogenschneider, B. (2021). *"Empirical Evidence on Robot Taxation: Literature Review and Technical Analysis", American University Business Law Review, Forthcoming*. Washington DC: NYU Stern School of Business.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers Proc. Am. Econ. Assoc.* 108, 43–47. doi: 10.1257/pandp.20181019

Bunn, D., Asen, E., and Enache, C. (2020). *Digital Taxation Around the World*. Washington, DC: Tax Foundation.

Carbonero, F., Ernst, E., and Weber, E. (2018). "Robots worldwide: the impact of automation on employment and trade", in *ILO Research Department Working Paper N. 36*.

CEBR (2017). *The Impact of Automation*. London: Centre for Economics and Business Research.

D'Adamo, I., González-Sánchez, R., Medina-Salgado, M. S., and Settembre-Blundo, D. (2021). E-commerce calls for cyber-security and sustainability: how European citizens look for a trusted online environment. *Sustain. MDPI* 13, 1–17. doi: 10.3390/su13126752

De Backer, K., DeStefano,. T., Menon, C., and Suh, J. R. (2018). "Industrial robotics and the global organisation of production," in *OECD Science, Technology and Industry Working Paper No. 2018/03*.

Dixon, J., Hong, B., and Wu, L. (2021). *The Robot Revolution: Managerial and Employment Consequences for Firms*. Oxford: Oxford Academic.

Ernst, E. (2019). "Big Data and its enclosure of the commons," in *Social Europe 12 June 2019*.

Ernst, E., Merola, R., and Samaan, D. (2019). The economics of artificial intelligence: implications for the future of work. *IZA J. Labour Policy* 9, 1–35. doi: 10.2478/izajolp-2019-0004

Frey, C. B., and Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast. Soc. Change* 114, 254–280. doi: 10.1016/j.techfore.2016.08.019

Georgieff, A., and Hyee, R. (2021). "Artificial intelligence and employment: New cross-country evidence". *OECD Social, Employment and Migration Working Papers, No. 265*. Paris: OECD Publishing.

Graetz, G., and Michaels, G. (2018). Robots at work. *Rev. Econ. Stat.* 100, 753–768. doi: 10.1162/rest_a_00754

Guerreiro, J., Rebelo, S., and Teles, P. (2020). "Should robots be taxed?," in *NBER Working Paper No. 23806*.

IMF (2021). Digitalization and Taxation in Asia. *Asia-Pacific and Fiscal Department Affairs DP/2021/017*.

Kelly, T., Liaplina, A., Tan, S. W., and Winkler Reaping, H. (2017). *Digital Dividends: Leveraging the Internet for Development in Europe and Central Asia*. Washington DC: World Bank.

Koch, M., Manuylov, I., and Smolka, M. (2019). "Robots and firms," in *CESifo Working Paper 7608*.

Koracev, R. J. (2020). A taxing dilemma: robot taxes and the challenges of effective taxation of AI, automation and robotics in the fourth industrial revolution. *Contemp. Tax J.* 9, 4. doi: 10.31979/2381-3679.2020.090204

Korinek, A., and Stiglitz, J. (2017). "Artificial intelligence and its implications for income distribution and unemployment," in *NBER Working Paper No. 24174*.

Lanier, J., and Weyl, G. (2018). "A blueprint for a better digital society," in *Harvard Business Review. September 2018*.

Méda, D. (2016). "The future of work: the meaning and value of work in Europe," in *ILO Research Paper No. 18*. Geneva: International Labour Office.

Nedelkoska, L., and Quintini, G. (2018). "Automation, skills use and training," in *OECD Social, Employment and Migration Working Papers, No. 202*. Paris: OECD Publishing. Retrieved from https://archive-ouverte.unige.ch/unige:94500

Oberson, X. (2017). Taxing robots? From the emergence of an electronic ability to pay to a tax on robots or the use of robots. *World Tax J.* 9, 2.

OECD/G20 Base Erosion and Profit Shifting Project (2021). *Statement on a Two-Pillar Solution to Address the Tax Challenges Arising From the Digitalisation of the Economy*.

Puntoni, S., Reczek, R. W., Giesler, M., and Botti, S. (2021). Consumers and artificial intelligence: an experiential perspective. *J. Market.* 85, 131–151. doi: 10.1177/0022242920953847

Saez, E., and Zucman, G. (2021). *"A Wealth Tax on Corporations' Stock", Forthcoming in Economic Policy*.

Villa, S., and Sassanelli, C. (2020). The data-driven multi-step approach for dynamic estimation of buildings' interior temperature. *Energies MDPI* 13, 1–23. doi: 10.3390/en13246654

World Bank (2016). *World Development Report*. Washington, DC: World Bank.

World Economic Forum (2018). *The Future of Jobs Report 2018*.

Conflict of Interest: RM was employed by the ILO.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Merola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/frai.2022.869282

How Are Patented AI, Software and Robot Technologies Related to Wage Changes in the United States?

Frank M. Fossen[1,2*], *Daniel Samaan*[3] *and Alina Sorgner*[2,4,5]

[1] Department of Economics, University of Nevada, Reno, NV, United States, [2] IZA, Bonn, Germany, [3] ILO, Geneva, Switzerland, [4] Department of Business Administration, John Cabot University, Rome, Italy, [5] Kiel Institute for the World Economy, Kiel, Germany

We analyze the relationships of three different types of patented technologies, namely artificial intelligence, software and industrial robots, with individual-level wage changes in the United States from 2011 to 2021. The aim of the study is to investigate if the availability of AI technologies is associated with increases or decreases in individual workers' wages and how this association compares to previous innovations related to software and industrial robots. Our analysis is based on available indicators extracted from the text of patents to measure the exposure of occupations to these three types of technologies. We combine data on individual wages for the United States with the new technology measures and regress individual annual wage changes on these measures controlling for a variety of other factors. Our results indicate that innovations in software and industrial robots are associated with wage decreases, possibly indicating a large displacement effect of these technologies on human labor. On the contrary, for innovations in AI, we find wage increases, which may indicate that productivity effects and effects coming from the creation of new human tasks are larger than displacement effects of AI. AI exposure is associated with positive wage changes in services, whereas exposure to robots is associated with negative wage changes in manufacturing. The relationship of the AI exposure measure with wage increases has become stronger in 2016–2021 in comparison to the 5 years before.

JEL Classification: J24, J31, O33.

Keywords: artificial intelligence, software, robots, wage dynamics, labor market

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Essex, United Kingdom

Reviewed by:
Luigi Aldieri,
University of Salerno, Italy
M. J. Cobo,
University of Granada, Spain

***Correspondence:**
Frank M. Fossen
ffossen@unr.edu

Specialty section:
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

Received: *04 February 2022*
Accepted: *16 May 2022*
Published: *14 June 2022*

Citation:
Fossen FM, Samaan D and Sorgner A (2022) How Are Patented AI, Software and Robot Technologies Related to Wage Changes in the United States? Front. Artif. Intell. 5:869282. doi: 10.3389/frai.2022.869282

INTRODUCTION

Recent literature on technological change and its consequences for labor markets has raised concerns that advances in artificial intelligence (AI) may result in a massive replacement of human labor with capital (e.g., Autor, 2015; Acemoglu and Restrepo, 2018a,b, 2019; Bessen, 2019; Acemoglu et al., 2020). Frey and Osborne (2017) influenced this discussion significantly by predicting that technological possibilities of digital technologies allow for the displacement of a large share of U.S occupations in the near future. This debate can be placed in the larger historical context of how technological change alters the demand for human labor. Profit-maximizing entrepreneurs would utilize a new technology if it is economically viable and choose a new capital-labor ratio in their production decision. This may be associated with an adjustment of labor demand and with an increase or decrease of the wage. Whether this has been the case for AI technologies is an empirical

question that we attempt to address in this paper. We are using a recent dataset on individual wages for the United States and combine it with a new measure for patented AI, software, and robot technologies as provided by Webb (2020).

As some economists have argued, AI is a general-purpose technology, which is neither linked to a certain type of physical device, nor to a specific application in a specific economic sector (e.g., Brynjolfsson and McAfee, 2014). Empirical measurement of the employability of AI throughout the economy, and of its current and future capabilities, is therefore not trivial. Presently, the most common approach in the literature is to directly compare currently existing human tasks carried out by labor with current or expected future capabilities of any AI-driven machine (see, e.g., Frey and Osborne, 2017; Brynjolfsson et al., 2018). Out of these considerations, a new important strand in the literature has arisen that aims at developing precise quantitative measures of different types of technologies' impacts on individual worker's tasks and occupations. This literature also aims at assessing the change in demand for certain types of work as evidenced by changes in employment and wages. In general, there seems to be agreement among researchers that an empirical estimation of the impact of AI on labor requires a metric that links the exposure of certain labor market variables such as human tasks, occupations or human skills to AI or other types of technologies.

As of today, there exist a variety of AI scores that all measure somewhat different aspects of "AI exposure" and therefore offer different economic interpretations. The four widely discussed measures of impacts of digital technologies are the occupational computerization probabilities by Frey and Osborne (2017), suitability of workers' tasks for machine learning (SML) by Brynjolfsson et al. (2018) as well as the within-occupation standard deviation of these SML scores, and AI Occupational Impact scores (AIOI) presented by Felten et al. (2019). Fossen and Sorgner (2022) use the four above measures to analyze heterogeneous effects of new digital technologies on individual-level wage and employment dynamics in the United States for 2011–2018. The authors employ data from the Current Population Survey (CPS) and its Annual Social and Economic Supplement (ASEC) to construct a panel. The results indicate that labor-displacing digital technologies (as captured by the computerization probabilities and the SML scores) are associated with slower wage growth and higher probabilities of switching one's occupation and becoming non-employed. In contrast, labor-reinstating digital technologies (as measured by the standard deviation of SML scores and AI occupational impact scores) improve individual labor market outcomes.[1] Workers with high levels of formal education are most affected by the new generation of digital technologies. Thus, it has been shown that existing measures of AI exposure capture different effects of AI on occupations. It is therefore crucial to understand how various existing and new measures of digital technologies are associated with individual labor market outcomes, as this will shed light on the following two important questions: First, does the technology already have observable associations with changes in individual workers' wages or other labor market outcomes? And second, what is the direction of these relationships?

The aim of this paper is to investigate—closely following the methodology of Fossen and Sorgner (2022)—the associations between new measures of patented technologies proposed by Webb (2020) and individual wage dynamics in the United States. Webb (2020) constructs three different metrics from patent data to measure the occupational exposure to three types of technologies: AI, software, and industrial robots. While we are particularly interested in the metric related to AI exposure, we include all three metrics in our empirical analysis to allow for a comparison between three different types of patented technologies.

Our results show that occupational exposure to AI is associated with increasing wages, whereas exposure to software and robots is associated with decreasing wages. They further indicate that the positive relationship of AI exposure with wage growth became stronger in 2016–2021 than in 2011–2015 and that it is stronger in services than in manufacturing. In contrast, exposure to robots is associated with wage decreases in manufacturing and became somewhat weaker over time. The results are robust to excluding the years of the Covid-19 pandemic. Fossen and Sorgner (2019) distinguish between "destructive" digitalization, when digital technology is used or can be used to replace labor, and "transformative" digitalization, when digital devices bring about changes in the way human work is performed, potentially an augmentation of work, without leading to a replacement of the activity. As we discuss in more detail below, our findings suggest that AI exposure can be cautiously interpreted as transformative digitalization, whereas exposure to non-AI software and robots can be interpreted as destructive digitalization in the sense that these technologies decrease labor demand.

The remainder of the paper proceeds as follows. Section Conceptual Background provides the theoretical background and highlights the need for developing new measures of workers' exposure to various types of technology. Section Data describes the measures of exposure to AI, software, and industrial robots proposed in Webb (2020). Our empirical strategy and results are presented in sections Methods and Empirical results. Section Discussion discusses the results and provides concluding remarks.

CONCEPTUAL BACKGROUND

Acemoglu and Restrepo (2018a) propose a task-based framework (the "AR model") in which new automation technologies lead to capital taking over tasks previously performed by human labor—if economically feasible. This displacement effect then results in a decrease in labor demand. The AR model implies several effects of automation that might countervail the displacement effect. These include, for instance, productivity effects that can increase the demand for tasks that cannot be

[1]Classification of these AI impact measures in terms of labor-displacing vs. labor-reinstating effects is based on empirical associations between these measures with individual labor market outcomes, as discussed in Fossen and Sorgner (2022). The authors of the measures themselves do not provide such a qualitative assessment for their impact scores.

automated (Autor and Dorn, 2013; Goos et al., 2014; Autor, 2015; Bessen, 2019); creation of new tasks for human workers; and an increase in the overall demand for human labor due to increases in capital accumulation.

Based on the AR model, we can distinguish a scale effect, triggered through higher productivity and the accumulation of capital, and a structural effect. The scale effect raises labor demand as such and can therefore lead to increases in employment or wages. The structural effect causes a re-allocation of tasks between humans and machines, whereby this re-allocation can result in a reduction of tasks for humans (displacement effect) or in an increase through new or altered tasks. Since occupations can be interpreted as bundles of tasks, the structural effect on occupations can be interpreted as being partly "transformative," that is, an occupation is altered but does not necessarily become obsolete, and as "destructive," i.e., an occupation is being partially destroyed by making some of the human tasks it consists of obsolete. Accordingly, Fossen and Sorgner (2019) distinguish transformative digitalization from destructive digitalization and categorize U.S. occupations along these lines. The resulting policy implications can be very different. A permanent or long-lasting destruction of a large number of occupations may justify entirely new economic policies, for example the institution of a universal basic income (UBI). If, instead, most occupations are transformed, then a strong focus needs to be put on training and re-skilling of the workforce.

An important implication from the AR model is that the labor market effects of technologies strongly depend on the type of technology and the purpose it was designed for. This makes it clear that there is a pronounced need for developing more precise measures of occupational exposure to different types of technologies and understanding how they are related to individual labor market outcomes. For instance, Fossen and Sorgner (2019) interpret the occupational computerization probabilities developed by Frey and Osborne (2017) as a measure of destructive digitalization and the AI occupational impact score introduced by Felten et al. (2019) as measure of transformative digitalization.[2] Since most existing measures are only available for U.S. occupations, Carbonero et al. (2021) propose a novel approach that allows to translate existing technology exposure scores that were developed for U.S. occupations into scores for occupations in other countries, including developing countries, and illustrate the method for the cases of Lao PDR and Viet Nam. In Carbonero et al. (2021), the authors use the SML ("suitability for machine learning") score developed by Brynjolfsson and Mitchell (2017) and Brynjolfsson et al. (2018) as a measure for destructive digitalization. The SML score is determined for work activities linked to U.S. occupations as reported in the O*NET database. The work activities, and hence the SML scores, can be aggregated on the occupational level. Carbonero et al. (2021) use the variance of the SML scores within an occupation as an indicator of transformative digitalization.

[2]Please note that this classification is based on empirical insights and is not defined ex ante by the authors who developed them.

In sum, it is not trivial to measure the exposure of different occupations to new AI-based technologies empirically. Therefore, it is important to develop new measures of occupational AI exposure that can grasp various aspects of AI capabilities, and then to empirically relate them with individual labor market outcomes. This will help better understand the size and the direction of technology impacts on workers' jobs.

DATA

Measures of Occupational Exposure to AI, Software, and Robots

Webb (2020) proposes a new approach to measure impacts of different types of digital technologies on occupations. In a nutshell, his method is based on the fact that patent data contain descriptions of the capabilities of the patented technologies. He links textual patent descriptions pertaining to a certain type of technology, such as AI, with the descriptions of tasks used in U.S. occupations from the O*NET database sponsored by the U.S. Department of Labor. O*NET provides for each existing occupation a list of tasks that are typically carried out by workers in this occupation, and it ranks the importance of each task. For example, "Document and maintain records of precision agriculture information" is one of the tasks that O*NET identifies for the occupation of agriculture technicians. To link the textual descriptions from O*NET with the description of an AI patent Webb extracts verb-noun pairs by means of a natural language processing algorithm and uses these verb-noun pairs to quantify the overlap between patents and tasks. In the previous example, such a pair would be "(maintain, records)." Basically, the algorithm would look for AI patents that are described by the same verb-noun pair. Each task is then assigned an exposure score that is based on the relative prevalence of the verb-noun pair in the total set of analyzed patents. Thus, the higher the task exposure score, the more patents were identified that describe a technology related to this task. To aggregate the task-level scores to the level of occupations, weights are used that are constructed as an average of the frequency, importance, and relevance of each task to the occupation, as specified in O*NET. The weights are scaled to sum to one. As source for patent information, Webb (2020) employs the Google Patents Public Data database. He does not impose a time restriction, but due to a strong increase in patents over time, few patents were filed before the 1990s. Most patents were filed in the 21st century, in particular in software and even more so in AI (Webb et al., 2018).

Webb (2020) constructs the exposure measure for three types of technologies: AI, software, and industrial robots. Hence, a distinction of AI from other digital technologies is possible through his method. Potentially, one can then disentangle heterogeneous effects of these technologies on wages. To restrict the set of patents to these three specific types of technologies, they had to be precisely defined. For example, only industrial robots that are used in the manufacturing sector are considered "robots" (according to the standardized definition, ISO 8373). Software are programs for which every action it performs has been specified in advance by a human, as opposed to AI, which is defined as all

forms of machine learning algorithms, supervised learning and reinforcement learning algorithms.

According to Webb (2020), labor market effects of robots and software are very different from those of AI since the occupational exposure to AI concerns different socioeconomic groups. Using census samples for the United States for the years 1980–2010, he finds that a change from the 25th to the 75th percentile of exposure to robots is associated with a decline in within-industry employment shares of between 9 and 18% and a decline in wages of between 8 and 14%. Male workers with lower education and lower wages are more exposed to robots than others. The results for software indicate that middle-wage occupations are most exposed to software. The exposure to software is also less sharply decreasing with education in comparison with robots. The direction of the effects of software on aggregated employment and wages is similar to that of robots, but the effects are smaller in size. For example, moving from the 25th to the 75th percentile of exposure to software is associated with a decline in within-industry employment shares of between 7 and 11% and a decline in wages of between 2 and 6%. Hence, Webb's (2020) finding for robots and software point toward what we coined "destructive effects" of digitalization for the United States. Overall, the replacement effect appears to dominate over labor-reinstating effects when robots and software are employed.

One limitation of Webb's analysis is that the effects of AI cannot be determined with the dataset 1980–2010 because many of the significant technological advances in AI occurred more recently. Webb analyzes the tasks that are related to the capabilities of AI and the corresponding occupations and shows that low-wage occupations are potentially among the least, and high-wage occupations are among the most exposed occupations. Highly educated individuals are more likely to be exposed to AI. Interestingly, the opposite pattern is observed for the occupational exposure to robots and software.

Based on his findings, Webb (2020) makes the assumption that the relationship between AI exposure and changes in wages has the same negative, approximately linear relationship as the relationship that existed between exposure to software and robots and changes in wages. To determine the likely impact of AI on the wage distribution, Webb (2020) runs a simulation and finds that AI could possibly compress wages in the middle of the distribution but expand inequality at the top. In the following section, we introduce a different dataset with very recent data on wages and individual worker characteristics to empirically test the associations of the technologies with wage changes. While we are primarily interested in estimating the relationship of AI technology with individual wage changes, we also use the other two metrics for occupational exposure to software and industrial robots to allow for comparison between these different types of technology.

Individual-Level Panel Data

To estimate associations of technology exposure with individual-level wage changes we use the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), a representative survey of households in the United States provided by the Census Bureau. Given that most recent advances and the diffusion of AI technologies only occurred over the last few years, we concentrate in the main estimations on the period 2016–2021. In supplemental estimations we use the longer period 2011–2021 and subperiods. The ASEC is always conducted in March and contains information on labor income. We use the IPUMS-CPS database provided by Flood et al. (2017), who match consecutive individual-level observations to construct rotating panel data, allowing us to link the March ASEC of two subsequent years for most respondents. We calculate hourly wage changes between t-1 and t for each respondent using information about the income and hours worked in the previous calendar year.

We merge Webb's three measures of technology exposure of occupations (exposure to AI, exposure to software, and exposure to robots) with the individual's occupation in the initial year, t-1, using a crosswalk of occupational codes. Some of the occupations coded in the ASEC combine more than one occupation in the more detailed SOC codes used by Webb (2020). In these cases, we aggregate the exposure scores by using their mean values weighted by the number of employees in the respective occupations in the United States as provided by the Bureau of Labor Statistics (2018). We can merge the exposure scores to 435 occupations in the ASEC. We standardize the exposure scores to facilitate interpretation of the regression coefficients.

METHODS

In our econometric analysis we follow closely the approach proposed by Fossen and Sorgner (2022). We regress wage growth on the three Webb measures and control variables based on the sample of working individuals:

$$\begin{aligned}\ln\left(wage_{i,t}\right) - \ln\left(wage_{i,t-1}\right) &= \delta'_1 \mathbf{techexp}_{j(i,t-1)} + \delta_2 switch_{i,t} \\ &+ \delta'_3 \mathbf{techexp}_{j(i,t-1)} \times switch_{i,t} \\ &+ \eta' \mathbf{v}_{i,t-1} + \xi' \mathbf{w}_{i,t} + \omega'_k \mathbf{z}_{j(i,t-1)} \\ &+ \theta^w_{year(t)} + \vartheta^w_{ind(i,t-1)} \\ &+ \mu^w_{j(i,t-1)} + \epsilon_{it}. \end{aligned} \quad (1)$$

The dependent variable is the relative change in hourly labor income of individual i between calendar years t-1 and t (log approximation). In year t-1 the individual worked in occupation $j(i,t-1)$. The three key explanatory variables summarized in the vector $\mathbf{techexp}_{j(i,t-1)}$ are the exposures of occupation j to patented AI, software, and robot technology using the measures developed by Webb (2020). We include the three exposure measures simultaneously in the regression such that the partial effect of each technology type is identified keeping the others constant. We also include a wide set of control variables that might affect individual wage growth.

The vector of coefficients δ_1 captures the effects of different types of technologies (robots, software, AI) on wage growth for individuals who do not switch their occupation. $switch_{i,t}$ is a dummy variable indicating whether a respondent i switched

TABLE 1 | Descriptive statistics.

	Means	Std. dev.	Correlation coefficients		
			Exposure to AI	Exposure to software	Exposure to robots
Technology exposure measures					
Exposure to AI	0.379	0.217	1.000		
Exposure to software	0.421	0.245	0.532	1.000	
Exposure to robots	0.498	0.629	0.026	0.503	1.000
Individual-level characteristics					
Annual wage growth	0.028	0.891	−0.011	0.018	0.022
Occupation switch	0.583		0.008	0.020	−0.027
Less than high school	0.063		−0.036	0.099	0.230
High school degree	0.271		−0.047	0.134	0.222
Some college	0.291		−0.030	0.022	−0.005
College degree	0.374		0.089	−0.193	−0.315
Male	0.514		0.169	0.134	0.187
Age	43.84	11.90	0.004	−0.038	−0.003
Metropolitan area	0.823		0.010	−0.047	−0.080
Married	0.620		0.045	−0.046	−0.067
Number of children in household	0.919		0.016	−0.011	0.005
White	0.824		0.022	−0.009	−0.027
Black	0.089		−0.026	0.021	0.052
Asian	0.058		0.007	−0.019	−0.037
Other race	0.029		−0.015	0.009	0.024
Self–employed (incorporated)	0.044		0.007	−0.059	−0.067
Self–employed (unincorporated)	0.057		−0.009	−0.018	0.012
Occupation–level characteristics					
Mean hourly wage in occupation	29.72	18.68	0.177	−0.224	−0.382
Share of women in occupation	0.486	0.295	−0.287	−0.227	−0.316
Self-employment rate in occ.	0.102	0.150	0.010	−0.119	−0.129
Offshorability score in occ.	1.810	1.317	0.271	0.071	−0.218
Routine cognitive task intensity in occ.	0.034	0.973	−0.034	−0.199	−0.414
Routine manual task intensity in occ.	−0.206	0.817	−0.005	0.369	0.503
High school diploma needed	0.358		0.032	0.013	−0.121
Postsecondary non-degree needed	0.077		−0.106	0.027	0.106
Some college needed	0.020		−0.037	0.019	−0.050
Associate's degree needed	0.027		0.071	0.006	−0.049
Bachelor's degree needed	0.260		0.253	−0.131	−0.317
Master's degree needed	0.023		0.017	−0.061	−0.089
Doctoral or prof. degree needed	0.043		−0.072	−0.140	−0.133

The table shows mean values, standard deviations for non-binary variables, and correlation coefficients. The exposure scores to AI, software and robots are not standardized here. Number of person-month observations: 58,394.
Source: Own calculations based on the ASEC 2016-21.

the main occupation between the years t-1 and t, identified by a change in the occupational code. The idea is that some individuals whose jobs are heavily affected by technologies could be able to switch to a different occupation, thereby preventing a possible wage decline. Interaction terms between this dummy variable and the three exposure measures (with coefficients δ_3) capture how much the impacts of the technologies in the previous occupation on the individual's wage growth change in case of an occupational switch.

$\mathbf{v}_{i,t-1}$ is a vector of 10 splines of the initial individual wage ($wage_{i,t-1}$) controlling for a potential general change in the income distribution. The vector also includes dummy variables indicating incorporated or unincorporated self-employment in t-1. The vector $\mathbf{w}_{i,t}$ contains further individual-level controls at time t: gender, age, age square, marital status, number of children in the household, ethnicity, highest educational attainment, residence in a metropolitan area, 8 dummies for the US Census regions, and a constant. We also include year dummies, $\theta^{w}_{year(t)}$, and 52 major industry dummies, $\vartheta^{w}_{ind(i,t-1)}$, to control for industry exposure to international trade in t-1. The occupational dummies $\mu^{w}_{j(i,t-1)}$ capture the 2-digit level of the occupation codes provided in t-1.

TABLE 2 | Relationship of technology exposure with annual wage growth (2016–2021).

	(1)	(2)	(3)	(4)
Occupation switch	−0.0393***	−0.0431***	−0.0459***	−0.0408***
	(0.0112)	(0.00966)	(0.0112)	(0.0144)
Exposure to AI	0.0268**	0.0501***	0.0595***	0.0796***
	(0.0115)	(0.0132)	(0.0136)	(0.0167)
Exposure to AI *x* occupation switch	−0.0371***	−0.0402***	−0.0353***	−0.0419***
	(0.0117)	(0.0116)	(0.0122)	(0.0138)
Exposure to software	−0.0326**	−0.0508**	−0.0531***	−0.0508**
	(0.0162)	(0.0201)	(0.0177)	(0.0230)
Exposure to software *x* occupation switch	0.0404**	0.0388**	0.0408**	0.0438**
	(0.0162)	(0.0167)	(0.0170)	(0.0193)
Exposure to robots	−0.0246**	−0.0307**	−0.0493***	−0.0542***
	(0.0115)	(0.0143)	(0.0125)	(0.0142)
Exposure to robots *x* occupation switch	0.0299**	0.0249**	0.0246*	0.0250*
	(0.0132)	(0.0122)	(0.0131)	(0.0134)
High school degree	0.0804***	0.0934***	0.102***	0.113***
	(0.0141)	(0.0141)	(0.0138)	(0.0150)
Some college	0.139***	0.156***	0.173***	0.181***
	(0.0153)	(0.0152)	(0.0156)	(0.0178)
College degree	0.296***	0.327***	0.373***	0.375***
	(0.0177)	(0.0173)	(0.0181)	(0.0215)
Male	0.146***	0.146***	0.154***	0.168***
	(0.00716)	(0.00720)	(0.00751)	(0.0101)
Age	0.0283***	0.0308***	0.0311***	0.0345***
	(0.00257)	(0.00233)	(0.00234)	(0.00300)
Age squared	−0.000295***	−0.000323***	−0.000328***	−0.000366***
	(0.0000298)	(0.0000271)	(0.0000271)	(0.0000356)
Marital status	0.0768***	0.0728***	0.0778***	0.0832***
	(0.00940)	(0.00870)	(0.00869)	(0.00881)
Number of children in household	0.00276	0.00106	−0.000225	−0.00269
	(0.00296)	(0.00286)	(0.00284)	(0.00303)
Metropolitan area	0.0740***	0.0753***	0.0754***	0.0745***
	(0.00882)	(0.00786)	(0.00809)	(0.00845)
Black	−0.0689***	−0.0744***	−0.0816***	−0.0759***
	(0.0131)	(0.0118)	(0.0117)	(0.0116)
Asian	−0.0254*	−0.0242*	−0.0235	−0.0190
	(0.0149)	(0.0142)	(0.0154)	(0.0171)
Other race	−0.0536**	−0.0371*	−0.0367*	−0.0423**
	(0.0208)	(0.0205)	(0.0206)	(0.0205)
Self-employed (unincorporated)	−0.157***	−0.168***	−0.171***	−0.172***
	(0.0275)	(0.0243)	(0.0227)	(0.0255)
Self-employed (incorporated)	−0.0418**	−0.0484**	−0.0326*	−0.0448**
	(0.0210)	(0.0189)	(0.0176)	(0.0203)
Hourly wage in occupation	0.00276***			
	(0.000778)			
Share of women in occupation	−0.0702**			
	(0.0330)			
Self–employment rate in occupation	−0.109**			
	(0.0503)			
High school needed	0.0935***			
	(0.0214)			
Post–secondary degree needed	0.0455			
	(0.0279)			

(Continued)

TABLE 2 | Continued

	(1)	(2)	(3)	(4)
Some college needed	0.0488			
	(0.0349)			
Associated degree needed	0.0413			
	(0.0425)			
Bachelor degree needed	0.104***			
	(0.0325)			
Master degree needed	0.179***			
	(0.0453)			
Doc. or professional degree needed	0.214***			
	(0.0470)			
Offshoreability score in occupation	−0.00197			
	(0.00669)			
Routine cognitive task intensity	0.0266***			
	(0.00732)			
Routine manual task intensity	−0.0252**			
	(0.0101)			
Constant	1.305***			
	(0.0918)			
Further individual controls, income splines	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	–
Occupation dummies (2 digits)	Yes	Yes	–	–
Occupation–level controls	Yes	–	–	–
Number of observations	58,394	69,434	69,434	70,650
R^2	0.304	0.297	0.288	0.274

*OLS regressions. The dependent variable is the growth rate in the hourly wage between two adjacent years in real US$ (logarithmic approximation). The exposure measures pertain to the first year of a 2-year pair. The switch dummy variable indicates that an individual switched to a new occupation between the 2 years. We interact this dummy variable with the exposure measures. The standard errors are clustered at the level of occupations. Stars (***/**/*) indicate significance at the 1/5/10% level.*
Source: Own calculations based on the ASEC 2016-21.

Additional occupation-level variables $\mathbf{z}_{j(i,t-1)}$ account for remaining variation within these 2-digit groups of occupations: the mean hourly wage rate, the self-employment rate, and the required degree of formal education at the entry level obtained from the Bureau of Labor Statistics (2018); the share of female workers in each occupation computed directly from our microdata; the measure of susceptibility of occupations for offshoring provided by Blinder and Krueger (2013); and the occupations' routine manual and routine cognitive task intensities that we create from O*NET following Acemoglu and Autor (2011). We cluster standard errors at the level of occupations.

EMPIRICAL RESULTS

All Workers

Table 1 shows sample means, standard deviations and correlation coefficients for the variables used in this analysis. Exposure to AI is positively correlated with software but less so with industrial robots. The correlations highlight the importance of including these technologies jointly in the regressions to estimate partial effects of each technology keeping the others constant. The raw correlation of wage growth with the AI exposure score is weakly negative and those with software and robot exposure are weakly positive (significant at the 5% level). As we will see below, these signs change in the multivariate regressions controlling for essential factors influencing wages at the individual and occupation levels. Exposure to robots is positively correlated with routine manual task intensity, confirming our expectations.

We present the main estimations (Equation 1) of the controlled associations of Webb's three different technology exposure measures with wage growth in **Table 2** based on the period 2016–2021. The four models include different sets of control variables whereby the preferred model (1) contains all variables discussed in the previous section with industry group, occupation group and time dummies, as well as all occupation-level controls. In models (2) – (4), the occupational variables and the occupation- and industry fixed effects are successively excluded from the estimation as robustness checks. These estimations are based on larger samples because they include observations with missing values in the occupation-level or industry variables.

We are mainly interested in the coefficients of the three technology exposure measures, i.e., exposure to AI, exposure to software, and exposure to industrial robots. We can see that the estimates of the coefficients are consistent in terms of signs and

TABLE 3 | Technology exposure and annual wage growth in different periods.

	2011–2021	2011–2015	2016–2019
Occupation switch	−0.0396***	−0.0387***	−0.0366***
	(0.00955)	(0.0101)	(0.0116)
Exposure to AI	0.0214**	0.0166**	0.0271**
	(0.00866)	(0.00835)	(0.0123)
Exposure to AI *x* occupation switch	−0.0331***	−0.0292***	−0.0332***
	(0.00905)	(0.00915)	(0.0125)
Exposure to software	−0.0331***	−0.0324***	−0.0376**
	(0.0119)	(0.0100)	(0.0179)
Exposure to software *x* occupation switch	0.0346***	0.0300***	0.0421**
	(0.0117)	(0.0108)	(0.0186)
Exposure to robots	−0.0278***	−0.0289***	−0.0259**
	(0.00983)	(0.0105)	(0.0115)
Exposure to robots *x* occupation switch	0.0312***	0.0311***	0.0339***
	(0.0111)	(0.0113)	(0.0128)
Further individual controls, income splines	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
Occupation dummies (2 digits)	Yes	Yes	Yes
Occupation–level controls	Yes	Yes	Yes
Number of observations	131,539	73,145	50,385
R^2	0.306	0.320	0.307

OLS regressions for different periods. The dependent variable is the growth rate in the hourly wage between two adjacent years in real US$ (logarithmic approximation). The exposure measures pertain to the first year of a two–year pair. The switch dummy variable indicates that an individual switched to a new occupation between the 2 years. We interact this dummy variable with the exposure measures. All control variables listed in model (1) of ***Table 2*** *are included in the regressions but not shown. The standard errors are clustered at the level of occupations. Stars (***/**/*) indicate significance at the 1/5/10% level.*
Source: Own calculations based on the ASEC 2011-21.

are significant at the 1 or 5 percent levels in all four models, indicating robustness of our estimation.

Exposure to software and robots is associated with a decrease in the growth rate of individual hourly labor income. In model (1) with full controls, a one standard deviation higher exposure to software is associated with a 3.26 percentage points lower annual wage growth, and a one standard deviation higher exposure to robots is related to a 2.46 percentage points lower annual wage growth. On the contrary, we find a *positive* association of wage growth with exposure to AI. We find a 2.68 percentage points higher wage growth for a one standard deviation increase in exposure to AI. The coefficients of the interaction terms with occupation switch are significantly different from zero for the three technologies and always have the opposite sign from the coefficient of technology exposure. Thus, by switching occupation, an individual mitigates or even overcompensates the effect the technology exposure in the original occupation has on the wage.

Did the strengths of the associations change over time? In **Table 3**, we repeat the estimation of the full model but using different time periods[3]. The first column shows the estimates for the prolonged period of 2011–2021 and the second for the first 5 years (2011–2015); these results can be compared to the main estimation using the last 5 years (2016–2021) in model (1) in **Table 2**. We can see that the positive association of AI exposure with wage changes is stronger in 2016–2021 than in the 5 years before. This observation might reflect that the diffusion of AI technologies has accelerated in the last 5 years. The relationship of exposure to software with wage dynamics remained unchanged over these 10 years while that of exposure to robots became somewhat weaker.

One might wonder if the results are driven by the Covid-19 pandemic, which changed work in dramatic ways including widespread shifts to remote work from home. To assess the sensitivity of our results, in the rightmost column of **Table 3** we exclude the years of the pandemic, 2020 and 2021, from our main sample, thus leaving the period 2016–2019[4]. The results are very similar to those in model (1) in **Table 2**, so we conclude that our findings are not driven by the COVID-19 pandemic.

[3]The full set of control variables is included in all regressions but the estimates are not shown in the table.

[4]On March 11, 2020, the World Health Organization declared COVID-19 a pandemic. On March 16, the San Francisco Bay Area imposed the first shelter-in-place restrictions in the United States followed by the State of California on March 19 and New York State the next day. Thus, the March 2020 ASEC collection is likely to reflect the early impacts of the pandemic, and the March 2021 ASEC was collected in the middle of the pandemic; on March 19, 2021, the first 100 million Covid vaccine doses were administered in the United States.

TABLE 4 | Technology exposure and annual wage growth by sector and employment status.

	Services	Manufacturing	Employees	Entrepreneurs
Occupation switch	−0.0312**	−0.0796***	−0.0449***	0.0893**
	(0.0126)	(0.0212)	(0.0108)	(0.0413)
Exposure to AI	0.0234*	0.0214	0.0247**	0.0626
	(0.0133)	(0.0172)	(0.0105)	(0.0467)
Exposure to AI *x* occupation switch	−0.0372***	−0.0274	−0.0417***	−0.0308
	(0.0140)	(0.0209)	(0.0113)	(0.0523)
Exposure to software	−0.0336**	−0.0266	−0.0296*	−0.171***
	(0.0167)	(0.0238)	(0.0156)	(0.0636)
Exposure to software *x* occupation switch	0.0387**	0.0427	0.0413***	0.138*
	(0.0180)	(0.0291)	(0.0154)	(0.0772)
Exposure to robots	−0.0174	−0.0601***	−0.0235**	0.000745
	(0.0122)	(0.0163)	(0.0117)	(0.0484)
Exposure to robots *x* occupation switch	0.0280*	0.0323	0.0221*	0.0884
	(0.0158)	(0.0200)	(0.0120)	(0.0601)
Further individual controls, income splines	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Occupation dummies (2 digits)	Yes	Yes	Yes	Yes
Occupation–level controls	Yes	Yes	Yes	Yes
Number of observations	46,265	11,425	52,494	5,900
R^2	0.307	0.332	0.319	0.350

OLS regressions for different sectors and by employment status. The dependent variable is the growth rate in the hourly wage between two adjacent years in real US$ (logarithmic approximation). The exposure measures pertain to the first year of a two-year pair. The switch dummy variable indicates that an individual switched to a new occupation between the 2 years. We interact this dummy variable with the exposure measures. All control variables listed in model (1) of ***Table 2*** *are included in the regressions but not shown. The standard errors are clustered at the level of occupations. Stars (***/**/*) indicate significance at the 1/5/10% level.*
Source: Own calculations based on the ASEC 2016-21.

Different Groups of Workers

Are occupational exposures to the technologies associated with stronger wage changes for certain groups of workers? In this section we split the sample by worker characteristics and run our preferred regression with the full set of controls, similar to model (1) in **Table 2**. We use the sector and type of worker in the initial year, *t*-1, to split the samples. Results in **Table 4** show that the relationships of exposure to AI and software with wage changes are strongest in the services sector, while the point estimates are insignificant in the manufacturing sector[5]. In contrast, exposure to robots is more strongly related to decreasing wages in manufacturing and unrelated to wage change in services. These links between the different technologies and sectors are consistent with expectations given the nature and use of the technologies, for example, the deployment of industrial robots in manufacturing, and underline the plausibility of our results. Moreover, employee's wage dynamics are mostly related to AI technologies and robots, while earnings of entrepreneurs are significantly associated only with software exposure. A possible explanation for this latter result could be that software may perform tasks that firms have previously subcontracted to entrepreneurs.

