frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 04 March 2025
DOI 10.3389/frai.2025.1538840

Augmented intelligence with voice assistance and automated machine learning in Industry 5.0

Alexandros Bousdekis[1]*, Mina Foosherian[2], Mattheos Fikardos[1], Stefan Wellsandt[2], Katerina Lepenioti[1], Enrica Bosani[3], Gregoris Mentzas[1] and Klaus-Dieter Thoben[2]

[1]Information Management Unit (IMU), Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA), Athens, Greece, [2]BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Bremen, Germany, [3]Beko Europe, Varese, Italy

OPEN ACCESS

EDITED BY
Amir Zadeh,
Wright State University, United States

REVIEWED BY
Evren Şadi Şeker,
Istanbul University, Türkiye
Dimitris Apostolou,
University of Piraeus, Greece

*CORRESPONDENCE
Alexandros Bousdekis
✉ albous@mail.ntua.gr

RECEIVED 03 December 2024
ACCEPTED 07 February 2025
PUBLISHED 04 March 2025

CITATION
Bousdekis A, Foosherian M, Fikardos M, Wellsandt S, Lepenioti K, Bosani E, Mentzas G and Thoben K-D (2025) Augmented intelligence with voice assistance and automated machine learning in Industry 5.0.
Front. Artif. Intell. 8:1538840.
doi: 10.3389/frai.2025.1538840

Augmented intelligence puts together human and artificial agents to create a socio-technological system, so that they co-evolve by learning and optimizing decisions through intuitive interfaces, such as conversational, voice-enabled interfaces. However, existing research works on voice assistants relies on knowledge management and simulation methods instead of data-driven algorithms. In addition, practical application and evaluation in real-life scenarios are scarce and limited in scope. In this paper, we propose the integration of voice assistance technology with Automated Machine Learning (AutoML) in order to enable the realization of the augmented intelligence paradigm in the context of Industry 5.0. In this way, the user is able to interact with the assistant through Speech-To-Text (STT) and Text-To-Speech (TTS) technologies, and consequently with the Machine Learning (ML) pipelines that are automatically created with AutoML, through voice in order to receive immediate insights while performing their task. The proposed approach was evaluated in a real manufacturing environment. We followed a structured evaluation methodology, and we analyzed the results, which demonstrates the effectiveness of our proposed approach.

KEYWORDS

digital intelligent assistant, automated machine learning, voice assistance, human-AI collaboration, artificial intelligence, smart manufacturing

1 Introduction

Industry 5.0 relies on placing human well-being at the center of manufacturing systems (Leng et al., 2022) and has recently been attracting the attention of researchers and practitioners in terms of both social and technological aspects (Leng et al., 2022). Human-centric manufacturing is a prerequisite for factories aiming at achieving flexibility, agility, and robustness against disruptions (Nguyen Ngoc et al., 2022; Wang et al., 2022; Bousdekis et al., 2020). From the technological perspective, enabling technologies, such as human-machine interaction, that combine the strengths of humans and machines as well as big data analytics for providing data-driven insights for advanced manufacturing systems, leading to actionable intelligence, are of outmost importance (Xu et al., 2021; Maddikunta et al., 2022).

As far as human-machine interaction is concerned, voice-enabled assistants have the potential to provide intuitive access to information and knowledge, minimize operators' cognitive workload, and support on-the-job training (de Assis Dornelles et al., 2022; Zheng et al., 2024). Voice assistants are intent-oriented support systems that make use of an infrastructure of digital services, i.e., they target the fulfillment of user intents expressed in natural language aiming at reducing the number of interaction steps of the user (Gärtler and Schmidt, 2021). As far as data analytics is concerned, in today's manufacturing environment, data-driven decision-making is enabled by Machine Learning (ML) algorithms, which aim to process large amounts of data in order to provide insights (Lepenioti et al., 2020). However, building an accurate ML model requires data science knowledge, which does not exist in the manufacturing workforce (Bangaru et al., 2019; Chaabi et al., 2022). Automated Machine Learning (AutoML) can overcome this challenge. AutoML aims at making ML accessible for non-ML experts (domain experts), by automating the configuration and execution of ML pipelines and models (Karmaker et al., 2021; Barbudo et al., 2023).

In the context of Industry 5.0, the integration of voice assistance and AutoML technologies can contribute to the achievement of augmented intelligence. Augmented intelligence puts together human and artificial agents to create a socio-technological system, so that they co-evolve by learning and optimizing decisions through intuitive interfaces, such as conversational, voice-enabled interfaces (Wellsandt et al., 2022). However, existing research works on voice assistants rely on knowledge management and simulation methods instead of data-driven algorithms that could take advantage of the large amounts of data existing in manufacturing enterprise systems (Bousdekis et al., 2021; Saka et al., 2023; Zheng et al., 2024; Gärtler and Schmidt, 2021). Even for these existing works, their adoption faces several barriers, thus leading to limited and unrealistic practical applications (Longo and Padovano, 2020). There is limited acceptance by operators (de Assis Dornelles et al., 2022), while such technologies require long setup periods and extensive training (Freire et al., 2022). Therefore, the practical application and evaluation in real-life scenarios are scarce and limited in scope (Longo and Padovano, 2020; Mirbabaie et al., 2021; Zheng et al., 2024), while there is a research gap on how to evaluate such solutions (Bernard and Arnold, 2019; Colabianchi et al., 2024). Only if the advantages of voice control for an efficient and secure production are sufficiently quantified, manufactures and users will consider applying such novel approaches as viable solutions (Norda et al., 2023).

The objective of this paper is to design and develop an integrated solution incorporating voice assistance technology and AutoML in order to enable the realization of the augmented intelligence paradigm in the context of Industry 5.0. AutoML automates the building and deployment of ML pipelines without requiring ML knowledge, while the voice interface exposes to the user the data analytics outcomes in an intuitive way in the context of dialogues. On the other hand, the user is able to interact with the assistant, and consequently with the ML models, through natural language in order to receive immediate insights while performing their task. The proposed approach is evaluated in a real manufacturing environment and follows a structured evaluation methodology to analyze the results.

The rest of the paper is organized as follows. Section 2 outlines the literature review on voice interfaces and AutoML approaches in manufacturing. Section 3 presents the proposed approach for augmented intelligence with voice assistance and AutoML in the frame of Industry 5.0. Section 4 implements the proposed approach in a manufacturing environment and Section 5 presents the evaluation results. Section 6 concludes the paper and presents our plans for future work.

2 Literature review

Voice-enabled assistants have the potential to provide intuitive access to information and knowledge, thus maximizing users' cognitive efficiency (de Assis Dornelles et al., 2022). Despite the emergence of voice assistants in everyday life or in the service sector, and the availability of technical frameworks to create custom human-centric applications, their exploitation in the manufacturing sector is still underexplored (Ludwig et al., 2023; Mukherjee et al., 2024; Norda et al., 2023; Afanasev et al., 2019). Among others, this is due to the fact that user acceptance of voice assistants is lower than GUI-based systems, while there is the need for more effort on the development side to have a robust system (Gärtler and Schmidt, 2021). However, voice assistance technology in manufacturing has the potential to tackle with the high cognitive load in the workspace and the shortage of highly skilled workforce (Linares-Garcia et al., 2022; Ionescu and Schlund, 2021), but at the same time, it faces some distinct challenges (Ghofrani and Reichelt, 2019). For example, manufacturing operations are complex, investments in equipment are expensive, the manufacturing environment can become dangerous, the shopfloor is noisy, and the users are experts (Norda et al., 2023). These challenges, together with negative experience with voice control in the service sector prevent manufacturers from adopting voice assistance technology, also given the fact that there are not detailed quantitative evaluation approaches and results in manufacturing environments (Norda et al., 2023). During the last years, there has been an increasing research interest in assistants for manufacturing. In the literature, they are mentioned as 'virtual assistants', 'Digital Intelligent Assistants' (DIA), 'voice assistants', or 'softbots'.

Jwo et al. (2021) proposed an assistant in order to facilitate the interaction between the user and the dashboard through natural language. Longo and Padovano (2020) developed a web application integrated to an ontology which adopts a flexible tree structure and a keyword labeling mechanism. Afanasev et al. (2019) presented a method and a prototype for the implementation of a voice assistant in manufacturing processes automation proposed a voice assistant as part of a Cyber-Physical System (CPS)in order to support data access automation. Li et al. (2023) presented an assistant for human-robot interaction aiming at managing several types of robots on the shop floor be embedding a pre-trained model to support the prediction of the intents. Li and Yang (2021) proposed an assistant for various manufacturing operations, such as order processing and production execution. Ruiz et al. (2023) proposed a question-answering system for supporting operators in getting access to relevant information.

Rabelo et al. (2018) presented a proof-of-concept for softbots for facilitating human-machine collaboration, sustainability of the manufacturing workforce, operational excellence, inclusiveness, satisfaction and motivation, safety, and continuous learning. Abner et al. (2020) proposed a softbot that uses data analytics and maturity models in order to support the decision-making of managers. Zambiasi et al. (2022) presented the concept of the resilient Operator 5.0, which aims at providing intuitive, human-centered, and cognitive working environments., by combining softbots and augmented reality for predictive maintenance.

Mukherjee et al. (2024) proposed a concept for voice assistants addressing the machine tools sector aiming at increasing the efficiency and safety of workers. To do this, they propose the use of speech-to-text pipelines. Norda et al. (2023) explored the use of voice interfaces in various scenarios in the CNC milling machines domain, either replacing or complementing existing touch control interactions. They found out that voice interfaces have the potential to contribute to time efficiency, especially for complex commands.

Wellsandt et al. (2020) examined a concept for a voice-enabled DIA for predictive maintenance. The authors identified the key functional modules for such an assistant as well as the requirements and constraints for its development. Wellsandt et al. (2022) proposed the adoption of a DIA for maintenance experts in order to enable the collection of feedback about the success of maintenance interventions. Colabianchi et al. (2024) proposed a DIA for manufacturing which utilizes Large Language Models (LLMs) targeted to assembly processes, and they assessed the technical robustness, the effect on operators' cognitive workload, and the user experience of their approach in a laboratory experiment.

The aforementioned research works are either conceptual or are based upon knowledge-based methods for structuring the acquired knowledge; the use of advanced data analytics algorithms and ML models has not been investigated (Abner et al., 2020; Karmaker et al., 2021). There are several challenges on providing ML-generated insights through a voice assistant in the context of dialogues, instead of visualization-based GUIs. Moreover, the configuration of ML pipelines is a laborious and costly activity, and it becomes even more difficult when integrating a voice-enabled interface. To this end, Bousdekis et al. (2021) proposed an augmented analytics framework for implementing quality analytics integrated into a voice interface for exposing the results to the user. However, their approach is still conceptual, and it does not dive into the specificities of ML models and algorithms.

On the other hand, AutoML has been gathering increasing research attention due to its capability of making ML accessible for non-ML experts by automating all the ML stages (Barbudo et al., 2023), and of usually outperforming conventional ML algorithms (Liang and Xue, 2023). The AutoML paradigm aims to automate the ML aspect of real-world applications through an end-to-end process (Karmaker et al., 2021). Karmaker et al. (2021) proposed six levels of automation for the different AutoML systems, each with varying automated tasks and accessibility to domain experts. Furthermore in the literature, there is a plethora of pipelines and methodologies that compromise AutoML. He et al. (2021) provide a pipeline that incorporates four main steps: (i) Data Preparation, (ii) Feature Engineering, (iii) Model Generation and (iv) Model Evaluation, where each one has multiple sub-tasks. The Model Generation step can be split into two sub-steps: the Search Space and the Optimization Methods. The former defines the design of the ML models (traditional or neural networks), while the latter handles hyperparameter optimization and architecture optimization (AO). These two sub-steps are the first to be automated from the aforementioned automation levels and are of great interest due to their increased computational and resource intensity. Researchers experimented with different methodologies to find optimal ML models and counterbalance the computational needs. For the hyperparameter optimization these methods adopt approaches such as Grid Search, Random Search, Bayesian Optimization, Early-stopping and Multi-fidelity Optimization (Yu and Zhu, 2020, Hutter et al., 2019). In addition, Nikitin et al. (2022) proposed FEDOT, which designs composite ML pipelines, through model and data operations that create a directed acyclic graph and automate them through an evolutionary approach and hyperparameter optimization. Despite the merits of automating the ML aspects, AutoML also has limitations regarding real-world applications, specifically with the available budget of time and computation. The hyperparameter optimization can be computationally expensive (He et al., 2021), and the whole AutoML framework adopted can result in a long waiting time to find solutions (Elshawi et al., 2019). The trade-off between time and computational resources creates strategic decision-making needs to balance the respective constraints (Azevedo et al., 2024). Nevertheless, the advantages of AutoML have made available several approaches and applications in the manufacturing realm. It is noteworthy that most of the existing research on AutoML in manufacturing deals with maintenance and quality operations. This is due to the fact that both operations include many non-value-adding activities that contribute significantly to the manufacturing firms' costs. However, although maintenance has largely adopted sensory technology in order to automate decision-making, quality procedures remain, to a large extent, manual in nature.

Schuh et al. (2022) provided a list of requirements and use cases that manufacturing companies should take into account when selecting an AutoML solution. Gerling et al. (2022) proposed an AutoML-based system that generates predictions about production faults and indications about the related root causes. Chaabi et al. (2022) applied AutoML methods in various manufacturing applications, such as quality control and predictive maintenance. Mallouk et al. (2023) presented an AutoML-based approach that aims at providing decision support to manufacturing experts. Krauß et al. (2020) performed a comparative analysis of AutoML frameworks in manufacturing, by also comparing it to manual processes for quality control. Zhai et al. (2023) proposed a domain-specific AutoML approach for anomaly detection and defect diagnosis in the semiconductor industry. Jayasurya et al. (2024) proposed a framework based on AutoML, employing tools such as AutoKeras, Sweetviz, NumPy, pandas, Streamlit, and PyCaret, in the context of predictive maintenance. Conrad et al. (2024) developed a workflow in the production engineering domain with the use of AutoML and compared it with the manual data mining process. Rooney et al. (2024) used AutoML for failure detection in additive manufacturing. Denkena et al. (2020) demonstrated the improvements achieved by AutoML to predict shape errors during milling for cold rolling procedures. Sousa et al. (2022) used the CRISP-DM methodology and AutoML to address production time prediction for metal containers production. They compared four open-source modern AutoML technologies: AutoGluon, H2O AutoML, rminer, and TPOT. Fikardos et al. (2022) proposed a framework architecture that utilizes AutoML in predictive quality.

AutoML has been proved to achieve a competitive performance; however, apart from technical challenges, related to, e.g., features selection and class imbalance (Zhai et al., 2024; Salehin et al., 2024), recent studies have shown that the lack of transparency and explainability of AutoML make the users reluctant to trust them (Crisan and Fiore-Gartland, 2021; Drozdal et al., 2020; Wang et al., 2021). Existing transparency and explainability approaches incorporate interactive visualization techniques (Zöller et al., 2023; Garouani et al., 2022; Amirian et al., 2021) and post-hoc explanation methods (Zhai et al., 2024). Although such approaches have been proved promising for GUI-based systems, they are not suitable when integrating AutoML to voice interfaces.

In the context of Industry 5.0, the integration of voice assistance and AutoML technologies can contribute to the achievement of augmented intelligence. However, such an approach has not been investigated, although it has the potential to enable and augment operators with data-driven insights in an intuitive, human-like, and 'hands-free' way, without requiring ML expertise, thus increasing the efficiency of their work.

3 The proposed solution for augmented intelligence with voice assistance and automated machine learning

The high-level architecture for augmented intelligence with voice assistance technology and AutoML is depicted in Figure 1. It embeds three main modules: Voice Assistant, Analytics Service, and Use Case Infrastructure. The Voice Assistant module corresponds to the DIA core and manages user interactions. This module first captures the user's message via the Android app and passes the message to the Conversational Agent Team (CAT), which is the technical integration of a Mycroft skill and a Rasa chatbot to minimize the disadvantages of the individual agents. We use Mycroft as the leading agent because users begin all their dialogues through it. The second agent is a Rasa chatbot—both exchange text messages, but only Mycroft responds to users directly. Both Mycroft and Rasa are open-source chatbot frameworks. However, Rasa facilitates sophisticated support in developing Natural Language Understanding (NLU) and Dialog Management (DM). Rasa uses a pipeline for NLU that can integrate various open NLU components, such as Spacy or Duckling. It applies a hybrid approach for DM, combining rules (reliable) with probabilistic models (flexible). Developers can train the latter based on real conversations and thus continuously improve the assistant. In our solution, Mycroft passes user requests to Rasa, where the request is mapped to one of the defined intents of the assistant, and a query is formulated. Rasa's responsibilities are to interpret the user input and generate the requests as queries for the Analytics Service, which responds with the insights needed. The Analytics Service communicates directly with the use case infrastructure, retrieving all the necessary data and information required to make the analyses and produce the outcomes. It also processes the received query, and the results of the requested analytics are encapsulated into a response sent back to the DIA. These modules are further detailed in the following sub-sections.