[5]In the column labeled "Manufacturing" we combine the primary and secondary sectors, but the agricultural sector accounts for only a very small employment share in Germany.

In **Table 5** we split the sample by demographic characteristics. Occupational exposure to robots is negatively related with wage growth of both male and female workers, although it is only statistically significant for males. The effects of AI and software exposure are comparable for both genders. Moreover, the associations of technology exposure with wage changes are statistically significant only for workers residing outside the core cities in the United States. The point estimates have the same signs within core cities, however, and the statistical insignificance there may be due to the smaller sample size of core city residents.

DISCUSSION

The aim of this paper is to investigate the relationships of three types of patented technologies, AI, software and industrial robots, with individual wage dynamics in the United States. To this end, we employ three measures of occupational exposure to these technologies developed by Webb (2020) that he constructs based on the textual descriptions of patents and of tasks that workers perform in their occupations. While Webb (2020) provides empirical evidence for how his measures of exposure to software and robots are associated with employment and wage dynamics at the level of occupations and industries during 1980–2010, we add to this evidence by focusing on the micro-level of individual workers in a more recent period from 2016 to 2021. Importantly,

TABLE 5 | Technology exposure and annual wage growth by demographics.

	Female	Male	Core city	Other areas
Occupation switch	−0.0414***	−0.0361**	−0.0566***	−0.0372***
	(0.0151)	(0.0143)	(0.0154)	(0.0126)
Exposure to AI	0.0283**	0.0319***	0.0130	0.0293**
	(0.0143)	(0.0121)	(0.0150)	(0.0133)
Exposure to AI *x* occupation switch	−0.0407***	−0.0375***	−0.0312*	−0.0366***
	(0.0157)	(0.0131)	(0.0163)	(0.0132)
Exposure to software	−0.0345*	−0.0277*	−0.0281	−0.0322*
	(0.0178)	(0.0161)	(0.0180)	(0.0187)
Exposure to software *x* occupation switch	0.0419**	0.0412***	0.0277	0.0443**
	(0.0199)	(0.0157)	(0.0201)	(0.0179)
Exposure to robots	−0.0231	−0.0341***	−0.0313	−0.0229*
	(0.0161)	(0.0128)	(0.0211)	(0.0131)
Exposure to robots *x* occupation switch	0.0351*	0.0267**	0.0228	0.0354***
	(0.0180)	(0.0133)	(0.0240)	(0.0116)
Further individual controls, income splines	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Occupation dummies (2 digits)	Yes	Yes	Yes	Yes
Occupation–level controls	Yes	Yes	Yes	Yes
Number of observations	28,358	30,036	14,530	34,353
R^2	0.334	0.293	0.310	0.303

OLS regressions by gender and in core cities vs. other areas. The dependent variable is the growth rate in the hourly wage between two adjacent years in real US$ (logarithmic approximation). The exposure measures pertain to the first year of a two–year pair. The switch dummy variable indicates that an individual switched to a new occupation between the 2 years. We interact this dummy variable with the exposure measures. All control variables listed in model (1) of ***Table 2*** *are included in the regressions but not shown. The standard errors are clustered at the level of occupations. Stars (***/**/*) indicate significance at the 1/5/10% level.*
Source: Own calculations based on the ASEC 2016-21.

using these recent data allows us to also estimate associations between wage changes and exposure of occupations to AI since the dissemination and implementation of AI technologies has accelerated considerably.

In a nutshell, we find, consistently with the AR model and previous empirical literature, that different types of technology are related to labor markets in different ways. Industrial robots and software are associated with decreasing individual wages, although the relationship has become weaker for robots in more recent years. In contrast, occupational exposure to AI technologies, which are defined as machine learning algorithms, supervised learning and reinforcement learning algorithms, is associated with a positive individual wage growth, controlling for other relevant factors. Remarkably, the strength of the relationship of AI exposure with wage growth has increased over the last decade, which may indicate that firms have started to implement these technologies at a larger scale. AI exposure is associated with wage dynamics in services and robots exposure with wage dynamics in manufacturing. Wages of individuals who switch their occupations are not affected by the exposure of their initial occupation to these technologies.

The opposite signs of the relationships with wage growth show that exposure to AI is very different from exposure to software and robots. Our results are consistent with the interpretation that software and robots entail a much stronger displacement effect on workers and hence exhibit destructive forms of digitalization, whereas AI rather transforms occupations and may make human workers more productive.

Our estimation results for 2016–2021 contrast with the simulation results by Webb (2020). By assuming that the relationship of wage changes with AI exposure will be the same negative relationship as it was with exposure to robots and software in the past, he predicts that AI exposure will decrease wages at the 90th percentile relative to the 10th percentile in the future. However, our estimation results suggest that this assumption is questionable because we find that, contrary to robots and software, the association between wage changes and AI exposure is positive.

When comparing Webb's AI exposure measure to other available measures of AI, the positive relationship with wage changes in recent US data is similar to results reported by Fossen and Sorgner (2022) for Felten et al.'s (2019) AI Occupational Impact scores, which reflect past progress in AI fields and therefore are likely to capture the transformative effects of AI on work. However, the positive relationship of Webb's AI measure contrasts with Brynjolfsson et al.'s (2018) measure of suitability of tasks for machine learning that was found to be negatively associated with individual wage growth in the US (Fossen and Sorgner, 2022), thus, indicating a destructive nature of this particular subfield of AI technologies. While Webb's measure includes machine learning technologies, it remains unclear why

the effect of this measure is contrary to the one by Brynjolfsson et al. (2018). A possible explanation could be that other subfields of AI that lead to productivity effects or create new tasks for human workers may overweigh the negative effect coming from ML technologies. If so, this calls for the development of new, more fine-grained measures of occupational exposure to various subfields of AI.

Our study is not without limitations. For instance, it is unclear to what extent the AI technologies reflected in the patents were already implemented in occupations in our estimation period 2016–2021. When these AI technologies will be more fully implemented in the future, the relationship with wage changes may change its direction, even though our findings for this rather early stage in this technology's lifecycle suggest otherwise. Moreover, we were not able to establish a causal relationship between the technology exposure scores and wage dynamics. Future research should try to find ways to estimate the causal impact of AI technologies on workers' economic outcomes. Another interesting question for future research would be to investigate how individuals use various risk mitigation strategies to deal with negative impacts of technologies on their jobs. Our study indicates that occupational switching is a potential route to minimize such risks, but it is certainly costly and not equally available to all affected workers. Last but not least, the Covid-19 pandemic has accelerated digital transformation processes, which might soon become observable in changing individual economic outcomes. Thus, future research could investigate the impact of the surge in digitalization of work processes on individual workers when these data become available.

In conclusion, exposure of occupations to patented AI technologies is positively associated with individual wage growth, as opposed to patented software and robot technologies. More research is needed on developing precise measures of specific AI technology impacts on workers' jobs and on assessing the labor market consequences.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://cps.ipums.org/cps/ and https://www.michaelwebb.co/.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We thank the guest editors of this Research Topic and participants at the 2021 ILO Workshop on Artificial Intelligence and the Future of Work: Humans in Control, especially our discussant, Alexandre Georgieff, and two reviewers for their valuable comments.

REFERENCES

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2020). *AI and Jobs: Evidence from Online Vacancies. NBER Working Paper 28257*. Available online at: http://www.nber.org/papers/w28257 (accessed February 04, 2022). doi: 10.3386/w28257

Acemoglu, D., and Autor, D. H. (2011). Skills, tasks and technologies: implications for employment and earnings. *Handbook Labor Econ.* 4, 1043–1171. doi: 10.1016/S0169-7218(11)02410-5

Acemoglu, D., and Restrepo, P. (2018a). The race between man and machine: implications of technology for growth, factor shares, and employment. *Am. Econ. Rev.* 108, 1488–1542. doi: 10.1257/aer.20160696

Acemoglu, D., and Restrepo, P. (2018b). "Artificial intelligence, automation, and work," in *The Economics of Artificial Intelligence: An Agenda,* eds A. K. Agrawal, J. Gans, and A. Goldfarb (NBER book). Available online at: https://www.nber.org/chapters/c14027 (accessed February 04, 2022).

Acemoglu, D., and Restrepo, P. (2019). Automation and new tasks: how technology displaces and reinstates labor. *J. Econ. Perspect.* 33, 3–30. doi: 10.1257/jep.33.2.3

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *J. Econ. Perspect.* 29, 3–30. doi: 10.1257/jep.29.3.3

Autor, D. H., and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *Am. Econ. Rev.* 103, 1553–1597. doi: 10.1257/aer.103.5.1553

Bessen, J. (2019). "Artificial Intelligence and jobs: the role of demand," in *The Economics of Artificial Intelligence: An Agenda*, eds A. K. Agrawal, J. Gans, and A. Goldfarb (NBER book). Available online at: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-agenda/artificial-intelligence-and-jobs-role-demand (accessed February 04, 2022).

Blinder, A. S., and Krueger, A. B. (2013). Alternative measures of offshorability: a survey approach. *J. Labor Econ.* 31, S97–S128. doi: 10.1086/669061

Brynjolfsson, E., and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY: W. W. Norton & Company, 1st ed.

Brynjolfsson, E., and Mitchell, T. (2017). What can machine learning do? *Workforce Implic. Sci.* 358, 1530–1534. doi: 10.1126/science.aap8062

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers Proc.* 108, 43–47. doi: 10.1257/pandp.20181019

Bureau of Labor Statistics (2018). *Occupational Employment Statistics, May 2018 Data*. Available online at: https://www.bls.gov/oes/home.htm (accessed February 04, 2022).

Carbonero, F., Davies, J., Ernst, E., Fossen, F. M., Samaan, D., and Sorgner, A. (2021). *The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam. IZA Discussion Paper No.* 14944. Available online at: https://docs.iza.org/dp14944.pdf (accessed February 04, 2022).

Felten, E. W., Raj, M., and Seamans, R. (2019). *The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization*. doi: 10.2139/ssrn.3368605. Available online at: https://ssrn.com/abstract=3368605 (accessed February 04, 2022).

Flood, S., King, M., Ruggles, S., and Warren, J. R. (2017). *Integrated Public Use Microdata Series, Current Population Survey: Version 5.0*. [dataset]. Minneapolis, MN: University of Minnesota,

Fossen, F. M., and Sorgner, A. (2019). Mapping the future of occupations: transformative and destructive effects of new digital technologies on jobs. *Foresight STI Governance* 13, 10–18. doi: 10.17323/2500-2597.2019.2.10.18

Fossen, F. M., and Sorgner, A. (2022). New digital technologies and heterogeneous wage and employment dynamics in the United States: evidence from individual-level data. *Technol. Forecast. Soc. Change* 175, 121381. doi: 10.1016/j.techfore.2021.121381

Frey, C. B., and Osborne, M.A. (2017). The future of employment: how susceptible are jobs to computerization? *Technol. Forecast. Soc. Change* 114, 254–280. doi: 10.1016/j.techfore.2016.08.019

Goos, M., Manning, A., and Salomons, A. (2014). Explaining job polarization: routine-biased technological change and offshoring. *Am. Econ. Rev.* 104, 2509–2526. doi: 10.1257/aer.104.8.2509

Webb, M. (2020). *The Impact of Artificial Intelligence on the Labor Market*. Working Paper, Stanford University. Available online at: https://www.michaelwebb.co/webb_ai.pdf (accessed February 04, 2022). doi: 10.2139/ssrn.3482150

Webb, M., Short, N., Bloom, N., and Lerner, J. (2018). *Some Facts of High-tech Patenting. NBER Working Paper No* (Stanford, CA). 24793. doi: 10.3386/w24793

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fossen, Samaan and Sorgner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/frai.2022.868382

An Occupational Safety and Health Perspective on Human in Control and AI

*Susanne Niehaus, Matthias Hartwig, Patricia H. Rosen and Sascha Wischniewski**

Unit Human Factors, Ergonomics, Federal Institute of Occupational Health and Safety (BAuA), Dortmund, Germany

The continuous and rapid development of AI-based systems comes along with an increase in automation of tasks and, therewith, a qualitative shift in opportunities and challenges for occupational safety and health. A fundamental aspect of humane working conditions is the ability to exert influence over different aspects of one's own work. Consequently, stakeholders contribute to the prospect of maintaining the workers' autonomy albeit increasing automation and summarize this aspiration with the human in control principle. Job control has been part of multiple theories and models within the field of occupational psychology. However, most of the models do not include specific technical considerations nor focus on task but rather on job level. That is, they are possibly not able to fully explain specific changes regarding the digitalization of tasks. According to the results of a large-scale study on German workers (DiWaBe), this seems to be the case to some extend: the influence of varying degrees of automation, moderated by perceived autonomy, on workers' wellbeing was not consistent. However, automation is a double-edged sword: on a high level, it can be reversely related to the workers' job control while highly autonomous and reliable systems can also create opportunities for more flexible, impactful and diverse working tasks. Consequently, automation can foster and decrease the factor of job control. Models about the optimal level of automation aim to give guidelines on how the former can be achieved. The results of the DiWaBe study indicate that automation in occupational practice does not always happen in line with these models. Instead, a substantial part of automation happens at the decision-making level, while executive actions remain with the human. From an occupational safety and health perspective, it is therefore crucial to closely monitor and anticipate the implementation of AI in working systems. Constellations where employees are too controlled by technology and are left with a high degree of demands and very limited resources should be avoided. Instead, it would be favorable to use AI as an assistance tool for the employees, helping them to gather and process information and assisting them in decision-making.

Keywords: human in control, AI-based systems, occupational safety and health (OSH), human factors, robotic systems, ICT

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Essex, United Kingdom

Reviewed by:
Petri Böckerman,
University of Jyväskylä, Finland
Farid Shirazi,
Ryerson University, Canada

***Correspondence:**
Sascha Wischniewski
wischniewski.sascha@baua.bund.de

Specialty section:
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

Received: *02 February 2022*
Accepted: *17 June 2022*
Published: *06 July 2022*

Citation:
Niehaus S, Hartwig M, Rosen PH and Wischniewski S (2022) An Occupational Safety and Health Perspective on Human in Control and AI. Front. Artif. Intell. 5:868382. doi: 10.3389/frai.2022.868382

INTRODUCTION

Due to digitalization, jobs and working tasks are continuously changing. The development of recent technologies, such as artificial intelligence (AI) or advanced robotics has established new possibilities for task automation and revived the debate on work-related psychosocial and organizational aspects and on workers' safety and health. Amongst other things, these new technologies have the capability to fundamentally change the workers' perceived level of autonomy (Arntz et al., 2020; Wang et al., 2020; Fréour et al., 2021). The reason lies within a key feature of modern AI, its ability to operate and adapt without human intervention, in other words, autonomously while the human is left with supervisory or ancillary activities. It should be noted that automation is not equivalent to functioning autonomously. AI is used to automate functions to a certain degree, often following pre-programmed rules which makes it necessary for an operator to be present and to perform certain tasks before or after. Only if the human is not required for input or guidance, the system is seen as autonomous. In most cases, a high level of automation is reversely related to the workers' freedom in how to perform a certain task and how or what to use while completely autonomous and reliable systems can create opportunities for more flexible, impactful and diverse working tasks (Parasuraman et al., 2000; Moore, 2019; Rosen et al., 2022). Therewith, AI-based systems hold the potential to a qualitative shift in opportunities and challenges for occupational safety and health (OSH). AI-based systems are not entirely new, however their availability, complexity, performance and scope of capabilities have been extremely enlarged by the increase in computational power within the last years (Hämäläinen et al., 2018). Definitions of AI have therefore been constantly changing as they are adapting to technological advances. The term has been defined in numerous ways and a universal definition of an AI-based system is not agreed upon. However, it can be helpful to look at the definitions of major stakeholders like the OECD (2019) and the European Commission (2021).

The OECD (2019) defines AI-based systems as follows:

> […]a machine-based system that is capable of influencing the environment by making recommendations, predictions or decisions for a given set of objectives. It uses machine and/or human-based inputs/data to: (i) perceive environments; (ii) abstract these perceptions into models; and (iii) interpret the models to formulate options for outcomes. AI systems are designed to operate with varying levels of autonomy. (OECD, 2019)

An expert group on artificial intelligence set up by the European Commission, presents the following definition:

> "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions—with some degree of autonomy—to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e. g., voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e. g. advanced robots, autonomous cars, drones or Internet of Things applications)." (EU, 2019)

Both concepts have in common that they include the varying degrees of autonomy in AI-based systems as well as their ability to perceive their environments in some way, analyze the information and act in response with different degrees of autonomy. It is therefore known that interacting with these systems often includes humans to rely on the machine's complex information-processing functions like sensory processing, information storage and analysis capabilities for, amongst others, decision-making (McCormick and Sanders, 1982; Kaber and Endsley, 1997; Parasuraman et al., 2000). With this, the implementation of AI can not only shift tasks from manual to more cognitive tasks, it also creates the risk of removing "operators from direct process control" and imposing high monitoring workload (Kaber et al., 2009). Moreover, highly automated systems have implemented algorithms that enable them to adapt, learn and function autonomously. This might curtail the workers' freedom as these systems have a low level of transparency that lowers the understandability and predictability of their actions. Therefore, it is difficult, if not impossible for the worker to understand how decisions are made or how to resist them (Ajunwa, 2020). Different stakeholders named both the principle of transparency and the **principle of the human being in control or preserving workers' autonomy** as the most important aspects when designing AI-based systems. The latter (human in control/preserving autonomy) is addressed within the principles presented by the EU Commission, ETUC, ETUI as well as in the European Social Partners Framework Agreement on Digitalization. This agreement is a shared commitment of the contributing partners "to optimize the benefits and deal with the challenges of digitalization in the world of work" (ETUC, 2020). It includes a chapter especially dedicated to "Artificial Intelligence (AI) and guaranteeing the human in control principle." The principle is related to OSH, especially to psychosocial risks, as a low level of autonomy can have negative effects on motivation, job satisfaction as well as on the employees' health and performance (Dwyer and Ganster, 1991; Melamed et al., 1995; Spector, 1998; Inoue et al., 2010; Rosen and Wischniewski, 2019; Arntz et al., 2020). The agreement demands the guaranteed control of humans over machines and AI in the workplace.

Our research questions in this study are twofold:

1. What Models Are Currently Employed to Estimate the Possible Role and Impact of Automation of Decisions on a Human-Centred Design Work?
2. What Is the Link Between Automation of Decisions at Work on Psychosocial Working Conditions of Employees? Two Answer These Research Questions, Theories and Models on Human in Control Are Presented Together With Recent Scientific Literature That Depicts Possible Effects of Digitalization and Automation on Workers' Wellbeing. Furthermore, the Results of the German Survey "Digitalization and Change in Employment (DiWaBe)" Will be Presented. The Study Intended Among Other Aspects to Investigate how

Workers Are Impacted by Automation Technologies Like ICT or Production Machines That Give Instructions to the Worker and With This, Possibly Decrease Worker Control. These Systems per se Are not Purely AI-Based, However the Ability to Give Instructions Is Already an Advanced Function Which can be Even Extended by the use of AI. This Section Will be Followed by a Discussion About the Applicability of Presented Theories and Models on AI-Based Systems and Concluding Remarks on the Design of These Systems From a Human Factors Perspective.

THEORIES AND MODELS ON HUMAN IN CONTROL

The term "human in control" can be viewed as a certain level of autonomy that a worker has, for example, about decision-making, timing control and used methods during a working task. Therefore, it is closely linked to the psychosocial working condition of job control that comprises different aspects like timing or method control or decision latitude that consists of decision authority and skill discretion. Another term that closely relates to the same concept is referred to as job autonomy or task autonomy. Within scientific literature, these terms are often used interchangeably albeit one might argue that there are slightly different nuances to them. However, the combining element is to exert influence over different aspects of one's own work (Semmer, 1990). The idea of this fundamental workplace resource can also be found in the human in control principle. The human in control principle, as was recently argued by the European Trade Union Confederation (ETUC), is one of the most important measures when designing artificial intelligence (AI) or machine learning systems in order to create the opportunity for good working conditions despite increasing levels of automation (ETUC, 2020). Research in the field of occupational psychology shows that in particular low levels of job control and a small extent of task variability can have negative effects on motivation, job satisfaction as well as on the employees' health and performance (Dwyer and Ganster, 1991; Melamed et al., 1995; Spector, 1998; Rosen and Wischniewski, 2019; Arntz et al., 2020). Job control or autonomy is therefore known as a fundamental task characteristic and has the potential to enhance job performance and increase motivation (Gagné et al., 1997; Morgeson et al., 2005; Ter Hoeven et al., 2016). However, technological developments and innovations, such as artificial intelligence, give rise to new possibilities for task automation that have the capability to fundamentally change the workers' perceived level of autonomy (Arntz et al., 2020; Wang et al., 2020; Fréour et al., 2021). Overall, it has been shown that automation can either benefit or decrement workers' performance and wellbeing, depending on the task itself, the organizational structure/environment, design implementation and the machine's level of autonomy (Wiener and Curry, 1980; Kaber and Endsley, 1997, 2004; Parasuraman et al., 2000; Arntz et al., 2020). Negative influences occur when automated systems have a low level of transparency and make humans rely on AI-based algorithms as they perform all complex information-processing functions. This can lead to out-of-the-loop (OOTL) performances that have been proven to be accompanied by negative effects such as vigilance decrements, complacency, loss of situation awareness and skill decay (Wiener and Curry, 1980; Kaber and Endsley, 1997, 2004; Endsley and Kaber, 1999; Gouraud et al., 2017). Nevertheless, the automation of routine tasks and the implementation of artificial intelligence can also decrease redundancy, improve safety conditions and create opportunities for more stimulating, challenging, and impactful working tasks (Moore, 2019; Rosen et al., 2022). In order to find modes in which the distributions of functions to a human or machine will increase performance while preventing the mentioned negative consequences, research has focused on presenting theories on levels of automation (LOAs) and degrees of automation (DOAs) (Kaber and Endsley, 1997; Parasuraman et al., 2000; Kaber et al., 2009; Wickens et al., 2010). Accordingly, the goal when designing AI is to develop methods for a human-machine interaction in which humans are not only in the loop but are enabled to be in control when making decision while aided by technology which goes in line with the human in control principle.

The following paragraphs will describe established models and theories that focus on the psychosocial working condition of job control as well as on degrees and levels of automation (DOAs; LOAs). Depending, selected scientific literature will be presented that depict the effects of automation and digitalization on workers' health, performance and sense of control over the working situation.

The Scope of Activity by Ulich

Ulich (2005) presents a theory on the effect of working conditions on people and focusses on job autonomy or job control. His theory is based on the assumption that job autonomy is a multidimensional construct and is comprised of three components that are equally important for human-centered and health-maintaining design of work: scope of action ("Handlungsspielraum"), the scope of variability/creativity ("Gestaltungsspielraum") and decision latitude ("Entscheidungsspielraum"). Ulich describes the **scope of action** as the degrees of freedom in the execution and temporal organization of work actions (flexibility). He further differentiates between the objective and subjective job autonomy. The former is described as the actual available choices while Ulich understands the latter as the *perceived* options of action. The **scope of variability/creativity** is described by Ulich as the extent to which the worker has the opportunity to independently design their work and procedures. The amount of variability of partial actions and partial activities thus creates differences in the present scope of creativity. **Decision latitude** is the third component in Ulich's theoretical framework and describes the extent of an employee's decision-making authority and autonomy to independently determine and delimit working tasks. According to Ulich, a higher occurrence of each of these components has a positive impact on the workers' health.

Job Characteristic Model

While Ulich structured and systematized a multidimensional construct to make general assumptions on the effect of job control or autonomy on employees health, the job characteristics model by Hackman and Oldham (1975) focusses on determinants of intrinsic job motivation. Hackman and Oldham provide a theoretical explanation for the level of intrinsic motivation, depending on work characteristics and workers' mental states. The core work characteristics in their model include skill variety, task variety, task significance, autonomy and feedback. These lead to the experience of meaningfulness at work, of responsibility for work outcomes and the knowledge of work results. They particularly emphasize the concept of autonomy and postulate that the possibility to influence the course of the work activity or of decision-making is a key factor for intrinsic work motivation. Moreover, their equation (see **Figure 1**) presupposes the presence of autonomy for any amount of work motivation, measured by the motivation potential score (MPS). Similar to the theory by Ulich (2005), Hackman and Oldham (1975) postulate a positive linear relationship between all five core characteristics and the outcome variables. In their model, **skill variety** is described as the extent to which a job requires different activities to carry out the work involving a number of different skills and talents of the person. **Task identity** is defined as to what extent a job is holistic and produces identifiable work results. **Task significance** represents the degree to which the activity that is carried out has a substantial impact on the life or work of other people. The core characteristic **autonomy** is specified as the scope of freedom, independence and discretion the human has regarding scheduling and procedures. The model supposes a linear relationship between autonomy and motivation as the authors claim that the more freedom, the stronger the employee's motivation will be. The last factor to influence work motivation and satisfaction in the equation by Hackman and Oldham is **feedback**, which is described as the extent to which an employee will get clear and direct information about their task performance. Besides autonomy, feedback is the only other factor that must be present in order to yield any motivation (see **Figure 1**).

Job-Demand-Control Model (JDC)

The Job-Demand-Control (JDC) model by Karasek (1979) and Karasek and Theorell (1990) focusses on the stress potential of different jobs. According to Karasek (1979), the perception of acute strain and stress in working situations depends on two dimensions, namely job demands and decision latitude. Hereby, the work-specific requirements account for the extent of perceived job demands while decision latitude is explained as the degree of task variety and decision autonomy. Karasek and Theorell (1990) understand control, that is, a high level of decision autonomy and task variety, as a requirement for good working conditions, which is in line with the before mentioned models. However, to characterize types of jobs with different stress potential, they also rely on the existing job demands. As a result, four possible types are postulated: the quiet job (low work requirements and large scope of decision latitude), the passive jobs (both dimensions are low), the stressful job (high work requirements with low levels of task variety or decision latitude) and the active job (both dimensions are high). The latter is seen as the job with optimal stress and as overall health promoting while the stressful and passive job causes health risks, over- or underload as well as a decline in abilities and activities (Karasek, 1979). Although the quiet job is not believed to be detrimental to the person's stress level, Karasek (1979) assumes that people will not add to their competency on the job and generally in life if the job demands are not matched with the skill or control they experience. Therefore, he supposes that more demanding jobs, which are accompanied by a high level of decision latitude or job control are the most desirable. An overview of the relationships postulated by the JDC model are shown in **Figure 2**.

Job-Demand-Resources Model

A broader scope of work-related stressors and resources compared to the Job-Demand-Control model is incorporated by the Job-Demand-Resources model by Demerouti et al. (2001). It does not only focus on job control, but includes a number of work-related demands and resources that can influence the development of motivation and occupational stress. Demerouti et al. (2001) include a wide range of working conditions that they classify as either resources that help achieving work goals, reduce job stressors and stimulate personal development, or as demands, which are factors that require sustained effort. The former category includes, for example, job control, social support, and task variety. The latter contains factors that increase the possibility for disengagement and exhaustion such as emotional pressure, workload, and time constraints. The Job-Demand-Resources model assumes that an accumulation of demands, without the worker having enough personal or environmental resources, leads to a health impairment process. However, this entails that strengthening the workers' resources can alleviate perceived stressors and both sides are always interacting with each other. Demerouti (2020) proposes that it is possible to turn automation into a resource rather than a stressor for workers when technology is designed to support decision autonomy and helping the worker with highly complex decisions while taking over redundant and heavy tasks. Moreover, Demerouti (2020) points out the importance of supporting employees through the implementation of new technological systems to diminish newly occurring demands such as changes in work routine or the acquirement of new knowledge. With this, Demerouti's model is highly applicable in today's digitalized world of work.

Vitamin-Model

In contrast to the other presented theories, the vitamin model by Warr (1987) differentiates between constant and decrement factors. That is, for some factors, Warr assumes not a linear but an inverted u-shaped or a saturation curve-relationship between their extent and mental health. Warr counts physical security, the availability of financial resources and a social position that favors self-esteem and recognition by others as constant effect factors. These can have a negative influence on workers' health if their occurrence is low but do not impact the worker positively if they exceed a sufficient level. That is, they hit a plateau (Warr, 1987). To the decrement factors, Warr denotes job control, the

$$\text{Motivating Potential Score (MPS)} = \left(\frac{Skill\ Variety + Task\ Variety + Task\ Significance}{3} \right) \times (Feedback) \times (Autonomy)$$

FIGURE 1 | Equation on Job Motivation.

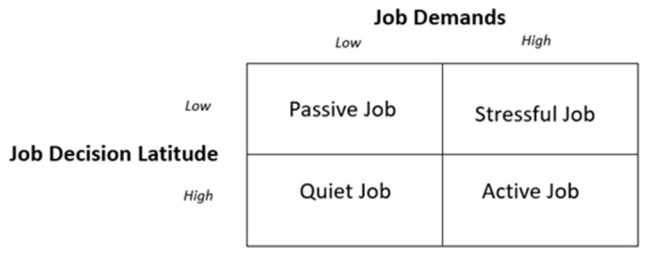


FIGURE 2 | Job-Demand-Control Model by Karasek (1979).

possibility of social contacts, the opportunity to develop and apply one's own skills, task variety (chance for new experiences) and the predictability and transparency of events. According to Warr, these follow an inverted u-shape. That is, the model predicts a negative impact on the worker's health if, for example, the level of job control is too low *or too high* (see **Figure 3**). However, the model lacks the specification of an optimal extent of autonomy. This uncertainty about the optimal level of autonomy is also present in theoretical considerations on LOAs as well. Nevertheless, most often a medium LOA is assumed to be beneficial which is more congruent with the assumption of a u-shaped relationship than a linear one.

From Psychological Models to Theories About Optimal Automation

As described before, automation refers to a set of functions that are performed automatically by technology. With a low degree of automation, the worker has overall control of the technology while transferring some of it, over a specific function, to the machine. However, automation can, in its varying degrees, lead to less interaction of the worker with the working task, leaving her with ancillary activities or supervisory control. With this, automation can have positive and negative effects on the workers' performance and wellbeing. According to the aforementioned models, the perceived level of control or job autonomy takes over a mediating role in this interplay. As stated, the continuous automation of tasks has the power to change the employees' level of job control, that is, possibilities to decide upon task variety, used methods and timing (Arntz et al., 2020; Wang et al., 2020; Fréour et al., 2021). Researchers have therefore tried to give guidelines on how automation can increase job performances and satisfaction instead of fostering skill decay, complacency, workload or OOTL performances. In the following, two fundamental models on LOAs will be described that have been the basis for most of the current research.

Ten-Level Model by Kaber and Endsley

The first model concerned with the optimal level of automation that will be described here, was put forth by Kaber and Endsley (1997). They developed a ten-level model of automation, ranging from manual control (Level 1) to full automation (Level 10) which gives a detailed description of who should be in charge of what function during the interaction. They present 10 levels of automation (LOAs) as well as ways in which the human and the machine could operate on different intermediate levels of automation that included shared control over the situation in order to identify scenarios beneficial for the human's situation awareness and for reducing workload (Kaber et al., 2009). They found that performance was best under low-intermediate levels of automation while higher levels of automation decreased the ability to recover from, and perform during, automation failures while manual control had a negative impact on performance and workload. One of the optimal scenarios included shared monitoring, planning and option selection with the final power of decision resting with the human. That is, Kaber and Endsley propose a medium LOA for positive effects on performance, situational awareness, operational safety and workload.

Theoretical Guideline on Automation by Parasuraman et al.

The model by Parasuraman et al. (2000) about types and levels of automation gives a detailed theoretical guideline to what kind of task should be automated in order to decrease mental workload and skill decay while not encouraging loss of vigilance, situation awareness or complacency. He supposes that the effect that automation has on workers depends on the kind of task that is automated as well as on the level of automation. Therefore, he established a model that systematically shows which tasks should be automated, and to what extent. With this, it is intended to assign the control between a human operator and machine in an optimal way. Parasuraman et al. differentiate four types of automation (acquisition, analysis, decision, and action) and a continuum of automation from high to low. The level of automation is then evaluated by the degree to which it influences certain human performance areas such as mental workload, complacency, reduced situation awareness and skill degradation which he describes as "potential costs" (Parasuraman et al., 2000). According to Parasuraman et al., OOTL performance problems arise if these costs are too high. The level can be adjusted in an iterative manner before secondary evaluative criteria are applied. These include automation reliability and

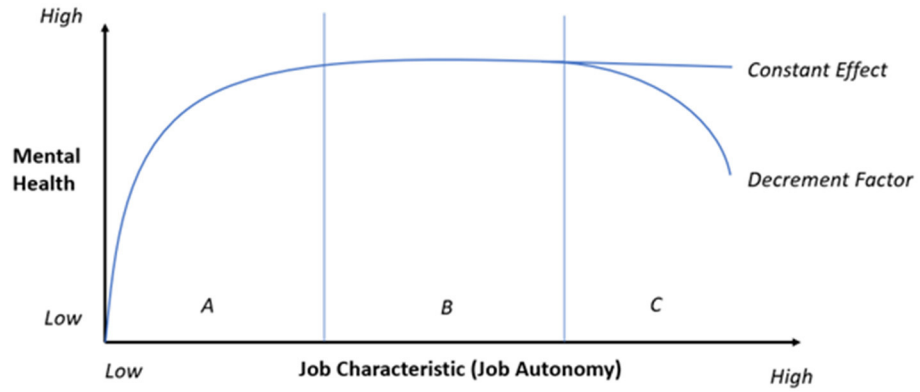


FIGURE 3 | Vitamin-Model. Positive effect of increasing autonomy **(A)**, plateau **(B)** and additional decrement **(C)**.

costs of decision or action outcomes. This process is repeated for all four types of automation. Parasuraman et al. also address the question, under which circumstances decision-making should be automated and in what scenarios it would not be suitable. As mentioned in previous paragraphs, a low decision latitude influences job satisfaction, motivation and, therewith, mental health, negatively. According to Parasuraman et al., this only occurs if the wrong tasks are automatized or if the level of automation is picked too high. Nevertheless, he notes that high-level automation and even full automation can be considered for decision-making if human operators are not required to intervene or take control under system failure as well as if they have time to respond (Parasuraman et al., 2000). Otherwise, high levels of automation would not be suitable since it would have a negative impact on mental workload, situation awareness and human performance.

Both theories on LOAs propose that assisting technologies that leave the action selection and the protocol development to humans and thus give workers control over the execution of tasks are more appropriate for tasks of great expertise while simple and redundant tasks can be performed by completely autonomous systems without negatively affecting the workers' autonomy.

Exemplary Studies on Automation and Human in Control

The described models show the importance of the level of autonomy, job control and decision-making for workers. Consequently, theories on LOA try to provide a framework to include an optimal level of these parameters within the changing nature of work. However, there is no consensus on the effects of automation on workers as the automation of tasks can be either perceived as a stressor (e.g., restriction of autonomy/control) or as a resource (e.g., ability expansion), depending on the task itself, the environment and the level of automation (Parasuraman et al., 2000; Robelski, 2016; Demerouti, 2020; Wang et al., 2020). Demerouti (2020), for example, proposes that automation can be a resource if heavy and redundant tasks are taken over by the technology while employees are assisted in dealing with their changing work environment. A changing work environment could for example refer to the implementation of new AI-based technologies and the increase of information processing, while being supported in decision-making, learning and personal development. Following this section, a large-scale study about the effects of automation on German workers will be described in detail.

Fréour et al. (2021) interviewed 3 types of employees (i.e., experts, managers, users) from an organization which has started a digitalization process and conducted a study on changing work characteristics. They assumed that the more instructions humans get from machines, the more their perceived level of autonomy diminishes. As shown in the review by Wang et al. (2020) a number of studies conducted in laboratory setting indicate a negative effect of ICT use on time pressure and workload. However, Fréour et al. (2021) showed that the workers' autonomy was not reduced when digital technologies executed repetitive tasks (Fréour et al., 2021). Moreover, their results indicated that technology that takes over action selection on tasks that require low human control and expertise enhances the workers' perceived level of autonomy by accomplishing less interesting tasks and giving the workers more time on tasks with added value. This is in line with the model by Parasuraman et al. who suppose that different situations can be more or less suitable for automation. Human autonomy should be favored if a large extent of expertise or variability is needed whereas automation is recommended for repetitive and predictable tasks or situations in which a quick reaction time is crucial. Wickens et al. (2010) conducted a meta-analysis of 18 experiments on the effect of varying LOA and included performance and workload as an outcome parameter. Again, automating redundant work had positive effects such as performance and decreased workload (if the system functioned properly). The ameliorating effects of both studies on working conditions find a theoretical basis in the models of Ulich (2005), Hackman and Oldham (1975) as well as Karasek (1979) and Demerouti et al. (2001) since task variability, significance, and (decision) autonomy were increased through the higher LOA,

resulting in an overall better condition. However, there are scenarios in which the automation of tasks increases mental workload and has a detrimental effect on situational awareness, the feeling of control as well as task variability (Kaber and Endsley, 1997; Endsley and Kaber, 1999; Weyer et al., 2015). This is often the case when a high LOA is implemented and the human is left with supervisory control over the system and only is expected to take control if the system fails (Wiener and Curry, 1980; Kaber and Endsley, 1997; Weyer et al., 2015; Gouraud et al., 2017). Parasuraman et al. (2000) mentioned that the reliability of the system is a key factor when it comes to lower the stress for the worker and impede overreliance on technology. A study about smart cars showed correspondingly that a higher level of automation increases satisfaction, but only if the malfunctions were low (Weyer et al., 2015). Other areas in which taking away human control can have positive effects are controlling the workers through occupational accident analysis, decision support systems or video surveillance for anomaly detection to prevent the occurrence of accidents and increase the workers' safety. Nevertheless, a study by Bader and Kaiser (2017) showed that ICT can foster the workers' feeling of being under control/surveillance and therewith curtail their freedom on working methods, scheduling of tasks and overall decision-making. Another negative consequence of these highly automated environments is the workers' loss of manual skills and the feeling to not be in control anymore (Berberian et al., 2012). The results by Berberian et al. (2012) suggest that the feeling of control is enlarged when action alternatives can be generated and selected as well as through greater involvement preceding an automated function. These conflicting arguments show the importance of a human-centered perspective when implementing AI or automating functions as well as the employees' opportunity to feel in control.

Overall, the studies suggest that the implementation of different LOAs can influence the employees' job autonomy and their sense of control. Although there are no clear results on the effect of specific LOAs on mental health, they do affect task variety and decision latitude as well as method and timing control, which in turn have been shown to influence the worker's perceived stress-level and overall health. Most findings suggest that automation is beneficial for redundant tasks that do not require the human to intervene in cases of system failure or if the takeover of manual control is easy. Negative effects occur if humans are left with supervisory control and redundant or ancillary activities. The "Ten-Level-Model" and the "Model for Types and Levels of Human Interaction with Automation" propose a medium level of automation for most tasks but clarify that multiple factors play into the decision on which tasks should be automated in order to influence performance and the worker's wellbeing positively. A key aspect of automation is the level of transparency that humans are able to experience when working with automated systems. Moreover, a high reliability should be given, as well as the possibility for the worker to take back control. In order to follow the human in control principle, it is necessary to take a human-centered approach and balance the degree of the system's autonomy with the level of desired control.

RESULTS OF DIGITALIZATION AND CHANGE IN EMPLOYMENT (DIWABE) SURVEY

The described models gave theoretical considerations on how much autonomy and job control are beneficial for the workers' wellbeing while the theories on LOAs and presented laboratory studies indicate that automation in itself influences the perceived level of human control and autonomy. However, until this day, studies on the actual situation in workplaces regarding the increasing automation and subsequent effects on task characteristics and the employees' wellbeing are rare. To fill this gap, the next paragraphs will describe in detail specific results of the German survey "Digitalization and Change in Employment (DiWaBe)." In this survey more than 8,000 employees answered questions on their working environment and conditions in order to find out how workers are impacted by automation technologies like ICT or machines. Of special interest are systems that give instructions to the workers and possibly reduce perceived job control. Moreover, the study assesses the current relation between decisions made by technologies, working conditions and mental health. The following paragraphs will include a short description of the survey and the results regarding the impact of technology in control.

The DiWaBe survey was jointly designed by the Federal Institute for Occupational Safety and Health (BAuA), the Federal Institute for Vocational Education and Training (BIBB), the Institute for Employment Research (IAB) and the Leibniz Centre for European Economic Research (ZEW) in 2019. The survey was conducted *via* telephone and included more than 8,000 employees from about 2,000 different German companies. These companies had already participated in a representative company survey (IAB-ZEW-Working World 4.0) in 2016 as a random sample stratified by region, company size and sector. Based on the population of all employees in these companies, participants in the DiWaBe study were also selected as a random sample stratified by age, gender and education level (for details, see Arntz et al., 2020). The questionnaire was specifically designed for the survey, including a differentiated assessment of working technologies, split up in the categories information and communication technologies (ICT) and machines/tools, which creates a unique data set. It also includes a wide array of questions regarding physical and psychological working conditions in form of stressors and resources, some of them oriented toward items in the Copenhagen psychosocial questionnaire (COPSOQ, see Kristensen et al., 2005) for comparability.

Sampling and Data Preparation

The overall response rate is 16.43%, and the distribution of the interviews deviates relatively clearly from the distribution of the gross sample. It is particularly noticeable that the utilization rates of the education group high are (as expected) significantly higher than those of the other two education groups (low and med), which was later corrected *via* weighting of the data. For a detailed description of sampling and composition, see also Arntz et al. (2020). The gathered data was subsequently compared with

administrative data and weighted by the variables mentioned above in order to be as representative as possible of the private sector in Germany. The individual weights were trimmed at the 95th percentile so that possible outliers would not have too much influence, possibly distorting the data. For the present analysis, the sample was restricted to currently employed individuals up to the age of 65 years (current age of retirement in Germany) with valid information on the main variables included. Moreover, persons with 200 or more days absent from work due to illness within the last 12 month were excluded because of potentially distorted answers after the prolonged absence. **Table 1** shows the resulting sample.

After assessing the technology use, the participants answered questions on how often technology makes decisions about their work process and gives instructions to the participant, addressing the automation of decision aspect of the Parasuraman model. The item wording was: "How often does it happen that the technology gives you instructions, e.g., about the next work step?" (1 = never, 5 = always). As work with ICT and machines differ substantially, the analysis was carried out separately for both technology classes.

Table 2 gives an overview on the mean of working instructions by technology for different sociodemographic groups. Regarding ICT, male participants report slightly more instructions by ICT than women. Among all groups, people aged 50 and over report a slightly higher level of instructions than the other age groups. Between the different qualification levels, there is a slight but continuous decrease in instructions through ICT as the qualification level increases. Employees in occupations with higher qualification requirements report, on average, fewer instructions than those with low qualification levels. Throughout the different occupational sectors, the most instructions through ICT are reported in the production manufacturing jobs. People in other economic service occupations report the second highest value.

In case of instructions given by machines, women report a slightly higher level on average than men. Among the different age groups, the lowest level of control by machines is seen in the group under 35 years of age. The other two groups report an almost identical mean. Surprisingly, a different picture emerges regarding the skill requirements for machines compared to ICT. The highest mean level of instructions by machines is reported by master craftsmen and technicians, the group with a rather higher level of qualification and typically associated with less standardized tasks. In terms of occupational sectors, people in other business services report the highest level of instruction by machines, while the other sectors are at a similar level.

To explore the potential impact of technology in control, linear regression in separate models was used to predict the impact of reported instruction by technology on several aspects of work intensity, job control and burnout indicators. These items are based on the Job-Demand-Resources model as key factors of potential stressors and beneficial resources at work. According to the Job-Demand-Control model by Karasek (1979) as well as the Job-Demand-Resources model by Demerouti et al. (2001), an unfavorable constellation of demands and low resources, especially in the long run, leads to a decrease in health associated variables. Methodologically, the use of parametric tests has advantages and disadvantages over non-parametric tests for likert scale-data, depending on the sample, the items and the research question. After weighing these factors, especially the sample size and the item design which does not include verbal gradations of the items, in this study we follow the argumentation of Norman (2010) and use linear regression as a robust parametric test method for calculation. **Table 3** shows the regression coefficients, standard error and standardized regression coefficients beta for instructions by ICT (left) and instructions by machines (right).

TABLE 1 | Sample description.

Sample	%	*n*
Total		6,153
Female	46.5	2,861
Age: 18–34	16.0	982
Age: 35–49	38.6	2,378
Age: ≥50	45.4	2,794
Qualification: No degree	6.5	399
Qualification: Apprenticeship/vocational	48.3	2,972
Qualification: Meister/Technician	14.3	881
Qualification: University degree	30.7	1,894
Working with ICT (at least rarely)	90.8	5,590
Working with machines (at least rarely)	49.2	3,026

More Physical Stress With Instructions by ICT

The statistical models prove that instructions by ICT are a significant predictor for multiple aspects of work intensity as well as all facets of job control. More specifically, regarding work intensity, a higher degree of instructions by ICT is associated with more physical stress (**Figure 4**). Surprisingly, higher levels of instructions by ICT are also connected with mildly less multitasking, which might indicate that work is more standardized and closely supervised with less parallel subtasks when automated. This would indicate that a high LOA regarding decision-making is implemented which leaves the human with more focus on (physical) action. This interpretation would be in line with the results regarding job control. Here, instructions by ICT predict all facets and high levels are associated with less freedom in organizing one's work, influencing the working speed, the possibility of choosing between different task approaches and influencing the amount of work. The strongest relation exists for repetition of working steps, where higher levels of instructions by ICT predict a substantial higher level of repetition of working steps (**Figure 5**). Regarding mental health, however, no relation is found between the instructions by ICT and indicators of burnout. This goes against the assumptions of Ulich (2005), Karasek (1979), Hackman and Oldham (1975) or Warr (1987) since they all propose a negative influence of low levels of job control on mental health. However, the model by Demerouti et al. (2001) could provide an explanation for the missing link

TABLE 2 | Sociodemographics.

	Instructions by ICT			Instructions by machines		
	Mean	**SD**	***n***	**Mean**	**SD**	***n***
Total	2.28	1.25	5,446	2.24	1.29	2,315
Gender: male	2.30	1.26	3,038	2.18	1.27	1,638
Gender: female	2.25	1.26	2,551	2.42	1.32	723
Age: 18–34	2.25	1.24	891	2.09	1.21	485
Age: 35–49	2.22	1.25	2,186	2.29	1.33	947
Age: 50–65	2.34	1.27	2,511	2.30	1.29	929
Qualification: no qualification	2.57	1.61	179	2.24	1.45	123
Qualification: apprenticeship/Vocational	2.37	1.33	2,608	2.23	1.32	1,205
Qualification: master craftsmen/technician	2.23	1.16	1,308	2.35	1.31	513
Qualification: university degree	2.11	1.11	1,350	2.15	1.14	473
Branch: manufacturing jobs	2.41	1.27	1,433	2.20	1.29	1,163
Branch: personal services	2.21	1.33	1,193	2.24	1.28	508
Branch: commercial company-related services	2.22	1.13	1,927	2.29	1.30	282
Branch: IT and scientific service professions	2.22	1.20	376	2.29	1.19	199
Branch: other economic services	2.30	1.50	515	2.40	1.38	161

TABLE 3 | Linear regressions.

Independent variable	**Instructions by ICT**			**Instructions by machines**		
	Regr. coeff	**Std. error**	**Beta**	**Regr. coeff**	**Std. error**	**Beta**
Dependent variable						
Work intensity						
Physical stress	0.109	0.014	0.101***	0.027	0.020	0.027
Multitasking	−0.030	0.010	−0.039**	0.016	0.016	0.021
Interruptions	−0.008	0.011	−0.010	0.035	0.016	0.045***
Information overload	−0.003	0.010	−0.004	0.094	0.015	0.131***
Job control						
Organizing work	−0.133	0.012	−0.147***	−0.054	0.020	−0.055**
Working speed	−0.094	0.012	−0.101***	−0.090	0.021	−0.090***
Task approach	−0.118	0.013	−0.125***	−0.064	0.020	−0.068**
Amount of work	−0.057	0.013	−0.058***	−0.058	0.020	−0.058**
Repetition of working steps	0.138	0.010	0.179***	0.139	0.015	0.187***
Burnout indicators						
Physical exhaustion	0.022	0.012	0.025	0.090	0.018	0.101***
Emotional exhaustion	−0.015	0.012	−0.017	−0.031	0.018	−0.036
Feeling drained	0.016	0.012	0.017	0.064	0.019	0.071**

2,341 < n < 5,592.
**p < 0.05, **p < 0.01, ***p < 0.001.*

as it suggest that existing resources can alleviate the negative effects of stressors, such as low job control. Possibly, employees that work with ICT have more personal resources or better social or management support. The models show no significant predictions for interruptions and information overload.

The presented models on LOAs can only be applied partly to these results, as they focus on the technical implementation and are task specific. They assume that an increase in workload after task automation is an indicator for an incorrect choice of task or for a level of automation that is picked too high.

More Information Overload With Instructions by Machines

When predicting work intensity regarding varying amounts of instructions by machines, a different pattern emerges. Higher levels of instructions by machines are associated with significantly more interruptions and more information overload (**Figure 6**). There was no significant prediction for physical stress or multitasking, however. Again, theories on LOAs would argue that these results are an indicator for a wrongly chosen task to be automatized

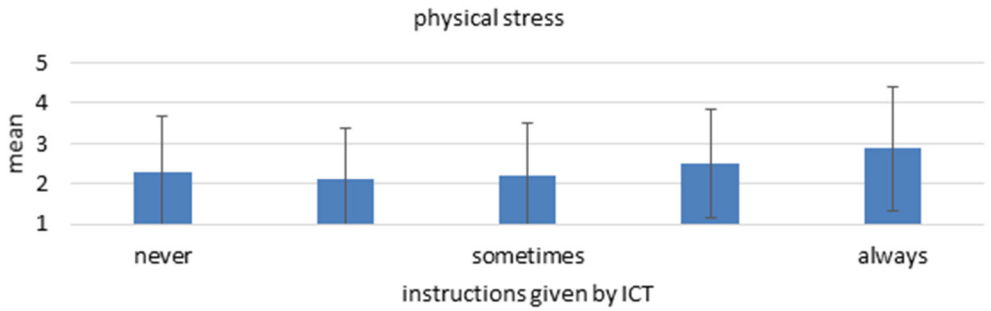


FIGURE 4 | Mean and standard deviation of the item "physical stress" among the different "Instructions by ICT" groups. $n = 5.586$, linear regression coefficient beta = 0.101, $p < 0.001$.

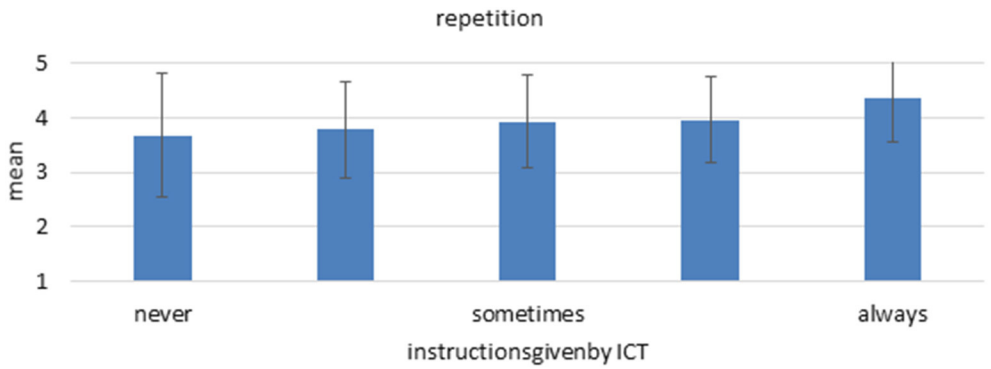


FIGURE 5 | Mean and standard deviation of the item "repetition" among the different "instructions by ICT" groups. $n = 5,587$, linear regression coefficient beta = 0.179, $p < 0.001$.

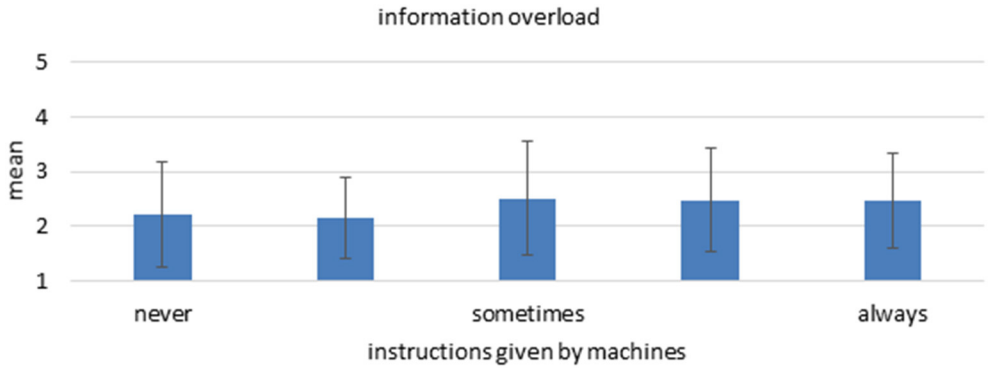


FIGURE 6 | Mean and standard deviation of the item "information overload" among the different "instructions by machines" groups. $n = 2,355$, linear regression coefficient beta = 0.129, $p < 0.001$.

or a higher level of automation than would be necessary or beneficial.

Regarding decision latitude and overall job control, results show an identical pattern between instructions by ICT and machines. Participants reporting more instructions by machines also report significantly less job control among all facets, although some predictions are weaker than those of instructions by ICT. The strongest relation is again found between instructions by machines and repetition, where more instructions are significantly associated with more repetition of working steps (**Figure 7**). Participants, who reported that they always receive instructions by machines, also reported usually executing the same subtask over and over again. It again highlights the assumptions in the Parasuraman model that while automation of *already* redundant tasks is beneficial for decision latitude and performance,

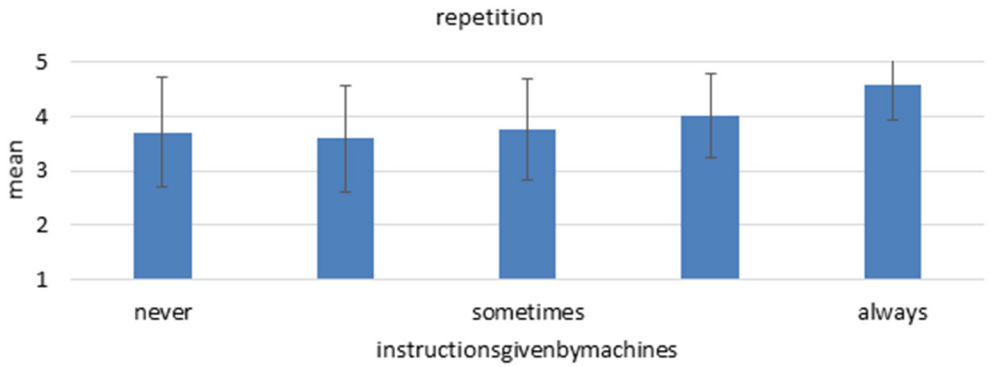


FIGURE 7 | Mean and standard deviation of the item "repetition" among the different "instructions by machines." $n = 2{,}361$, linear regression coefficient beta $= 0.184$, $p < 0.001$.

automation of decision-making often does not lead to better working conditions.

Instructions by machines proved also to predict two out of three burnout screening items. High level of instructions by machines were associated with more feeling of being drained and more physical exhaustion (**Figure 8**). As a higher level of instructions is correlated with job control, the shown negative impact on mental health is according to the presented models by Ulich (2005), Karasek (1979), Demerouti et al. (2001) and Hackman and Oldham (1975). The vitamin model by Warr (1987) predicts that too much control also can have negative effects, which can be partly seen in the present data. Furthermore, the authors of theories on LOAs founded their models on the fact that redundant working conditions and less decision autonomy has detrimental effects on workers. Therefore, the present results go in line with these theoretical considerations as well as with other studies in laboratory settings (Kaber and Endsley, 1997; Endsley and Kaber, 1999; Parasuraman et al., 2000; Weyer et al., 2015). However, according to Parasuraman et al. (2000) as well as Kaber and Endsley (1997), these negative effects occur only in case of weak technical reliability, wrong task selection or overly high level of automation.

In sum, the analysis provides a broad, explorative overview of the extent to which technologies currently exert control over employees' work activities and what working conditions go hand in hand with this. Overall, it shows that partial control of employees by automation technologies is already part of everyday working life. On average, the participants state that they receive "rarely" to "sometimes" instructions by technology about their next work steps. Older and lower-skilled employees are on average affected by instructions through technology slightly more often than other workers. With regard to the correlations of control through technology with relevant working conditions and indicators of mental health, a distinction must be made between the basic types of automation technologies. Different patterns can be found for ICT used for the automation of information-related tasks compared to technologies like production machines used for processing physical objects. While control by ICT systems is associated with a higher degree of physical stress and less multitasking, higher control by machines predicts more interruptions and an increase in information overload. In contrast, the correlations with job control such as facets freedom of action and degree of repetition in the activity as a whole, are similar. Here, more control by technology in both classes is associated with a decrease in job control. The results imply that control by technology does not only substitute control that was previously exercised by humans. Instead, control by technology seems either adding to existing control by superiors, or it seems to be associated with tighter instructions. It is striking that some facets of job design are already rated in the lower (autonomy and decision latitude) or upper (degree of repetition) range of the scale, and are rated even more extremely by participants that report more intensive decisions through technology.

These results are important, as a minimum level of autonomy is a relevant factor for the mental health of employees and represents a long-term risk for mental illness. Regarding control by machines, results indeed indicate a connection between the degree of control and screening facets of burnout symptoms, where more instructions by machines are associated with more physical exhaustion and more feelings of being drained. Overall, the results therefore point to a worsening of working conditions rather than an improvement, if more decisions are made by technology.

Limitations and Evaluations of the Results

Regarding limitations of the study, several should be noted. Firstly, all empirical results are based on subjective data and are therefore prone to specific measurement error, for example due to biased personal perception of the situation. Within large samples, one can assume that random measurement error is somewhat nullified by large numbers. However, there is always a risk for a systematic error to distort the results, as employees are not randomly assigned to workplaces. Therefore, certain groups of employees might answer the questions systematically different than other subgroups due to confounding variables beyond the ones that were included as control variables, for example personal work motivation. There are approaches to handle this possible

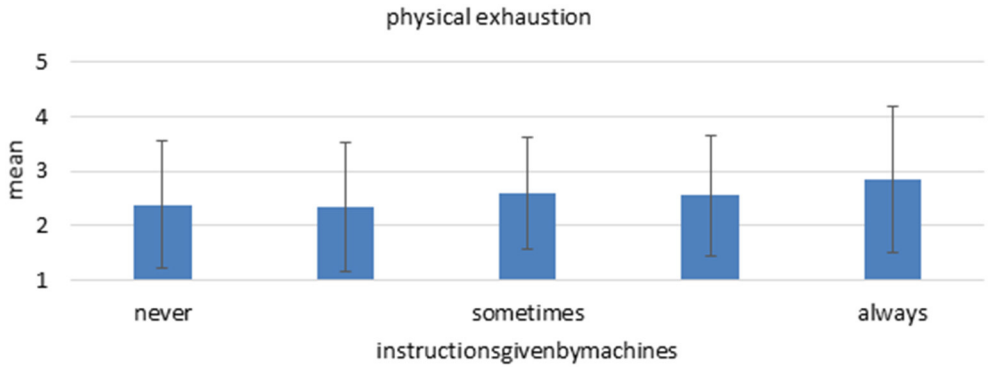


FIGURE 8 | Mean and standard deviation of the item "physical exhaustion" among the different "instructions by machines." $n = 2{,}356$, linear regression coefficient beta $= 0.101$, $p < 0.001$.

error by including employees wages and personal work histories (for example Böckerman et al., 2012). However, as the data set include thousands of employees, we did not use indirect methods to control for variables that were not directly in the data.

Additionally, the results cannot make a clear statement about the extent to which the correlations are due to decision-making by machines *per se*, or due to the specific design and implementation of technologies. Also, due to the exploratory nature of the analysis, a more in-depth investigation of individual subgroups of employees effects was not yet undertaken and will be subject of future research. For example, correlations might be substantially different for different subgroups regarding age, health, qualifications level, company, personality traits and so on. Due to these aspects, we rate the external validity of the data as medium. The big sample, careful sampling and weighting of the data leads to generally high global validity, limited by a non-randomized setup and subjective data. In addition, due to the high abstraction level and the therefore very heterogenic sample, individual results in a specific setting might differ substantially.

DISCUSSION

The digital transformation of work is apparent across all sectors and therewith, entails fundamental changes for the world of work and society. Multiple aspects of today's work, including task characteristics, work environment, health and safety can profit from digitalization and automation in terms of increased productivity, more creative freedom in organizing work and new job opportunities. However, this shift in digitalization can also pose risks and challenges for workers when they are not included in the process and changes are not anticipated correctly. Due to the extraordinary increase in computational power, AI-based systems get more available, complex and capable day by day and therefore hold the potential to qualitatively impact occupational safety and health. AI-based systems are built to automate certain tasks and are even able to work autonomously with little human control, which can be a threat to job autonomy. Theories and models from the field of occupational psychology have argued that a decrease in the factor of job control which involves the possibility to choose tasks, working methods and procedure as well as decision autonomy has detrimental effects on workers' wellbeing. Consequently, stakeholders contribute to the prospect of maintaining the workers' autonomy albeit increasing automation and summarize this aspiration with the human in control principle. The mentioned models agree on the fact that autonomy is a fundamental aspect of good working conditions and is crucial to ensure motivation, job satisfaction and mental health. However, the models are not AI-specific and do not include any specific technical considerations nor focus on task but rather on job level. That is, they are possibly not able to fully explain the changes in the world of work regarding the digitalization of tasks. As automation can foster and decrease the factor of job control, the influence of varying degrees of automation, moderated by perceived autonomy, and on workers' wellbeing and mental health might not be directly visible. As seen in the large-scale study on German workers (DiWaBe), this seems to be the case. More instructions by ICT were correlated with lower levels of perceived job control but did not influence mental health factors. Interestingly, more instructions by machines affected the feeling of control negatively and as predicted by the mentioned models, mental health. Therefore, models on the factor of job control can only partly explain the influence of AI on employees in actual working situations in the present survey. There are other factors such as task significance, feedback or social support that contribute to the overall working conditions and have not been included in the survey which could explain the missing link to mental health factors. The Job-Demand-Resources model by Demerouti et al., 2001 would support this assumption as it proposes that other interacting work conditions can function as resources, which have the ability to balance out demands, such as the decrease on job control. However, this would not account for the differences between instruction by ICT or machines. When looking at the models on LOAs that focus on the technical implementation of automation, a clear focus on task specific automation becomes apparent. They do not differentiate between different types of AI-based systems or sectors but rather dedicate attention to the specific human ability that is automated. According to these models, it is highly

important which subtasks are automated in order to foresee the impact on workers. Both models assume that the automation of redundant tasks influences working conditions positively when the technology is reliable while taking away control from the human in tasks of expertise, has negative impacts. Overall, they only take away the decisional power from the human on the highest level, that is, under full automation. For all other levels, the human remains with a certain degree of decisional power. These models portray the optimal way of using automation in order to foster human performance while decreasing the negative effects it can have, such as a lower perceived level of job autonomy and control.

Results indicate that automation in occupational practice does not happen fully in line with this postulated model of automation. Instead, a substantial part of automation happens at the decision-making level, while executive actions remain with the human. The question remains why this process has led to significant effects on mental health factors when instructions came from machines, compared to instructions by ICT. According to all mentioned models, the reduction of perceived job control should have influenced mental health factors in both cases negatively if there are no other positive factors for workers that got instructions by ICT which would alleviate the impact of a reduced feeling of control. A possible explanation might be that work with ICT is accompanied with higher average levels of job control, so that a reduction by more instructions by technology does not lead to a critical level. This also emphasizes the application of the presented theories and models not on a broad overall level, but when considering the specific working task.

CONCLUSION

Models and theories on human in control draw on well-established research in occupational psychology. In sum, literature has proven that less control and autonomy has negative effects on workers' job satisfaction, performance and mental health. These models clearly show the importance of the factor job control, as well as other factors, such as task significance, feedback and task variety. Due to more automation in the world of work and overall higher degrees of digitalization, automation technologies often take over different subtasks from humans. This happens on varying levels, sometimes leaving the human with supervisory tasks or simply following instructions. This transformation has led to the justified fear of loss of control in workers. Indeed, recent studies showed that a higher degree of automation can have detrimental effects such as loss of control, complacency, reduced situational awareness and task variety. Models on LOAs have therefore taken on the challenge to create an optimal pattern for task automation in which humans can remain in control while aided by technology to increase performance and optimize workload and the mentioned effects. However, they are very task specific and entail multiple loops to evaluate the degree to which the automation influences human performance. Unfortunately, they do not give specific guidelines for different tasks or sectors so that each task with a change in the degree of automation has to pass through the complete theoretical framework in order to have positive implications. The results of the DiWaBe study on German workers shows the large scope of digitalization as more than 90% of people are already working with ICT and nearly 50% with machines. These changes have made it important for stakeholders to highlight the principle of the **human being in control or preserving workers' autonomy** when designing AI-based systems (Rosen et al., 2022).

Although the assumed influence of a decrease in job control on mental health factors seen in the models by Ulich (2005), Karasek (1979) and Demerouti et al., 2001 as well as Hackman and Oldham (1975) cannot be seen consistently in the DiWaBe results, they are visible for workers who get more instructions by machines. This might be due to a higher average level of job control among (knowledge based) ICT-Work than machine work, preventing the demand-resources balance to reach critical levels. As literature emphasizes automation is a double-edged sword, it is crucial to closely monitor changes in automation from an objective point of view, taking productivity, reliability and profitability into account while also looking at automation from a worker's perspective in detail to face challenges for occupational health and safety. Furthermore, fostering positive work conditions such as good social support, feedback as well as opportunities for learning and personal development could provide a higher chance to turn automation into a resource (Demerouti et al., 2001; Demerouti, 2020). The technical models by Parasuraman et al. (2000) and Kaber and Endsley (1997) describe optimal ways when implementing automation, leaving control and supervisory subtasks with human while automating physical subtasks and information gathering.

Unfortunately, results indicate that automation in occupational practice does not happen in line with the models of optimal automation. Instead, there is a substantial level of decision-making by technology, which then exercises control on human employees. In addition, results show that this development is accompanied by a more unfavorable change in terms of demands and resources. Regarding the current rapid development of artificial intelligence, the possibilities to further automate decision-making within work processes will be increased massively, with the risk of more unfavorable working conditions. Therefore, it is of utmost importance from an occupational safety and health perspective to closely monitor and anticipate the implementation of AI in working systems. These results should then be considered continuously by policy making for workplace design, for example regarding in standardization procedures. The goal here is to avoid constellations where employees are too controlled by technology and are left with a high degree of demands and very limited resources. Instead, it would be favorable to use AI as an assistance tool for the employees, helping them to gather and process information and assisting them in decision-making.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: It is planned on a medium-term to publish the data of the DiWaBe survey as weakly

anonymized dataset within the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Germany. Requests to access these datasets should be directed to MH, hartwig.matthias@baua.bund.de.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SW and PR conceptualized the paper's content and structure, and revised the manuscript. MH performed the data analysis. SN and MH wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

The DiWaBe survey was supported by the Interdisciplinary Social Policy Research Funding Network of the German Federal Ministry of Labour and Social Affairs.

REFERENCES

Ajunwa, I. (2020). The "black box" at work. *Big Data Soc.* 7, 205395172096618. doi: 10.1177/2053951720938093

Arntz, M., Dengler, K., Dorau, R., Gregory, T., Hartwig, M., Helmrich, R., et al. (2020). *Digitalisierung und Wandel der Beschäftigung* (DIWABE): *Eine Datengrundlage für die interdisziplinäre Sozialpolitikforschung*. ZEW Dokumentation.

Bader, V., and Kaiser, S. (2017). Autonomy and control? How heterogeneous sociomaterial assemblages explain paradoxical rationalities in the digital workplace. *Manag. Revue* 28, 338–358. doi: 10.5771/0935-9915-2017-3-338

Berberian, B., Sarrazin, J.-C., Le Blaye, P., and Haggard, P. (2012). Automation technology and sense of control: a window on human agency. *PLoS ONE* 7, e34075. doi: 10.1371/journal.pone.0034075

Böckerman, P., Bryson, A., and Ilmakunnas, P. (2012). Does high involvement management improve worker wellbeing? *J. Econ. Behav. Organ.* 84, 660–680. doi: 10.1016/j.jebo.2012.09.005

Demerouti, E. (2020). Turn digitalization and automation to a job resource. *Appl. Psychol*. doi: 10.1111/apps.12270 [Epub ahead of Print].

Demerouti, E., Bakker, A. B., Nachreiner, F., and Schaufeli, W. B. (2001). The job demands-resources model of burnout. *J. Appl. Psychol.* 86, 499. doi: 10.1037/0021-9010.86.3.499

Dwyer, D. J., and Ganster, D. C. (1991). The effects of job demands and control on employee attendance and satisfaction. *J. Organ. Behav.* 12, 595–608. doi: 10.1002/job.4030120704

Endsley, M. R., and Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics* 42, 462–492. doi: 10.1080/001401399185595

ETUC (2020). *Social Partners Agreement on Digitalisation (FAD)*. Brussels: European Trade Union Confederation. Available online at: https://www.etuc.org/en/document/eu-social-partners-agreement-digitalisation (accessed December 2, 2021).

EU (2019). *High-level Expert Group on Artificial Intelligence*. Brussels: European Commission. Available online at: https://ec. europa. eu/digital-single

European Commission. (2021). *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. Available online at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

Fréour, L., Pohl, S., and Battistelli, A. (2021). How digital technologies modify the work characteristics: a preliminary study. *Span. J. Psychol.* 24, S.1–21. doi: 10.1017/SJP.2021.12 (accessed December 14, 2021).

Gagné, M., Senecal, C. B., and Koestner, R. (1997). Proximal job characteristics, feelings of empowerment, and intrinsic motivation: a multidimensional model 1. *J. Appl. Soc. Psychol.* 27, 1222–1240. doi: 10.1111/j.1559-1816.1997.tb01803.x

Gouraud, J., Delorme, A., and Berberian, B. (2017). Autopilot, mind wandering, and the out of the loop performance problem. *Front. Neurosci.* 11, 541. doi: 10.3389/fnins.2017.00541

Hackman, J. R., and Oldham, G. R. (1975). Development of the job diagnostic survey. *J. Appl. Psychol.* 60, 159. doi: 10.1037/h0076546

Hämäläinen, R., Lanz, M., and Koskinen, K. T. (2018). "Collaborative systems and environments for future working life: towards the integration of workers, systems and manufacturing environments," in *The Impact of Digitalization in the Workplace* (Springer), 25–38. doi: 10.1007/978-3-319-63257-5_3

Inoue, A., Kawakami, N., Haratani, T., Kobayashi, F., Ishizaki, M., Hayashi, T., et al. (2010). Job stressors and long-term sick leave due to depressive disorders among Japanese male employees: findings from the Japan Work Stress and Health Cohort study. *J. Epidemiol. Commun. Health* 64, 229–235. doi: 10.1136/jech.2008.085548

Kaber, D. B., and Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Saf. Prog.* 16, 126–131. doi: 10.1002/prs.680160304

Kaber, D. B., and Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theor. Issues Ergon. Sci.* 5, 113–153. doi: 10.1080/1463922021000054335

Kaber, D. B., Stoll, N., Thurow, K., Green, R. S., Kim, S.-H., and Mosaly, P. (2009). Human–automation interaction strategies and models for life science applications. *Hum. Factors Ergon. Manuf. Serv. Ind.* 19, 601–621. doi: 10.1002/hfm.20156

Karasek, R., and Theorell, T. (1990). Healthy work: stress, productivity, and the reconstruction of working life. *Natl. Prod. Rev.* 9, 475–479. doi: 10.1002/npr.4040090411

Karasek, R. A. Jr. (1979). Job demands, job decision latitude, and mental strain: implications for job redesign. *Adm. Sci. Q.* 285–308. doi: 10.2307/2392498

Kristensen, T. S., Hannerz, H., Høgh, A., and Borg, V. (2005). The Copenhagen Psychosocial Questionnaire-a tool for the assessment and improvement of the psychosocial work environment. *Scand. J. Work Environ. Health* 31, 438–449. doi: 10.5271/sjweh.948

McCormick, E. J., and Sanders, M. S. (1982). *Human Factors in Engineering and Design*. New York, NY: McGraw-Hill Companies.

Melamed, S., Ben-Avi, I., Luz, J., and Green, M. S. (1995). Objective and subjective work monotony: effects on job satisfaction, psychological distress, and absenteeism in blue-collar workers. *J. Appl. Psychol.* 80, 29. doi: 10.1037/0021-9010.80.1.29

Moore, P. V. (2019). "OSH and the future of work: benefits and risks of artificial intelligence tools in workplaces," in *International Conference on Human-Computer Interaction. Symposium Conducted at the Meeting of Springer*. doi: 10.1007/978-3-030-22216-1_22

Morgeson, F. P., Delaney-Klinger, K., and Hemingway, M. A. (2005). The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *J. Appl. Psychol.* 90, 399. doi: 10.1037/0021-9010.90.2.399

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ. Theory Pract.* 15, 625–632. doi: 10.1007/s10459-010-9222-y

OECD (2019). *AI Principles Overview*. Available online at: https://www.oecd.ai/wonk/a-first-look-at-the-oecds-framework-for-the-classification-of-ai-systems-for-policymakers (accessed December 14, 2021).

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern.: Syst. – Part A: Syst. Hum.* 30, 286–297. doi: 10.1109/3468.844354

Robelski, S. (2016). *Psychische Gesundheit in der Arbeitswelt Mensch-Maschine-Interaktion*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Rosen, P. H., Heinold, E., Fries-Tersch, E., Moore, P., and Wischniewski, S. (2022). *Advanced Robotics, Artificial Intelligence and the Automation of Tasks: Definitions, Uses, Policies, Strategies and Occupational Safety, and Health*. Report commissioned by the European Agency for Safety and Health at Work (EU-OSHA).

Rosen, P. H., and Wischniewski, S. (2019). Scoping review on job control and occupational health in the manufacturing context. *Int. J. Adv. Manuf. Technol.* 102, 2285–2296. doi: 10.1007/s00170-018-03271-z

Semmer, N. K. (1990). "Stress und Kontrollverlust," *Das Bild Der Arbeit*, eds F. Frei, and I. Udris (Huber), 190–207.

Spector, P. E. (1998). "A control theory of the job stress process," in *Theories of Organizational Stress* 153–169.

Ter Hoeven, C. L., van Zoonen, W., and Fonner, K. L. (2016). The practical paradox of technology: the influence of communication technology use on employee burnout and engagement. *Commun. Monogr.* 83, 239–263. doi: 10.1080/03637751.2015.1133920

Ulich, E. (2005). *Arbeitspsychologie. 6. überarbeitete und erweiterte Auflage*. Stuttgart: Schäffer-Poeschel.

Wang, B., Liu, Y., and Parker, S. K. (2020). How does the use of information communication technology affect individuals? A work design perspective. *Acad. Manag. Ann.* 14, 695–725. doi: 10.5465/annals.2018.0127

Warr, P. (1987). *Work, Unemployment, and Mental Health*. Oxford: Oxford University Press.

Weyer, J., Fink, R. D., and Adelt, F. (2015). Human–machine cooperation in smart cars. An empirical investigation of the loss-of-control thesis. *Saf. Sci.* 72, 199–208. doi: 10.1016/j.ssci.2014.09.004

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., and Sarter, N. B. (2010). "Stages and levels of automation: an integrated meta-analysis," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Symposium Conducted at the Meeting of Sage Publications* (Los Angeles, CA: Sage). doi: 10.1177/154193121005400425

Wiener, E. L., and Curry, R. E. (1980). Flight-deck automation: promises and problems. *Ergonomics* 23, 995–1011 doi: 10.1080/00140138008924809

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Niehaus, Hartwig, Rosen and Wischniewski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/frai.2022.869114

Politics by Automatic Means? A Critique of Artificial Intelligence Ethics at Work

Matthew Cole, Callum Cant, Funda Ustek Spilda and Mark Graham*

Social Sciences Division, Oxford Internet Institute, University of Oxford, Oxford, United Kingdom

Calls for "ethical Artificial Intelligence" are legion, with a recent proliferation of government and industry guidelines attempting to establish ethical rules and boundaries for this new technology. With few exceptions, they interpret Artificial Intelligence (AI) ethics narrowly in a liberal political framework of privacy concerns, transparency, governance and non-discrimination. One of the main hurdles to establishing "ethical AI" remains how to operationalize high-level principles such that they translate to technology design, development and use in the labor process. This is because organizations can end up interpreting ethics in an *ad-hoc* way with no oversight, treating ethics as simply another technological problem with technological solutions, and regulations have been largely detached from the issues AI presents for workers. There is a distinct lack of supra-national standards for fair, decent, or just AI in contexts where people depend on and work in tandem with it. Topics such as discrimination and bias in job allocation, surveillance and control in the labor process, and quantification of work have received significant attention, yet questions around AI and job quality and working conditions have not. This has left workers exposed to potential risks and harms of AI. In this paper, we provide a critique of relevant academic literature and policies related to AI ethics. We then identify a set of principles that could facilitate fairer working conditions with AI. As part of a broader research initiative with the Global Partnership on Artificial Intelligence, we propose a set of accountability mechanisms to ensure AI systems foster fairer working conditions. Such processes are aimed at reshaping the social impact of technology from the point of inception to set a research agenda for the future. As such, the key contribution of the paper is how to bridge from abstract ethical principles to operationalizable processes in the vast field of AI and new technology at work.

Keywords: artificial intelligence, labor, work, ethics, technological change, collective bargaining, industrial relations, job quality

OPEN ACCESS

Edited by:
Phoebe V. Moore,
University of Essex, United Kingdom

Reviewed by:
Alexander Nikolaevich Raikov,
V. A. Trapeznikov Institute of Control Sciences (RAS), Russia
Kokil Jaidka,
National University of Singapore, Singapore

****Correspondence:***
Matthew Cole
matthew.cole@oii.ox.ac.uk

Specialty section:
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

Received: *03 February 2022*
Accepted: *30 May 2022*
Published: *15 July 2022*

Citation:
Cole M, Cant C, Ustek Spilda F and Graham M (2022) Politics by Automatic Means? A Critique of Artificial Intelligence Ethics at Work. Front. Artif. Intell. 5:869114. doi: 10.3389/frai.2022.869114

INTRODUCTION

The advent of a new era of innovation in machine learning AI and its diffusion has prompted much speculation about how it is reshaping society (Gentili et al., 2020). As well as seeing it as an opportunity to advance common social goals, many have also identified how such developments may pose significant risks, particularly to actors who are already disempowered and discriminated against. Consequently, much thought has gone into the risks and opportunities of AI,

and the creation of principles for its ethical development and deployment. However, this thought tends to be at the intersection of the instrumental-economic and abstract ethics (Algorithm Watch, 2020), with operationalization generally restricted to the domain of privacy concerns, transparency and discrimination. Questions around workers' fundamental rights, job quality (see Cazes et al., 2015) and working conditions more generally have not received as much attention.