3.1 Voice assistant

The interface with which the user interacts, as well as the link between the analytics component and the user, is the DIA core, which covers the Android App, the message bus manager, Mycroft, the "Talk to Rasa" skill (in Mycroft), the Rasa chatbot, and the data exchange. The information flow among the various components of the DIA core is depicted in Figure 2.

Users interact with Mycroft through the Android app, where Google Speech-To-Text (STT) transcribes audio to text. The Mycroft core is accessible via a reverse proxy supporting secure connections managed by security mechanisms. Integration with Keycloak ensures that only authorized users can access Mycroft, with the app exchanging valid tokens with Keycloak for authentication. HiveMind, running as a separate Docker container, adds a security and management layer, enabling multi-user interactions and extending Mycroft core to various devices, including those not running Mycroft natively.

Mycroft's Natural Language Understanding (NLU) identifies user intents and entities from transcribed utterances and tries to match the intent with a suitable skill. Here we have a customized "Talk to Rasa" skill that facilitates message exchange between Mycroft and Rasa. This way, complex dialogs are managed by Rasa, with a configurable NLU pipeline, a dialog manager based on rules and probabilistic models, and a fulfillment server performing custom actions via Python code. Rasa fulfills user intents by executing code and accessing external data sources. Rasa action server uses the data exchange services responses (i.e., semi-structured data from analytics service) to build a human-readable response message, which is sent to the user via Mycroft. The

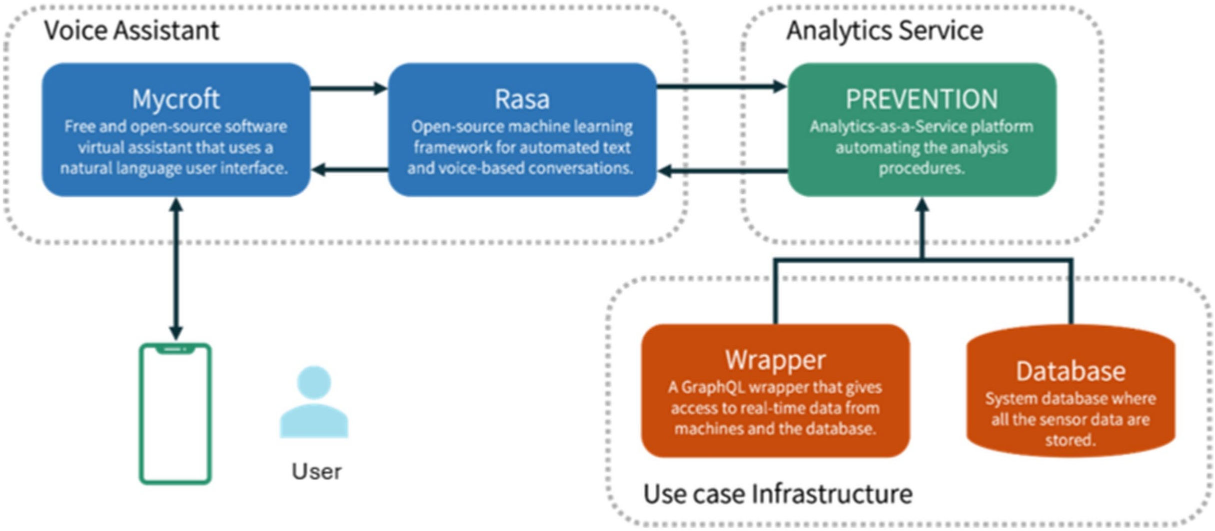

Voice Assistant
Mycroft
Free and open-source software virtual assistant that uses a natural language user interface.
Rasa
Open-source machine learning framework for automated text and voice-based conversations.
Analytics Service
PREVENTION
Analytics-as-a-Service platform automating the analysis procedures.
User
Wrapper
A GraphQL wrapper that gives access to real-time data from machines and the database.
Database
System database where all the sensor data are stored.
Use case Infrastructure

FIGURE 1
High-level architecture.

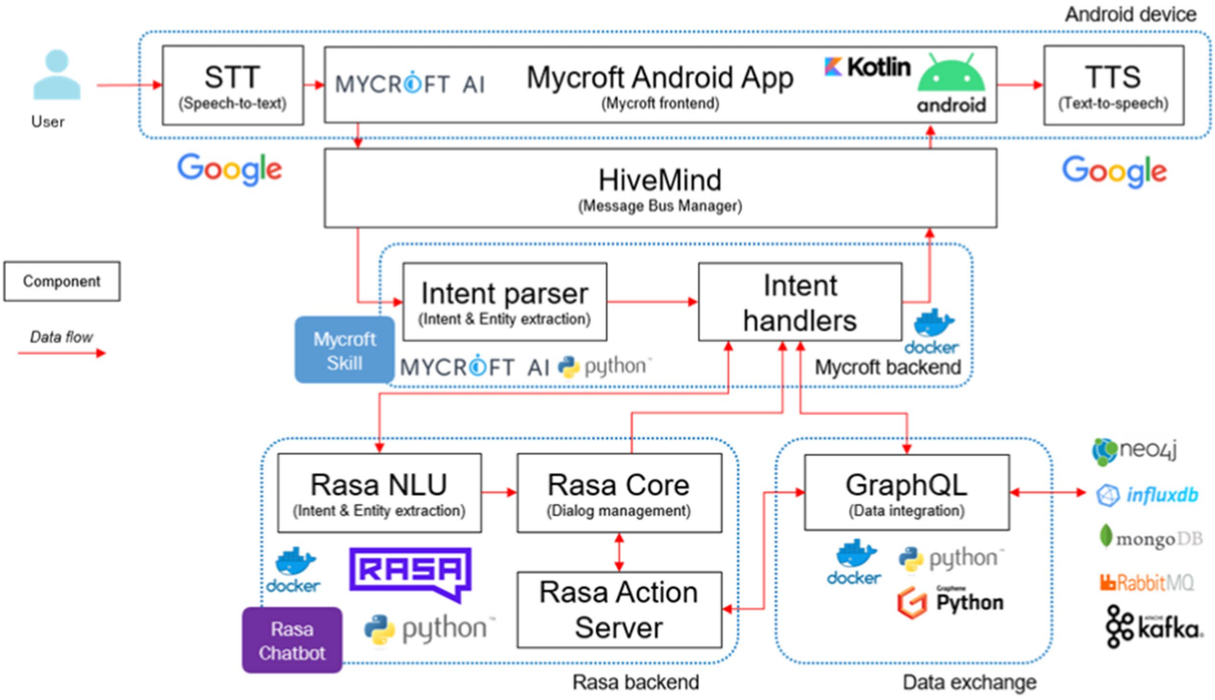


FIGURE 2
DIA core architecture.

Google Text-to-Speech (TTS) service generates audio from text in the Android app.

3.2 Analytics service

The Analytics Service takes advantage of the AutoML process in order to minimize human intervention while constructing and configuring ML within specific computational limits. In this way, it tackles effectively with the challenge of developing appropriate ML models for the problem at hand since the ML model suitability depends on the available dataset, its preprocessing, as well as the configuration of algorithms' parameters. Further, these computing skills and ML knowledge do not usually exist in the manufacturing workforce. The architecture of the Analytics Service is distinguished to the Design Phase and the Runtime Phase, as depicted in Figure 3.