Given that technologies tend to be path-dependent (MacKenzie and Wajcman, 1999), embedding a set of concrete principles and benchmarks from the outset of technological diffusion is an important way to control their social effects as it supports regulation. There is an urgent need to create a set of evaluation mechanisms that directly address the impact of AI on working conditions, and that can feed into regulation of these technologies. However, research on this topic is limited. A Scopus query for the term "AI ethics" retrieves 2,922 results. When "work" is added to the search string, this number drops by more than half, to 1,321 results. Of these, 309 are publications in the social sciences, indicating limited engagement of our field with the topic. When analyzed in detail, we see that only 58 of them discuss AI ethics pertaining to work and employment. Most of these focus on digital wellbeing (Burr et al., 2020) or worker wellbeing (Nazareno and Schiff, 2021), the impact of algorithms on decision-making in government, employment agencies (Kuziemski and Misuraca, 2020), predictive policing (Asaro, 2019; Yen and Hung, 2021) and bias in algorithmic decision-making (Hong et al., 2020; Metaxa et al., 2021). The studies that are specifically on work and employment target recruitment (Yam and Skorburg, 2021), human resources management (Bankins, 2021) or workplace management (Jarrahi et al., 2021). There is, thus, a clear gap in the literature concerning how AI ethics relates to fairness, justice and equity in the context of work and employment.

Against this background, this paper sets out a critique and a research agenda to address this gap. However, the pathway from high-level principles to enforceable regulation on working conditions has yet to be clearly defined. As noted by Wagner (2018, 2019), the current focus on AI ethics is simply a watered-down version of regulation—especially when technology companies opt for voluntary codes of practice that they've shaped themselves. As Algorithm Watch (Thiel, 2019) notes, most existing AI ethics guidelines are non-binding, and they operate on an opt-in basis. AI ethics can therefore be something companies congratulate themselves on for their good intentions, without the need to turn these so-called ethics into actual practice. Hence, there is an urgent need to move from abstract principles to concrete processes that ensure compliance. This is a necessary step, irrespective of emerging regulations on the issue.

In this article, we provide a critique of how AI systems are shaping working conditions before identifying ways in which it can foster fairer work (see Section Proliferating Principles). We first review a selection of AI guidelines, ethics and meta-analyses using Boolean search, and outline four critiques that cut across the recent proliferation of ethical guidelines. These are summarized by the four headings: (1) Not everything is a trolley problem (ethical critique); (2) AI is not that special (socio-technical critique); (3) The problem with automatic politics (ethico-political critique); (4) Big Tech Ethics is Unilateral (a socio-political critique). These critiques set up methodological basis for the University of Oxford's AI for Fairwork project (supported by the Global Partnership on AI), which aims to produce a set of AI guidelines that avoid these pitfalls and contribute to fairer uses of AI at work. These guidelines, a draft of which are presented in Section Proposed AI for Fairwork Standards below, are not exclusively intended to assist in either risk mitigation or opportunity maximization. Instead, we view those two goals as inseparably linked. By shifting our attention from mere negative outcomes of technological development to the processes of technological innovation and design, we aim to embed fairness into the very technologies that get built, instead of attempting to fix problems once and as they arise.

We position our understanding of fairness as both as an ethical absolute that should be strived for, but also as a virtue that is context dependent to time, space and conditions. As such, we treat fairness not as a static and unchanging category or end point in itself, but rather as a process that involves continual revision relative to material circumstances. To use the language of Silicon Valley: making things fairer is an iterative process. Agents are required to constantly attempt to move toward a horizon of fairness that they can't quite reach. This will likely be the case for a long time to come, as we can foresee no final point at which any work environment could be declared completely fair—at least, not under this economic and political system.

FROM ETHICS TO FAIRNESS

Proliferating Principles

The proliferation of real (or speculative) AI use-cases and corresponding national industrial strategies (HM Government, 2021), has provoked a swath of voluntarist ethical guidelines from an array of actors, from the OECD to the European Parliament, Microsoft, and even the Pope. Governments, supra-national institutions and non-governmental organizations have all shown an interest in understanding and regulating AI systems. In this section, we review a selection of such principles that are most relevant to our research and the development of Fairwork principles for AI. By investigating a selection of these principles more closely, we can lay the groundwork for our subsequent critique.

The OECD (2019) Principles on Artificial Intelligence were the first AI ethics guidelines signed up to by governments. They complement existing OECD standards in areas such as privacy, digital security risk management, and responsible business conduct (see **Table 1**). The G20 also adopted "human-centered AI principles" that drew on these principles. In a similar vein, the European Parliament (European Parliament, 2019) has drawn up a code of voluntary ethics guidelines for AI and robotics engineers involving seven key requirements (see **Table 1**). Such requirements informed the 33 policy and investment recommendations that guide the proposal for "Trustworthy AI" put forward by the EU High-Level Expert

TABLE 1 | Summary of four illustrative AI principles.

Principle author	Stated values	Specific discussion of work
OECD	(1) Regular engagement of multiple external and internal stakeholders; (2) mechanisms for independent oversight; (3) transparency around decision-making procedures; (4) justifiable standards based on evidence; (5) clear, enforceable legal frameworks and regulations.	Affirms the importance of international labor rights. Suggests that workers should be aware of their interactions with AI systems. Encourages governments to prepare for "labor market transition" through skill development social dialogue, and promoting increases in safety and job quality.
UNESCO	(1) Proportionality and "do no harm"; (2) safety and security, fairness and non-discrimination; (3) sustainability, right to privacy and data protection; (4) human oversight and determination; (5) transparency and explainability; (6) responsibility and accountability, awareness and literacy; (7) multistakeholder and adaptive governance and collaboration.	Encourages governments to implement impact assessments that monitor, amongst other things, the effect of AI on labor rights, Strongly emphasizes the need for skill development, retraining and "fair transition" for at-risk employees. States the need for ongoing research on the impact of AI systems on work.
European Parliament	(1) Human agency and oversight; (2) robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, non-discrimination and fairness; (6) societal and environmental well-being; (7) accountability	Notes concern about impact on labor market and describes workers as one of nine relevant stakeholder groups.
President of the United States	(1) Lawful and respectful of our Nation's values; (2) purposeful and performance-driven; (3) accurate, reliable, and effective; (4) safe, secure, and resilient; (5) understandable; (6) responsible and traceable; (7) regularly monitored; transparent; (8) accountable.	None.

Group on Artificial Intelligence (AI HLEG) and their self-assessment checklist (High-Level Expert Group on AI, 2020). The European Commission wants "Trustworthy AI" that puts "people first" (European Commission, 2020). However, the EU's overall approach emphasizes the commercial and geopolitical imperative to lead the "AI revolution", rather than considering in detail the technological impact on workers and work. It has been noted that members of the AI HLEG have already condemned the results as an exercise in ethics washing (Metzinger, 2019).

Following this trend, UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence' also emphasize the production of "human-centered AI" around 10 principles. UNESCO also proposes a set of 11 policy areas aligned with these fundamental principles for member states to consider. The President of the United States issued an "Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government", that provided "Principles for the use of AI in Government" (White House, 2020).

The US executive order was likely in response to the Algorithmic Accountability Act proposed on April 10, 2019 in the United States Congress, which aimed to legislate rules for evaluating highly sensitive automated systems. The Act was never taken to a vote, but a new version has recently been introduced on March 1, 2022 that aims "To direct the Federal Trade Commission to require impact assessments of automated decision systems and augmented critical decision processes, and for other purposes" (United States Congress, 2022). This legislation aims to increase certain kinds of transparency with regard to automated decisions affecting US citizens from the use to algorithms. It requires both the firm that builds the technology enabling the automation as well as the company using it to make the decision to conduct impact assessments for a range of factors including bias, effectiveness, and security. Furthermore, the bill aims to provide a benchmark requirement which stipulates that companies assess the impacts not only of new automated decision-making processes, but also already-existing ones. It mandates that the Federal Trade Commission (FTC) creates regulations that standardizes assessment and reporting, requires auditing of impact-assessment and the FTC to publish an annual anonymized report on trends and provide a public dataset of automation decision technologies.

Ethical principles have proliferated to such a degree that there are now multiple databases cataloging them online. One such inventory of AI Ethics Guidelines is crowd sourced and maintained by the NGO Algorithm Watch. The database currently identifies 173 sets of principles "of how systems for automated decision-making (ADM) can be developed and implemented ethically". These fall into three broad categories: binding agreements (8), voluntary commitments (44), and recommendations (115). Similarly, the OECD maintains a live database showing over 700 initiatives related to AI policy from 60 countries, territories and the EU.[1] In a recent study, Jobin et al. (2019) identified 84 different ethical AI standards, produced by a range of private companies, government agencies, research institutions, and other organizations. They identified 11 overarching principles, namely (in order of popularity): transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity. Only transparency, justice and fairness, non-maleficence, responsibility, and privacy appeared in most of the standards.

These various efforts to track the ongoing proliferation of guidelines is a useful starting point for thinking about how effective they might be in practice. Mittelstadt et al. (2016) identified six key issues raised by the use of algorithms (which they define in such a way as to include much of what we might call AI): inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, transformative effects, and

[1] https://oecd.ai/en/dashboards.

traceability. These concerns appear to have stayed relatively consistent over time (Roberts et al., 2021), which is somewhat problematic given the limited set of stakeholder perspectives contained in the field of principles (Hickok, 2021). We could summarize that the last 5 years have seen the same people raising the same issues, with limited evidence of progress or widening participation in the discussion.

To these concerns about perspectival limitations, we would also add concerns about the ideological limitations of these principles. The current debate on AI ethics in the literature tends to be limited by ontological, epistemological and political assumptions drawn from classical liberal thought—namely around rights and privacy. The horizon of these principles takes certain conditions as given: private ownership of means of production, capitalist social relations, the institutional reproduction of such relations, the individualist perspective on decision-making and responsibility and the embeddedness of technologies within this context. As a result, many participants in the debate have only been able to consider courses of action which fall within these limitations. An example of how this limited frame raises problems is Oren Etzoni's (2018) "Hippocratic oath for artificial intelligence practitioners". The oath—an attempt to model a framework for AI ethics analogous to that underpinning the medical profession— reads:

> *I will consider the impact of my work on fairness both in perpetuating historical biases, which is caused by the blind extrapolation from past data to future predictions, and in creating new conditions that increase economic or other inequality.*

But as Mittelstadt (2019) argues: the lack of an analogous institutional context to medicine means that the Hippocratic principle-based model of ethical regulation doesn't map well to AI. Indeed, AI development lacks "common aims and fiduciary duties, professional history and norms, proven methods to translate principles into practice, robust legal and professional accountability mechanisms" (Mittelstadt, 2019, p. 1). This problem with simply mapping one domain onto another—and assuming it will work in that new context—points to a broader concern with the impact of guidelines, particularly in the context of working life.

The central question here is how to translate ethical principles into ethical practice (Hagendorff, 2020, 2021). The difficulty of providing a robust answer has been repeatedly identified. Hagendorff (2020) examines whether certain principles have a real-world impact on the ethics of process and outcomes in AI-mediated work and concluded "No, most often not" (Hagendorff, 2020, p. 99). Floridi (2019) identifies risks in the transition from *what* to *how* that include: digital ethics shopping, "bluewashing" (i.e., digital greenwashing), lobbying, ethics dumping (outsourcing to other actors), and shirking. Morley et al. (2021) argue that, while principles are important in defining the normative values against which AI is evaluated, the translation of broad principles into concrete action is difficult. Following the metaphor of cloud computing, they envision a hybrid institutional arrangement that can offer "ethics as a service".

Despite the breadth and depth of work on AI ethics, there remains a profound blind spot in terms of implementation, since organizations are left largely to interpret and enact ethical guidelines themselves and then assess if they are abiding by them. This exposes workers to potential abuse of AI technologies not only in terms of digital Taylorism, but also the degradation of work by reproducing biases and inequalities, intensifying work and denying collective control. Mitchell (2019), (p. 152) highlights the diversity of ethical issues vying for the attention of regulators:

> *Should the immediate focus be on algorithms that can explain their reasoning? On data privacy? On robustness of AI systems to malicious attacks? On bias in AI systems? On the potential "existential risk" from superintelligent AI?*

Yet questions of work and employment are conspicuously absent from both this set of questions and the ethics guidelines mentioned above. Indeed workers and employees are rarely cited when lists of relevant AI stakeholders are listed.

While some legislation relating to AI transparency in the workplace has been passed in certain countries such as Spain and France, further steps are needed to ensure that laws and requirements of this type are enforceable and effective (Algorithm Watch, 2020). Marx [1887] (1976) referred to the sphere of production as "the hidden abode" in order to point out how the purported values of liberalism were restricted to operating only in the market. So far, the field of AI ethics has, by and large, also refused to venture into this black box. Rather than deal with the contentious social relations which structure production and the workplace, the current debate so far has focused its attention on how AI impacts its users in their roles as citizens and consumers, but not as workers.

What Is "Fair" Anyway?

From Aristotle to Rawls, from Fraser to Nussbaum and Sen, fairness and its broader counterpart, justice, have acquired multiple meanings when seen from different philosophical standpoints and in different practical contexts. In all these different interpretations, however, issues of justice emerge in circumstances of scarcity, when there are then potentially conflicting claims to what each person is entitled to, or how institutions can administer fair allocation of resources (Miller, 2021). Thus, fairness for whom, and fair according to what/whose criteria remain as two key questions. In other words, we would not need fairness or justice, if we had unlimited resources and as individuals we had unlimited skills and capabilities. We need fairness and justice because there are limited resources and as humans we have limited capacities (individually). Following from this, in answering how to be a virtuous person for instance, Aristotle counts justice as one of the four seminal virtues a person should have, and notes that it is thought to be "another's good" because it is defined always in relation to another individual, another status and positionality, and as such he conceptualizes justice as proportionality (Aristotle, 2000, p. 73).

Rawls' theory of justice, which remains by far the most referenced theory on the topic, aims to solve the dilemma of

establishing justice in a society where different individuals are all seeking to advance their own interests (e.g., utilitarian, modern, capitalist, and so on). While ultimately Rawls tries to reconcile the freedom of choice for individuals with fair outcome for all (as in a world of scarce resources, the choices of individuals may not always be guaranteed), Rawls presents two informational constraints for individuals in making that choice. He imagines individuals behind a "veil of ignorance" that deprives them of any knowledge of personal characteristics which might make some of the choices more available, more favorable or more easily attainable for some individuals. This ignorance of personal characteristics, skills and capabilities ultimately serves to make individuals base their choices on an impartial principle of what would be fair for everyone. Here, Rawls also suggests that this impartiality can be benchmarked by assuming one must make a choice for the worst off in society. This person, in a hypothetical context, can be the individual making the choice for others (Rawls, 1993).

For Rawls, then fairness implies some level of equal distribution in society of opportunities and resources. Scanlon (1998) argues that individuals will never realistically be able to perform a veil of ignorance because we are all aware of our own relative positions, wants and needs. Instead, he argues for a theory of justice which no one could reasonably reject, even when they are given a right to veto, should they see it as unfair. Philosophers have rightly commented that giving everyone a right to veto will ultimately create a deadlock as anyone can reject a principle which does not treat them favorably (Miller, 2021). However, Scanlon emphasizes that this will not be the case, if the principle of reasonable rejection applies, as individuals will be able to weigh if the current principle seems unfair, if an alternative would involve someone else fairing worse still (Miller, 2021). Scanlon also notes that the right to veto is significant for individuals because if a principle treats them unfairly, such as faring well for some but not others and for arbitrary reasons; individuals should be in a position to reject this (Miller, 2021), unlike in Rawls' theory, whereby individuals would not be in a position to judge whether arbitrariness played any role in individuals' decisions.

In contrast to Rawls and Scanlon, who both argue for a contractual theory of justice, Sen, for instance proposes a more distributive form of justice with the capability approach, explaining that what we need is not a theory that describes a utopian ideal of justice, but one that helps us make comparisons of injustice, and guide us toward a less unjust society (Robeyns and Byskov, 2021). In this regard, Sen (and also to an extent Nussbaum) proposes that the intention of a theory of justice is not necessarily to identify and only aim for the ideal of fairness, but rather identifying and then equipping individuals with the capabilities they would need to strive for lesser injustice, and less unfair societies. Some philosophers have argued that the capability approach overcomes some of the inflexibilities inherent to Rawlsian (or indeed Aristotelean) theories of justice, because it takes into account the different needs, circumstances and priorities of different individuals (Robeyns and Byskov, 2021).

In this paper, we define fairness not by its unchanging absoluteness, but conditionality, contextuality and proportionality based on the circumstances of individual and institutional decision-makers. In this regard, fairness influences the whole decision-making process from ideation to development and execution of AI-based systems, rather than one final goal that can be achieved once and for all. Hence, we focus more on increasing individual and organizational capabilities to guide us toward a less unfair society.

Generally, individuals will bring their own expectations to bear on the meaning of fairness, such that two people may consider the same set of working conditions fair or unfair. In order to overcome such confusion, we use "fair" in the sense of the capability approach outlined above. At the abstract level, we define fairness as direction of travel toward a more just society. When power asymmetries are being undermined through democratization, when opportunities and outcomes are being equalized, when access to self-determination and positive freedom are being opened to a wider range of people, then we consider work to be getting fairer.

Concretely, standards and benchmarks of fairness have a significant role to play as waypoints along this journey. While what qualifies as decent or good quality work can vary between and among different workers, stakeholders and policy-makers, most standards (from the ILO to OECD and Eurofound) involve six key dimensions of job quality: pay and other rewards; intrinsic characteristics of work; terms of employment; health and safety; work–life balance; and representation and voice (Warhurst et al., 2017). In this regard, we begin from the baseline standards of decent work developed by the Fairwork Project, which include fair wages, conditions, contracts, management and representation (Heeks et al., 2021). These principles have evolved over years of action-research and broadly align with the wide-range of job-quality metrics in contemporary academic research.

FOUR CRITIQUES OF ARTIFICIAL INTELLIGENCE ETHICS

Not Everything Is a Trolley Problem (Ethical Critique)

Current AI ethics approaches present a mix of various schools of thoughts in ethics. Sometimes we find schools that have long been in conflict with one another combined to suit the needs of the particular parties who are building the principles. The two most common schools are *consequentialist ethics* (a version of which is utilitarian ethics) and *deontological ethics*. Consequentialist ethics examines the consequences of ethical decisions and asks the ethical agent to make an ethical judgment based on the consequences that are important to her (Sen and Williams, 1982). Utilitarianism (following Bentham) suggests that the most ethical decision would be the one that provides the greatest utility for the greatest number of people. Of course, defining both "utility" and a "number of people affected" are both complex questions. In contrast, deontological ethics disregards the consequences of any ethical decision or the intentions that lead to it, but focuses entirely on the principles instead (Anscombe, 1958). Principles such as "Thou shall not kill" hold irrespective of individual circumstances and particular intentions

of the ethical agent. Finally, virtue ethics (stemming from Aristotle's *Nichomachean Ethics)* argues that the only road to *eudaimonia* (or personal happiness, *flourishing*) is through living in accordance to fulfilling one's virtues (Annas, 2006).

Much of the current ethical thinking with respect to AI ignores the important differences between consequentialist, deontological and virtue ethics, and instead follows a mix and match approach, as it fits the questions and desired outcomes. Most commonly, consequentialism mixed with a touch of deontological ethics based on the assumption of a virtuous actor (e.g., developer, entrepreneur, and investor) in the field of AI is imagined and proposed. In this imaginary, the ethical proposition is done in a way that it does not conflict with or hinder the intention to "move fast and break things" (Ustek-Spilda, 2018).

Consequentialism dominates the discussions also because, in comparison to de-ontological ethics or virtue ethics, it can be seemingly easily translated to decisions that are taken in technology development. This suggests that when the consequences of a decision cannot be predicted fully, then the best option is to hope that the principles that guide that process will avoid the worst possible outcomes. We might very well ask: "worst for whom" and "worst in accordance to what criteria"? Note that this is not a call for subjectivism—that is, the ethical position that all values change from person. To person and there are no objective or absolute values, but simply to note the serious need for identifying how principles can facilitate ethical decision-making, amidst this uncertainty.

For example, the "trolley problem" is used to unpack particular issues identified in AI ethics. This thought experiment concerns a runaway trolley that will kill someone—but where a person can choose between alternative outcomes. In the version developed by MIT Media Lab, the person thinking through the problem is asked to choose between killing young children or the elderly, a small number of disabled people or a higher number of able-bodied people, an overweight person or a fit person.[2] The problem is that by adapting this thought experiment to the AI development context, it simplifies complex decisions into either/or questions. It doesn't allow any room for the possibility of *no one* (for example) being killed. So, there is no room to discuss one of the central questions with AI—whether or not it should actually be built in the first place. Or should a problem which can be fixed with AI, actually be fixed with AI, or should it perhaps not be fixed at all (Penn, 2021).

There are many examples of AI reproducing and/or amplifying patterns of social discrimination, and the thinking used in the trolley problem being extended to solving these problems too. In 2018, for instance, Reuters published a story revealing that Amazon had been forced to scrap an AI recruiting tool that was intended to analyze CVs and score applicants from one to five. Since the tool was trained using training data taken from the hiring process at Amazon over the last 10 years, it faithfully reproduced the bias against women it found therein (Dastin, 2018). Other famous cases of discriminatory AI such as the ProPublica investigation into racial bias and the COMPAS risk assessment software used in bail, probation and sentencing decisions across the US (Angwin et al., 2016) have demonstrated the potential of serious social harms from automated discrimination. However, what is notable in many discussions of these cases is that they focus on how to design *better* hiring and risk assessment software—rather than asking if decisions of this kind should be automated at all. In the workplace context, the failure to ask serious questions about the ethical integrity of decisions to deploy AI can lead to very significant negative consequences for workers. The complexity of questions regarding issues like hiring demands more from us than a mishandled application of the Trolley Problem, or ethical theories being thrown in together just to fit in with the desired framework and outcome of a particular AI ethics project.

AI Is Not That Special (Socio-Technical Critique)

Since its inception by John McCarthy in 1956, the academic field of Artificial Intelligence has been premised on the creation of a software program that can solve not just one narrow kind of problem, but that can apply its capacity for calculation to any kind of problem (Wooldridge, 2021). Such a truly general AI does not currently exist. While the most advanced forms of AI created to date, such as GPT-3 and AlphaGo, can outperform humans in some very limited tasks, they still have near-zero general applicability, and lack the ability to think in a manner which at all reflects the human brain (Chui et al., 2018).

Despite this, AI is often treated as an *exceptional* technology with a universalist or unbounded horizon. Indeed, despite not yet having achieved real AI, the assumption among many is that that is the direction of travel. So, rather than treating AI as a technical field concerned with advanced, non-symbolic, statistical methods to solve specific, bounded problems (facial recognition, natural language processing, etc.), AI positivists identify the field as something unprecedented. AI ethics therefore begins to become orientated toward a hypothetical future scenario, rather than the reality of our present moment.

2012 marked the start of a sea change in how AI practitioners go about their work, and it was enabled by increases in the volume of data, dataset-creating labor and computing power available for the development of AI (Cole et al., 2021). From natural language to game playing and visual object recognition, the turn to deep convolutional neural network and machine learning has allowed for significant progress across the various subfields that make up AI and is the basis for the latest surge in funding and media attention. Justified celebration of these developments has gone hand-in-hand with unjustified hyperbole about the future. In 2017, Ray Kurzweil, Google's Director of Engineering, famously claimed that the "technological singularity" would be achieved by 2045, as we "multiply our effective intelligence a billion-fold by merging with the intelligence we have created" (Reedy, 2017). Such predictions are characteristic of the past decade of AI hype.

A significant number of recent studies have countered this AI hype in fields such as translational medicine (Toh et al., 2019), multidisciplinary medical teams (Di Ieva, 2019), radiology (Rockall, 2020), COVID-19 (Abdulkareem and Petersen, 2021),

[2]See https://www.moralmachine.net/.

machine vision (Marquardt, 2020), management (Holmström and Hällgren, 2021), and interaction design (Liikkanen, 2019). The advances of the last decade have indeed been significant, but AI is only capable of performing well on narrow tasks for which they can be trained over an extended period with a significant amount of relevant data (and significant number of humans working on labeling this data). The disconnect between the specialist capabilities of a neural network which has learned a specific task and the general capabilities of AI to perform a range of tasks remains significant.

Maclure (2020) has described the tendency to make unsupported claims about the speed of technological progress as "AI inflationism". He argues that inflationism focuses our collective ethical energies on the wrong problems. The close attention required to apply a set of abstract ideas to a concrete situation necessarily results in a selective approach. Even an ethics based on the broadest deontological principles becomes selective when those principles must be applied to a particular dilemma: the deontologist answering the trolley problem must necessarily be thinking about the trolley's direction of travel. AI inflationism risks concentrating our ethical energies on issues which are not yet relevant, at the expense of those which are. Inflationist approaches which center ethicists' attention on topics such as the best response to the singularity indirectly prevent us from paying attention to the issues that impinge upon the wellbeing of people who interact with significantly less advanced AI right now.

We draw two lessons from this critique. First, we should avoid expending our finite ethical energy on speculative digressions and ensure that our focus is on applying ethics to the most salient issues. Second we should conduct research into AI on the basis of a fundamental continuity with wider studies of technology in the capitalist workplace. As such, we advocate a deflationary approach which, in line with Maclure (2020), attempts to look past the AI hype to identify the risks and opportunities raised by the current deployment of AI in the workplace.

AI inflationism leads to a perception of technological exceptionalism. Because AI is unlike any previous technology, the thinking goes, all historical ways of thinking about technology are irrelevant. Such exceptionalist narratives can contribute to the degradation of ethical standards in academic AI research. For example, as Metcalf and Crawford (2016) have argued, research that uses large quantities of data in higher education contexts in the US have often lacked the kind of ethical control in place in other disciplines. Despite using datasets generated by human subjects, they are not classified as human subject research—often because the data contained within is publicly available. Whereas, an equivalent study in the social sciences would be required to pass ethical review, no such requirement applies in AI research—partly because of its evolution out of disciplines without such institutions in place (computer science, statistics, etc.). One study claimed to use a neural network to identify gang-related crimes with only four data points (Seo et al., 2018). This neural net was trained on Los Angeles Police Department data, which is heavily influenced by a CalGang database that has since been shown to be both inaccurate and deliberately manipulated by LAPD officers (Davis, 2020). Despite the potential harm caused by a neural network reproducing failures of the database and intensifying the patterns of systematic discrimination present in LAPD policing practices, there was no ethical review of the research. Issues such as bias and racism went completely uncommented upon in the paper itself. As Crawford (2021, p. 116–117) argues, "the responsibility for harm is either not recognized or seen as beyond the scope of the research." The exceptional framing of AI contributes to the absence of ethical standards.

We have already noted the lack of attention to labor in the literature; AI exceptionalism risks exacerbating this further. Instead, we argue, the long history of thinking about technology in the workplace—from Smith's (1776) *Wealth of Nations* and Ure's (1835) *Philosophy of Manufactures* to Marx [1887] (1976) *Capital* and beyond—has much to tell us about the way in which AI operates in our context today. For example, by emphasizing a continuity-based analysis of technology, Steinberg's (2021) work on the automotive lineage of the platform economy presents an analysis of a supposedly novel technology (labor platforms) that is situated in the historical tendency to outsource labor costs and mine data from labor processes. AI is best understood in context and as a distinct development in a lineage of technology. Rather than being generated *ex nihilo*, AI is a product the same mode of production that gave us the spinning jenny. The social relations that shape AI are familiar to us and existing theories of work technology have much to teach us about AI's development. Analysis must strike a balance between what is old and what is new so that it can accurately represent the degree of continuity and discontinuity in technological change. Ethical approaches which fail to understand this basic point and buy into AI inflationism and exceptionalism tend toward making three kinds of errors: (A) focusing on potential ethical challenges that may arise in the future rather than existing problems of the present, or postponing dealing with the ethical challenges until they become a major problem that cannot be ignored (Ustek-Spilda, 2019); (B) abdicating or deferring responsibility for creating robust ethical standards and regulations due to the perceived speed of AI's development, and the curious assumption of ethics potentially being in conflict with technology development; and (C) failing to see the fundamental continuity of AI with a longer lineage of technological development (Law, 2004) which can help contextualize our current ways of thinking and acting. Hence we argue that a deflationary approach to AI must insist it is *not* exceptional to similar historical waves of technological change, and the continuities between past and present are more important to explore than the unique aspects of AI for developing ethical principles and practices.

The Problem With Automatic Politics (Ethico-Political Critique)

The drive for accumulation is inherent to capitalism and with it "an autonomous tendency for the productive forces to develop" (Cohen and Kymlicka, 1988, p. 177). How these forces develop in relation to capital's imperative to control them, however, is socially shaped by the regulation of markets, finance, state power, geopolitics and the power of organized labor. As Noble (1984) noted, technologies alone do not determine

changes in social relations but rather tend to reflect such changes. There is a dominant view among AI positivists that technological innovation always constitutes social progress. Yet this view ignores the politics of design and production. Sabel and Zeitlin (1985) argue that politics determines technology design and implementation at work, rather than an inherent capitalist drive toward efficiency. At the same time, technology design also tends to require or strongly encourage particular forms of social organization (for a discussion of the machine-determined "intelligence" in AI see Moore, 2020). This tension forms a dialectic that shapes the boundaries of control. The accountability mechanisms (or lack thereof) that stem from a particular politics—whether at the level of the state or the workplace—will ultimately determine the impact of technology design on workers.

This dialectic is observable in the recent emergence of information and communications technologies (ICT). In her example of the introduction of mobile phones for managers, Orlikowski (2007) takes up a soft version of Winner's (1980) thesis that artifacts have politics. Mobile phones did not simply make communication more convenient—they changed that nature of communication itself. Similarly, cloud technologies have not simply enabled greater connectivity, they have changed what connectivity means. Recent legislation such as the "right to disconnect" was introduced in France to limit the political impact of mobile and cloud connectivity on workers, and the push toward always being available for work, blurring the distinction between work and home, private and public sphere and online and offline hours.

The tension between technology development and the politics of the labor process is further illustrated by the "labor extraction problem", i.e., the broad range of factors employers use to minimize unit labor costs (Edwards and Ramirez, 2016). There is always a trade-off between positive rewards for performance and negative punishments for failing to meet standards—all of which depends on work culture, supervision costs, social protections for workers, and the power of organized labor. Technology doesn't sit independently of these factors; it is always already socially embedded. The ways technologies like AI are developed is inextricably bound with the ways in which companies' direct innovation, diffusion and application to the tasks most attuned to the profitable extraction of surplus-labor. The degree of labor effort is integral to this extraction and requires the development of different organizational strategies. From tightly controlled and fragmented tasks performed on a continuous, mechanized production line to complex, team-driven and capital-intensive production systems—extraction requires different levels of discipline (Edwards and Ramirez, 2016). In complex production systems, the absence of just one worker could disrupt the entire network of labor, thus increasing workers' bargaining power vis-à-vis capital. However, this degree of power may induce management to reduce the complexity of tasks and substitute machinery for labor, depending on the costs and benefits of control. Another factor concerns the costs of work performance monitoring. If it is expensive due to the need for human supervisors and workers are in a strong bargaining position due to labor protection, unions and/or high technical knowledge; employers will tend to use positive incentives to elicit greater labor effort. If, on the contrary, worker performance monitoring is cheap and workers can be disciplined, dismissed and replaced easily (though using self-employment or temporary contracts, for example), then negative incentives will tend to be used (Edwards and Ramirez, 2016).

Digital labor platforms (a prevalent use-case of AI) are illustrative of how this labor extraction takes place. As Stanford (2017) points out, technological changes that do not require large amounts of direct capital investment (such as cloud-based AI-powered platforms), enable the decentralization of production through mobile tracking, surveillance and algorithmic management without necessarily sacrificing the element of direct employer control. For example, platforms such as Uber use facial-recognition AI to verify user identity and rely on customer ratings and real-time movement tracking in their app to manage a global workforce of drivers. Ratings and automated tracking essentially outsource performance monitoring and keep management costs low—yet this has real costs for workers (Moore, 2018). For example, racial bias in facial-recognition AI has led to the deactivation of many non-white Uber drivers, because the technology would not recognize their face (Kersley, 2021). This caused considerable disruption to workers' livelihoods. Similarly, unfair or inaccurate customer reviews can reduce drivers' earning capacity, and in the worst cases lead to deactivation with no opportunity for formal mediation by a trade union (Temperton, 2018). If it is legal to terminate workers' contracts based on an algorithmic decision without any transparency or formal process of contestation, managers can simply defer to the "black box" of the AI system rather than being held to account for the design of such systems themselves.

As noted above, whilst both employers and trade unions might agree on the need for fairness in applications of workplace AI, the question of what each party considers "fair" is likely to differ significantly. Rather than building agreement, such statements of principle simply identify the values over which conflict will take place between actors with opposing interests. One potential solution to this conflict would be for individual stakeholders to produce documents determining the meaning of ethical practice in isolation. IBM's *Everyday Ethics for Artificial Intelligence* (2019), for example, achieves a higher level of concrete detail than we might otherwise expect by using a hypothetical example of a hotel implementing an AI virtual assistant service into its rooms to demonstrate how five particular areas of ethical focus (accountability, value alignment, explainability, fairness and user data rights) might be applied in practice. That said, when it comes to defining what it means by "fairness", the document only identifies the need to guard against algorithmic bias, ignoring other potential negative impacts such as undermining workers' decision-making capacity, deskilling, or even jobs destruction. Furthermore, broader issues in the sector such as low wages, insecure employment and lack of collective bargaining are not considered, implying that such concerns somehow lie outside the realm of technology ethics.

In this context, we might detour Crawford (2021) notion that "AI is politics by other means" and posit that AI is

politics by automatic means. Multi-stakeholder agreements involving high-level principles can hide profound differences in political assumptions and the divergent interests of labor and capital. Without independent accountability mechanisms aimed at more equitable social outcomes, AI will simply deepen existing inequities.

Big Tech Ethics Is Unilateral

One of the principal issues with AI ethics frameworks is that the development of self-assessment and voluntary guidelines involves a conflict of interest. As Bietti (2019) notes, tech companies tend to deploy ethics frameworks to avoid statutory regulation and serve as a defense mechanism for criticism from wider society. Lack of disclosure, regulation and protection increases the autonomy of capital and increases a range of public threats from automated hacking (Veiga, 2018) to political disinformation and deep fakes (Westerlund, 2019). In this context, self-regulation is a direct attempt to avoid any real accountability to the public and inevitably serves the interests and objectives of capital and companies themselves. External mechanisms are the only way that the public can exercise power over AI companies and hold them to account.

The case of Google is instructive on how ethics unilateralism fails. Since 2017 Google has attempted to implement an AI ethics strategy through top-down internal policies, in response to a backlash of criticism by both Google employees and the public. The backlash was first provoked by the revelation that Google had partnered with the US Department of Defense who were using their TensorFlow AI system in military drone programs known as Project Maven. Numerous other internal strategies were pursued by Google, such as setting up an Advanced Technology External Advisory Council (ATEAC), whose mission was to consider the "most complex changes that arise under (Google's) AI principles" (Google, 2019). It was quickly disbanded a week later as members resigned over the failure of the company to live up to its political principles (Phan et al., 2021). Google persisted in other attempts to facilitate more ethical AI by consulting with academics and community-based, non-profit leaders, and recruiting ethicists as part of the Google Research Ethical AI team (Google, 2020). Yet regardless of how refined and well-considered any resulting principles might be, there is virtually no enforcement, and no consequences for breaching them by any statutory body. Google employees continue to be fired for speaking out against the company (Ghaffary, 2021).

Other voluntarist initiatives aimed at Fairness, Accountability, and Transparency (FAccT) in AI and ML such as AI Fairness 360 by IBM, Google Inclusive ML, and Microsoft FairLearn[3] have been developed in collaboration with universities (Phan et al., 2021). The development of these products allows firms to claim they have solved the problem of bias and revised their customer-facing brand identity along ethical lines. Yet the development of ethics frameworks through tech-company funded University research projects largely serves the interests of the private sector, and therefore capital. The individualized, privatized, and voluntarist nature of these initiatives also poses a fundamental limit to the scale and scope of enforcement.

Indeed, efforts to debias AI never seem to consider the bias of capital, i.e., the interests of shareholders over workers, accumulation over distribution, and private exchange-value over social use-value. The prevailing restriction of the AI ethics discussion to the classical liberal principles of property and privacy also takes effect in discussions of bias. The issue has primarily been framed as one of poorly trained algorithms acting in a way which illegitimately penalizes individuals or groups. But bias in the form of specific inaccuracies is less concerning than the broader reproduction of existing patterns of social inequality *via* AI (Eidelson, 2021). The majority of AI developed in the private sector has, for instance, systematically biased the interests of shareholders and managers over the interests of workers, and placed the private accumulation of capital over the public accumulation of social goods (Crawford, 2021). Such considerations are (unsurprisingly) not within the remit of Microsoft, IBM, or Google's FAccT programs. This glaring omission highlights how inadequate the self-regulation of such biases will likely prove in the long term. In proposing technological solutions to social problems, these initiatives mask the wider social and economic context in which they are operating. Unilateral ethical commitments tend to avoid the difficult areas where the interests of the party writing that commitment contrast with ethical practice—and therefore fail to address the areas of greatest risk.

The obvious solution is to develop and apply ethical principles through collaborative multilateral processes which involve a variety of stakeholders. Many sets of ethical principles have embedded a commitment to social dialogue, but often this commitment remains largely non-binding and non-specific, and it rarely goes beyond the immediate discursive bubbles of those setting up the dialogue. What is needed, if a set of principles are to actively foster the kind of multistakeholder engagement that can turn ideas in to practice, is a concrete set of accountability and enforcement mechanisms that can allow for negotiated agreement over areas of conflicting interest.

ETHICS VS. ACCOUNTABILITY

Statutory vs. Non-statutory Accountability Mechanisms

It is to this question of accountability mechanisms which we will now turn. Hagendorff (2020) has demonstrated that most of the 100+ ethical AI statements of principle generated in the last decade have had minimal practical impact. Stakeholders who want to support the development of ethical AI therefore face an uphill battle. Our argument is that if the jump from ethical theory to practice is to be successfully made, then the field of AI ethics must progressively replace the dominant pattern of seeking consensus through increased abstraction with negotiating multistakeholder agreements through progressively greater levels of detail.

[3]See IBM https://aif360.mybluemix.net/, Google https://cloud.google.com/inclusive-ml and Microsoft https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_whitepaper.pdf respectively.

Regulators and the public are entitled to clear explanations of the rules and choice criteria of AI technologies, despite their proprietary nature, and voluntarist ethical guidelines will be useless if the algorithm remains a black box (Karen and Lodge, 2019). Some claim that the complexity of the technology presents serious barriers to explaining how a particular function was carried out and why a specific result was achieved (Holm, 2019). However, there already exists a mechanism enshrined in workers' statutory rights through which *accountability* (if not explainability *per se*) can be carried out through multi-stakeholder negotiation—collective bargaining. At both the sector and enterprise level, collective bargaining offers stakeholders a way to agree the concrete details of ethical AI implementation in the workplace, with the introduction of new AI ideally being negotiated beforehand, not retrospectively, if optimal translation of principles to practices is to be achieved (De Stefano and Taes, 2021).

Early case studies of how collective bargaining operates to produce ethical outcomes are beginning to emerge. Workers represented by the German union ver.di expressed concerns over the use of RFID technology and algorithmic management by multinational retail corporation H&M. The risk of negative impacts such as deskilling, work intensification, unwarranted increases in managerial control, workforce segmentation, and increases in precarity were significant for them. Using their works council, retail workers were able to delay the introduction of the new technology pending further negotiations over risk mitigation measures (López et al., 2021). Here, the unilateral implementation of new forms of AI in the face of ethical concerns was avoided because collective worker power was exercised to assert co-determination rights.

It is indicative of the managerial bias of the AI ethics literature so far that collective bargaining has rarely been mentioned as an essential part of the translation between theory and practice. But it is by no means inevitable that the representatives of capital should rigidly oppose collective bargaining. Indeed, robust collective bargaining has historically facilitated forms of partnership between labor and capital in Northern European economies. At the level of the firm, it has tended to reduce industrial conflict and employee turnover and increase trust and cooperation. On the national level, it has frequently been one factor in reducing strike rates, increasing productivity, and controlling the pace of wage growth (Doellgast and Benassi, 2014). The desirability of these outcomes for workers themselves is debatable, yet opposition to collective bargaining is by no means a necessary position for the representatives of capital. Any employer seriously interested in the ethical application of AI in the workplace should proactively respect workers' rights to organize and ensure workers' perspectives are represented as far as possible pre-union.

Statutory regulations around the use of technology, including AI, in the labor process have been developed, introduced and enforced in many countries, and this process will gradually see broad theoretical principles about AI ethics translated into legislation. This is to be welcomed, but the ability to introduce and shape legislation tends to be restricted to a small range of actors, locking out many interested parties from direct mechanisms through which they can support that translation process. As a result, there remains a significant need for forms of non-statutory regulation which can be designed and implemented by civil society actors acting outside of (and often in opposition to) governments. For example, the Living Wage Foundation's non-statutory identity was used by the UK Government to market their statutory changes to the minimum wage.

Positive examples of non-statutory regulation are already abundant in the world of work. As shown by the Fairwork project,[4] objective monitoring of labor standards in the platform economy by researchers can contribute to raising standards across 27 different countries. For example, following low scores for fairness in Ecuador and Ghana, food delivery platform Glovo consulted Fairwork on the creation of a "Courier Pledge" that aimed to introduce a set of basic standards.[5] Not all of Fairwork's suggestions were implemented, but Glovo *did* introduce a living wage guarantee for all the hours couriers were logged into the app; the provision of health and safety equipment for couriers; the creation of a formalized appeal process for disciplinary action with access to a human representative and a mediator system; a commitment to introduce channels of the improvement of collective workers' voice; and the institutionalization of anti-discrimination policies.

This crisis of ethical impact that Hagendorff (2020) identified is not an inherent feature of AI as a technology. While statutory solutions offer the best accountability mechanisms, there remains a place for non-statutory mechanisms. With the right models for translating principles into practices, there are ways for non-statutory regulation based on statements of ethical principle to shape the way in which AI is implemented in the workplace. In line with our critique above, however, this approach to AI ethics should not just look like a repetition of what has come before. As well as changes to the content of principles, AI ethics should be open to new modes of translation. The example of Fairwork demonstrates that non-statutory regulation will have to be both willing to take a potentially adversarial stance toward AI developers and employers who use their products, while also be willing to prioritize collective worker voice and participation if it is to start forcing profit-motivated private companies to act more in the interests of society at large.

Proposed AI for Fairwork Standards

We identified the important gap of omitting workplace, employment and labor concerns from AI ethics. We also noted that in order for ethical principles to be implemented into practices, we need the organizations to be not merely committing to them voluntarily, but actually be held accountable to them. Working in partnership with the Global Partnership on AI (GPAI), the authors are involved in an ongoing "AI for Fair Work" project to create a set of principles and an associated non-statutory implementation scheme which can deliver on this goal.

[4]https://fair.work.

[5]Conflict of interest statement: None of the researchers have any connection with any of the platforms, the work undertaken received no funding or support in kind from any platform or any other company, and we declare that there is no conflict of interest.

TABLE 2 | Nine draft principles for the GPAI's "Fair Work for AI" project.

1	Guarantee decent work	The right to decent work has been extensively established. The introduction of AI to a labor process is no excuse for undermining basic labor standards. We also cannot assume that decent work conditions are going to be provided *de facto* in new working arrangements and can be taken for granted. Regardless of changes in workplace technology, this right must be upheld.
2	Build fair supply chains	AI development is not conducted in isolation. The requirement to pursue fair conditions must apply across the supply chain, and organizations have a responsibility to use their procurement power toward that end and should be held accountable of the practices of the parties they subcontract parts of their work.
3	Promote explainability	Workers have a right to understand how the use of AI impacts their work. Organizations must respect this right and provide detailed, understandable resources to allow workers to exercise it.
4	Strive for equity	The way AI is produced means that it is never purely objective. So, the values used to design AI need to be open for discussion and evaluation with the goal of minimizing both algorithmic bias and patterned inequality.
5	Make fair decisions	The automation of decision making can lead to a loss of accountability, but mere human oversight over decision making doesn't guarantee fair decisions either. By combining a strong right of appeal with a process to implement lessons learned, organizations can create a robust system which harnesses the power of AI whilst delivering fairer decisions that take into account limitations to resources and socio-economic opportunities, but aims to reduce injustices in their allocation as much as possible
6	Use data fairly	The concentration of data can create risk both for individual persons and groups, so limits must be put on collection (i.e., data minimization) and processes created for subjects to access their personal data in a comprehensive and explainable format. There should be opportunities for individuals to learn and increase their understanding about potential data risks, so that they are able to question and when necessary, reject, decisions made about them.
7	Enhance safety	The right to healthy, safe working environments must be protected. Advances in algorithmic management have increased the risks of work intensification and surveillance. Organizations should seek to actively improve health and safety through their technology.
8	Create future-proof jobs	The introduction of workplace AI can cause specific risks such as job destruction and deskilling. These risks can be avoided by treating the introduction of AI as an opportunity to engage in a participatory and evolutionary redesign of work. This approach should mitigate the risks above and look to use the advantages conferred by the use of AI to increase job quality.
9	Advance collective worker voice	By facilitating collective bargaining, stakeholders can create the conditions for productive negotiation to determine how to turn ethical principles into ethical practice. This also guarantees the principles to be embraced by a larger group of the society, and the developers and users of AI to be held accountable.

The 10 initial principles developed in the project which aim to address the gaps identified in the earlier sections of this paper are summarized in **Table 2**.

The full detail of these principles and their associated measurable benchmarks will be available in a report in 2023, following the conclusion of the consultation process. However, we believe it will be of value to discuss how our critique of the existing AI ethics literature has informed the drafting of these principles, even in advance of the full results of the project being available.

These principles refuse the narrow liberalism inherent to much of current AI ethics debate, which tends to remain in the classical frame of property and privacy. Instead, this project accepts the need to deal directly with the often-suppressed issues of power and control in the workplace. The values encoded in the social relations of production are not an epiphenomenon of ethical discussion that is more properly conducted in the purely conceptual terrain: instead, these values are often determined by the balance of forces between groups of agents and their ability to advance their respective interests. Where the interests of labor and capital do come into conflict, two choices are available: either a retreat toward unilateral principle statements made by individual stakeholders in isolation, or a mechanism to negotiate that conflict in order to achieve improvements in ethical practice.

Collective bargaining is a crucial a mechanism to negotiate the conflict between capital and labor, though it varies hugely across different global contexts. Not only does the absolute number of workers covered by an agreement differ from country to country, so too does the dominant kind of agreement: whilst some cover entire sectors, some are only relevant for specific employers or sites of employment. This diversity necessitates a certain degree of adaptability in how the principles can be applied. As a result, the principles also contain a draft provision for an anonymous consultation process which can be applied in workplaces where there is no trade union presence—whilst emphasizing the need for organizations to respect the right to organize of all workers and not in any way seeking to circumvent union organization. Taken together, these principles attempt to avoid the pitfalls identified in the discussion above and identify a route through which stakeholders can work toward the implementation of fairer workplace AI that mitigates the risks and maximizes the opportunities associated with this ongoing process of technological development.

CONCLUSION

To paraphrase James Ferguson in his critique of "development", what do existing ethical AI principles do besides fail to make AI ethical? It's not just that they are ineffective, it's that they can provide a screen to all manner of unethical behavior and practice. We have argued in this paper that ethics must be focused on the concrete to make them useful. The principles we have presented hone in on the immediate challenge presented

by AI in the workplace. In part, the draft of the principles has drawn on pre-existing standards and understandings of rights in the workplace, but it also goes beyond them. The work-centered critique of existing principles and the proposed new standards set outs a research agenda and is the primary contribution of this article to the burgeoning literature on AI and work.

Worker voice has been significantly neglected in debates around AI, and so we have paid particular attention to those critiques leveled from the perspective of workers on hegemonic ethical values as they apply to the workplace. As part of adopting a deflationist attitude to AI, this has often meant looking back at historic theories of technological change. For example, Braverman's analysis of the deskilling tendencies of Taylorism is the major theoretical background to principle nine: increase job quality. This historical perspective also emphasizes the need for stakeholders to begun to formulate rules that govern the operation of technologies before path dependencies can block off potentially emancipatory or liberatory routes for development.

As a result, the principles emphasize the need for external normative values to be imposed on field of possibilities created by tech. This inevitably means that we don't just need workers as stakeholders—we also need governments. The regulatory turn is now well underway with respect to AI, and the end goal of any discussion of normative values must be to feed into that process of development. By involving representatives from global governments in the consultations conduct over the principles, we aim to link these discussions into concrete programs of action at the legislative level.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MC lead the research and writing of the paper and organized the argument, contributed the most time in research and writing, particularly around the literature review, the second two critiques, the abstract and framing, and provided detailed edits and feedback. CC contributed the second most time researching and writing the paper, is the lead in GPAI, and contributed to two seconds as well as writing up the principles which all four authors contributed to developing. FU contributed primarily around ethics questions. MG had the idea to develop the paper and provided editorial and conceptual work. All authors contributed to the article and approved the submitted version.

FUNDING

This paper was made possible by funding from the Global Partnership on Artificial Intelligence.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2022.869114/full#supplementary-material

REFERENCES

Abdulkareem, M., and Petersen, S. E. (2021). The promise of AI in detection, diagnosis, and epidemiology for combating COVID-19: beyond the hype. *Front. Artif. Intell.* 4, 652669. doi: 10.3389/frai.2021.652669

Algorithm Watch (2020). *AI Ethics Guidelines Global Inventory*. Available online at: https://inventory.algorithmwatch.org (accessed November 19, 2021).

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). *Machine Bias*. Publica. Available online at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed April 27, 2022).

Annas, J. (2006). "Virtue ethics," in *The Oxford Handbook of Ethical Theory*, ed. D. Copp (Oxford University Press). doi: 10.1093/0195147790.001.0001

Anscombe, G. E. M. (1958). Modern moral philosophy. *Philosophy* 33, 1–19. doi: 10.1017/S0031819100037943

Aristotle (2000). *Nicomachean Ethics*. Batoche Books: Kitchener. Available online at: http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3314407 (accessed April 29, 2022).

Asaro, P. M. (2019). AI ethics in predictive policing: from models of threat to an ethics of care. *IEEE Technol. Soc. Mag.* 38, 40–53. doi: 10.1109/MTS.2019.2915154

Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: a decision-making framework. *Ethics Inform. Technol.* 23, 841–854. doi: 10.1007/s10676-021-09619-6

Bietti, E. (2019). "From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy," in *Proceedings to ACM FAT* Conference (FAT* 2020)*. Available online at: https://papers.ssrn.com/abstract=3513182 (accessed November 19, 2021).

Burr, C., Taddeo, M., and Floridi, L. (2020). The ethics of digital well-being: a thematic review. *Sci. Eng. Ethics.* 26, 2313–2343. doi: 10.1007/s11948-020-00175-8

Cazes, S., Hijzen, A., and Saint-Martin, A. (2015). *Measuring and Assessing Job Quality: The OECD Job Quality Framework*. Paris: OECD doi: 10.1787/5jrp02kjw1mr-en

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., et al. (2018). *Notes from the AI Frontier: Applications and Value of Deep Learning*. London: McKinsey Global Institute Available online at: https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning (accessed December 25, 2021).

Cohen, G. A., and Kymlicka, W. (1988). Human nature and social change in the marxist conception of history. *J. Philos.* 85, 171–191. doi: 10.2307/2026743

Cole, M., Radice, H., and Umney, C. (2021). "The political economy of datafication and work: a new digital taylorism?," in *Socialist Register 2021: Beyond Digital Capitalism: New Ways of Living* (New York, NY: Monthly Review Press). Available online at: https://socialistregister.com/index.php/srv/article/view/34948 (accessed December 25, 2021).

Crawford, K. (2021). *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. Yale; London: Yale University Press.

Dastin, J. (2018). *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*. Reuters. Available online at: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (accessed April 27, 2022).

Davis, Z. (2020). *LAPD Chief Says Its Gang Database Abuse Scandal Now Has "Criminal Aspects."* Reason.com. Available online at: https://reason.com/2020/01/15/lapd-chief-says-its-gang-database-abuse-scandal-now-has-criminal-aspects/ (accessed November 15, 2021).

De Stefano, V., and Taes, S. (2021). *Algorithmic Management and Collective Bargaining*. Brussels: European Trade Union Institute. Available online at: https://www.etui.org/publications/algorithmic-management-and-collective-bargaining (accessed January 25, 2022).

Di Ieva, A. (2019). AI-augmented multidisciplinary teams: hype or hope? *Lancet* 394, 1801. doi: 10.1016/S0140-6736(19)32626-1

Doellgast, V., and Benassi, C. (2014). "Collective bargaining," in *Handbook of Research on Employee Voice*, eds. A. Wilkinson, J. Donaghey, T. Dundon, and R. Freeman (Edward Elgar Publishing). Available at: https://www.elgaronline.com/view/9780857939265.00023.xml (accessed January 25, 2022).

Edwards, P., and Ramirez, P. (2016). When should workers embrace or resist new technology? *New Technol. Work Empl.* 31, 99–113. doi: 10.1111/ntwe.12067

Eidelson, B. (2021). Patterned inequality, compounding injustice and algorithmic prediction. *Am. J. Law Equal.* 1, 252–276. doi: 10.1162/ajle_a_00017

Etzoni, O. (2018). *A Hippocratic Oath for Artificial Intelligence Practitioners*. TechCrunch. Available online at: https://social.techcrunch.com/2018/03/14/a-hippocratic-oath-for-artificial-intelligence-practitioners/ (accessed January 27, 2022).

European Commission (2020). *On Artificial Intelligence - A European Approach to Excellence and Trust*. Brussels: European Commission. Available online at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed December 25, 2021).

European Parliament (2019). *EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation*. European Parliament. Available online at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf (accessed December 25, 2021).

Floridi, L. (2019). Translating principles into practices of digital ethics: five risks of being unethical. *Philos. Technol.* 32, 185–193. doi: 10.1007/s13347-019-00354-x

Gentili, A., Compagnucci, F., Gallegati, M., and Valentini, E. (2020). Are machines stealing our jobs? *Camb. J. Regions Econ. Soc.* 13, 153–173. doi: 10.1093/cjres/rsz025

Ghaffary, S. (2021). *Big Tech's Employees are One of the Biggest Checks on Its Power*. Vox. Available online at: https://www.vox.com/recode/22848750/whistleblower-facebook-google-apple-employees (accessed January 25, 2022).

Google (2019). *An External Advisory Council to Help Advance the Responsible Development of AI*. Google. Available online at: https://blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/ (accessed January 25, 2022).

Google (2020). *An Update on Our Work on AI and Responsible Innovation*. Google. Available online at: https://blog.google/technology/ai/update-work-ai-responsible-innovation/ (accessed January 25, 2022).

Hagendorff, T. (2020). The ethics of AI ethics: an evaluation of guidelines. *Minds Mach.* 30, 99–120. doi: 10.1007/s11023-020-09517-8

Hagendorff, T. (2021). *AI Virtues – The Missing Link in Putting AI Ethics Into Practice*. ArXiv201112750 Cs. Available online at: http://arxiv.org/abs/2011.12750 (accessed October 13, 2021).

Heeks, R., Graham, M., Mungai, P., Van Belle, J.-P., and Woodcock, J. (2021). Systematic evaluation of gig work against decent work standards: the development and application of the Fairwork framework. *Inform. Soc.* 37, 267–286. doi: 10.1080/01972243.2021.1942356

Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI Ethics* 1, 41–47. doi: 10.1007/s43681-020-00008-1

High-Level Expert Group on AI (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment*. European Commission. Available online at: https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (accessed February 4, 2021).

HM Government (2021). *National AI Strategy*. Office for Artificial Intelligence. Available online at: https://www.gov.uk/government/publications/national-ai-strategy (accessed December 25, 2021).

Holm, E. A. (2019). In defense of the black box. *Science* 364, 26. doi: 10.1126/science.aax0162

Holmström, J., and Hällgren, M. (2021). AI management beyond the hype: exploring the co-constitution of AI and organizational context. *AI Soc.* doi: 10.1007/s00146-021-01249-2

Hong, J.-W., Choi, S., and Williams, D. (2020). Sexist AI: an experiment integrating CASA and ELM. *Int. J. Hum. Comp. Interact.* 36, 1928–1941. doi: 10.1080/10447318.2020.1801226

Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., and Sutherland, W. (2021). Algorithmic management in a work context. *Big Data Soc.* 8, 20539517211020332. doi: 10.1177/20539517211020332

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat. Mach. Intellig.* 1, 389–399. doi: 10.1038/s42256-019-0088-2

Karen, Y., and Lodge, M. (eds.) (2019). *Algorithmic Regulation*. *First Edn*. New York, NY: Oxford University Press.

Kersley, A. (2021). *Couriers Say Uber's 'Racist' Facial Identification Tech Got Them Fired*. Wired UK. Available onlie at: https://www.wired.co.uk/article/uber-eats-couriers-facial-recognition (accessed January 25, 2022).

Kuziemski, M., and Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommun. Policy* 44, S0308596120300689. doi: 10.1016/j.telpol.2020.101976

Law, J. (2004). *After Method: Mess in Social Science Research*. London: Routledge. Available online at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=nlebk&AN=115106&site=ehost-live&authtype=ip,uid (accessed June 27, 2022).

Liikkanen, L. A. (2019). "It AIn't Nuttin' new – interaction design practice after the AI Hype," in *Human-Computer Interaction – INTERACT 2019 Lecture Notes in Computer Science*, eds. D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris (Cham: Springer International Publishing), 600–604.

López, T., Riedler, T., Köhnen, H., and Fütterer, M. (2021). Digital value chain restructuring and labour process transformations in the fast-fashion sector: evidence from the value chains of Zara and HandM. *Global Netw.* doi: 10.1111/glob.12353

MacKenzie, D. A., and Wajcman, J. (1999). "Introductory essay: the social shaping of technology," in *The Social Shaping of Technology*, eds. D. A. MacKenzie and J. Wajcman (Buckingham: Open University Press), 3–27.

Maclure, J. (2020). The new AI spring: a deflationary view. *AI Soc.* 35, 747–750. doi: 10.1007/s00146-019-00912-z

Marquardt, E. (2020). Künstliche Intelligenz in optischen Mess- und Prüfsystemen: Chance oder Hype? *Z. Für Wirtsch. Fabr.* 115, 731–733. doi: 10.1515/zwf-2020-1151019

Marx, K. (1976). *Capital, Volume I: A Critique of Political Economy*. Harmondsworth: Penguin in Association With New Left Review.

Metaxa, D., Gan, M. A., Goh, S., Hancock, J., and Landay, J. A. (2021). An image of society: gender and racial representation and impact in image search results for occupations. *Proc. ACM Hum. Comput. Interact.* 26, 23. doi: 10.1145/3449100

Metcalf, J., and Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data Soc.* 3, 2053951716650211. doi: 10.1177/2053951716650211

Metzinger, T. (2019). *Ethics Washing Made in Europe*. Tagesspiegel. Available online at: https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html (accessed October 13, 2021).

Miller, D. (2021). "Justice," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Metaphysics Research Lab, Stanford University). Available online at: https://plato.stanford.edu/archives/fall2021/entries/justice/ (accessed April 28, 2022).

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. London: Pelican Books.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat. Mach. Intellig.* 1, 501–507. doi: 10.1038/s42256-019-0114-4

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data Soc.* 3, 205395171667967. doi: 10.1177/2053951716679679

Moore, P. (2020). The Mirror for (Artificial) Intelligence: In Whose Reflection? *Comparative Labor Law and Policy Journal*. doi: 10.2139/ssrn.3423704

Moore, P. V. (2018). *The Quantified Self in Precarity: Work, Technology and What Counts*. Abingdon, Oxon: Routledge.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., and Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds Mach.* 31, 239–256. doi: 10.1007/s11023-021-09563-w

Nazareno, L., and Schiff, D. S. (2021). The impact of automation and artificial intelligence on worker well-being. *Technol. Soc.* 67, 101679. doi: 10.1016/j.techsoc.2021.101679

Noble, D. F. (1984). *Forces of Production: A Social History of Industrial Automation*. New York, NY: Knopf.

OECD (2019). *The OECD Artificial Intelligence (AI) Principles*. Available online at: https://www.oecd.ai/ai-principles (accessed January 24, 2021).

Orlikowski, W. J. (2007). Sociomaterial practices: exploring technology at work. *Org. Stud.* 28, 1435–1448. doi: 10.1177/0170840607081138

Penn, J. (2021). Algorithmic silence: a call to decomputerize. *J. Soc. Comput.* 2, 337–356. doi: 10.23919/JSC.2021.0023

Phan, T., Goldenfein, J., Mann, M., and Kuch, D. (2021). Economies of Virtue: The circulation of 'ethics' in big tech. *Sci. Cult.* 31, 121–135. doi: 10.1080/09505431.2021.1990875

Rawls, J. (1993). "Political Liberalism," in *Justice: The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (New York, NY: Columbia University Press). Available online at: https://plato.stanford.edu/archives/fall2021/entries/justice/ (accessed April 28, 2022).

Reedy, C. (2017). *Kurzweil Claims That the Singularity Will Happen by 2045*. Futurism. Available onnline at: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045 (accessed January 25, 2022).

Roberts, H., Cowls, J., Hine, E., Mazzi, F., Tsamados, A., Taddeo, M., et al. (2021). Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US. *Science and Engineering Ethics*. p. 27. doi: 10.1007/s11948-021-00340-7

Robeyns, I., and Byskov, M. F. (2021). "The capability approach," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Metaphysics Research Lab, Stanford University). Available online at: https://plato.stanford.edu/archives/win2021/entries/capability-approach/ (accessed April 28, 2022).

Rockall, A. (2020). From hype to hope to hard work: developing responsible AI for radiology. *Clin. Radiol.* 75, 1–2. doi: 10.1016/j.crad.2019.09.123

Sabel, C., and Zeitlin, J. (1985). Historical alternatives to mass production: politics, markets and technology in nineteenth-century industrialization. *Past Pres.* 133–176. doi: 10.1093/past/108.1.133

Scanlon, T. (1998). "What we owe to each other," in *Justice: The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (Cambridge, MA London: Belknap Press of Harvard University Press). Available online at: https://plato.stanford.edu/archives/fall2021/entries/justice/ (accessed April 28, 2022).

Sen, A., and Williams, B. (eds.) (1982). *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Seo, S., Chan, H., Brantingham, P. J., Leap, J., Vayanos, P., Tambe, M., et al. (2018). "Partially generative neural networks for gang crime classification with partial information," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (New Orleans, LA: ACM), 257–263.

Smith, A. (1776). *Wealth of Nations*. Oxford: Oxford World Classics.

Stanford, J. (2017). The resurgence of gig work: Historical and theoretical perspectives. *Econ. Labour Relat. Rev.* 28, 382–401. doi: 10.1177/1035304617724303

Steinberg, M. (2021). *From Automobile Capitalism to Platform Capitalism: Toyotism as a Prehistory of Digital Platforms - Marc Steinberg, 2021*. Organisation Studies. Available online at: http://undefined/doi/full/10.1177/01708406211030681 (accessed November 18, 2021).

Temperton, J. (2018). *The Biggest Legal Crisis Facing Uber Started With a Pile of Vomit*. Wired UK. Available online at: https://www.wired.co.uk/article/uber-employment-lawsuit-gig-economy-leigh-day (accessed January 25, 2022).

Thiel, V. (2019). *"Ethical AI Guidelines": Binding Commitment or Simply Window Dressing?* AlgorithmWatch. Available online at: https://algorithmwatch.org/en/ethical-ai-guidelines-binding-commitment-or-simply-window-dressing/ (accessed February 4, 2021).

Toh, T. S., Dondelinger, F., and Wang, D. (2019). Looking beyond the hype: applied AI and machine learning in translational medicine. *EBioMedicine* 47, 607–615. doi: 10.1016/j.ebiom.2019.08.027

United States Congress (2022). *H.R.6580 - 117th Congress (2021-2022): Algorithmic Accountability Act of 2022*. Available online at: https://www.congress.gov/bill/117th-congress/house-bill/6580 (accessed April 19, 2022).

Ure, A. (1835). *The Philosophy of Manufactures, or, An Exposition of the Scientific, Moral, and Commercial Economy of the Factory System of Great Britain. Second Edn*. London: C. Knight.

Ustek-Spilda, F. (2018). *A Conceptual Framework for Studying Internet of Things: Virtue Ethics, Capability Approach and Care Ethics – VIRT-EU*. VIRT-EU. Available online at: https://blogit.itu.dk/virteuproject/2018/11/05/a-conceptual-framework-for-studying-internet-of-things-virtue-ethics-capability-approach-and-care-ethics/ (accessed April 29, 2022).

Ustek-Spilda, F. (2019). *Do-ers v. Postpon-ers: How do IoT Developers Respond to Ethical Challenges? – VIRT-EU*. VIRT-EU. Available online at: https://blogit.itu.dk/virteuproject/2019/02/08/do-ers-v-postpon-ers-how-do-iot-developers-respond-to-ethical-challenges/ (accessed April 29, 2022).

Veiga, A. P. (2018). Applications of artificial intelligence to network security. *arXiv [Preprint]*. arXiv: 1803.09992. doi: 10.48550/arXiv.1803.09992

Wagner, B. (2018). "Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?," in *Being Profiling*. Cogitas ergo sum, ed M. Hildebrandt (Amsterdam: Amsterdam University Press).

Wagner, B. (2019). "Algorithmic accountability - towards accountable systems," in *The Oxford Handbook of Intermediary Liability Online* (Oxford: Oxford University Press).

Warhurst, C., Wright, S., and Lyonette, C. (2017). *Understanding and Measuring Job Quality. Chartered Institute of Personnel and Development and Warwick Institute for Employment Research*. Available online at: https://www.cipd.co.uk/Images/understanding-and-measuring-job-quality-3_tcm18-33193.pdf (accessed December 25, 2021).

Westerlund, M. (2019). The emergence of deepfake technology: a review. *Technol. Innov. Manag. Rev.* 9, 39. doi: 10.22215/timreview/1282

White House, T. (2020). *Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government – The White House*. Available online at: https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/ (accessed April 18, 2022).

Winner, L. (1980). Do artifacts have politics? *Daedalus* 109, 121.

Wooldridge, M. (2021). *The Road to Conscious Machines: The Story of AI*. London: Penguin.

Yam, J., and Skorburg, J. A. (2021). From human resources to human rights: impact assessments for hiring algorithms. *Ethics Inform. Technol.* 23, 611–623. doi: 10.1007/s10676-021-09599-7

Yen, C.-P., and Hung, T.-W. (2021). Achieving equity with predictive policing algorithms: a social safety net perspective. *Sci. Eng. Ethics* 27, 1–16. doi: 10.1007/s11948-021-00312-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cole, Cant, Ustek Spilda and Graham. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 25 August 2022
DOI 10.3389/frai.2022.888817

Check for updates

OPEN ACCESS

EDITED BY
Phoebe V. Moore,
University of Essex, United Kingdom

REVIEWED BY
Jose Ramon Saura,
Rey Juan Carlos University, Spain
Marco Briziarelli,
University of New Mexico,
United States

*CORRESPONDENCE
Jamie Woodcock
jamie.woodcock@googlemail.com

SPECIALTY SECTION
This article was submitted to AI in Business, a section of the journal Frontiers in Artificial Intelligence

RECEIVED 03 March 2022
ACCEPTED 10 August 2022
PUBLISHED 25 August 2022

CITATION
Woodcock J (2022) Artificial intelligence at work: The problem of managerial control from call centers to transport platforms. *Front. Artif. Intell.* 5:888817. doi: 10.3389/frai.2022.888817

COPYRIGHT
© 2022 Woodcock. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence at work: The problem of managerial control from call centers to transport platforms

Jamie Woodcock*

Department of People and Organisations, The Open University, Milton Keynes, United Kingdom

There has been much recent research on the topic of artificial intelligence at work, which is increasingly featuring in more types of work and across the labor process. Much research takes the application of artificial intelligence, in its various forms, as a break from the previous methods of organizing work. Less is known about how these applications of artificial intelligence build upon previous forms of managerial control or are adapted in practice. This paper aims to situate the use of artificial intelligence by management within a longer history of control at work. In doing so, it seeks to draw out the novelty of the technology, while also critically appraising the impact of artificial intelligence as a managerial tool. The aim is to understand the contest at work over the introduction of these tools, taking call centers and transport platforms as case studies. Call centers are important because they have been a site of struggle over previous forms of electronic surveillance and computation control, providing important lessons for how artificial intelligence is, or may, be used in practice. In particular, this paper will draw out moments and tactics in algorithmic management has been challenged at work, using this as a discussion point for considering the possible future of artificial intelligence at work.

KEYWORDS

artificial intelligence, algorithmic management, labor process, call centers, platform work, gig economy

Introduction

Artificial intelligence is a broad category of digital technologies that involve intelligence demonstrated by computers and machines. The definition of intelligence used here is broad, covering examples of search engines, recommendations of what to watch next on streaming services, all the way up to the artificial general intelligence of science fiction. As Cook (2018) has argued, "many people are misrepresenting AI in order to make it appear more intelligent than it is." In part, this is due to the aggressive marketing of new technology to both investors and consumers. Similarly, Taylor (2018) has coined the term "fauxtomation", explaining how "automated processes are often far less impressive than the puffery and propaganda surrounding them imply—and sometimes they are nowhere to be seen."

There has been much discussion in the context of platform work on the role of artificial intelligence and algorithms (Srnicek, 2017). More widely, there has been research interest in artificial intelligence, robotics, and other advanced technologies for use in the workplace, with one paper finding 13,136 potentially relevant studies (Vrontis et al., 2022). Following the increasing popularity of research on the topic in general (Pasquale, 2015; Kitchin, 2017; O'Neil, 2017; Turow, 2017; Eubanks, 2019), research on Uber has often focused on algorithm management (Lee et al., 2015; Rosenblat and Stark, 2016; Scholz, 2017; Rosenblat, 2018). For some in the literature, this is seen as a new attempt by management to overcome worker resistance (Veen et al., 2019; Mahnkopf, 2020), while others have drawn attention to the experience of workers struggling against these new techniques (Waters and Woodcock, 2017; Fear, 2018; Briziarelli, 2019; Cant, 2019; Gent, 2019; Leonardi et al., 2019; Cant and Mogno, 2020; Tassinari and Maccarrone, 2020).

The aim of this article is to engage with the topic of artificial intelligence at work over a longer history of supervision and control at work, drawing on empirical and conceptual research. It starts by considering this history and the lessons that can be taken from call center work. The article then moves on to discuss labor process theory and systems of control in factories, call centers, and transport platforms. This provides the theoretical framing for the argument that is explored through the case studies and the exploration of algorithmic management in practice. The article ends by considering how this can shape our understanding of the strengths and limits of artificial intelligence in general, and specifically algorithmic management, at work. The intention is to develop an argument about the significance of artificial intelligence, not as a general technology, but as a form of surveillance and control in the workplace. This is important for both its implementation, but also for understanding struggles against its use.

Approach

This is a conceptual paper that draws on existing research on call centers and platforms. The author has conducted substantial ethnographic fieldwork in call centers (Woodcock, 2017) as well as with the transport platforms (Woodcock and Graham, 2019; Woodcock, 2021) that contribute empirical data toward the argument in this article. The approach taken here is an attempt to draw these findings, as well as those from other research in the field, into a conceptual argument about the role of artificial intelligence at work. This involved the synthesizing of findings from traditional factory settings, call centers, and transport platforms, to conceptualize the role of systems of control within the labor process. This draws primarily from the literature in labor process theory, combined with critical research on algorithms and power more widely.

Call centers, surveillance, and control

There is a long history of control at work. From the moment that bosses started buying workers' labor-power, there have been successive attempts to watch and control what workers are doing at work. Taylor (1967) identified this as the fear of "soldiering" in his theory of scientific management, the belief that workers would deliberately work slower than they could. This managerial fear was not limited to Taylor or Taylorism and is present throughout many kinds of work.

Before the emergence of platform work, call centers were a focus for debates on technological surveillance and control. These debates are useful to revisit in the context of artificial intelligence, particularly as many of the forms of electronic surveillance and outsourcing developed in call centers laid the basis for the technical organization of platform work. Call centers were an important focus of debates on technology, control, and resistance (Woodcock, 2017). The technical arrangements of the labor process made call center work particularly susceptible to early attempts at electronic surveillance and control. As the phone systems were integrated with computers, this provided the possibility to use new technologies in a way that would have been harder to achieve in other forms of low paid work.

Call centers provide an important early example of work that could be digitally legible (see Woodcock and Graham, 2019) that allows it to be measured through discrete data points. Through the integration of telephones and computers, facilitated by the development of automatic call distributors, the modern call center was established. This took away the control from call center workers, automating the process of dialing and speeding up the work. It created the experience of an "assembly line in the head" for call center workers (Taylor and Bain, 1999, p. 103). The new technology also provides a way to electronically supervise the labor process. The computerisation of the process involved developing the capacity to measure each part of the labor process: how many calls made, successful sales, length of calls, time between calls, breaks, and other metrics. Given work in a call center requires a range of clear quantitative indicators, these could now be collected automatically. As I found in a call center in the UK, these "quantitative variables are context free; not something that can be debated, considered instead as the evidence base for rewards or discipline" (Woodcock, 2017). The scale of this data collection is impressive: it "allows an unprecedented level of surveillance; every call encounter is permanent, every mistake could be punishable in the future. It operates like the ability to recall every commodity produced on an assembly line and to be able to retrospectively judge the quality of its production" (Woodcock, 2017).

There are many studies of call centers that have confirmed similar findings (Taylor and Bain, 1999; Bain et al., 2002; Kolinko, 2002; Mulholland, 2002). However,

there is also evidence that call centers had aggressive management techniques that preceded the development of these new technological methods. For example, as an interviewee explained:

> There were all sorts of rules right. I mean for instances hanging coats on the back of your chair was banned, little things like that. Constantly listing things that people couldn't do. I've seen people being chased into toilets because they have their phones on them and stuff like that! All these things you can do with or without the computers (quoted in Woodcock, 2017).

It is therefore important to remember that new technological forms of surveillance and control are developed and implemented within the existing social relations of the workplace—even if they then go on to transform them further.

There is a broad existing literature on call centers that has produced detailed understandings of "work organization, surveillance, managerial control strategies and other central concerns of labor process analysis" (Ellis and Taylor, 2006, p. 2). The key debates within the literature centers around the extent and implications of new technological forms of control. On one side of the debate were academics who argued that call centers were becoming organized like an "electronic panopticon." For example, Fernie and Metcalf (1997, p. 3) claims that the "possibilities for monitoring behavior and measuring output are amazing to behold—the 'tyranny of the assembly line' is but a Sunday school picnic compared with the control that management can exercise in computer telephony." This notion of an electronic panopticon—which draws heavily on both Foucault (1991) and Bentham (1995) and the architectural model of a prison—has similarities with some of the contemporary debates on algorithmic management. However, on the other side of the debates, McKinlay and Taylor (1998, p. 75) argued that the comparison fails to take into account that "the factory and the office are neither prison nor asylum, their social architectures never those of the total institution." Indeed, as Taylor and Bain (1999, p. 103) argue, the "dynamic process of capital accumulation" that takes place in the workplace means that Foucauldian approaches drawing on the panopticon analogy "understates both the voluntary dimension of labor and the managerial need to elicit commitment from workers." This has important implications for theorizing work, particularly that it can "disavow the possibilities for collective organization and resistance" (Taylor and Bain, 1999, p. 103).

These debates can be revisited in a more productive way today, particularly tracing the development from factory supervision, call centers, and then to contemporary platforms (Woodcock, 2020). The claims about the novelty or scope of technological changes today can be reinterpreted through these older debates, providing important theoretical grounding, as well as reminder about the continuing dynamics of work. For example, Taylor and Bain (1999) argument reminds us that technological methods of control cannot solve the problems of management. In the call center, vast quantities of data are collected, but human supervisors are still required to interpret the data and act upon any insights. There are 1-2-1 meetings, coaching, training, and "buzz sessions" that attempt to elicit motivation from workers on the call center floor (Woodcock, 2017). In the context of call center work, there is "no electronic system can summon an agent to a coaching session, nor highlight the deficiencies of their dialogue with the customer." Instead, as Taylor and Bain (1999, p. 108-109) continue, call centers "rely on a combination of technologically driven measurements and human supervisors", which nevertheless "represents an unprecedented level of attempted control which must be considered a novel departure."

From call centers to platforms

In order to apply these lessons to our understanding of artificial intelligence at work, it is therefore necessary to return to the concerns of labor process theory (both in the call center and more widely) to understand the implications of these new management techniques. A "common feature of all digital labor platforms is that they offer tools to bring together the supply of, and demand for, labor" (Graham and Woodcock, 2018). Regardless of whether the legal categorization is employment or self-employment (De Stefano and Aloisi, 2019), these platforms involve work. The labor process is coordinated *via* a digital platform and in the case of transport platforms, often involves a smartphone and GPS. The rapid growth of food delivery and private hire driving platforms has been facilitated by the digital legibility of the labor process, involving discrete data points of start and end journeys.

The concerns of labor process theory involve understanding what happens in the workplaces after the purchase of workers labor-power by capital. This involves the "indeterminacy of the labor process" that requires managing in practice. For example, Edwards (1979, p. 12) argues that:

> conflict exists because the interests of worker and those of employers collide… control is rendered problematic because unlike the other commodities involved in production, labor power is always embodied in people, who have their own interests and needs and who retain their power to resist being treated like a commodity.

The act of mediating these relationships on a platform does not remove the different interests or make the distributed workplace any less of a "contested terrain." Edwards (1979, p. 18) provides a three-part framework for understanding the "system of control" in the workplace. The first is "direction", or the ways in which the tasks that workers have to do are specified.

The second is "evaluation", or how the employer supervises and assesses the workers performance. The third is "discipline", or what methods are used "to elicit cooperation and enforce compliance with the capitalist's direction of the labor process."