In the Design phase, the data analyst selects and configures the appropriate AutoML frameworks. In the Configuration component of the Design phase, the data analyst applies Data Processing Algorithms, retrieved from the Algorithms Library, on the dataset in order to pursue data cleaning and feature engineering. Then, he/ she defines and configures the AutoML framework to be used with regards to model parameters, evaluation metrics, and termination conditions. Any further configurations, such as model acceptance conditions and output formats, can be defined in the Model Specifications.

In the Runtime phase, the AutoML process is executed as soon as the data analyst defines new configurations or as soon as new data becomes available for the already configured AutoML models. In the first case, the algorithm evaluates several models and optimizes the candidate ones. The finally selected model feeds into the Model Management component in order either to be stored in the Model Warehouse or to be discarded according to the acceptance conditions that have been configured in the Design phase. In the second case, the models are automatically retrained or optimized by incorporating the new data that has become available. They are retrieved from the Model Warehouse in order to feed into the AutoML process. The new model feeds into the Model Management process in order to compare its performance with the one of the previous model. The best-performing model is stored, and the other one is discarded.

3.3 Use case infrastructure

The Voice Assistant and Analytics Service technologies need to be integrated to the use case infrastructure in order to store and retrieve data that are required for the analyses. This data can be accessed through APIs or directly from the database.

4 Deployment in quality control operations

In this Section, we describe the use case under examination, which is the quality control in the home appliances industry (Section 4.1), and we present some indicative demonstration scenarios of the proposed solution (Section 4.2).

4.1 Home appliances use case

The use case under examination concerns the quality control procedures of Whirlpool, one of the leading companies in the home

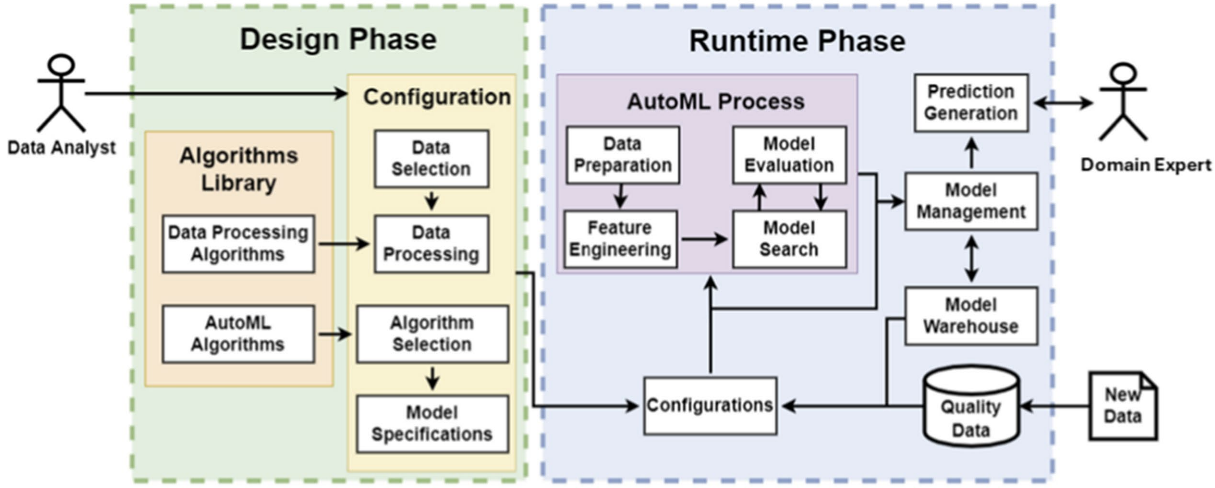


FIGURE 3
Analytics service architecture for AutoML.

appliances industry, with around 92,000 employees and over 70 manufacturing and technology research centers worldwide. The use case addresses the end-of-line quality, which involves quality testing in order to ensure a high standard level of product quality to final customers. To these testing actions, the Whirlpool Production system also adds some statistical quality check actions that are applied both on internal production parts on quality critical processes (statistical process control stations) and on finished goods after the packaging process. In particular, this last testing, called Zero Hour Testing (ZHT) (Figure 4) refers to the Statistical Quality Control applied in a dedicated laboratory out of production flow on some finished products retrieved from the quantities ready to be delivered to the markets. The main objectives of ZHT are to measure the quality level of the outgoing product from an aesthetic, functional, and normative point and to measure the effectiveness of process control. These tests are executed in a dedicated laboratory environment, created in each production site, and following a specific operating procedure. This testing method is designed to replicate the customer approach to the product, simulating the normal product usage conditions at the final customer's first usage.

4.2 Demonstration scenarios

In this sub-section, we demonstrate indicative scenarios in the context of the aforementioned use case. The interaction can be done via voice or text input, and the dialogues are displayed on a tablet screen in order to, among others, investigate additional product-related information and provide visualization capabilities, e.g., in case the user needs additional explanations on the generated outcomes of the algorithms. Figures 5, 6 demonstrate indicative dialogues between the user and the DIA during the end-of-line testing procedures, including the quality testing process and various analytics questions. The end-of-line quality testing process is designed to support operators with three expertise levels: novice, intermediate, and expert. Novice users receive more detailed instructions to perform the test, whereas experts receive high-level instructions. At each step, users can report a defect, and the DIA makes suggestions to assist with the defect recording process (Figure 5A).

In Tables 1, 2, we also present two example queries that facilitate the interaction of AutoML with the voice interface. In Table 1, to predict the number of orders that will manifest some defect in the following day, the user triggers the implemented analysis with the goal "PREDICT_NUMBER_OF_DEFECTS." This analysis has been implemented using AUTOKERAS (created and trained) AutoML library. The prediction of the number of orders with defects in the following day can be used by the workers to get a glimpse of the work ahead. It can be used as an alarm mechanism for a failure in the production line, causing a greater number of defects than expected. An example conversation with the DIA using this query is presented in Figure 6B. In Table 2 the user wants to get information about the products affected by the defect with defect ID 3035, and the digital intelligent assistant is expected to return the list of the SKUs of the affected products. This scenario has been implemented by the analysis with analytics goal "GROUP_DEFECTS_BY_DEFECT_ID_AND_SKU," using the historical data of the recorded defect occurrences. In this case, the resultRequest query with the appropriate request and filters will return the list of affected products, where each record is described by three fields: the DefectID, the FGNum (or SKU), and the count. All the records have the requested Defect ID, a unique FGNum, and the number of recorded defects for the product (count). Based on the answer, the user can understand that multiple products are affected by the given Defect ID with varying recorded occurrences. An example conversation with the DIA using this query is presented in Figure 5C.

5 Evaluation in real-life manufacturing scenarios

In this Section, we describe the evaluation procedure and results in real-life manufacturing scenarios, in the context of the aforementioned home appliances industry. More specifically, we present our evaluation methodology for voice-enabled AI solutions