As Table 1 illustrates, systems of control can be broken down into the three aspects to develop a more specific understanding of how control is operating in practice. The first thing to note is that elements of automation are present throughout each example. Automation is not the preserve of algorithms, nor is it a simple binary. From the moment that workers started to use tools and machines at work, parts of the labor process began to become automated. It is rare that tasks are ever completely automated, instead the element of human labor becomes decreased—sometimes drastically. For example, as factories have developed since the industrial revolution, the individual productivity of workers has increased by huge amounts. Yet there are still workers in factories. Even in so-called "lights out" factories, workers are required for setting up manufacturing tombstones, quality assurance checks, and the repair and maintenance of machinery.

Table 1 shows how the traditional operation of factories involves aspects of automation, but relies upon a layer of supervisors who monitor, assess, and intervene in the labor process. It builds on the classical Taylorist division of labor and the separation of the conception of tasks from their execution. This involves management attempting to take control away from the workplace, directing workers to complete tasks in specific ways and within set times. It is also worth noting that before the theory was applied to call centers, Foucault (1991, p. 174) wrote about supervision in a factory context. He argued that it involved:

> an intense, continuous supervision; it ran right through the labor process; it did not bear – or not only—on production… It became a special function, which had nevertheless to form an integral part of the production process, to run parallel to it throughout its entire length. A specialized personnel became indispensable, constantly present and distinct from workers.

The obsession with measurement and supervision that begins in the factory becomes applied to an increasing range of work.

Call centers represent a significant development from this model of control. The separation of conception and execution is developed through a form of computational Taylorism and scripting of the phone calls that workers made (Woodcock, 2017). The integration of computers and telephones the collection and digital storage of a range of quantitative metrics, as well as recordings of calls. However, this data requires supervisors to interpret and intervene in order for it to be productive in the workplace. This is not a straightforward process in call centers, with many having high levels of turnover. Instead, disciplinary actions are combined with attempts to motivate workers and the use of monetary bonuses. The role of supervisors develops from the factory floor, particularly in relation to handling abstract data on the labor process, but remains a key interface between workers and capital.

The shift to transport platforms involves the development of control across each of the three component parts. However, one of the key differences is that there is no longer a formal employment arrangement. This means that many of the tools that are available in other kinds of work cannot be used, less the platform risks workers being reclassified away from self-employment (Woodcock and Cant, 2021). With transport platforms there are clear start and end points, with points of contact with either other workers or customers. The work is suitable for metrics in a way that would be harder for other forms of low paid work like cleaning or care. The majority of the metrics are quantitative (how long did the task take) rather than qualitative (how well was the task completed). Similarly, this form of work organization has developed alongside a specific

TABLE 1 Systems of control.

	Factory	Call center	Transport platform
Direction	Taylorist separation of conception and execution of work, workers given specific instructions. Assembly line automatically sets central pace	Separation of conception and execution of work with scripting. Automatic dialing of calls increases pace	Separation of conception and execution of overall work on platform. Workers receive direction through smartphone, but can have discretion with route choices
Evaluation	Supervisors assess the labor process on the factory floor, quality assurance of outputs	Quantitative metrics from electronic supervision, qualitative evaluation by supervisors	Automated evaluation of the labor process with quantitative metrics. Customer evaluation in some cases
Discipline	Supervisors encourage performance, bonuses can be used to increase output. Sanctions for poor performance	Supervisors encourage performance, bonuses used to increase output. Sanctions for missing targets	Use of bonuses to encourage engagement at peak times. Automated interventions based on automated evaluation ("deactivations")

form of contractual relationship: independent contractor or self-employment status.

However, across each case, the aim of the process is to elicit motivation for workers to complete tasks in the labor process. In the factory and the call center, this means trying to overcome the indeterminacy of the labor process, ensuring that capital gets the full value (or, at least, as much as it can) from the purchase of workers labor-power. The problem with this, as Thompson (1983, p. 123) reminds us, is that "complications arise when attempts are made to specify how control is acquired and maintained." Workers want to have energy left after a shift ends—and often there is no benefit to working harder. The widespread use of bonuses can be seen as one solution to this problem, as well as the development of increasing complex methods of supervision and surveillance. Control can mean, in "an absolute sense, to identify those 'in control', and in a relative sense, to signify the degree of power people have to direct work" (Thompson, 1983, p. 124). That degree of power can shift with the use of new techniques and technologies. Indeed as Goodrich (1975) notes, there is always a dynamic "frontier of control" in the workplace that pushes back and forth between the different interests of workers and capital.

Artificial intelligence as technology of workplace control

To talk about artificial intelligence in general terms in the workplace is not meaningful. It involves, as noted early, many forms of simple and more complex artificial intelligence that are proliferating throughout work. At the core, algorithms involve "sets of defined steps structured to process instructions/data to produce an output" (Kitchin, 2017, p. 14). In more complex iterations, this can involves very large or rapid processes, meaning the operation can be obscured as if they operate like a "black box" (Pasquale, 2015). In many cases, algorithms do not shift the frontier of control between capital and labor in any substantial way. For example, autocomplete options in emails are not likely to effect widespread changes in the balance of power in the workplace. However, automated decision making over shift bookings can have a tangible impact on the experience of work.

The development of platform work has provided an important "laboratory for capital" (Cant, 2019), experimenting with new uses for artificial intelligence and automation in the organization of delivery work. However, it has also involved the specific contractual relationships noted above. Instead of entering into formal employment contracts with workers, platforms instead seek to engage workers as self-employed contractors. This misclassification of workers means that platforms can evade the protections and liabilities they would other have to take on with conventional employment models. This model has facilitated the rapid expansion of platforms, particularly in transportation, but it also prevents platforms from acting like employers in some instances. Given the challenges to employment status in many jurisdictions, some platforms have responded by limiting training and communication to ensure they will not fail employment status tests (see, with Deliveroo, Woodcock and Cant, 2021).

Without the traditional forms of workplace control, platforms rely upon algorithmic management to manage a dispersed workforce. Due to the employment status issues, physical supervision is no longer an option, removing interventions like calling workers in for disciplinary meetings or performance improvement meetings, while limiting communication across the platform. One of the basic functions of supervision—telling workers to work harder—is therefore more complicated to achieve in practice. Instead, platforms can use Service Level Agreements and other contractual tools, setting targets in the hope that workers will try to meet them. Instead of direction supervision, this involves a wider set of practices that seek to "seduce, coerce, discipline, regulate and control: to guide and reshape how people… interact with and pass through various systems" (Kitchin, 2017, p. 19). One example of this is the bonus structure, including "boosts" for deliveries during busy periods or adverse weather conditions. This incentivizes workers to log onto the platform, rather than requiring it through strict scheduling. Other strategies are more direct. For example, the use of "deactivation" or firing workers who do not meet performance targets—or some other algorithmically determined reason. The strengths and weaknesses of this approach are considered in Table 2.

As can be seen in Table 2, algorithmic systems of control at work have both strengths and weaknesses. In the case of food delivery platforms, this has involved the removal of a supervisors or managerial layer from the work, instead relying upon automated decision-making processes. This has proven to be a successful model for organizing work—at least for the majority of the time. However, this "platform management model" is contested by workers in practice (Moore and Joyce, 2019). The weaknesses of the approach can be seen when workers actively resist platform control, particularly during strike action. It is during these moments that the lack of managerial intervention (disciplinary or otherwise) shows that there are two kinds of precariousness at Deliveroo, both for the workers involved, but also for the management of the platform (Woodcock, 2020).

Building from the arguments about the "electronic panopticon" (Fernie and Metcalf, 1997), the metaphor can also be used to make sense of algorithmic management (Woodcock, 2020). Unlike the physical architecture of the prison, it is possible to see how the dynamics of the panopticon operate on a platform like Deliveroo. The work involves discrete tasks that increase in frequency during peak times, particularly lunch and dinner. The role of supervision, algorithmic or otherwise, involves trying to ensure that the purchased labor-power is used most effectively. As Foucault (1991, p. 150) noted in

TABLE 2 Algorithmic systems of control.

	Transport platform	Strengths	Weaknesses
Direction	Separation of conception and execution of overall work on platform. Workers receive direction through smartphone, but can have discretion with route choices	In a straightforward task with clear start and finish this is an effective way of distributing instructions	If there are problems during the labor process, there are few options available to workers to negotiate the process. It is not effective with complex tasks or those without clear start or end points
Evaluation	Automated evaluation of the labor process with quantitative metrics. Customer evaluation in some cases	The labor process creates data on locations and timings that is straightforward to track. Customer feedback can be quickly collected	It is difficult to accurately evaluate qualitative aspects of the labor process
Discipline	Use of bonuses to encourage engagement at peak times. Automated interventions based on automated evaluation ("deactivations")	Bonuses can encourage workers to work. Threat of "deactivations" can play a disciplinary function	Bonuses increase the cost of labor-power and may not achieve aim of the labor process. Workers can find ways to game the system. There are no intermediate disciplinary actions before "deactivation"

the context of the factory: "to assure the quality of the time used: constant supervision, the pressure of supervisors, the elimination of anything that might disturb or distract; it is a question of constituting a totally useful time." While this has developed significantly from hiring human supervisors to prowl the workplace, it still involves finding ways to discipline time, as "time measured and paid must also be a time without impurities or defects; a time of good quality, throughout which the body is constantly applied to its exercise" (Foucault, 1991, p. 150).

This point about time is important, as it underlined the original arguments for the panopticon. Bentham (1995, p. 80) argued that the panopticon could find uses beyond the prison: "whatever be the manufacture, the utility of the principle is obvious and incontestable, in all cases where the workmen are paid according to their time." The panopticon was therefore also considered as a potential solution to the problem of the indeterminacy of labor power. Bentham continued to argue that the panopticon could be combined with a piece rate payment scheme, as "there the interest which the workman has in the value of his work supersedes the use of coercion, and of every expedient calculated to give force to it" (Bentham, 1995, p. 80). Foucault, of course, took this further, arguing that the panopticon as an "architectural apparatus should be a machine for creating and sustaining a power relation independent of the person who exercises it; in short, that the inmates should be caught up in a power situation of which they are themselves the bearers" (Foucault, 1991, p. 201).

In the context of the call center, this meant the constant threat of supervisors listening in to calls—as well as being able to recall recordings of all previous calls that had been made. Clearly, no supervisor could be listening to all calls taking place at one time in the call center, but it created the sense that they could be. This experience led to Fernie and Metcalf (1997, p. 3) applying the metaphor of the "electronic panopticon", as discussed above. In many call centers, this is combined with bonus structures, but rarely with payment that is entirely piece rate.

With platform work, the attention is usually on the technology, software, or algorithmic management. These are the "new" features of the work that have gathered substantial attention. Indeed, Foucault (1991, p. 173) discusses how:

> the perfect disciplinary apparatus would make it possible for a single gaze to see everything constantly. A central point would be both the source of light illuminating everything, and a locus of convergence for everything that must be known: a perfect eye that nothing would escape and a center toward which all gazes would be turned.

Given the claims made about the potential of algorithmic management, it is easy to see how the automation of these processes looks increasingly like the metaphor of the panopticon. Recent research has used more general terms for the role of algorithms at platforms like Deliveroo. For example, Muldoon and Raekstad (2022) use the concept of "algorithmic domination" to refer to the "dominating effects of algorithms used as tools of worker control." They argue that "bosses can employ systems of algorithmic domination to control a more flexible labor force."

There is a risk of considering algorithmic management as a general solution to the problem of controlling the labor process. Much less attention is paid to the fact that much of this work is organized around piece rate payment. The first struggle at Deliveroo in London was organized in response to the platform moving away from payment per hour to only payment by drop (Waters and Woodcock, 2017). Muldoon and Raekstad

(2022) consider this in terms of "dynamic pricing", but the focus quickly returns to the role of algorithms. There is a long history of piece rates being used in many industries, which can provide a challenge, but have definitely not prevented workers collectively organizing.

The while there are a range of practices that algorithmic control can entail, as noted earlier by Kitchin (2017, p. 19), it is also worth considering the role of "seduction" in more detail. Foucault identified the "form of power which makes individual subjects", both "a form of power which subjugates and makes subject to" (Foucault, 1982, p. 781). This implies a level of consent, albeit produce through the seduction of algorithmic practices, in the labor process. There are similarities here with the argument of "manufacturing consent" (Burawoy, 1979). While this is secondary to the processes unfolding, it remains a consistently present feature of platform work, often seen in the subjectivity that develops around freedom and flexibility. Algorithmic control, therefore, builds on a relation of power developed between platform and worker. In a Foucauldian sense:

> it incites, it seduces, it makes easier or more difficult, in the extreme it constrains or forbids absolutely. it is nevertheless always a way of acting upon an acting subject or acting subjects by virtue of their acting or being capable of action. A set of actions upon other actions (Foucault, 1982, p. 789).

This can be seen across Tables 1, 2 with the use of different actions, from the direction, evaluation, and discipline, now transformed away from the direct managerial prerogative of a conventional workplaces through platform technology.

The general surveillance of algorithmic management represents something new, but it does not necessarily mean that workers are now dominated by algorithms. Platforms use technologies that subject workers to new forms of surveillance and attempted control. However, the Foucauldian argument sees workers "become the principle of" their "own subjection" (Foucault, 1991, p. 203). This is the risk of talking about control—or indeed domination—in general terms. Foucault (1991, p. 174) recognized that "the disciplinary gaze did, in fact, need relays… it had to be broken down into smaller elements, but in order to increase its productive function: specify the surveillance and make it functional." In the call center I studied, workers found ways to oppose surveillance and make it less functional. Mulholland (2004, p. 711) notes that general accounts claim that "management is triumphant, and it is suggested that discipline has replaced conflict, when seductive discourses make workers the captives of organizational values." The workplace is not a prison and involves different social relations (McKinlay and Taylor, 1998). This is what makes call centers an interesting example, that the innovations of capital at the time represented "an unprecedented level of attempted control" (Taylor and Bain, 1999, p. 109). Due to the different interests in the labor process, management cannot achieve totalising aims, because "control mechanisms embodied significant levels of managerial coercion and therefore attracted varying levels of resistance" (van den Broek, 2004).

Algorithmic management takes this at least one step further than the call center. Instead of the physical supervision at the center of the prison, instead there is an automated collection of data that runs throughout the entire labor process. As I found in my research with Deliveroo riders, the algorithmic process goes beyond measurement, but rely upon illusions of control and freedom. The threat of algorithmic management is not total and has many gaps and issues in practice. Workers find these through their day-to-day engagement with the platform. The illusion of control can operate relatively effectively in the regular operation of the platform, but suffers when workers struggle against control (Woodcock, 2020). For example, during wildcat strikes which have become a frequent form of protest on platforms (Joyce et al., 2020), there a few options left to the platform, other than introducing boosts to the piece rate.

Struggles over technology

One of the important things that is missing from the panopticon metaphor, either in the call center or with platforms, is that it tends to hide the planner of the system. Artificial intelligence is not neutral and is instead designed for particular purposes. As with the automation of factories, the choices made about the kinds of technologies used and how they are implemented is about more than just efficiencies at work (Noble, 1978).

There are many examples of ways in which workers have circumvented algorithmic control in practice in platform work (Woodcock, 2021), but we know less about the choices that happen inside these companies to implement the technology. However, as Braverman (1998, p. 137) reminds us, capital became built into the machinery of factories:

> Thus, as the process takes shape in the minds of engineers, the labor configuration to operate it takes shape simultaneously in the minds of its designers, and in part shapes the design itself. The equipment is made to be operated; operating costs involve, apart from the cost of the machine itself, the hourly costs of labor, and this is part of the calculation involved in machine design. The design which will enable the operation to be broken down among cheaper operators is the design which is sought be management and engineers who have so internalized this value that it appears to them to have the force of natural law or scientific necessity.

Historically, the introduction of machines has been part of a concerted attempt to undermine workers power in the

workplace. For example, "machinery offers to management the opportunity to do by wholly mechanical means that which it had previously attempted to do by organizational and disciplinary means" (Braverman, 1998, p. 134). Machines provided the opportunity to set and control the pace of work centrally, shifting the balance of power away from workers. The application of technology is not only about efficiency, but also as an attempt at control.

In order to understand the implications of artificial intelligence at work, any analysis needs to consider how this new application of technology builds upon previous interventions in the labor process over a long history of struggles at work. First, artificial intelligence needs to be interrogated, rather than taken for granted. Research needs to critically unpack the relationships involved in the development, use, and resistance to new applications. Second, there are a wide range of forms that artificial intelligence can take. If they are involved in controlling—or attempting to control—work, these can be unpacked further by considering what role they play within the control of the labor process: direction, evaluation, and/or discipline. No system of control at work can operate without bringing these components together and they often rely on human manager/supervision intervention at some point within or across these. This involves understanding how data collection, no matter how complex the data are or how rapidly it can be achieved, is only one part of the process. Data needs to be acted on to become and attempt at control. Third, arguments about artificial intelligence at work need to be put into conversation with the theoretically and empirically rich traditions of labor process theory.

Future research is needed on how specific applications of artificial intelligence are operating in practice in different kinds of work. As the examples of the call center and transport platforms show, the reality of using technology within the labor process is far from straightforward. Empirical studies provide an important way to move our understanding of the implications of different kinds of artificial intelligence at work forward, particularly moving beyond the claims or marketing that are associated with them. Rather than general research, what is needed is critical research that searches for the contradictions, conflicts, and struggles along the supply chains of artificial intelligence. This is part of situating artificial intelligence as a technology that emerges from, and is used within, the existing social relations at work and in society.

Future research can also benefit from analyzing the different types of struggles against power. For example, Foucault notes that there can be struggles "either against forms of domination; against forms of exploitation which separate individuals from what they produce; or against that which ties the individual to himself and submits him to others in this way" (Foucault, 1982, p. 781). Understanding struggles against artificial intelligence at work can be understood through these different types. Is a struggle aimed at domination, exploitation, or against forms of subjectivity and submission more widely? For example, Moore (2022) recent research on data subjects points toward this with emerging struggles for subjectivity. While some may herald artificial intelligence as driving change within the contemporary world, attention needs to be drawn to the interests it serves and the relationships of power, as well as how other interests can struggle against this too.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bain, P., Watson, A., Mulvey, G., Taylor, P., and Gall, G. (2002). Taylorism, targets and the pursuit of quantity and quality by call centre management. *New Tech. Work Employ.* 17, 170–185. doi: 10.1111/1468-005X.00103

Bentham, J. (1995). *The Panopticon Writings*. London: Verso.

Braverman, H. (1998). *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York, NY: Monthly Review Press.

Briziarelli, M. (2019). Spatial politics in the digital realm: the logistics/precarity dialectics and Deliveroo's tertiary space struggles, *Cult. Stud.* 33, 823–840. doi: 10.1080/09502386.2018.1519583

Burawoy, M. (1979). *Manufacturing Consent*. Chicago, IL: University of Chicago Press.

Cant, C. (2019). *Riding for Deliveroo: Resistance in the New Economy*. Cambridge: Polity.

Cant, C., and Mogno, C. (2020). Platform workers of the world, unite! The emergence of the transnational federation of couriers', *South Atl. Q. 119*, 401–411. doi: 10.1215/00382876-8177971

Cook, M. (2018). *A Basic Lack of Understanding*. Available online at: https://notesfrombelow.org/article/a-basic-lack-of-understanding (accessed March 2, 2022).

De Stefano, V., and Aloisi, A. (2019). "Fundamental labour rights, platform work and protection of non-standard workers", in *Labour, Business and Human Rights Law,* eds J. R. Bellace and B. Haar (Cheltenham: Edward Elgar Publishing), 359–379. doi: 10.4337/9781786433114.00033

Edwards, R. (1979). *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. New York: Basic Books.

Ellis, V., and Taylor, P. (2006). "'You don't know what you've got till it's gone": re-contextualising the origins, development and impact of the call centre.' *New Tech. Work Employ.* 21, 107–122. doi: 10.1111/j.1468-005X.2006.00167.x

Eubanks, V. (2019). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fear, C. (2018). "'*Without Our Brain and Muscle not a Single Wheel Can Turn": The IWW Couriers Network.*' Available online at: https://notesfrombelow.org/article/without-our-brain-and-muscle (accessed March 2, 2022).

Fernie, S and Metcalf, D. (1997). *(Not) Hanging on the Telephone: Payment Systems in the New Sweatshops*. London: Centre for Economic Performance at the London School of Economics and Political Science.

Foucault, M. (1982). 'The subject and power', *Crit. Inq.* 8, 777–795. doi: 10.1086/448181

Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison*. London: Penguin.

Gent, C. (2019). *The Politics of Algorithmic Management: Class Composition and Everyday Struggle in Distribution Work*. Coventry: University of Warwick.

Goodrich, C. L. (1975). *The Frontier of Control: A Study in British Workshop Politics*. London: Pluto Press.

Graham, M., and Woodcock, J. (2018). 'Towards a fairer platform economy: introducing the fairwork foundation', *Alter. Routes* 29, 242–253.

Joyce, S., Neumann, D., Trappmann, V., and Umney, C. (2020). 'A global struggle: worker protest in the platform economy.' *ETUI Policy Brief* 2, 1–6. doi: 10.2139/ssrn.3540104

Kitchin, R. (2017). 'Thinking critically about and researching algorithms.' *Inform. Commun. Soc.* 20, 14–29. doi: 10.1080/1369118X.2016.1154087

Kolinko (2002). *Hotlines: Call Centre, Inquiry, Communism*. Oberhausen: Kolinko.

Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. (2015). "Working with machines: the impact of algorithmic, data-driven management on human workers," in *Proceedings of the 33rd Annual ACM SIGCHI Conference,* eds B. Begole, J. Kim, K. Inkpen, and W. Wood (New York: ACM Press), 1603–1612. doi: 10.1145/2702123.2702548

Leonardi, D., Murgia, A., Briziarelli, M., and Armano, E. (2019). 'The ambivalence of logistical connectivity: a co-research with Foodora Riders.' *Work Organ. Labour Global.* 13, 155–171. doi: 10.13169/workorgalaboglob.13.1.0155

Mahnkopf, B. (2020). 'The future of work in the era of "digital capitalism."' *Socialist Register* 56, 111–112.

McKinlay, M., and Taylor, P. (1998). "Foucault and the politics of production," in *Management and Organization Theory,* eds A. McKinlay and L. Starkey (London: Sage), 1–37.

Moore, P. V. (2022). '*Problems in Protections for Working Data Subjects: Becoming Strangers to Ourselves*.' Zemki Communicative Figurations, Working Paper No. 41. doi: 10.2139/ssrn.4050564

Moore, P. V., and Joyce, S. (2019). 'Black box or hidden abode? The expansion and exposure of platform work managerialism.' *Rev. Int. Econ.* 27, 926–948. doi: 10.1080/09692290.2019.1627569

Muldoon, J., and Raekstad, P. (2022). Algorithmic domination in the Gig economy. *Eur. J. Political Theory* 147488512210820. doi: 10.1177/14748851221082078

Mulholland, K. (2002). 'Gender, emotional labour and teamworking in a call centre.' *Pers. Rev.* 31, 283–303. doi: 10.1108/00483480210422714

Mulholland, K. (2004). 'Workplace resistance in an Irish call centre: slammin', scammin' smokin' an' leavin." *Work Employ. Soc.* 18, 709–724. doi: 10.1177/0950017004048691

Noble, D. F. (1978). 'Social choice in machine design: the case of automatically controlled machine tools, and a challenge for labor.' *Politics Soc.* 8, 313–347. doi: 10.1177/003232927800800302

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674736061

Rosenblat, A. (2018). *Uberland: How Algorithms are Rewriting the Rules of Work*. Oakland: University of California Press. doi: 10.1525/9780520970632

Rosenblat, A., and Stark, L. (2016). 'Algorithmic labor and information asymmetries: a case study of uber's drivers', *Int. J. Commun.* 10, 3758–3784.

Scholz, T. (2017). *Uberworked and Underpaid: How Workers are Disrupting the Digital Economy*. Cambridge: Polity.

Srnicek, N. (2017). *Platform Capitalism*. Cambridge: Polity.

Tassinari, A., and Maccarrone, V. (2020). 'Riders on the storm: workplace solidarity among gig economy couriers in Italy and the UK.' *Work Employ. Soc.* 34, 35–54. doi: 10.1177/0950017019862954

Taylor, A. (2018). 'The Automation Charade'. Logic Magazine, no. *ue* 5.

Taylor, F. W. (1967). *The Principles of Scientific Management*. New York: Norton.

Taylor, P., and Bain, P. (1999). 'An assembly line in the head': work and employee relations in the call centre', *Industrial Relat. J.* 30, 101–117. doi: 10.1111/1468-2338.00113

Thompson, P. (1983). The Nature of Work: An Introduction to Debates on the Labour Process. London: Macmillan. Available online at http://books.google.com/books?id=qwzGAAAAIAAJ (accessed February 28, 2022).

Turow, J. (2017). *The Aisles Have Eyes: How Retailers Track Your Shopping, Strip Your Privacy, and Define Your Power*. New Haven, CN: Yale University Press.

van den Broek, D. (2004). "'We have the values": customers, control and corporate ideology in call centre operations.' *New Tech. Work Employ.* 19, 2–13. doi: 10.1111/j.1468-005X.2004.00124.x

Veen, A., Barratt, T., and Goods, C. (2019). 'Platform-capital's "app-etite" for control: a labour process analysis of food-delivery work in Australia', *Work Employ. Soc.* 3, 388–406. doi: 10.1177/0950017019836911

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., and Trichina, E. (2022). 'Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review', *Int. J. Hum. Resour. Manage.* 33, 1237–1266. doi: 10.1080/09585192.2020.1871398

Waters, F., and Woodcock, J. (2017). *Far From Seamless: A Workers' Inquiry at Deliveroo*. Viewpoint Magazine. Available online at: https://www.viewpointmag.com/2017/09/20/far-seamless-workers-inquiry-deliveroo/ (accessed March 2, 2022).

Woodcock, J. (2017). *Working the Phones: Control and Resistance in Call Centres*. London: Pluto. doi: 10.2307/j.ctt1h64kww

Woodcock, J. (2020). 'The algorithmic panopticon at deliveroo: measurement, precarity, and the illusion of control.' *Ephemera* 20, 67–95.

Woodcock, J. (2021). *The Fight Against Platform Capitalism: An Inquiry into the Global Struggles of the Gig Economy*. London: University of Westminster Press. doi: 10.2307/j.ctv1ktbdrm

Woodcock, J., and Cant, C. (2021). 'Platform worker organising at Deliveroo in the UK: from wildcat strikes to building power.' *J. Labor Soc.* 1, 1–17. doi: 10.1163/24714607-bja10050

Woodcock, J., and Graham, M. (2019). *The Gig Economy: A Critical Introduction*. Cambridge: Polity.

frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 19 October 2022
DOI 10.3389/frai.2022.886561

Check for updates

OPEN ACCESS

EDITED BY
Hamed Zolbanin,
University of Dayton, United States

REVIEWED BY
Xiaoyan Liu,
University of North Texas,
United States
Ali Shirzadeh Chaleshtari,
University of Massachusetts Boston,
United States

*CORRESPONDENCE
Ekkehard Ernst
ernste@ilo.org

SPECIALTY SECTION
This article was submitted to
AI in Business,
a section of the journal
Frontiers in Artificial Intelligence

RECEIVED 28 February 2022
ACCEPTED 12 August 2022
PUBLISHED 19 October 2022

CITATION
Ernst E (2022) The AI trilemma: Saving the planet without ruining our jobs. *Front. Artif. Intell.* 5:886561. doi: 10.3389/frai.2022.886561

COPYRIGHT
© 2022 Ernst. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The AI trilemma: Saving the planet without ruining our jobs

Ekkehard Ernst*

International Labour Organization, Department of Research, Geneva, Switzerland

Digitalization and artificial intelligence increasingly affect the world of work. Rising risk of massive job losses have sparked technological fears. Limited income and productivity gains concentrated among a few tech companies are fueling inequalities. In addition, the increasing ecological footprint of digital technologies has become the focus of much discussion. This creates a trilemma of rising inequality, low productivity growth and high ecological costs brought by technological progress. How can this trilemma be resolved? Which digital applications should be promoted specifically? And what should policymakers do to address this trilemma? This contribution shows that policymakers should create suitable conditions to fully exploit the potential in the area of network applications (transport, information exchange, supply, provisioning) in order to reap maximum societal benefits that can be widely shared. This requires shifting incentives away from current uses toward those that can, at least partially, address the trilemma. The contribution analyses the scope and limits of current policy instruments in this regard and discusses alternative approaches that are more aligned with the properties of the emerging technological paradigm underlying the digital economy. In particular, it discusses the possibility of institutional innovations required to address the socio-economic challenges resulting from the technological innovations brought about by artificial intelligence.

KEYWORDS

sustainability, artificial intelligence, inequality, productivity, jobs

1. Introduction

The past decade has seen an explosion of applications powered by artificial intelligence (AI). With the ubiquity of large, unstructured databases ("Big data") and a rapid fall in computing costs over the past four decades, AI applications using non-linear statistical and machine learning methods have gained renewed prominence after falling out of favor for long periods since the inception of the field of AI properly speaking. This has triggered both fears about a robo-apocalypse with machines dominating the world as well as enthusiastic techno-scenarios where humanity can solve most of its current global challenges, be they related to climate change, poverty or diseases (Brynjolfsson and McAfee, 2014; Frey and Osborne, 2017; Frey, 2019; Ford, 2021). Yet, none of these scenarios seem to materialize right now. Rather, we see specific challenges arising from the wide-spread use of AI, in particular when it comes to the use of social media. Also, the rising ecological footprint of digital tools—and specifically AI-powered

applications—notably as regards cryptocurrencies[1] and foundation models[2], has raised concerns about the sustainability of these developments (Robbins and van Wynsberghe, 2022). At the same time, enhancements to our way of life have been equally limited, mostly concentrated around improvements in digital navigation or the rapid rise in online shopping and delivery. At the back of these rather limited effects looms a more concerning trend: the rise in economic power of a few dominant technological companies that increasingly seems to add to inequalities already prevalent before the rise in AI.

By now, all three challenges resulting from the rise of AI are well documented, whether they concern limited productivity gains (Gordon, 2021), worsening inequalities (Bessen, 2020) or rising ecological costs (van Wynsberghe, 2021). This paper argues that these three challenges are interrelated and need to be understood as resulting from an "AI trilemma:" Following its current path the technological paradigm taken by AI will worsen its ecological footprint and deepen economic inequalities without delivering better living standards for all. Using the concept of a technological paradigm as developed by Dosi (1982) and Nightingale et al. (2008), I will argue that at the heart of this trilemma lies a particular way of how this technology develops, related to both technical and economic aspects of its current paradigm. I will also argue that these developments are not inevitable as specific policy interventions and institutional changes can modify this paradigm in such a way as to deliver positive contributions to our way of life without worsening or even with improving on its ecological and social costs to become a truly sustainable paradigm. This point is similar to the one raised by Acemoglu (2022) in as much as the unfettered technological development under the current paradigm is unlikely to deliver the benefits expected from AI; in contrast, I argue that identifying a direction of technological change that delivers these benefits requires to understand the inherent trade-offs between inequality, ecological costs and productivity growth that comes with the current paradigm.

Many researchers and observers focus their analyses of AI on its applications in the world of work, which initially rose fears of wide-spread technological unemployment (Frey and Osborne, 2017; Balliester and Elsheikhi, 2018; Frey, 2019). Whether autonomous taxis, fully automated logistics centers, the Robo-Hotel concierge Pepper or the Bar Tender Tipsy Robot; in more and more areas machines seem to be able to replace us. This is especially true in those areas where we ourselves have been convinced of being irreplaceable: In artistic or intellectual activities (Muro et al., 2019). Calls for a universal basic income or some other unconditional forms of government transfers abound in order to secure all those masses of employees falling out of work and providing them some minimum way of life. In the meantime, however, it seems that (technological) unemployment should be the least of our concerns with these new digital technologies, at least in advanced economies (Carbonero et al., 2018). Indeed, if anything, unemployment has declined in OECD countries during the past decade up until the outbreak of the Covid-19 pandemic (see Figure 1).

Part of the reason why AI-powered applications have so far not led to a job-less future relates to the vary narrow range of applications that are currently being developed by industry (Ernst and Mishra, 2021), affecting only a small percentage of the workforce. Indeed, over the past decade most applications have been centered around business process robotisation, autonomous driving, e-commerce and digital platforms, which together accounted for more than 40 per cent of all applications developed between 2010 and 2020 (see Figure 2). In particular, business process robotisation—such as applications in accounting and compliance—seem to have been developed partly as a reaction to rising compliance cost and regulatory overhead, rather than to substitute employment. Some researchers have even highlighted that many of these applications are likely to prove labor augmenting rather than replacing, possibly leading to job enrichment, which, in principle, should allow workers to command higher incomes and firms to enjoy higher productivity (Fossen and Sorgner, 2019).

Yet, these more positive conclusions also do not seem to have materialized. Productivity growth has continued its secular decline over the 2010s (Ernst et al., 2019) and does not seem to have accelerated with the onset of the recovery as we are gradually moving out of the pandemic. Despite much touted benefits from working-from-home and the further growth in e-commerce, apparent hourly labor productivity growth in the OECD has not increased (see Figure 3), with the possible exception of the United States that saw a gradual increase since the mid-1990s, albeit well below levels achieved in decades prior to the second oil shock in the early 1980s.

Meanwhile the rising ecological cost of developing and using AI has become an important concern. This has become most

1 Cryptocurrencies and blockchain applications more broadly are not strictly relying on artificial intelligence. However, many of their applications do, including latest developments around Decentralized Autonomous Organization applications (DAO) that execute certain functions autonomously or market trading applications to anticipate price movements in these currencies. Many of the ecological implications discussed in this article relative to AI applications do carry over to other digital tools such as the use of blockchains.

2 Foundation models, a term first coined by the Stanford Institute for Human-Centered Artificial Intelligence, are generic models trained on a large set of unlabeled data that can be re-purposed for a specific set of tasks. For example in natural language processing, the Bidirectional Encoder Representations from Transformers (BERT) model has been trained on a large corpus of the English language; the model can then be refined for specific tasks to recognize English sentences in technical applications, such as to identify the similarity in the description of skills in different classification systems (see, for instance, Fossen et al., 2022).

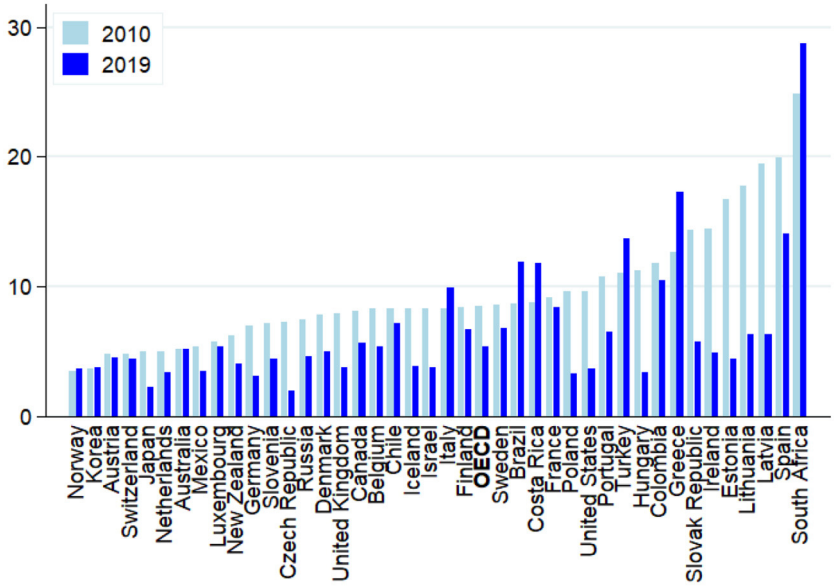


FIGURE 1
Evolution of unemployment: OECD and selected G20 countries (2010 vs. 2019; in per cent of total labor force). Source: OECD, Stats Portal.

visible in the area of cryptocurrencies where the particular security concept behind Bitcoin, for instance, has led to an explosion in the use of electricity, up to the point that several countries have restricted or outright banned its use (e.g., China, Kosovo). Other areas of the digital economy have also experienced increasing constraints. Some large digital companies have started experimenting placing its cloud computing servers in deep sea water for cooling. Large-scale neural networks such as the natural language processing network GPT-3, currently one of the largest and most powerful tools in this area, is reported to cost US$ 12 million on a single training run, making it very costly to correct training errors (for instance due to biased data) and effectively preclude a more wide-spread application of this tool, especially by smaller companies (OECD, 2021). What is more, as these tools become more complex and presumably more precise, their economic and energetic costs explode and do not scale up linearly (Thompson et al., 2020). In the meantime, a call for "Green AI" or sustainable AI has emerged, focusing on how to lower the carbon-footprint of these tools and ensure their (low-cost) accessibility of a large range of researchers and users (Robbins and van Wynsberghe, 2022). Various possible technological improvements have been suggested but, so far, none of them seems promising enough to contribute significantly to a solution as we will discuss in more detail below. Presumably, the rise in renewables in the energy mix would bring down the carbon footprint of AI but only to the extent that its use does not continue the exponential rise observed over the past decade, which seems unlikely.

Interestingly, those areas where AI neither replaces nor (directly) complements work have not received much attention. In economic terms, new technologies can affect productivity at three levels: labor, capital or total factor productivity. The latter typically refers to technologies that help combine both production factors in more efficient ways, for instance through re-organization of work processes. More broadly, technologies to manage networks more efficiently, for example in transport and logistics, in electricity and waste management or in information exchange, are prime candidates for improvements in total factor productivity (UN DESA, 2018). Modern urban traffic control systems can use flexible traffic management to direct individual and public transport in such a way that the traffic volume is managed optimally and efficiently. AI will also become increasingly important in the area of electricity network control, especially where more and different energy sources (e.g., renewables) have to be connected as economies are transiting toward sustainable energy supply. Similarly, as economies are trying to reduce their overall ecological burden, waste management will become more important together with an increasing role played by the circular economy. Such (complex) supply chains remain beyond the purview of human intervention and require high-speed control by machines.