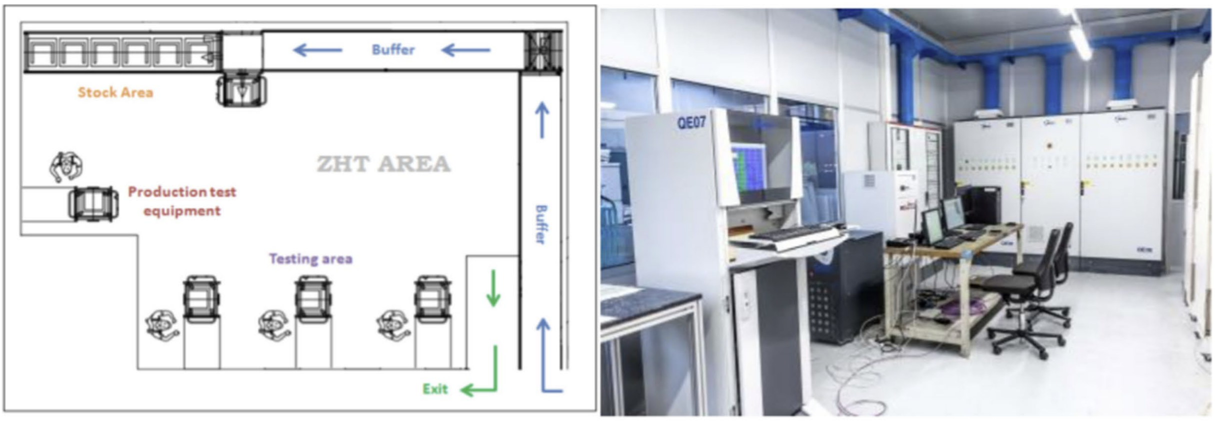
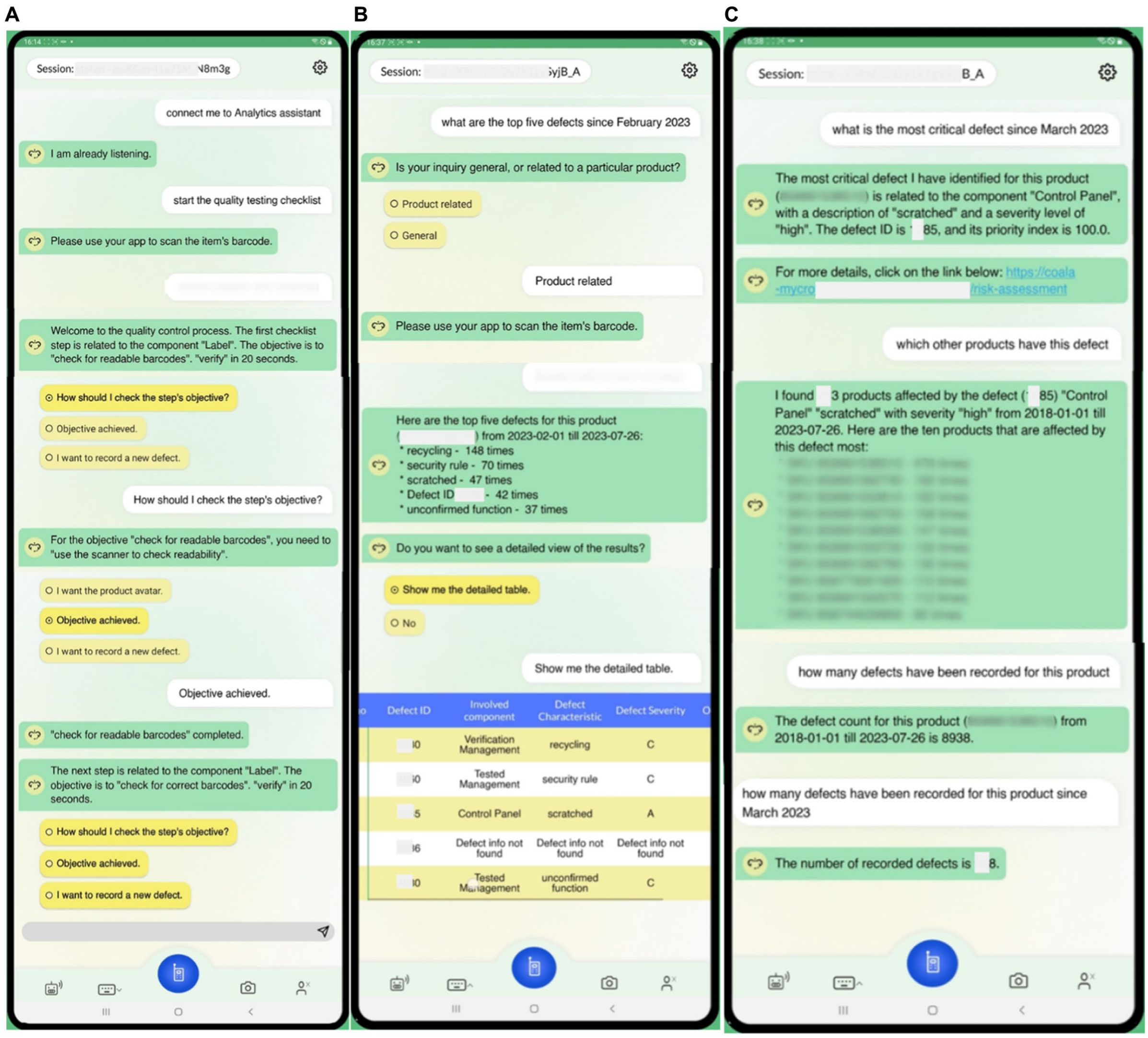


FIGURE 4
Zero hour testing (ZHT) laboratory at whirlpool.



Session: [redacted]N8m3g

connect me to Analytics assistant

I am already listening.

start the quality testing checklist

Please use your app to scan the item's barcode.

[redacted]

Welcome to the quality control process. The first checklist step is related to the component "Label". The objective is to "check for readable barcodes". "verify" in 20 seconds.

● How should I check the step's objective?

○ Objective achieved.

○ I want to record a new defect.

How should I check the step's objective?

For the objective "check for readable barcodes", you need to "use the scanner to check readability".

○ I want the product avatar.

● Objective achieved.

○ I want to record a new defect.

Objective achieved.

"check for readable barcodes" completed.

The next step is related to the component "Label". The objective is to "check for correct barcodes". "verify" in 20 seconds.

○ How should I check the step's objective?

○ Objective achieved.

○ I want to record a new defect.

Session: [redacted]yjB_A

what are the top five defects since February 2023

Is your inquiry general, or related to a particular product?

○ Product related

○ General

Product related

Please use your app to scan the item's barcode.

[redacted]

Here are the top five defects for this product ([redacted]) from 2023-02-01 till 2023-07-26:
- recycling - 148 times
- security rule - 70 times
- scratched - 47 times
- Defect ID [redacted] - 42 times
- unconfirmed function - 37 times

Do you want to see a detailed view of the results?

● Show me the detailed table.

○ No

Show me the detailed table.

Defect ID	Involved component	Defect Characteristic	Defect Severity
[redacted]0	Verification Management	recycling	C
[redacted]0	Tested Management	security rule	C
[redacted]5	Control Panel	scratched	A
[redacted]6	Defect info not found	Defect info not found	Defect info not found
[redacted]0	Tested Management	unconfirmed function	C

Session: [redacted]B_A

what is the most critical defect since March 2023

The most critical defect I have identified for this product ([redacted]) is related to the component "Control Panel", with a description of "scratched" and a severity level of "high". The defect ID is [redacted]85, and its priority index is 100.0.

For more details, click on the link below: https://coala-mycro[redacted]/risk-assessment

which other products have this defect

I found [redacted]3 products affected by the defect ([redacted]85) "Control Panel" "scratched" with severity "high" from 2018-01-01 till 2023-07-26. Here are the ten products that are affected by this defect most:

[redacted]

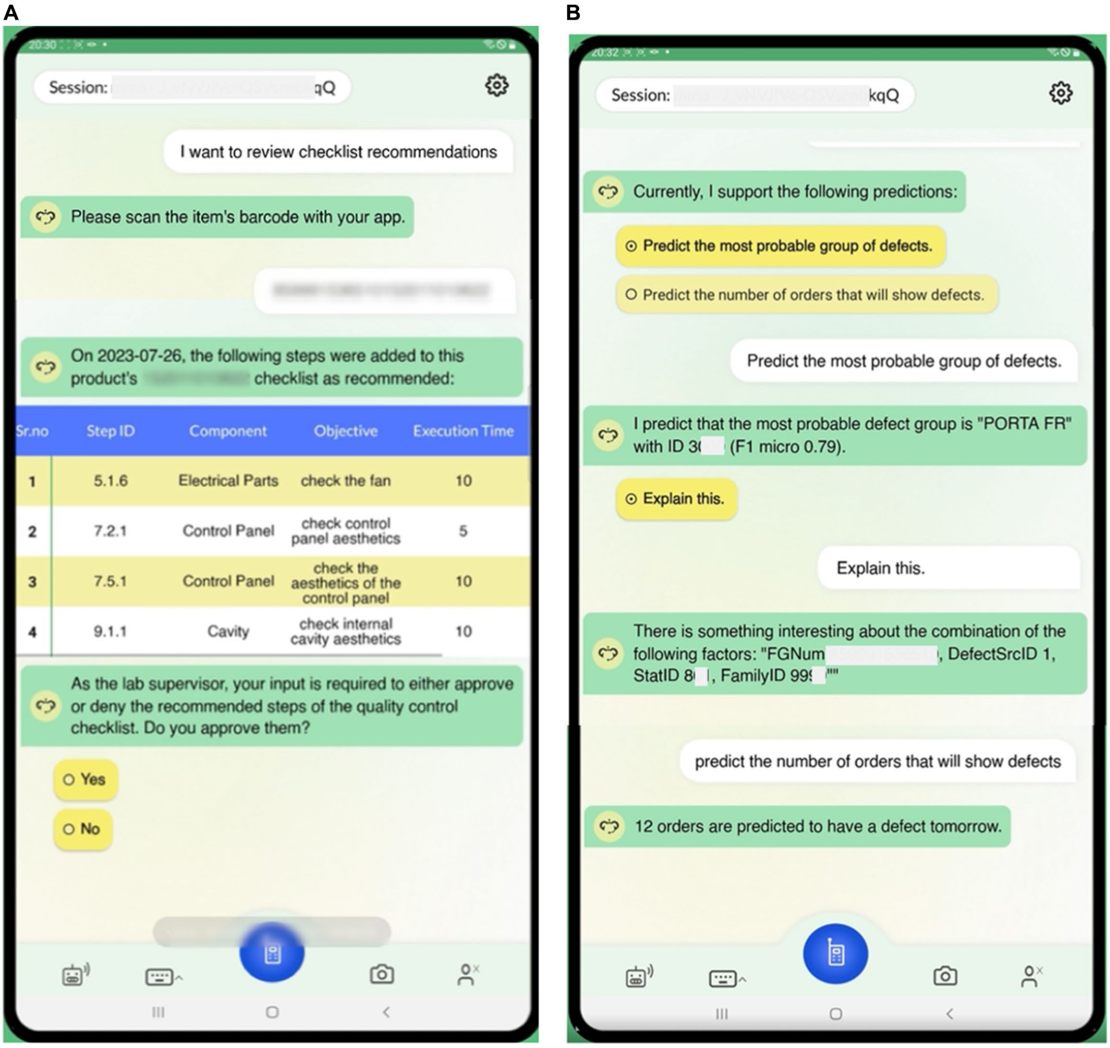
how many defects have been recorded for this product

The defect count for this product ([redacted]) from 2018-01-01 till 2023-07-26 is 8938.

how many defects have been recorded for this product since March 2023

The number of recorded defects is [redacted]8.

FIGURE 5
Demonstration of dialogues for end-of-line testing procedures (1). **(A)** End-of-line testing process execution in the left image, **(B)** Requesting top 5 defects with SKU and date filter in the middle image, **(C)** Requesting most critical defect with SKU and date filter, requesting products affected by a specific defect, requesting number of defects with SKU filter in the right image.



Session: [redacted]qQ

I want to review checklist recommendations

Please scan the item's barcode with your app.

[redacted]

On 2023-07-26, the following steps were added to this product's [redacted] checklist as recommended:

Sr.no	Step ID	Component	Objective	Execution Time
1	5.1.6	Electrical Parts	check the fan	10
2	7.2.1	Control Panel	check control panel aesthetics	5
3	7.5.1	Control Panel	check the aesthetics of the control panel	10
4	9.1.1	Cavity	check internal cavity aesthetics	10

As the lab supervisor, your input is required to either approve or deny the recommended steps of the quality control checklist. Do you approve them?

○ Yes

○ No

Session: [redacted]kqQ

Currently, I support the following predictions:

⊙ Predict the most probable group of defects.

○ Predict the number of orders that will show defects.

Predict the most probable group of defects.

I predict that the most probable defect group is "PORTA FR" with ID 3[redacted] (F1 micro 0.79).

⊙ Explain this.

Explain this.

There is something interesting about the combination of the following factors: "FGNum[redacted], DefectSrcID 1, StatID 8[redacted], FamilyID 99[redacted]""

predict the number of orders that will show defects

12 orders are predicted to have a defect tomorrow.

FIGURE 6
Demonstration of dialogues for end-of-line testing procedures (2). **(A)** End-of-line testing process update recommendations based on critical defect occurrences in the left image, **(B)** Requesting predictions of the most probable defect group and number of orders with defect in the next day in the right image.

(Section 5.1), the evaluation setup (Section 5.2), the evaluation results (Section 5.3), and a discussion on the generalizability criteria that need to be considered (Section 5.4).

5.1 Evaluation methodology

Our evaluation methodology was based on the evaluation methodology for voice-enabled AI solutions, proposed by Bousdekis et al. (2022). In our adaptation, the methodology is structured across five dimensions: AI trustworthiness, system usability, cognitive workload, technical robustness, and lessons learned. Table 3 shows the methods and tools that were used to address each dimension. Below, we briefly present the methods that address these dimensions.

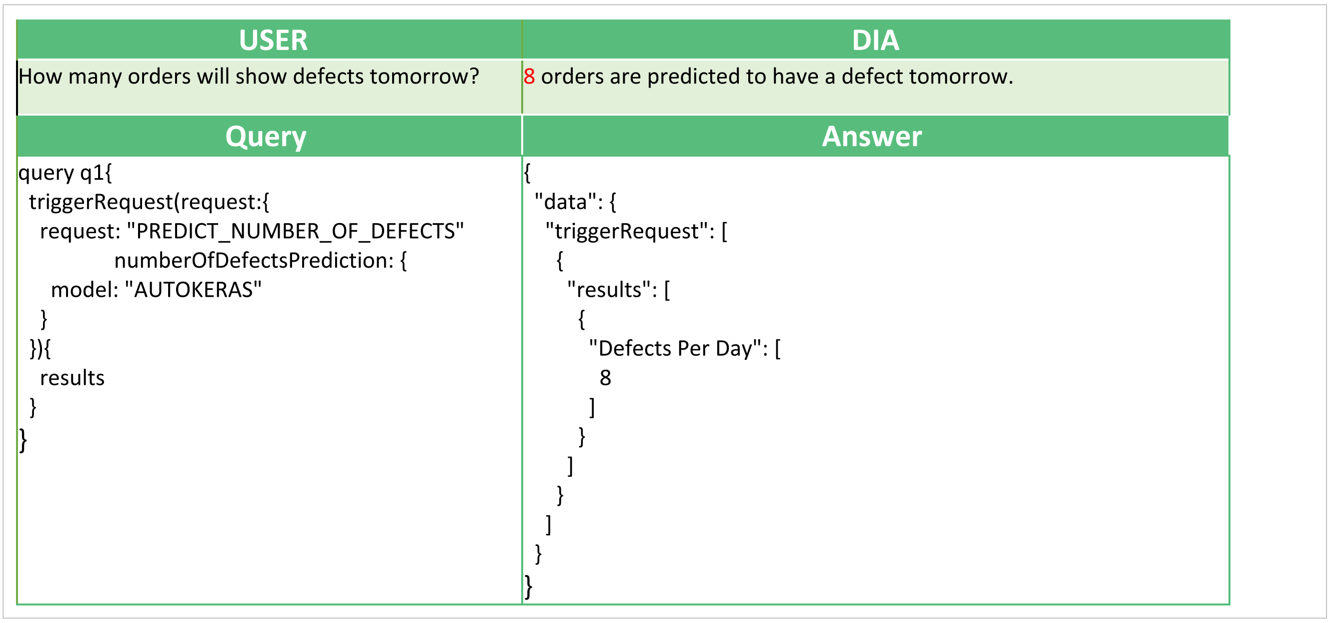
5.1.1 AI trustworthiness

The concept of Trustworthy AI (TAI) dictates that humans, organizations, and societies will achieve the full potential of AI if trust can be established in its development, deployment, and use (Mentzas et al., 2024; Bousdekis et al., 2022). In this realm, the High-Level Expert Group on Artificial Intelligence (AI-HLEG) has created the Assessment List for Trustworthy Artificial Intelligence (ALTAI) tool that helps organizations to self-assess the trustworthiness of their AI systems through a questionnaire (Artificial Intelligence, 2019). The ALTAI is structured according to seven high-level requirements for TAI: Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, non-discrimination and fairness; Societal and environmental well-being; Accountability.

5.1.2 System usability

The usability of voice-enabled AI solutions is evaluated through System Usability Scale (SUS) (Brooke, 1996) and Voice Usability Scale (VUS) (Murad et al., 2019). Both tools are needed due to the distinct characteristics of voice interfaces, which include: understanding of pauses during a conversation (Zwakman et al.,

TABLE 1 Example next day number of orders with defect prediction.



USER	DIA
How many orders will show defects tomorrow?	8 orders are predicted to have a defect tomorrow.