So far, however, none of these applications seem to play an important role in the discussion among economists and social scientists about how transformative this technology potentially can be. As I will argue below, this has to do with the particular way the technology business operates and requires a conscious effort to redirect (partly) our efforts in developing innovations

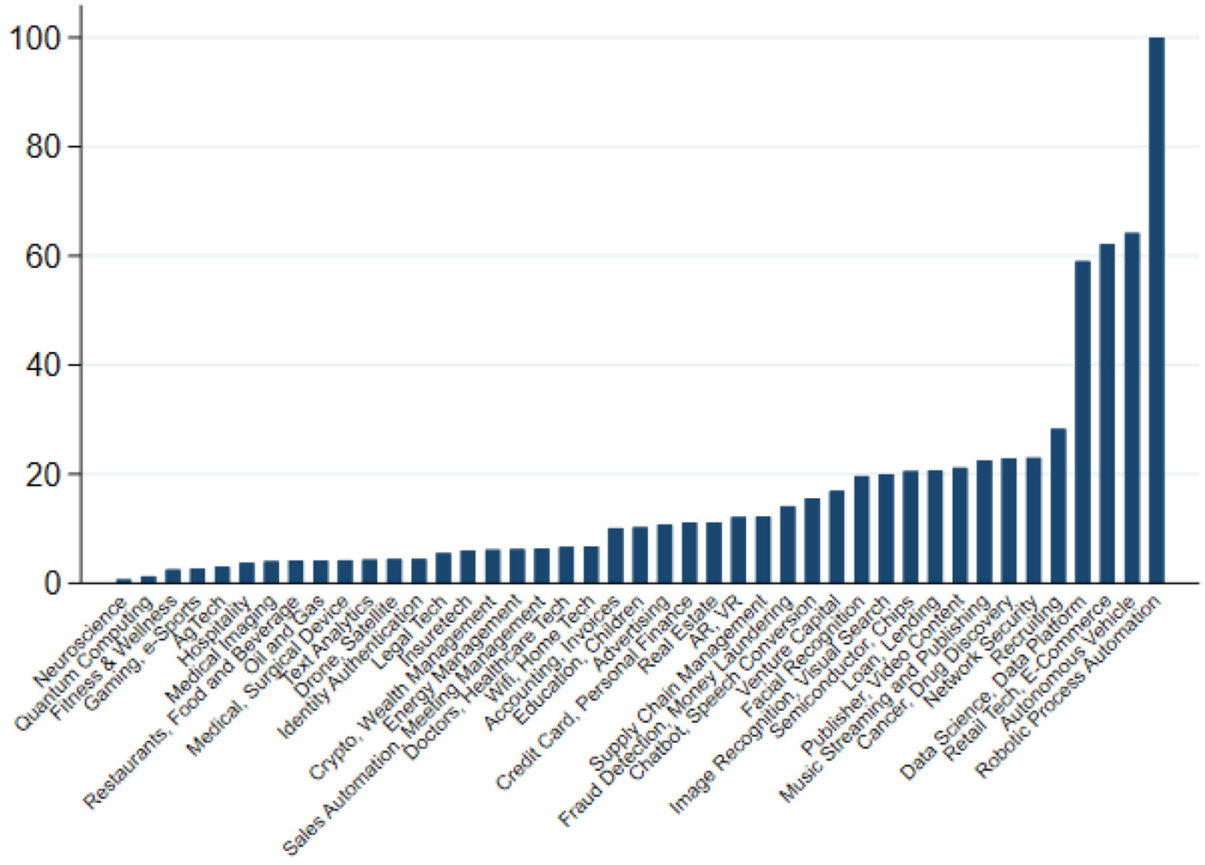


FIGURE 2
Main areas of AI development (2010–2020, cumulative investment in US$, global; Robotic Process Automation = 100). The chart depicts the cumulative, global investment in US$ over the period 2010 to 2020 in various AI-applications. Investments have been scaled such that total investment in Robotic Process Automation = 100. Source: Ernst and Mishra (2021) based on the Stanford AI Vibrancy index.

in this area. I will start with some methodological considerations before presenting the AI trilemma in a nutshell, highlighting the key mechanisms underlying it. I will then delve into its three main components: lack of productivity growth, rising inequality and market concentration, and a worsening ecological footprint. In Section 4, I demonstrate several areas in which technological progress in the digital work can indeed contribute to address the AI trilemma and present some policy proposals on how to instigate such a change. A final section concludes.

2. The AI trilemma in a nutshell: A technological paradigm

2.1. Technological paradigms

Underlying the understanding of the AI trilemma is the concept of a technological paradigm as a socio-technological interaction between technological capabilities, economic conditions and social structures that determine the future development of the productive forces of an economy (Dosi, 1982; Nightingale et al., 2008). A technology here refers to a set of combinations between labor, capital and ideas to produce a certain economic output. At its most basic level, technological development then can be either autonomously driven by scientific progress ("ideas")—the scientific supply push paradigm—or determined by economic conditions under which firms operate on both labor and capital markets—the demand pull paradigm. As such, the concept of a technological paradigm expands on Kuhn's scientific paradigms as one that applies more broadly even outside academic communities.

As highlighted by Dosi (1982), the two ideal forms do not uniquely reflect the dynamics of technological progress, which will inevitably navigate between the available scientific knowledge of any particular era and the specific socio-economic conditions under which firms operate. One shall add to this that either of the two forces will be influenced by institutional and regulatory conditions, such as laws and regulations governing

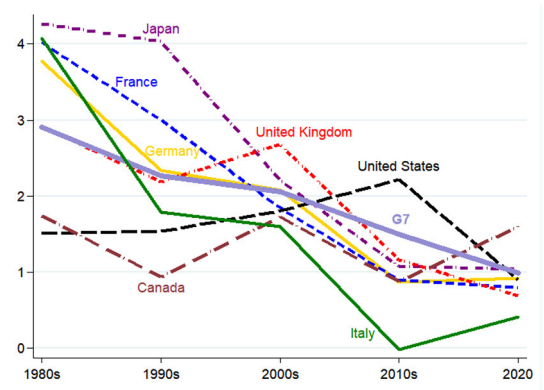


FIGURE 3
Hourly labor productivity growth (in %, decade averages, 1980-2020, G7 countries). The chart depicts the average hourly labor productivity growth between 1980 and 2020 for G7 countries. Decade averages only available for 1980–1989; 1990–1999; 2000–2009; 2010–2019. Last observation is for 2020 and might be biased due to the effect of the COVID-19 pandemic on national statistics. Source: OECD, Statistics Portal, available at: https://stats.oecd.org/.

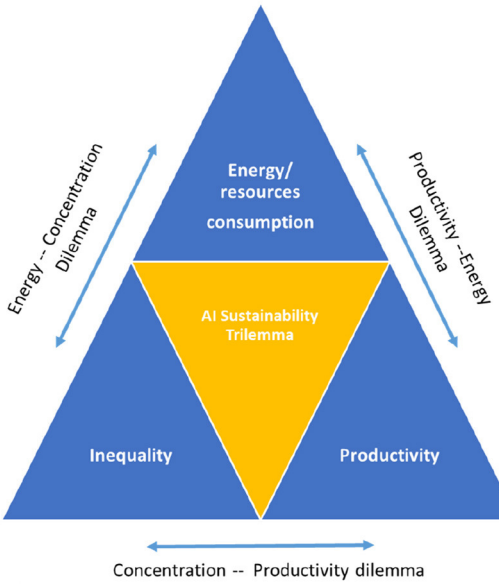


FIGURE 4
The AI sustainability trilemma in a nutshell.

intellectual property rights, tax regimes or government subsidies for R&D, among others.

It is against this concept of technological paradigms that the AI trilemma will be developed in this paper: I will explore how the current scientific and technological development of digital tools in general and AI in particular interacts with the institutional and regulatory regime on labor and capital markets. I will then analyse the specific socio-economic outcomes this interaction produces to show that certain undesirable properties cannot be overcome within this prevailing technological paradigm. As such the AI trilemma is not a logical impossibility to achieve more desirable outcomes with the currently available technologies but rather a contextual trilemma that can be overcome with the right institutional and regulatory adjustments.

To develop my argument, I start by reviewing the technological characteristics of what is typically dubbed "machine intelligence" and compare it with our current understanding of human cognitive processes. Specifically, I will show how the current wave of machine intelligence is correlated with significant scale effects that make economic concentration a prerequisite for further technological development. Through an overview of empirical studies I will demonstrate the extent to which such concentration effects can already be observed and discuss the specific underlying mechanisms. Based on this analysis, I will argue that this tendency for economic concentration has some other, undesirable consequences from a macro-economic standpoint, including a slowdown in technological diffusion and a deceleration of productivity growth. My argument, therefore, consists in considering the current technological paradigm around AI as one of "supply push," driven predominantly by technological considerations, rather than one of "demand pull," oriented by policy goals regarding the development of productive forces and sustainable societies.

The way the AI trilemma is being developed in the following relies on an extensive review of the available evidence as it is being brought together by computer scientists, economists and policy experts to form a new, coherent understanding of the current difficulties that help understanding the apparent contradiction of a seemingly accelerating technological progress and a manifest difficulty to detect this progress in improvements in economic and social indicators.

2.2. The AI trilemma as a supply push paradigm

Figure 4 summarizes the key message of the AI sustainability trilemma: We cannot have low inequality, high productivity and ecological sustainability simultaneously, at least not when pursuing the current technological paradigm underlying the development of AI-powered automated decision making systems. As such, the AI trilemma is composed of three, interrelated dilemmata of which only two can be solved simultaneously at the expense of the third one.

Specifically, the AI trilemma consists of the following three interrelated dilemmata:

- The productivity-energy dilemma (Figure 4, upper-right leg): Rising (labor) productivity can only be achieved through the replacement of human labor by machines at the expense of higher use of energy (electricity). This is not specific to the AI revolution. In the case of the digital economy, it implies that human cognitive work is being substituted by machine intelligence. In the next section we will see more closely that this often means that the energy efficiency of decision-making processes actually declines rather than improves. From the storage of data in cloud computing centers, to data analysis by high-performance computers, to the power consumption of even the smallest mobile digital devices needed to stay connected, the digital economy is already using up more than 6 percent of average electricity consumption. And the trend is accelerating. Without major efficiency gains, electricity consumption is expected to rise to over 20 percent in 2030 (Jones, 2018). This dilemma could only be overcome if productivity were to rise beyond what humans could achieve with the same amount of energy expended. As we argue, this is currently not the case.
- The energy-economic concentration dilemma (Figure 4, upper-left leg): For energy efficiency to increase rather than to fall, data concentration needs to grow further in order to exploit the variation of information in large samples. This is the logic currently underlying the development of approaches such as Large Language Models that exploit almost the entire (English-speaking) library. Given the network externalities involved in data collection (which we will discuss in more detail below), market concentration is bound to worsen, at least within the small segment of data collection and algorithm training. Such concentration of data collection can indeed enhance energy efficiency and hence yield productivity gains but only at the level of individual companies. At the aggregate level, this concentration worsens economic inequalities. This dilemma could only be overcome if access to data were regulated as a public good that allows strong competition among data users. In Section 4, we will discuss different options how this could be achieved.
- The concentration-productivity dilemma (Figure 4, bottom leg): Higher income inequality, especially in mature economies, is associated with lower productivity gains. As incomes are getting more concentrated at the top, aggregate demand grows more sluggishly, slowing down embodied technological change, i.e., that part of technological progress that requires investment in new machines. Whether higher productivity growth increases or declines inequality, on the other hand, depends on whether and how quickly new technologies diffuse throughout the economy. Highly specialized technologies that benefit only few sectors might permanently lift inequality when other sectors of the economy cannot from its advantages. In contrast, General Purpose Technologies are thought to "lift all boats," albeit sometimes with a long delay, creating a J-curve effect (Brynjolfsson et al., 2021) with increases in unemployment in the short run and faster job growth in the long run (Chen and Semmler, 2018). In Section 4, we discuss possible ways of addressing the growth-depressing consequences of higher economic concentration.

There is indeed some debate regarding whether a J-curve effect is relevant in understanding why major economies have not yet seen productivity improvements commensurate of what has been expected from the latest wave of technological advancements. Depending on how flat the "J" is, the effect can take several decades, related to major sectoral restructuring and work-process re-organization. Ernst (2022) argues that because of the rise in inequality triggered by the specific conditions under which digital technologies evolve, it is rather unlikely to see a fast diffusion of these new applications spreading through the economy. In the worst case, these benefits might never materialize broadly. In other words, it is increasing market concentration of digital companies and widening income differentials that prevent stronger growth for all. Digital growth is not inclusive and—depending on the application—it is not resource efficient.

What explains this AI sustainability trilemma? This paper argues that the trilemma—low growth, greater inequality and high energy consumption despite rapid technological progress—is mainly due to the specific technological regime in which the digital economy currently operates: Under the current regime of intellectual property rights, energy efficiency of silicon-based information processing tools can only be achieved through high degrees of data concentration, preventing economy-wide productivity spillovers while generating significant economic inequalities. In other words, it is a supply-push technological paradigm driven by the specific conditions under which technological companies develop their applications. This "weightless economy" now occupies the largest place and leads to market distortions that have so far received insufficient attention (Haskel and Westlake, 2017). Moreover, AI-powered tools trigger various forms of inequality beyond the failure to diffuse its benefits more widely. Indeed, at the micro level, too, problems are emerging that perpetuate existing inequalities. The use of historical data, for instance, necessary to train AI routines, often reflects discrimination, specifically of women or ethnic minorities in the labor market. If an AI-routine is fed with such data without a corresponding filter, the disadvantages will be perpetuated, for instance through continued discrimination in hiring processes. Several major tech companies have already experienced this to their disadvantage. Taken together, the specific institutional and technological characteristics of artificial intelligence and AI-based innovations cause and perpetuate the AI sustainability trilemma. In order to offer possible ways out, however, we first need to better

understand what is driving these three different elements of the trilemma in the next section.

3. Understanding the mechanisms of the trilemma

3.1. Why are brains so much more efficient than computers?

A core assertion of the AI trilemma is that computers are highly energy intensive. Therefore, their massive use in the current digital transformation of our economies comes at a significant cost for the environment, specifically in form of the use of electricity and its related carbon footprint. Looking at it from a total factor productivity perspective—i.e., considering all input factors, labor, capital and energy—we start by exploring the first axis of the AI trilemma: the trade-off between using computing power vs. brain power in the drive toward higher levels of productivity. This first section starts by looking into the reasons why digital tools in general—and machine learning in particular, at least as it is currently being conceived—are high consumers of energy. I discuss key differences between brains and computers, arguing that despite many broad similarities, their underlying architecture and information processes show remarkable differences that explain much of why brains are much more efficient than computers. I also discuss how recent changes in the way computer algorithms have evolved have integrated ideas inspired by neurological research, producing remarkable improvements in computing performance. My core argument in this section is that the way computers are currently being used is unsustainable from an ecological point of view. That is not to say that a different kind of use could not prove beneficial for society, but it would require reorienting our current technological paradigm away from trying to substitute for human cognition toward a paradigm where computers and brains are complements[3].

Key for the argument in this section will be to understand the trade-offs involved between the functioning of a computer in comparison to the brain. This might come as a surprise for some as computers are often being seen and modeled following the architecture of the brain. Indeed, seeing the computer as the (better) version of the brain has a long history, going back to the early beginnings of the computer age (Cobb, 2020, ch. 12). Yet, there are fundamental differences in the working of a computer and a brain, beyond the physical characteristics of both (inorganic vs organic matter).

What adds to confounding both—computers and brains—is the fact that key components with similar function are present in both: Memory and circuits, i.e., structured connections between elementary units that can recall previously stored information—using transistors in the case of computers and neurons in the case of brains. Both elements have been shown to be essential for information processing. Indeed at a fundamental level, all mathematical functions can be represented by a suitable connection of basic logical gates, represented as neural networks, which makes the comparison of computers and brains particularly appealing (Hornik et al., 1989). Moreover, progress in computing performance over the past decades has been driven to a non-negligible part by improvements in algorithm design, often inspired by a better understanding of some of the key principles behind the workings of the brain. The exponential development and use of neural networks, for instance, was responsible for vast improvements over and above what simple hardware developments would have made possible (Sherry and Thompson, 2021).

As a consequence, many researchers consider that a convergence of computers toward brains is underway. Moreover, the rapid growth in applications around artificial intelligence suggests that computers would eventually not only work in a fashion similar to brains, they would even follow the same information process, making predictions based on limited information inputs (Friston, 2010; Agrawal et al., 2018). And yet, a direct comparison reveals significant differences in terms of performance and efficiency (see Table 1). In particular, a trade-off becomes apparent regarding the energy consumption and the precision/speed at which calculations are being carried out: individual human neurons are rather slow and imprecise when it comes to processing information. At the same time, they turn out to be much more powerful than transistors in computers, displaying much more complex patterns of activity than a simple binary activation potential (Gidon et al., 2020). On the other hand, computers can calculate at a significantly higher speed and precision, even though most of them dispose of less transistors and connections with much simpler activation patterns[4]. Moreover, this higher precision and speed comes at a significant price tag in the form of higher energy consumption.

Similarly, computers are significantly better at long-term storage of information (memory), which can span several decades, depending on the physical characteristics, the rate of technological obsolescence and processes to transfer information from one (digital) medium to another. In contrast, humans have difficulties in recollecting precisely even personal information, can easily be manipulated in what they remember

3 My argument is different from a general backlash against technological progress and rather stresses the comparative advantages each cognitive technology brings, see, for instance, the criticism expressed here: https://datainnovation.org/2022/01/innovation-wars-episode-ai-the-techlash-strikes-back/.

4 The hardware evolution continues to add significant amounts of transistors every year. At the time of writing, the largest computer, the Chinese-built supercomputer Sunway TaihuLight counted around 400 trillion transistors across all its CPUs. https://en.wikipedia.org/wiki/Transistor_count.

TABLE 1 Comparing (traditional) computers and brains.

Properties	Computer	Human brain
Number of basic units	Up to 114 billion transistors	∼100 billion neurons; ∼100 trillion synapses
Speed of basic operation	20 teraflops/s.	<1,000/s
Precision	1 in 18.4 quintillion (for a 64-bit processor)	∼1 in 100
Power consumption	up to 215 watt	∼10 watt
Information processing mode	mostly serial with 20 cores	serial and massively parallel
Input/output for each unit	1–3	∼1,000
Signaling mode	digital	digital and analog

Note: Based on an Apple M1 Ultra chip in 2022. Flops, floating-point operations per second.
Source: https://support.apple.com/en-am/HT213100 Herculano-Houzel (2009), Luo (2015).

(Shaw, 2016) and "suffer" systematically from forgetting due to the plasticity of the brain that adjusts to external input, something a computer cannot do (Ryan and Frankland, 2022).

Several architectural differences between computers and brains seem to explain a large part of the observed differences in performance, albeit computer scientists are keen in trying to close the gap regarding some of them. The question then becomes: if the architectural differences can be closed, would computers still perform better than brains where they currently have their comparative advantages? In other words: would it not be preferable to improve computers along the dimension where they currently have an advantage rather than trying to emulate the brain? At least from an economic point of view, such a trade-off would call for a more careful assessment of the use of digital tools depending on where their comparative advantages lie. In the following, I focus on four differences that are relevant from an efficiency point of view[5].

A first difference, as noted in Table 1, stems from the parallel structure of the brain in comparison to the mostly serial way a computer functions. The massive expansion of machine learning approaches in computer science demonstrates that enormous efficiency gains can be achieved by parallelizing calculations in the computer. Essentially, neural networks that lie at the heart of recent progress in artificial intelligence use layers of parallel nodes stacked one upon each other, similar to the structure found in the brain, at least to a first order. Researchers increasingly recognize, however, that it is not only the parallel structure but also the specific way in which neurons are connected that explains performance differences (Luo, 2021). Indeed, the importance of a particular network topology in explaining this network's function is currently an active area of research and some of the insights are already being reflected in the way neural networks are being set up in order to further enhance their performance (Zambra et al., 2020). Related, the brain seems to be hardwired for particular tasks that are important for our social experience. For instance, our capacity to recognize faces (Alais et al., 2021) or letters (Turoman and Styles, 2017) seem to be hard-wired in our brains, whereas computers need to learn this. Similarly, we all seem to benefit from a universal grammar that allows us to learn language even without ever being exposed to the full richness of a language, a point made long ago by Noam Chomsky[6]. Such "pre-training," although increasingly used in ML-applications makes our brain particularly energy-efficient if only less flexible.

A second difference lies with the particular way memory is structured in the brain. For one, memory loss as discussed before seems to play a significant role in enhancing a brain's energy efficiency by gradually removing information no longer needed (Li and van Rossum, 2020). Moreover, rather than having a fixed-size memory chip that stores all our information, memory is distributed and stored dynamically. Information, therefore, does not need to be shifted around and read out but is accessible exactly where it is needed. This has inspired recent research to develop integrated memory-computing circuits that allow information being stored where calculations are taken place, so called "mem-resistors" (Zahedinejad et al., 2022). So far, this remains experimental and has not yet been successfully implemented in large-scale computing but shows that significant efficiency gains even in hardware design are still available.

A third, and for our argument most decisive difference lies in the way information is being recorded in neurons in comparison to computer bytes. Indeed, computers process information in the form of small, fixed-sized chunks, so called bytes, in binary format. Regardless of the specific computer type, at any point in time during the operation, a significant number of the individual bits that compose each byte are active. In other words, computers use "dense representation" of information. More importantly, every time such a bit loses its action potential through a computing operation, energy is being released. In contrast, neurons have been shown to operate with sparse representations, where individual dendrites of a neuron are being activated when a certain (small) percentage of a large set of potential links is active, often less than 5 per cent (Ahmad and Scheinkman, 2016; Hawkins and Ahmad, 2016; Hole and Ahmad, 2021). Not only do operations on sparse representations use much less energy than those

5 There are further differences that are less relevant for our argument, such as embodiment. A good overview of the differences between the brain and how artificial intelligence is being set up, see https://www.technologyreview.com/2021/03/03/1020247/artificial-intelligence-brain-neuroscience-jeff-hawkins/.

6 https://thebrain.mcgill.ca/flash/capsules/outil_rouge06.html, https://theconversation.com/our-ability-to-recognise-letters-could-be-hard-wired-into-our-brains-83991

of dense ones—most operations involve zeros—they are also particularly robust against errors: Calculations by Hawkins and Ahmad (2016) demonstrate that for typical synapses error rates can reach 50 per cent without neurons losing their capacity to properly identifying underlying patterns. Such robustness against errors is an additional contributor for energy efficiency as it avoids costly error correction of calculations that need to be done on standard computing devices, in particular for critical hardware.

Finally, while these architectural differences primarily point to differences in the hardware, sparsity is also an important issue regarding algorithmic differences between computers and brains. As highlighted by Kahneman (2011), humans dispose of two main modes of decision making: slow, optimizing and calculating decision processes and fast, heuristic routines. The latter might come with cognitive biases but allow for quick decisions, in particular relevant in periods of stress and high threats. Heuristics are typically domain-specific, which is why their application to other domains induce cognitive biases by not considering all relevant options (Gigerenzer et al., 2011). At the same time, they are fast and energy-efficient. A role performed by the brain in this regard is to identify the specific situation and to mobilize the relevant resources for each decision problem. In contrast, algorithms currently employed in computers will systematically mobilize all available resources for any problem. Integrating these considerations there are shifts toward the use of more specialized CPUs that focus on particular tasks with more efficiency. So far, however, this more modular and specialized set-up has not reached the level of sophistication of the brain.

Taken together, the specific advantage of computers lies with fast, high precision calculations, such as those needed to design high-tech devices or to search quickly through the available library of human knowledge (or protein folding for that matter). In contrast, human brains have evolved to respond to particular challenges posed by our social environment in which empathy and understanding social settings play a fundamental role. Here, coordination, collaboration and adaptability to changing (social) circumstances are key for (collective) success, a task that is difficult for a computer to achieve as it is programmed for a (fixed) number of tasks. A first result of this comparison of the relative performance of computers vs brains, therefore, is the complementarity rather than substitutability of brain vs. computing power. This ties nicely with other research indicating the importance of AI as a transformative force rather than a disruptive one (Fossen and Sorgner, 2019; Carbonero et al., 2021). It also implies that current attempts to generate productivity gains by massively substituting labor for computers will not lead to the expected outcomes. Rather it will lead to a worsening of the energy bill of those companies that rely on such technologies.

As a consequence, technological developments of digital devices in general and AI-powered tools in particular suggest an exponential rise in the ecological footprint under the current technology paradigm (Jones, 2018; Thompson et al., 2020). A simple projection of the growth in model size that are driven by rising demands for precision shows that both the economic and ecological costs would quickly become unsustainable (see Table 2). As noted by the authors, this projection is a simple illustration and the trajectory unlikely to be followed literally as economic, financial and ecological constraints would prevent it from happen. One area, where this can already be observed regards applications around cryptocurrencies where several jurisdictions have issued restrictions or outright bans for so-called "mining" of currencies on their territory, mostly for reasons related to the rising energy costs (with knock-on effects on other activities in these countries).

Regardless of the limits to growth for specific applications, a key challenge in promoting more efficient computing procedures and in assessing which tasks can better be carried out by humans rather than machines remains the proper assessment of the energy consumption involved over the entire computing value chain, from data collection, storage to machine learning and data use (García-Martín et al., 2019; Henderson et al., 2020).

3.2. Information rules: Consequences for market structure

A direct consequence of this high energy consumption is a rising market concentration among AI producing companies and a concentration of AI applications around the most promising—i.e., most profitable—applications as shown in Figure 2. One of the direct consequences of the rising economic costs implied by the exponential increase in energy consumption is a "narrowing of AI research" (Klinger et al., 2022). As highlighted by the authors, this narrowing of AI research is linked to a focus on data- and computational-intensive approaches around deep learning at the expense of other approaches in artificial intelligence that might be more easily accessible by smaller research outlets and academic researchers. Indeed, the ubiquitous availability of large, unstructured databases and the exponential fall in computing costs since the 1980s have contributed researchers to focus on a particular branch of AI development, namely statistical and machine learning at the expense of earlier attempts using symbolic AI to program expert systems which are potentially more easily accessible by a wider group of developers.

Related, narrowing AI research and rising ecological and economic cost lead to market concentration, both in the development and training of new (large) models (Bender et al., 2021) and in related digital applications such as blockchain applications in cryptocurrency markets, where similar tendencies to oligopolistic concentration can be observed (Arnosti and Weinberg, 2022). This should not come as a

TABLE 2 Computational costs of deep learning.

Benchmark	Error rate	Polynomial			Exponential		
		Computations required	Environmental cost (CO_2)	Economic cost ($)	Computations required	Environmental cost (CO_2)	Economic cost ($)
ImageNet	Today: 11.5%	10^{14}	10^{6}	10^{8}	10^{14}	10^{6}	10^{6}
	Target 1: 5%	10^{19}	10^{10}	10^{11}	10^{27}	10^{19}	10^{19}
	Target 2: 1%	10^{28}	10^{20}	10^{20}	10^{120}	10^{112}	10^{112}
MS COCO	Today: 46.7%	10^{14}	10^{6}	10^{6}	10^{15}	10^{7}	10^{7}
	Target 1: 30%	10^{23}	10^{14}	10^{15}	10^{29}	10^{21}	10^{21}
	Target 2: 10%	10^{44}	10^{36}	10^{36}	10^{107}	10^{99}	10^{99}
SQuAD 1.1	Today: 4.621%	10^{13}	10^{4}	10^{5}	10^{13}	10^{5}	10^{5}
	Target 1: 2%	10^{15}	10^{7}	10^{7}	10^{23}	10^{15}	10^{15}
	Target 2: 1%	10^{18}	10^{10}	10^{10}	10^{40}	10^{32}	10^{32}
CoLLN 2003	Today: 6.5%	10^{13}	10^{5}	10^{5}	10^{13}	10^{5}	10^{5}
	Target 1: 2%	10^{43}	10^{35}	10^{35}	10^{82}	10^{73}	10^{74}
	Target 2: 1%	10^{61}	10^{53}	10^{53}	10^{181}	10^{173}	10^{173}
WMT 2014 (EN-FR)	Today: 54.4%	10^{12}	10^{4}	10^{4}	10^{12}	10^{4}	10^{4}
	Target 1: 30%	10^{23}	10^{15}	10^{15}	10^{30}	10^{22}	10^{22}
	Target 2: 10%	10^{43}	10^{35}	10^{35}	10^{107}	10^{99}	10^{100}

Computations required in Gflops.
Source: Thompson et al. (2020), p. 14.

surprise as any market that requires large fixed investment to enter will show signs of concentration. This does not need to be a problem if alternative products and services are available that are (close) substitutes, a situation of monopolistic competition, which is at the heart of many models of economic growth (Aghion and Howitt, 1992). However, the narrowing of AI research suggests that the offer of such potential close substitutes is also declining, which would indeed lead to a concentration of the market as a whole. Indeed, the tendency of digital technologies to lead to superstar firms that dominate their market with knock-on effects on both down- and upstream market power is increasingly well documented (Coveri et al., 2021; Rikap, 2021).

But there is another force that pushes the data economy toward concentration: the network externalities of data collection (Jones and Tonetti, 2020). Indeed, individual data has three characteristics that distinguish it from standard goods and services: (1) its provision is (almost) costless and often done as a byproduct of other activities (such as purchasing a good online; Arrieta-Ibarra et al., 2018); (2) once provided it can be shared and re-used without costs; and (3) finally, its individual value is almost negligible other than in some extreme cases (e.g., rare diseases). Only as part of a larger database will individual data generate some economic value, for instance in order to determine customer profiles or applicants' characteristics (Varian, 2018). Such network externalities are known to lead to concentration effects as has become obvious with the rising share of only a small number of platform and social media providers on global stock exchanges.

In principle, concentration due to network externalities can be productivity enhancing, provided that the productivity gains generated from data concentration are being shared with platform users. This can happen, for instance when platforms are price-regulated, a principle that has been applied with previous network monopolies in telecommunication or electricity distribution. In the case of data monopolies, this is almost never possible as the use of many of these digital tools is not priced and users pay these services through alternative means, more difficult to regulate (e.g., exposure to commercials). Alternatively, stiff competition by alternative platform providers could help share these productivity gains more widely, but many of the incumbent platforms have grown so big that they either pre-emptively purchase potential competitors (e.g., Instagram in the case of Facebook) or use predatory pricing strategies against possible newcomers in order to limit their growth or reduce entry altogether (as in the case of Amazon, see Khan, 2017). As a consequence, productivity gains remain highly concentrated among a few, ever larger firms that see their evaluations sky-rocket. In contrast, the average company in OECD countries

has barely experienced any (productivity) growth despite an ever larger investment in digital assets (Andrews et al., 2015; Haskel and Westlake, 2017). Indeed, a simple calculation can show that these gains represented in the form of rising stockmarket evaluations have macro-economic proportions: If the entire stockmarket value of the five largest digital companies were to be paid out as an indefinite annuity, US GDP would grow by almost 1.1 per cent, a significant improvement[7].

Distributional aspects of the rising use of AI do not only appear at the macro-economic level, they also arise at the micro- and the meso-economic level. A direct consequence of the increased capacity of algorithms to treat large databases is the possibility for much refined pricing strategies, so-called individual pricing (or price discrimination). Such approaches redistribute welfare gains from consumers to producers, which can, under certain circumstances, be welfare-enhancing to the extent that they allow to increase the overall volume of production. Indeed, it can be shown that these circumstances arise fairly easily, which would argue for a more relaxed stance on such price discriminating strategies (Varian, 1985, 2010). On the other hand, research increasingly demonstrates that with the scaling of AI, these welfare-enhancing output expansion is exactly what is lacking: Instead, customer discrimination is being used to exclude certain socio-demographic categories from being served. This is particularly problematic in applications for human resources management, for instance, where automated hiring tools often seem to apply overly strict criteria for selection, thereby excluding large parts of the applicant pool ("hidden workers," Fuller et al., 2021). Often, this is being discussed as algorithmic discrimination due to biases in historical databases upon which these algorithms are being trained. More profoundly, however, the reason for these welfare-reducing effects of AI in such cases lies in the legal prerogatives to prevent open discrimination, thereby setting incentives for firms to restrict services to certain groups only.

Recently the debate has started to focus on the distributional impact of algorithms at the meso-economic level, specifically on issues arising from algorithmic collusion (OECD, 2017; Calvano et al., 2020). In a traditional setting, pre-agreement is often necessary in a market with only few players in order to move from the welfare maximizing price level (the "Bertrand oligopoly") to a profit-maximizing but welfare-reducing higher price level with lower output (the "Cournot oligopoly"). Anti-trust regulators, therefore, spend significant effort in documenting such written or oral commitments to compete on quantities rather than on prices. In a world where prices can be adjusted almost instantaneously and through algorithms, such agreements are no longer necessary: algorithms would learn from each others behavior and tacitly agree on profit-maximizing pricing strategies (Ezrachi and Stucke, 2020).

There is substantial disagreement, however, as to whether such tacit collusion has already been observed or could even become a serious threat not only to income distribution but to efficiency-gains to be obtained from AI (Dorner, 2021)[8]. Evidence is available primarily from online platforms, such as online drug sellers or airline ticket pricing (Brown and McKay, forthcoming) but also retail gasoline market where prices adjust frequently and increasingly through the use of algorithms (Assad et al., 2020). Whether markets are prone to algorithmic collusion might depend on the characteristics of the product or service sold, including the frequency of trades, the degree of transparency and the homogeneity of products, besides the availability of algorithms that could exploit such opportunities (Bernhardt and Dewenter, 2020). Regardless of how widespread the phenomenon is today, however, traditional anti-trust regulation will have difficulties to identify such cases, precisely because of their tacit nature. There is, therefore, a risk that scaling up the use of AI in determining prices (and wages) will not only lead to further concentration and rent seeking behavior, it will also significantly reduce efficiency regardless of any labor displacement effects these technologies might have. Some options exist to regulate firm behavior through appropriate setting of fines and divestitures but current examples involving social media platforms suggest that such regulatory activism is likely met with strong resistance (Beneke and Mackenrodt, 2021).

3.3. Why do we not see more productivity growth?

The last aspect of our AI trilemma looks at the low and declining productivity growth observed in most advanced countries and major emerging economies. As noted above, economists have long noted a productivity puzzle between the apparent acceleration in technological progress, specifically around digital technologies, and the lack of observed productivity gains, at least at the national level (Brynjolfsson et al., 2019). To understand this puzzle, national productivity growth needs to be broken down into its

7 At the end of 2021, the fifth largest digital companies were (by stockmarket valuation): Apple ($2.91 T), Microsoft ($2.53 T), Alphabet/Google ($1.92 T), Amazon ($1.69 T), and Tesla ($1.06 T). Assuming an annuity with an infinite time horizon paid out at the historical average real return for US treasury bonds (around 2.47 per cent p.a.), this would lead to a total annual pay-out of around $250 B or slightly less than 1.1 per cent of US GDP in 2021 ($23 T). Stockmarket valuations are taken from https://companiesmarketcap.com/, US GDP comes from the Bureau of Economic Analysis: https://www.bea.gov/news/2022/gross-domestic-product-fourth-quarter-and-year-2021-second-estimate and historical (real) treasury bond rates have been calculated on the basis of Jordà et al. (2019).

8 See also https://www.autoritedelaconcurrence.fr/sites/default/files/algorithms-and-competition.pdf.

components: Indeed, aggregate increases in productivity are the product of productivity improvements at the firm or factory level and the spread of these gains across the economy. Simply put, productivity = innovation times diffusion. The question therefore becomes twofold: Is the lack of observed productivity gains due to a failure of the digital economy and AI to push productivity at the individual firm level or is it related to a failure of such gains to diffuse through the economy more broadly. The answer researchers have given so far is: problems reside at both ends and are possibly linked.

At the firm level, the introduction of new technologies in general and AI in particular has always been confronted with a necessary re-organization of work processes (Dhondt et al., 2021). As such re-organization takes time and energy, a J-curve effect arises: Each new technology requires upfront costs in the form of restructuring that might actually depress productivity and firm profitability. Once these adjustments have successfully taken place, however, productivity will rise above the level at the start of the adjustment process (Brynjolfsson et al., 2021). At the firm level, evidence is indeed emerging that the recent surge in patenting around artificial intelligence and robotisation has led to a global increase in firm level productivity, especially among SMEs and in services (Damioli et al., 2021). Research specifically for the United States seems to suggest, however, that effects of AI are particularly strong in large firms that patent significantly (Alderucci et al., 2020). Looking at productivity spillovers, on the other hand, Venturini (2022) suggests that at least during the early periods of the transition toward automation based on AI and robotics, significant spillovers might have contributed to the observed productivity increases. In other words, despite increases in productivity at both the firm and the sectoral level that were driven by AI and robotization, aggregate apparent labor productivity growth decelerated, suggesting that other factors must have been holding back the possible positive contribution of AI on growth.

One possible factor might lie in the restructuring of production chains. Indeed, as highlighted by McNerney et al. (2022), as economies mature, production chains normally become longer, which increases their capacity to generate aggregate productivity growth from individual, firm-level or sectoral improvements in productivity. However, over the last 15 years, global trade growth has stalled, suggesting at least a stagnation if not shrinking of the length of production chains, which would suggest a loss in the capacity of AI to generate productivity growth at the aggregate level. Unfortunately, the evidence in McNerney et al. (2022) stops in 2009 but suggests that some of these dampening effects of slow global trade growth might indeed have started to appear toward the end of their observation period.

Closer to the argument developed here, the narrowing of AI research suggests another possibility, following Zuboff (2019): Indeed, the rapid increase in AI applications might be concentrated around surveillance software and human resources management tools that impact workplace organization more than it contributes to overall productivity increases. Part of the restructuring induced by such software impacts not so much the overall innovative capacity of firms but rather the type of innovation carried out, with little impact on firm profitability and employee output. In other words, rising investment in this type of AI focused on HR management helps more with overall information processing and incentive provisions than it does for value creation, which is why firm level studies suggests that only some firms seem to benefit from these tools.

At the macro level, another factor limiting aggregate productivity gains from AI is explored by Gries and Naudé (2020) expanding on Acemoglu and Restrepo (2019) and analyzing an endogenous growth model. The authors analyse the impact of AI-induced automation of tasks rather than entire jobs, demonstrating that regardless of the elasticity of substitution between AI and human labor, the aggregate labor income share falls, with adverse consequences for aggregate demand and productivity growth. When the elasticity of substitution is high, the displacement effect is always greater than the reinstatement effect of new tasks (Acemoglu and Restrepo, 2019). However, Gries and Naudé (2020) show that even in the case when the elasticity of substitution is low, the reinstatement effect fails to compensate for labor displacement in an endogenous growth setting provided that the benefits from AI are heavily concentrated among capital owners, a direct consequence from the distributional aspects of AI discussed in the previous section. In contrast to previous waves of automation, therefore, the data economy generates highly concentrated benefits that do not generate enough demand spillovers to push up growth on a broad basis.

A last factor, intimately related to the distributional consequences of the data economy concerns its impact on the degree of market competition, a point stressed by Aghion et al. (2021). Indeed, Schumpeterian rents arising from innovation such as AI need to be gradually eroded through the entry of new producers of highly substitutable goods and services in order to allow for a wide diffusion of productivity gains. This is the essence of Aghion and Howitt (1992)'s original work on creative destruction and subsequent empirical evidence. As demonstrated by Hidalgo and Hausmann (2009) and Pinheiro et al. (2021) when such growth models are prevalent in a large range of unrelated sectors they lead countries on a path of high and persistent economic development. In this case, monopolistic competition coupled with creative destruction ensures the continued upgrading of productivity across a broad range of sectors, a model that was followed broadly during the first two waves of industrial revolutions. However, with the arrival of digital capitalism and data markets, the data rents generated by platform providers and AI innovators only partly diffuse through the economy, thereby lowering labor income shares

and aggregate demand, a trend observed since the arrival of the computing revolution in the 1980s that continues until today.

This ties well with another observation that has puzzled economists for some time: The decline in business creation and start-up activity over the past two decades (Bessen, 2022). Indeed, the trend toward rising market power across the globe is well documented, following directly from a lack of market contestability by smaller, younger firms (Eeckhout, 2021). As Bessen (2020) demonstrates, this trend toward industry concentration can be directly linked to the rise in the data economy and the related growth in proprietary information technology. Such industry concentration, even if driven by innovative products and services, are not without adverse consequences for aggregate productivity growth (De Loecker et al., 2020).