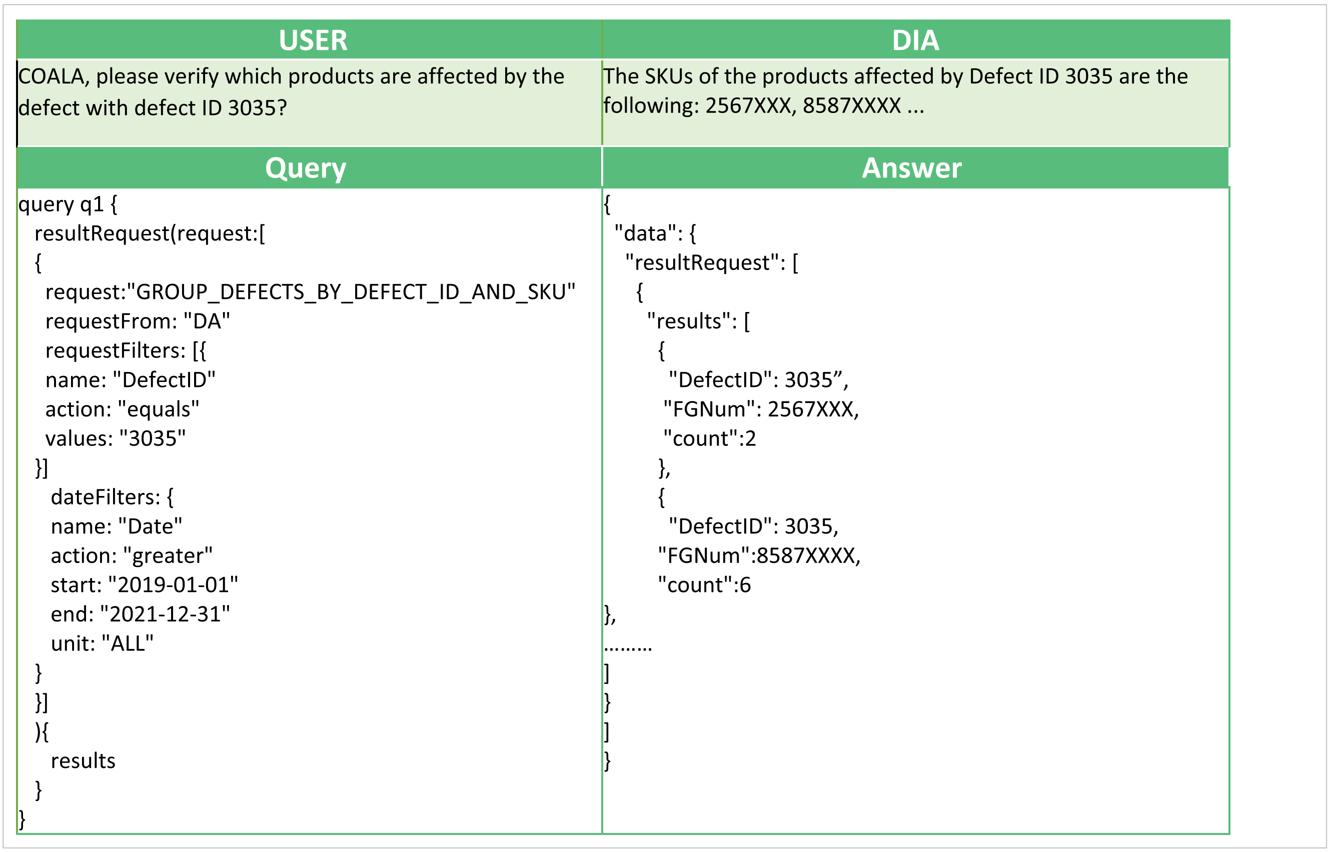
Query

```
query q1{
 triggerRequest(request:{
  request: "PREDICT_NUMBER_OF_DEFECTS"
         numberOfDefectsPrediction: {
   model: "AUTOKERAS"
  }
 }){
  results
 }
}
```

Answer

```
{
 "data": {
  "triggerRequest": [
   {
    "results": [
     {
      "Defects Per Day": [
       8
      ]
     }
    ]
   }
  ]
 }
}
```

TABLE 2 Example requesting the products affected by a specific defect.



USER	DIA
COALA, please verify which products are affected by the defect with defect ID 3035?	The SKUs of the products affected by Defect ID 3035 are the following: 2567XXX, 8587XXXX ...

Query

```
query q1 {
  resultRequest(request:[
  {
   request:"GROUP_DEFECTS_BY_DEFECT_ID_AND_SKU"
   requestFrom: "DA"
   requestFilters: [{
   name: "DefectID"
   action: "equals"
   values: "3035"
  }]
    dateFilters: {
    name: "Date"
    action: "greater"
    start: "2019-01-01"
    end: "2021-12-31"
    unit: "ALL"
  }
  }]
  ){
    results
  }
}
```

Answer

```
{
 "data": {
  "resultRequest": [
   {
    "results": [
     {
      "DefectID": 3035”,
     "FGNum": 2567XXX,
     "count":2
     },
     {
      "DefectID": 3035,
     "FGNum":8587XXXX,
     "count":6
},
.........
]
}
]
}
```

2021), limitations to back-and-forth navigation (Holmes et al., 2019), not having a visualization of results (Cowan et al., 2017), user expectations about the structure of dialogues (Murad et al., 2019), absence of familiarity with synthetic voice (Babel et al., 2014). Both tools include 10 items having declarative statements of opinion to which the participants will respond with their rate of agreement on a Likert scale.

5.1.3 Cognitive workload

The extent of human cognitive resources utilization is called cognitive load (Ninomiya et al., 2024). The emergence of AI technologies in the manufacturing domain dictates the capability of adaptation to the new processes in terms of, among others, the efficient management of workload (Matt et al., 2015). The cognitive workload can be measured through subjective measures in order to overcome the challenges in assessing operators' performance in complex and automated environments (Rubio et al., 2004). This dimension is addressed by the NASA-TLX (Hart, 1988), a widely used subjective method, which includes six dimensions: Mental Demand; Physical Demand; Temporal Demand; Overall Performance; Effort; Frustration Level.

5.1.4 Technical robustness

Technical robustness is evaluated for the two main modules of the proposed solution, i.e., Analytics Service and Voice Assistant. Regarding the Analytics Service, its AutoML models are validated by using appropriate performance metrics. It should be noted that the selection of the evaluation metrics according to the algorithm used can be automated by the AutoML process. Regarding the Voice Assistant, its performance is measured based on intent recognition accuracy for the recorded conversations.

5.1.5 Lessons learned

Based on workshops among the involved users and bilateral interviews, including also the management, in this dimension, we collect qualitative feedback at the end of the evaluation procedure in order to gather the business perspective and to conclude with lessons learned for the technological solution and its adoption by the manufacturing firm.

TABLE 3 Method/Tool per each evaluation dimension.

Dimension	Method/Tool
AI Trustworthiness	ALTAI
System usability	SUS, VUS
Cognitive workload	NASA-TLX
Technical robustness	ML performance metrics, intent recognition accuracy
Lessons learned	Workshops, interviews

5.2 Evaluation setup

The evaluation procedure started on December 2022 and was being performed in parallel with the integration activities in order to continuously provide early feedback on the technical improvements. We separated the evaluation procedure in 4 rounds, engaging various roles of testers in the Whirlpool use case, as it is shown in Figure 7. During the 3rd round, we captured the last points to improve the solution, such as errors in response translations, data accuracy and quality in the quality control checklist database, etc., and we also provided training to the operators. It should be noted that the number of participants and their roles were subject to restrictions derived from the factory operations. The final evaluation was performed with 2 lab supervisors, 3 expert operators and 1 intermediate operator.

During the third and fourth evaluation rounds, the procedure consisted of several key steps. Initially, participants were required to sign a printed consent form to acknowledge their participation. Subsequently, in the case of the third evaluation round, participants underwent training, which involved the use of training materials and a personalized training session using the app. The participants were asked to go through an online questionnaire that involved the following sections: user role, a set of tasks to perform based on the user's role, NASA-TLX questionnaire, SUS questionnaire, VUS questionnaire, personal factors questionnaire including items such as orientation towards using new technologies, gender, age, education, occupation, and work experience, two additional open questions regarding trust in the technological solution. Figure 8 depicts the average scores on the Likert scale (1–5) of personal factor questions per user role.

5.3 Evaluation results

In this sub-section, we present the evaluation results per dimension of the evaluation methodology, i.e., AI trustworthiness (Section 5.3.1), system usability (Section 5.3.2), cognitive workload

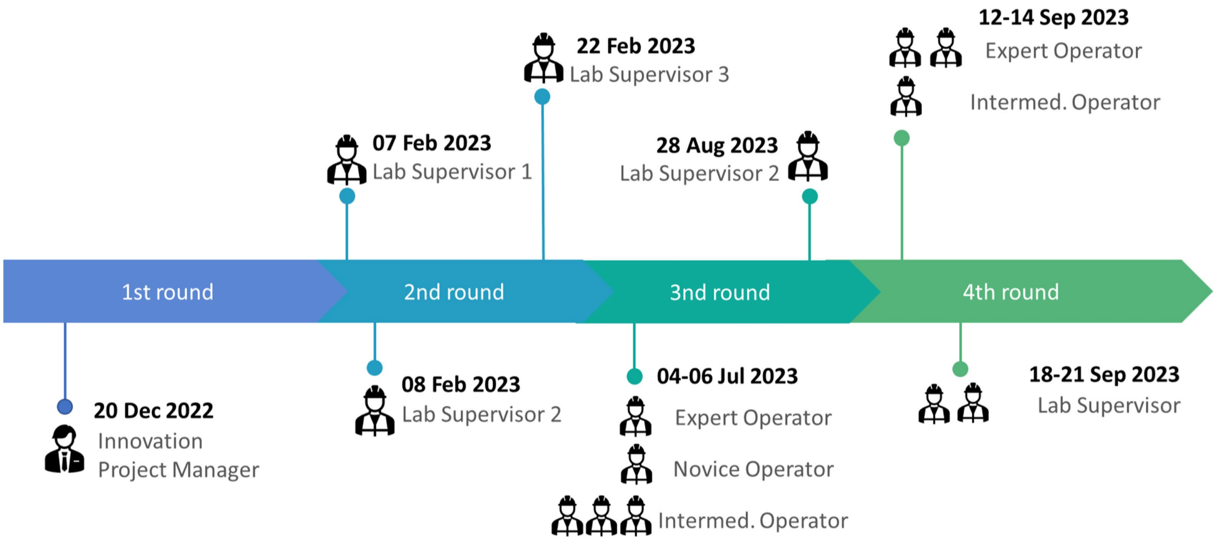


FIGURE 7
The timeline of the evaluation procedure in Whirlpool.

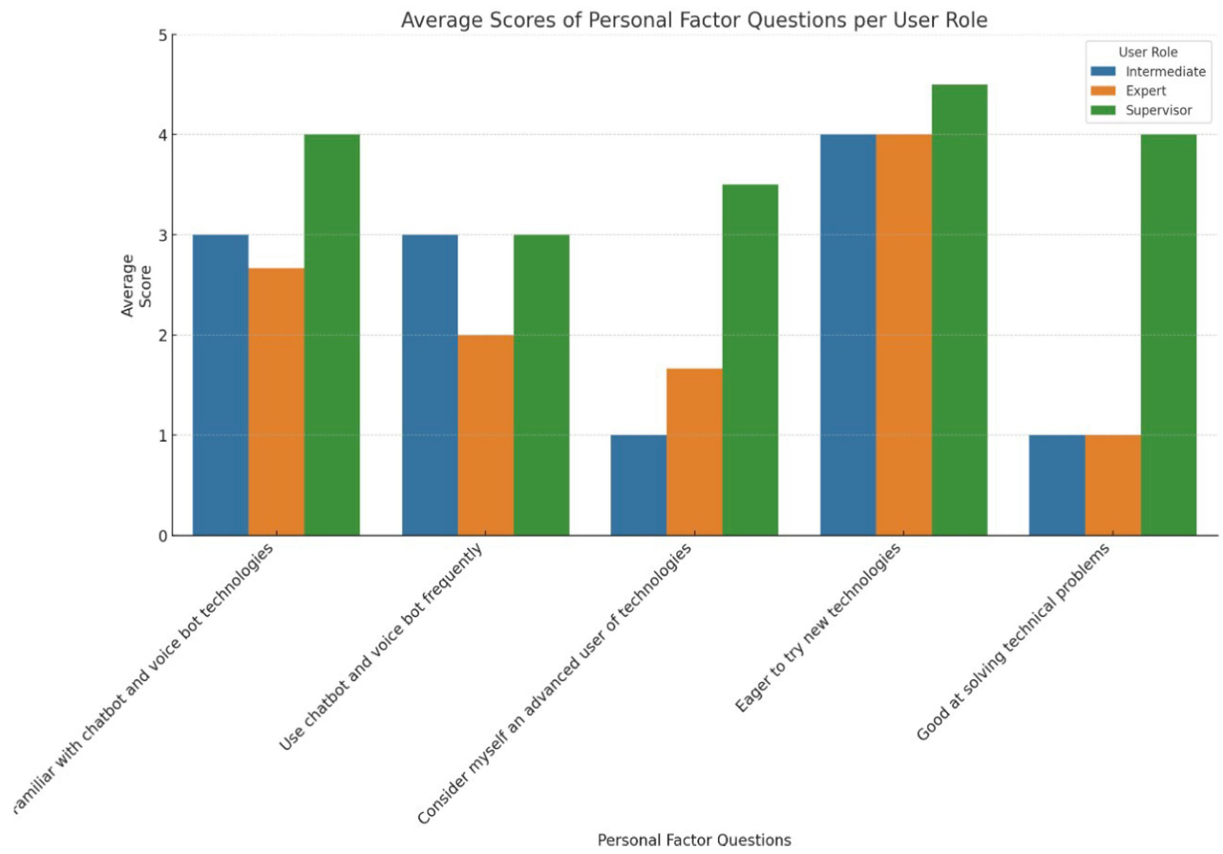


FIGURE 8
Personal factors questions per user role in Whirlpool.

(Section 5.3.3), technical robustness (Section 5.3.4), and lessons learned (Section 5.3.5).

5.3.1 AI trustworthiness

AI trustworthiness was assessed by a multidisciplinary team of people with both business and technical background by using the web-based tool developed by EC.[1] After completing the web-based questionnaire, based on the responses, the tool extracts a visualization of the self-assessed level of adherence of the AI system with the TAI requirements, and recommendations. Figure 9 depicts the results in the form of a Polar diagram for the seven requirements for Trustworthy AI, while Supplementary Table S1 presents the resulting recommendations per requirement.

Overall, the solution excels in "Privacy and Data Governance," while it performs very well with regards to "Accountability," "Technical Robustness and Safety," "Transparency," and "Diversity, Non-discrimination and Fairness." The dimensions "Human Agency and Oversight" and "Societal and Environmental Well-being" gather a lower score, although no recommendations are generated by the tool. This may be caused by some limitations of ALTAI, having been identified in the literature (Radclyffe et al., 2023; Stahl and Leach, 2023), such as the fact that some questions are not applicable in all the application domains or the candidate responses that are provided do not accurately represent its status, but they affect the resulting score.

Regarding the recommendations that ALTAI provides, one should take into account its limitations that have been already mentioned in the literature (Bousdekis et al., 2024). ALTAI has been designed for end products; it does not address the various phases of software development lifecycle. It incorporates generic questions aiming at addressing every AI system; however, the AI system may refer to a business environment with expert and qualified users. In addition, ALTAI considers as "AI system" the software and does not treat it as a socio-technical system, potentially leading to disregard of unforeseen challenges. For some questions, no alternative response is accurate, while its current structure hinders its applicability.

5.3.2 System usability

The assessment of system usability was undertaken by determining the SUS and VUS scores, which were derived from the questionnaires completed by the users. Each scale encompassed a section of 10 questions, which are cited in Table 4. The findings for each scale are depicted through three bar plots representing the average scores per user, per question, and per user role. The plots that illustrate the scores per question are scaled from 1 to 5, whereas the remaining plots feature percentile scores ranging from 0 to 100.

Figure 10 presents the average scores per user. The findings reveal that the system earned an 'Excellent' rating (>80.3) from two users and

[1] https://altai.insight-centre.org/