This then closes the loop of the AI trilemma. Despite the potential of creating substantial productivity gains at the firm level and some evidence for productivity spillovers, the potential for a broad-based increase in aggregate productivity is limited by the adverse distributional consequences of the way the data economy functions. Empirically, this shows up in a widening productivity gap between frontier firms and the rest (Andrews et al., 2015). At the same time, the high energy consumption not only limits the societal benefits of this technology; it is itself partly responsible for the high concentration of AI providers and a narrowing of AI applications. In this regard, the suggestions put forward by some observers to alter the regulatory environment of the data economy, for instance by modifying current regulation on intellectual property rights might not be sufficient to address the trilemma as presented here (e.g., Karakilic, 2019). We will see in the next section that solving the AI trilemma requires a more encompassing approach that targets the specific benefits that a widespread adoption of AI can have by mitigating its adverse ecological and social costs.

4. Solving the AI trilemma

Dissecting the underpinnings of the AI trilemma allows an understanding of how to address it. Key to any policy or regulatory intervention is that the trilemma is specific to the current technological paradigm under which the digital economy develops, not an inherent characteristic of the technology. Such paradigms are subject not only to the physical characteristics of a specific technology but also to the institutional framework under which the technology is being developed (Dosi, 1982; Bassanini and Ernst, 2002; Nightingale et al., 2008). Specifically, as argued in the previous section, the current technological paradigm is one of a supply-push, where technology develops mostly through individual company strategies. In this section, I argue that to overcome the AI trilemma a switch to a demand-pull technological regime is necessary where technology develops through a deliberate shift in the institutional framework geared toward applications beneficial from a societal perspective.

In the following, I offer three approaches to address the AI trilemma, each one targeting one specific axis of the trilemma as highlighted by Figure 4. What follows from the discussion in the previous section is that breaking the trilemma requires one of three things: an orientation of technological development toward complementary, efficiency enhancing innovations; a more equitable distribution of innovation rents; or a more widespread diffusion of productivity gains through restoration of competitive markets.

A first approach uses standard public economics: If the current technological regime under which AI development operates produces externalities (environmental, social, etc.), these need to be internalized through regulatory or institutional changes, for instance through changes in the corporate tax code or by strengthening labor market institutions. A second approach considers direct interventions to orient technological development through policy action into applications with high societal value that can lift productivity growth sufficiently to justify the additional energy consumption, i.e., an approach that will lead to an overall reduction in total resource consumption. A final approach focuses on the concentration dilemma, addressing the public goods problem of the current regime of digital technologies. The following Figure 5 summarizes the solutions for solving the AI trilemma that are being discussed in the following.

4.1. Solving the energy-concentration dilemma: Shifting the energy-labor balance

Addressing the AI trilemma faces two interconnected challenges: (i) steering technological progress into a direction that is at least neutral and ideally complementary to jobs (Mazzucato, 2021) so that the introduction of new machines strengthens the demand for labor; and (ii) ensuring that technological progress in general—and the increasing use of AI in particular—reduces its ecological footprint rather than to increase it (Acemoglu et al., 2012). However, as discussed in much of the literature on environmental transition, these two objectives often conflict, not least because the investment in new, environmental technologies requires time to allow for resources to be fully re-allocated. Moreover, many jobs in industries that have a heavy ecological footprint are often well-paying jobs for workers with less than graduate degrees (Montt et al., 2018). In other words, our AI trilemma induces a policy trade-off between better jobs and more energy efficiency, with both transitions possibly coming at the cost of a—at least—temporary slow-down or even reduction in productivity growth.

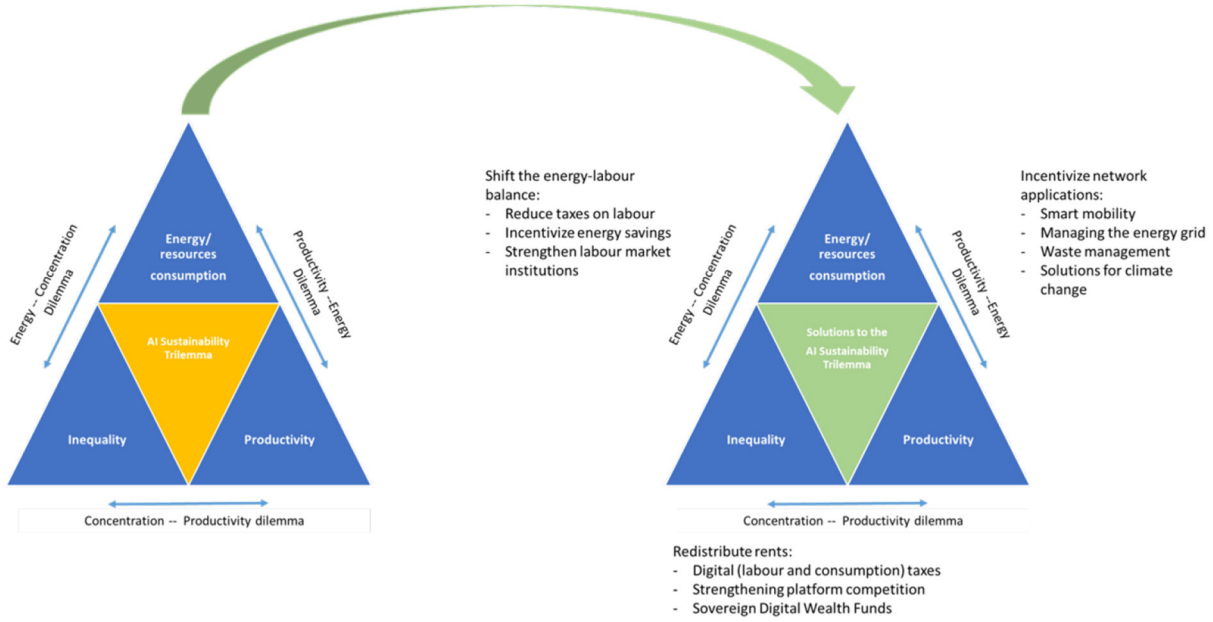


FIGURE 5
Solving the AI trilemma.

Here we want to suggest an alternative adjustment path that can tackle these problems directly and still solve the AI trilemma. This is made possible by the particular characteristics of AI, which did not exist to the same extent with previous forms of technological change, including technologies such as robots. For this, we need to extend our view on aggregate production to not only include energy but also organizational capital broadly understood:

$$Y = A\,(K, L, E, O)$$

where input factors are noted as K for capital, L for labor, E for energy, and O for organizational capital. Such extensions have a long history in economics, especially in firm-level empirical analysis (see, for instance, Atkeson and Kehoe, 2005). At the macro-economic level, however, improvements in organizational capital, O, are typically subsumed under the heading of "total factor productivity," without clarifying whether these occur at the micro-, meso- or macro-economic level.

Current conceptualisations focus on AI as a technology that either replaces or complements jobs similar to previous waves of automation (Fossen and Sorgner, 2019). Notwithstanding the fact that economic analysis has increasingly focused on the impact of technology not on the individual job but on the underlying tasks that are being performed by a job (Autor et al., 2003, 2006; Autor, 2013; Acemoglu and Restrepo, 2019), AI is not considered to be distinct from previous forms of technological progress in this respect. However, as discussed in the opening part of the previous section, one specificity of AI is its capacity to process information in order to make predictions, for instance regarding the dynamics of a particular system. At the micro-economic level, such predictions can help an individual worker, for instance, in respecting a certain order in which to process the workflow by giving recommendations about the next step. Similarly, in a research environment, AI has been used to facilitate the discovery process of new drugs, thereby improving the productivity of the innovation process. At the sectoral level, AI can and has been used for dynamic pricing purposes (Calvano et al., 2020). Both can be thought of being complementary to labor, in the sense of a traditional production function. At the macro-economic level, these considerations add a new dimension. Here, applications exist that are not readily interpretable as either complements or substitutes for labor. For instance, AI tools are increasingly being used to improve the management of waste and electricity networks or help with improving the use and utilization of transport systems, including through inter-modal connectivity (see also the discussion in the next sub-section). None of these activities are directly linked to human labor (unless, for instance, one considers the commute to and from work as part of the aggregate production function, which typically it is not). Most of these applications of AI would, therefore fall into the category of innovations to improve total factor productivity.

Such innovations focused at improving resource efficiency are unlikely to have any direct employment effects but might impact comparative advantages of different sectors as they

impact the way capital and labor is being used. Applications to improve waste management (e.g., in Barcelona), to help municipal officials to identify more rapidly infrastructure shortcomings (e.g., Amsterdam) or to improve the management of traffic systems (e.g., Delhi, Kuala Lumpur) reduces overhead costs. As such, they do not substitute for any current or future jobs (other than the engineers developing the software). However, to the extent that applications help to improve resources efficiency in particular industries or sectors, with effects on the comparative advantages of this industry both domestically and internationally, resources will be reallocated across sectors with implications for jobs and growth (Rentsch and Brinksmeier, 2015). Similarly, to the extent that cities benefit from AI differently, more advanced municipalities are likely to attract new businesses and jobs, leading to a geographical reallocation of resources. For the moment and to our knowledge, however, there is no good empirical understanding of the extent to which AI can help in improving resource efficiency in the aggregate, at the sectoral level or spatially, which precludes a proper quantitative assessment of this particular dimension of improvements in AI.

Such indirect effects of efficiency improvements on labor markets can be complemented by specific interventions that help strengthening labor to be complementary rather than a substitute. In particular, there are three areas where policy makers and social partners alike can help to steer technological change to become complementary to workers rather than substitutes:

- A first and most direct way of intervening to prevent excessive automation is *via* R&D incentives and tax credits: As highlighted in Figure 2, investment in AI is highly concentrated among a few areas, mostly associated with excessive automation (Acemoglu and Restrepo, 2019, 2020). Such interventions are always possible and might bring about a more balanced developed as regards the evolution of AI and its social impact. However, from the discussion of the AI trilemma, it follows that a broad-based support of advances in AI that are complementary to labor might not necessarily solve the energy problem at the same time. Rather, as with previous waves of technological progress, automation can come at the cost of excessive use of energy. In other words, direct interventions for AI development need to focus simultaneously on their resource-efficiency and labor-complementarity aspect in order to be effective when trying to address the AI trilemma.
- A second intervention works through reducing the tax burden on labor that has specifically in the US led to strong incentives for automation (Acemoglu et al., 2020). Instead, a shift of the tax burden away from labor toward energy consumption can address both the adverse resource and labor impact of AI. Indeed, as discussed by Ciminelli et al. (2019) an often overlooked channel of a revenue-neutral tax reform toward consumption taxes is that it strengthen labor supply incentives at the lower end of the income distribution, thereby partly correcting for its regressive income effect.
- Finally, the most indirect and challenging way to steer the degree to which a resource-efficient evolution of AI can produce positive outcomes on jobs and working conditions is by strengthening labor market institutions, such as work's councils that influence technological choices at the firm level (El-Ganainy et al., 2021). Such institutional arrangements have been shown to affect the way in which technologies are being applied and implemented at the workplace level. In the scenario envisaged here, activities would develop in sectors and occupations that would benefit from both AI-triggered resource efficiency improvements and institutional comparative advantages in favor of cooperative labor relations (Ernst, 2005).

A first approach to address the AI trilemma, therefore, lies with the necessity to steer AI developments in the direction of improving total factor productivity as an aspect for which AI is particularly suited and where its potential to substitute for labor is minimized, simply because so far none of these network functions are fulfilled by human labor. Complementary interventions are needed, however, to address possible adverse effects of resource-efficiency enhancing AI applications in labor-intensive occupations and sectors. In the following, we discuss how the particular network complementarities implied by AI might challenge such an approach.

4.2. Solving the productivity-energy dilemma: Incentivize the use of network applications

Not all AI applications are affected to the same extent by the AI-trilemma. Especially the already mentioned network applications have the potential to perform particularly well when it comes to lower resource consumption and improve inclusivity. Well-trained AI routines, for example regarding electricity management or water consumption in agriculture already reduce the burden on the environment today and offer possibilities to address climate change effectively (see, for instance, Rolnick et al., 2023). Digital technologies are likely to play a key role in helping our societies to adapt to rising climate risks by making critical infrastructure more resilient (Argyroudis et al., 2022). Furthermore, such solutions also offer opportunities for cost-effective knowledge transfer to developing countries, where there is still a great need to catch up on modern technologies adapted to local conditions. Companies such as Google and Microsoft have already discovered this

need and have begun to establish their own research centers in some developing countries. And local solutions, especially in agriculture, also show potential productivity gains in these countries (Ernst et al., 2019). In the following, we briefly discuss three areas where the network management of AI tools can prove of particular support: energy management, traffic management and remote work.

Energy management is particularly high on the agenda for AI applications. Managing complex electricity grids across different jurisdictions (particularly acute in Europe) and diverse energy sources as energy production is increasingly ensured by renewables pose formidable challenges to grid management. Failure for proper management and anticipation of external (weather) events can lead to grid outage, as experienced in Texas during the winter of 2020/21, for instance. Combining Internet of Things devices and smart meters into smart grids has been a focus of development in the energy industry (Ahmad et al., 2022). Beyond grid management, preventive maintenance and smart consumption are also major areas of research and development that can help both in reducing risks of outage and overall consumption[9]. Power consumption management, in particular, has become an active area of research for tech companies in their attempt to reduce their own carbon footprint and is likely to contribute to a substantial reduction of the energy-intensity of AI models[10].

Mobility management as part of a smart city policy is another area of high potential for digital tools to address the AI trilemma. Logistics management is an area where modern communication networks and complex supply chain management is already making use of AI-powered tools[11]. Similarly, applications regarding modal interconnectivity for individual transportation receive increasing attention, especially in areas where transport supply elasticity is limited. These applications are meant to facilitate personal traffic in dense urban settings that provide alternative modes of transportation for the same route. Managing such traffic networks through AI-powered tools will allow to improve traffic fluidity and manage limited infrastructure capacity more effectively (Nepelski, 2021).

A final area to be considered here is the role AI can play in our current transition to a higher share of remote work. Advanced economies, in particular, have demonstrated surprising resilience with respect to requirements to work from home that came with the pandemic-induced lockdowns in 2020/21. Dubbed "potential capital," the large share of digital infrastructure and personal computing devices allowed a large part of the workforce to continue their economic activities and limit the economic outfall of the health crisis (Eberly et al., 2021). As economies are recovering from this shock, remote work will remain a reality at least for part of the workforce, creating challenges in terms of scheduling, information sharing and networking (Kahn, 2022). In particular the development and maintenance of personal and professional ties that are important for economic advancement have been shown to be critically affected by remote work (Yang et al., 2022). So far, gains from going remote have been meager. Both business leaders and employees are still trying to figure out how best to make use of the new flexibility that working from home offers (Cappelli, 2021). Here again, AI tools can prove an important answer to solve this challenge at least partially, developing complex scheduling software and helping to maintain information integrity across highly distributed networks of employees.

Taken together such applications make use of the potential of AI tools to directly address questions of aggregate resource efficiency rather than substituting capital for labor, thereby bringing us closer to resolving the AI-trilemma.

4.3. Solving the concentration-productivity dilemma: Redistribute rents

As the previous discussion makes clear, these changes require adjustments not only in the way technology is being developed but also in the institutional and policy settings under which innovators and businesses operate. In concluding this section, three approaches are being discussed that have the potential to address both the technological and the distributional aspects of the AI trilemma:

A first, traditional answer is to try to use taxes to better capture capital gains, while at the same time shifting the tax pressure from labor back toward capital. This has often been discussed in connection with a robot tax (Merola, 2022). On the one hand, it would allow the enormous profits of digital companies to be captured. On the other hand, tax fairness would be restored, which could relieve the factor labor and ease the pressure toward rationalization and job losses. However, in a global economy, governments have tight limits on how much they can tax internationally operating companies. Attempts to extend taxation to the consumption of digital services instead of profits are being resisted by those countries that are home to a myriad of large, digital companies. Moreover, as mentionned before, the tax burden needs to shift away from labor and toward energy consumption if the trilemma is to be properly addressed.

A second, more innovative approach is to ensure greater competition between digital enterprises, for instance by making it easy to transfer data between platforms using uniform

9 https://www.xcubelabs.com/blog/applications-of-ai-in-the-energy-sector/

10 https://ai.googleblog.com/2022/02/good-news-about-carbon-footprint-of.html

11 https://www.technologyreview.com/2021/10/20/1037636/decarbonizing-industries-with-connectivity-and-5g/?mc_cid=98f3a8206d&mc_eid=59ed455432

standards and protocols. Some solutions also propose data ownership in order to provide a monetary incentive for those who make their data available by using the platforms. So far, however, none of these solutions are fully developed and practicable yet. Moreover, only very few users can derive relatively large profits from such approaches, while the vast majority of them would have little to expect. The incentive to switch platforms or to reap monetary rewards would be too low to solve the AI trilemma.

A final, little debated solution is to set up a sovereign digital wealth fund that participates widely in the digital economy. Currently, sovereign wealth funds (SWF) have been set up in relation with tangible public goods such as natural resources. Leaving the exploitation of such resources to private companies, sovereign wealth funds invest in these activities to the benefits of a public shareholder, such as the government. This allows the benefits of such public goods to be passed on to a broad group of people. However, instead of feeding off oil wells (as in the case of Saudi Arabia, Norway) or fish stocks (as in Alaska), a Sovereign Digital Wealth Fund would be financed by taxes and new debt, in order to generate returns by investing in a broad fund of innovative digital companies. At the same time, such a fund, provided it invests deeply enough, would also be able to directly influence the operative business in market-dominant companies in order to prevent the exploitation of such positions. Similarly, the fund could also aim at exerting influence at the micro level to ensure that ethical and ecological standards are met when using AI. Existing SWFs have increasingly invested in technology sectors, without, however, taking an active stance as regards the technological development nor the economic impact of the companies they have invested in Engel et al. (2020).

None of the solutions outlined here will be sufficient in themselves to resolve the AI trilemma. National solutions often do not provide sufficient guarantee that all market participants will actually be offered the same conditions. International approaches, especially in the area of taxation, are slowly gaining acceptance, but often only at the lowest common denominator. Innovative solutions such as data ownership require institutional changes, which will most likely take some time to be established and enforced. However, an approach that addresses all three proposed solutions should make it possible to find initial answers to the AI trilemma while at the same time offering new, individualized proposals that optimize the potential that AI holds for jobs, income and inclusiveness. The future of work demands not only technological innovations, but also political and institutional ones.

5. Conclusion

The article introduces and discusses the AI sustainability trilemma, the impossibility to achieve ecological sustainability, (income) equality and productivity growth under the current technological paradigm. It presents arguments as to why the energy-intensive nature of current computing capabilities combined with strong network externalities leads to market concentration, narrow AI research and weak (aggregate) productivity gains. The paper also discusses possible answers to this trilemma, demonstrating the potential for directed technological change toward network applications, for instance in electricity and mobility management, as a way to improve total factor productivity that will lead to a lower overall ecological footprint and higher aggregate productivity without worsening inequality. Such directed technological change requires, however, both technological and institutional changes to take place in order to reduce the tendancy of the digital economy toward market concentration.

Much of the potential to overcome the AI trilemma remains speculative at this stage, simply because the overall impact of directed technological change has not been tested or implemented at scale. Some of the institutional shifts required are likely to be resisted by strong incumbents that might lose their market dominant positions. At the technological level, individual applications show the potential to address the shortcomings of the current direction of technological change but real-world examples are lacking at the time of writing of this article. As new applications are being developed and implemented at scale, careful empirical research is necessary to assess the extent to which they can truly address the AI trilemma and possible additional policy changes required to fully benefit from the technological evolution around digital tools and artificial intelligence. Policy shifts that encourage less resource use and reduces (tax) penalties on hiring labor can help induce the development of more socially beneficial digital tools. A more active stance, for instance, *via* the establishment of Sovereign Digital Wealth Funds similar to existing models on natural resources management should be used to accelerate the transition toward a new technological paradigm that overcomes the AI trilemma. The switch from a supply-push to a demand-pull technological regime as argued for in this paper requires further analysis regarding the specific applications that can help overcome the trilemma. In particular, beyond the technological feasibility of these changes, the specific political and institutional roadblocks need to be carefully identified and addressed, opening yet another interesting research avenue.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author wishes to gratefully acknowledge very helpful comments from Bart van Ark, Iain Begg, Tim Leunig, and three referees.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acemoglu, D. (2022). "Harms of AI," in *The Oxford Handbook of AI Governance*, eds J. Bullock, B. Zhang, Y.-C. Chen, J. Himmelreich, M. Young, A. Korinek, and V. Hudson (Oxford: Oxford University Press).

Acemoglu, D., Aghion, P., Bursztyn, L., and Hemous, D. (2012). The environment and directed technical change. *Am. Econ. Rev*. 102, 131–166. doi: 10.1257/aer.102.1.131

Acemoglu, D., Manera, A., and Restrepo, P. (2020). "Does the us tax code favor automation?" in *Brookings Papers on Economic Activity* (Washington, DC), 231–285.

Acemoglu, D., and Restrepo, P. (2019). Automation and new tasks: how technology displaces and reinstates labor. *J. Econ. Perspect*. 33, 3–30. doi: 10.1257/jep.33.2.3

Acemoglu, D., and Restrepo, P. (2020). The wrong kind of AI? artificial intelligence and the future of labor demand. *J. Regions Econ. Soc*. 13, 25–35. doi: 10.1093/cjres/rsz022

Aghion, P., Antonin, C., and Bunel, S. (2021). *The Power of Creative Destruction. Economic Upheaval and the Ealth of Nations*. Cambridge, MA: Harvard University Press.

Aghion, P., and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica* 60, 323–351. doi: 10.2307/2951599

Agrawal, A. K., Gans, J. S., and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston, MA: Harvard Business Review Press.

Ahmad, S., and Scheinkman, L. (2016). *How can we be so dense? the benefits of using highly sparse representations*. Technical report, Numenta, Redwood City.

Ahmad, T., Zhu, H., Zhang, D., Tariq, R., Bassam, A., Ullah, F., et al. (2022). Energetics systems and artificial intelligence: applications of industry 4.0. *Energy Rep*. 8, 334–361. doi: 10.1016/j.egyr.2021.11.256

Alais, D., Xu, Y., Wardle, S. G., and Taubert, J. (2021). A shared mechanism for facial expression in human faces and face pareidolia. *Proc. R. Soc. B Biol. Sci*. 288, 1954. doi: 10.1098/rspb.2021.0966

Alderucci, D., Branstetter, L., Hovy, E., Runge, A., and Zolas, N. (2020). "Quantifying the impact of ai on productivity and labor demand: evidence from U.S. census microdata," in *American Economic Association Meeting* (San Diego, CA).

Andrews, D., Criscuolo, C., and Gal, P. N. (2015). *Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries*. Paris: OECD Future of Productivity.

Argyroudis, S. A., Mitoulis, S. A., Chatzi, E., Baker, J. W., Brilakis, I., Gkoumas, K., et al. (2022). Digital technologies can enhance climate resilience of critical infrastructure. *Clim. Risk Manag*. 35, 10387. doi: 10.1016/j.crm.2021.100387

Arnosti, N., and Weinberg, S. M. (2022). Bitcoin: a natural oligopoly. *Manag. Sci*. 68, 4755–5555. doi: 10.1287/mnsc.2021.4095

Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., and Weyl, E. G. (2018). Should we treat data as labor? Moving beyond "free". *AEA Papers Proc*. 108, 38–42. doi: 10.1257/pandp.20181003

Assad, S., Clark, R., Ershov, D., and Xu, L. (2020). *Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market, Working Paper 8521*. Munich: CESifo.

Atkeson, A., and Kehoe, P. J. (2005). Modeling and measuring organization capital. *J. Polit. Econ*. 113, 1026–1053. doi: 10.1086/431289

Autor, D. H. (2013). The "task approach" to labour markets: an overview. *J. Labour Market Res*. 46, 185–199. doi: 10.1007/s12651-013-0128-z

Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The polarization of the us labor market. *Am. Econ. Associat. Papers Proc*. 96, 189–194. doi: 10.1257/000282806777212620

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: an empirical exploration. *Q. J. Econ*. 118, 1279–1333. doi: 10.1162/003355303322552801

Balliester, T., and Elsheikhi, A. (2018). *The Future of Work: A Literature Review, Research Department Discussion Paper 29*. Geneva: ILO.

Bassanini, A., and Ernst, E. (2002). Labour market regulation, industrial relations and technological regimes: a tale of comparative advantage. *Ind. Corporate Change* 11, 391–426. doi: 10.1093/icc/11.3.391

Bender, E. M., Gebru, T., McMillan-Major, M., and Shmitchell, S. (2021). "On the dangers of stochastic parrots:can language models be too big?" In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Beneke, F., and Mackenrodt, M.-O. (2021). Remedies for algorithmic tacit collusion. *J. Antit. Enforcement* 9, 152–176. doi: 10.1093/jaenfo/jnaa040

Bernhardt, L., and Dewenter, R. (2020). Collusion by code or algorithmic collusion? When pricing algorithms take over. *Eur. Compet. J*. 16, 312–342. doi: 10.1080/17441056.2020.1733344

Bessen, J. (2020). Industry concentration and information technology. *J. Law Econ*. 63, 531–555. doi: 10.1086/708936

Bessen, J. (2022). *The New Goliaths: How Corporations Use Software to Dominate Industries, Kill Innovation, and Undermine Regulation*. New Haven, NJ: Yale University Press.

Brown, Z. Y., and McKay, A. (forthcoming). Competition in pricing algorithms. *Am. Econ. J. Macroecon*.

Brynjolfsson, E., and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY: W. W. Norton & Company.

Brynjolfsson, E., Rock, D., and Syverson, C. (2019). "Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics," in *The Economics of Artificial Intelligence. An Agenda*, eds A. Agrawal, J. Gans, and A. Goldfarb (Chicago: University of Chicago Press), 23–60.

Brynjolfsson, E., Rock, D., and Syverson, C. (2021). The productivity j-curve: how intangibles complement general purpose technologies. *Am. Econ. J. Macroecon*. 13, 333–372. doi: 10.1257/mac.20180386

Calvano, E., Calzolari, G., Denicolò, V., and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *Am. Econ. Rev*. 110, 3267–3297. doi: 10.1257/aer.20190623

Cappelli, P. (2021). *The Future of the Office: Work from Home, Remote Work, and the Hard Choices We All Face*. Upper Saddle River, NJ: Wharton School Press.

Carbonero, F., Davies, J., Ernst, E., Fossen, F. M., Samaan, D., and Sorgner, A. (2021). *The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam, Discussion Paper 14944*. Bonn: Institute for the Study of Labor.

Carbonero, F., Ernst, E., and Weber, E. (2018). *Robots worldwide: The impact of automation on employment and trade, Research Department Working Paper 36*. Geneva: ILO.

Chen, P., and Semmler, W. (2018). Short and long effects of productivityon unemployment. *Open Econ. Rev*. 29, 853–878. doi: 10.1007/s11079-018-9486-z

Ciminelli, G., Ernst, E., Merola, R., and Giuliodori, M. (2019). The composition effects of tax-based consolidation on income inequality. *Eur. J. Polit. Econ*. 57, 107–124. doi: 10.1016/j.ejpoleco.2018.08.009

Cobb, M. (2020). *The Idea of the Brain. A History*. London: Profile Books.

Coveri, A., Cozza, C., and Guarascio, D. (2021). *Monopoly Capitalism in the Digital Era, LEM Working Paper Series 33*. Pisa: Scuola Superiore Sant'Anna Institute of Economics.

Damioli, G., Van Roy, V., and Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Econ. Rev*. 11, 1–25. doi: 10.1007/s40821-020-00172-8

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Q. J. Econ*. 135, 561–644. doi: 10.1093/qje/qjz041

Dhondt, S., Kraan, K. O., and Bal, M. (2021). Organisation, technological change and skills use over time: a longitudinal study on linked employee surveys. *New Technol. Work Employ*. doi: 10.1111/ntwe.12227. [Epub ahead of print].

Dorner, F. E. (2021). *Algorithmic collusion: A critical review, Computers and Society 2110.04740*. Zurich: arXiv.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Res. Policy* 11, 147–162. doi: 10.1016/0048-7333(82)90016-6

Eberly, J. C., Haskel, J., and Mizen, P. (2021). *"Potential Capital", Working from Home, and Economic Resilience, Working Paper 29431*. Cambridge, MA: National Bureau of Economic Research.

Eeckhout, J. (2021). *The Profit Paradox: How Thriving Firms Threaten the Future of Work*. Princeton, NJ: Princeton University Press.

El-Ganainy, A., Ernst, E., Merola, R., Rogerson, R., and Schindler, M. (2021). "Labor markets," in *How to Achieve Inclusive Growth, Chapter 3*, eds V. Cerra, B. Eichengreen, A. El-Ganainy, and M. Schindler (Oxford: Oxford University Press).

Engel, J., Barbary, V., Hamirani, H., and Saklatvala, K. (2020). "Sovereign wealth funds and innovation investing in an era of mounting uncertainty," in *Global Innovation Index. World Intellectual Property Organisation* (Geneva).

Ernst, E. (2005). "Financial systems, industrial relations and industry specialization: an econometric analysis of institutional complementarities," in *The Transformation of the European Financial System*, ed H. Schubert (Vienna), 60–95.

Ernst, E. (2022). "Artificial intelligence: productivity growth and the transformation of capitalism," in *Platforms and Artificial Intelligence*, ed A. Bounfour (Cham: Springer), 149–181.

Ernst, E., Merola, R., and Samaan, D. (2019). The economics of artificial intelligence: implications for the future of work. *IZA J. Labor Policy* 9, 1–35. doi: 10.2478/izajolp-2019-0004

Ernst, E., and Mishra, S. (2021). AI efficiency index: identifying regulatory and policy constraints for resilient national AI *ecosystems*. Available online at: https://papers.ssrn.com/sol3/papers._cfm?abstract_id=3800783

Ezrachi, A., and Stucke, M. E. (2020). Sustainable and unchallenged algorithmic tacit collusion. *Northwestern J. Technol. Intell. Property* 17, 217–259.

Ford, M. (2021). *Rule of the Robots. How Artificial Intelligence Will Transform Everything*. New York, NY: Basic Books.

Fossen, F., and Sorgner, A. (2019). Mapping the future of occupations: transformative and destructive effects of new digital technologies on jobs. *Foresight STI Govern*. 13, 10–18. doi: 10.17323/2500-2597.2019.2.10.18

Fossen, F. M., Samaan, D., and Sorgner, A. (2022). How are patented ai, software and robot technologies related to wage changes in the united states? *Front. Artif. Intell*. 5, 869282. doi: 10.3389/frai.2022.869282

Frey, C. B. (2019). *Technology Trap: Capital, Labor, and Power in the Age of Automation*. Princeton, NJ: Princeton University Press.

Frey, C. B., and Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast Soc. Change* 114, 254–280. doi: 10.1016/j.techfore.2016.08.019

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci*. 11, 127–138. doi: 10.1038/nrn2787

Fuller, J. B., Raman, M., Sage-Gavin, E., and Hines, K. (2021). *Hidden Workers Untapped Talen. How Leaders Can Improve Hiring Practices to Uncover Missedtalent Pools, Close Skills Gaps, and improve Diversity*. Cambridge, MA: Harvard Business Review.

García-Martín, E., Rodrigues, C. F., Riley, G., and Grahn, H. (2019). Estimation of energy consumption in machine learning. *J. Parallel Distrib. Comput*. 134, 75–88. doi: 10.1016/j.jpdc.2019.07.007

Gidon, A., Zolnik, T. A., Fidzuinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., et al. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* 367, 83–87. doi: 10.1126/science.aax6239

Gigerenzer, G., Hertwig, R., and Pachur, T. (2011). *Heuristics. The Foundations of Adaptive Behaviour*. Oxford: Oxford University Press.

Gordon, R. J. (2021). "Productivity and growth over the years at bpea," in *Brookings Papers on Economic Activity* (Washington, DC).

Gries, T., and Naudé, W. (2020). *Artificial intelligence, income distributionand economic Growth*. IZA Discussion Paper No. 13606.

Haskel, J., and Westlake, S. (2017). *Capitalism Without Capital: The Rise of the Intangible Economy*. Princeton, NJ: Princeton University Press.

Hawkins, J., and Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Front. Neural Circ*. 10, 23. doi: 10.3389/fncir.2016.00023

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., and Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *J. Mach. Learn. Res*. 21, 1–43.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci*. 3, 2009. doi: 10.3389/neuro.09.031.2009

Hidalgo, C. A., and Hausmann, R. (2009). The building blocks of economic complexity. *Proc. Natl. Acad. Sci. U.S.A*. 106, 10570–10575. doi: 10.1073/pnas.0900943106

Hole, K. J., and Ahmad, S. (2021). A thousand brains: toward biologically constrained ai. *SN Appl. Sci*. 3, 743. doi: 10.1007/s42452-021-04715-0

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw*. 2, 359–366. doi: 10.1016/0893-6080(89)90020-8

Jones, C. I., and Tonetti, C. (2020). Nonrivalry and the economics of data. *Am. Econ. Rev*. 110, 2819–2858. doi: 10.1257/aer.20191330

Jones, N. (2018). How to stop data centres from gobbling up the world's electricity. *Nature* 561, 163–167. doi: 10.1038/d41586-018-06610-y

Jordà, O., Knoll, K., Kuvshinov, D., Schularick, M, and Taylor, A. M. (2019). The rate of return on everything, 1870-2015. *Q. J. Econ*. 134, 1225–1298.

Kahn, M. E. (2022). *Going Remote*. Berkeley, CA: University of California Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. London: Penguin.

Karakilic, E. (2019). Rethinking intellectual property rights in the cognitive and digital age of capitalism: an autonomist marxist reading. *Technol. Forecast Soc. Change* 147, 1–9. doi: 10.1016/j.techfore.2019.06.007

Khan, L. M. (2017). Amazon's antitrust paradox. *Yale Law J*. 126, 710–805.

Klinger, J., Mateos-Garcia, J., and Stathoulopoulos, K. (2022). A narrowing of ai research? *arXiv*. doi: 10.48550/arXiv.2009.10385

Li, H. L., and van Rossum, M. C. (2020). Energy efficient synaptic plasticity. *eLife* 9, e50804. doi: 10.7554/eLife.50804

Luo, L. (2015). *Principles of Neurobiology, 2nd Edn*. Boca Raton, FL: Garland Science.

Luo, L. (2021). Architectures of neuronal circuits. *Science* 373, eabg7285. doi: 10.1126/science.abg7285

Mazzucato, M. (2021). *Mission Economy: A Moonshot Guide to Changing Capitalism*. New York, NY: Harper Business.

McNerney, J., Savoie, C., Caravelli, F., Carvalho, V. M., and Farmer, J. D. (2022). How production networks amplify economic growth. *Proc. Natl. Acad. Sci. U.S.A*. 119, e2106031118. doi: 10.1073/pnas.2106031118

Merola, R. (2022). Inclusive growth in the era of automation and AI: how can taxation help? *Front. Artif. Intell*. 5, 867832. doi: 10.3389/frai.2022.867832

Montt, G., Wiebe, K. S., Harsdorff, M., Simas, M., Bonnet, A., and Wood, R. (2018). Does climate action destroy jobs? An assessment of the employment implications of the 2-degree goal. *Int. Labour Rev*. 157, 519–556. doi: 10.1111/ilr.12118

Muro, M., Whiton, J., and Maxim, R. (2019). *What jobs are affected by AI? Better-paid, better-educated workers face the most exposure, Metropolitan Policy Program*. Washington, DC: Brookings Institution.

Nepelski, D. (2021). *Ai Watch. AI Uptake in Smart Mobility*. Technical report, Joint Research Centre, Sevilla.

Nightingale, P., von Tunzelmann, N., Malerba, F., and Metcalfe, S. (2008). Technological paradigms: Past, present and future. *Instit. Corporate Change* 17, 467–484. doi: 10.1093/icc/dtn012

OECD (2017). *Algorithms and Collusion. Competition Policy in the Digital Age*. Paris: OECD.

OECD (2021). *The Digital Transformation of SMEs*. Paris: OECD Studies on SMEs and Entrepreneurship.

Pinheiro, F. I., Hartmann, D., Boschma, R., and Hidalgo, C. A. (2021). The time and frequency of unrelated diversification. *Res. Policy* 51, 104323. doi: 10.1016/j.respol.2021.104323

Rentsch, R., and Brinksmeier, C. H. E. (2015). Artificial intelligence for an energy and resource efficient manufacturing chain design and operation. *Procedia CIRP* 33, 139–144. doi: 10.1016/j.procir.2015.06.026

Rikap, C. (2021). *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered*. London: Routledge.

Robbins, S., and van Wynsberghe, A. (2022). Our new artificial intelligence infrastructure: becoming locked into an unsustainable future. *Sustainability* 14, 4829. doi: 10.3390/su14084829

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2023). Tackling climate change with machine learning. *ACM Comput. Surveys* 55, 1–96. doi: 10.1145/3485128

Ryan, T. J., and Frankland, P. W. (2022). Forgetting as a form of adaptive engram cell plasticity. *Nat. Rev. Neurosci*. 23, 173–186. doi: 10.1038/s41583-021-00548-3

Shaw, J. (2016). *The Memory Illusion: Remembering, Forgetting, and the Science of False Memory*. Toronto, ON: Doubleday Canada.

Sherry, Y., and Thompson, N. C. (2021). How fast do algorithms improve? *Proc. IEEE* 109, 1768–1777. doi: 10.1109/JPROC.2021.3107219

Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F. (2020). The computational limits of deep learning. *arXiv 2007.05558*. Cambridge, MA: MIT.

Turoman, N., and Styles, S. J. (2017). Glyph guessing for 'oo' and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *R. Soc. Open Sci*. 4, 170882. doi: 10.1098/rsos.170882

UN DESA (2018). *World Economic and Social Survey 2018: Frontier Technologies for Sustainable Development*. New York, NY: United Nations Department of Economic and Social Affairs.

van Wynsberghe, A. (2021). Sustainable ai: AI for sustainability and the sustainability of AI. *AI Ethics* 1, 213–218. doi: 10.1007/s43681-021-00043-6

Varian, H. R. (1985). Price discrimination and social welfare. *Am. Econ. Rev*. 75, 870–875.

Varian, H. R. (2010). Computer mediated transactions. *AEA Papers Proc*. 100, 1–10. doi: 10.1257/aer.100.2.1

Varian, H. R. (2018). "Artificial intelligence, economics, and industrial organization," in *The Economics of Artificial Intelligence: An Agenda, Chapter 16*, eds A. Agrawal, J. Gans, and A. Goldfarb (Chicago: The University of Chicago Press).

Venturini, F. (2022). Intelligent technologies and productivity spillovers: Evidence from the fourth industrial revolution. *J. Econ. Behav. Organ*. 194, 220–243. doi: 10.1016/j.jebo.2021.12.018

Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., et al. (2022). The effects of remote work on collaboration among information workers. *Nat. Hum. Behav*. 6, 43–54. doi: 10.1038/s41562-021-01196-4

Zahedinejad, M., Fulara, H., Khymyn, R., Houshang, A., Dvornik, M., Fukami, S., et al. (2022). Memristive control of mutual spin hall nano-oscillator synchronization for neuromorphic computing. *Nat. Mater*. 21, 81–87. doi: 10.1038/s41563-021-01153-6

Zambra, M., Maritan, A., and Testolin, A. (2020). Emergence of network motifs in deep neural networks. *Entropy* 22, 2024. doi: 10.3390/e22020204

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. London: Profile Books.

Frontiers in
Artificial Intelligence

Explores the disruptive technological revolution of AI

A nexus for research in core and applied AI areas, this journal focuses on the enormous expansion of AI into aspects of modern life such as finance, law, medicine, agriculture, and human learning.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact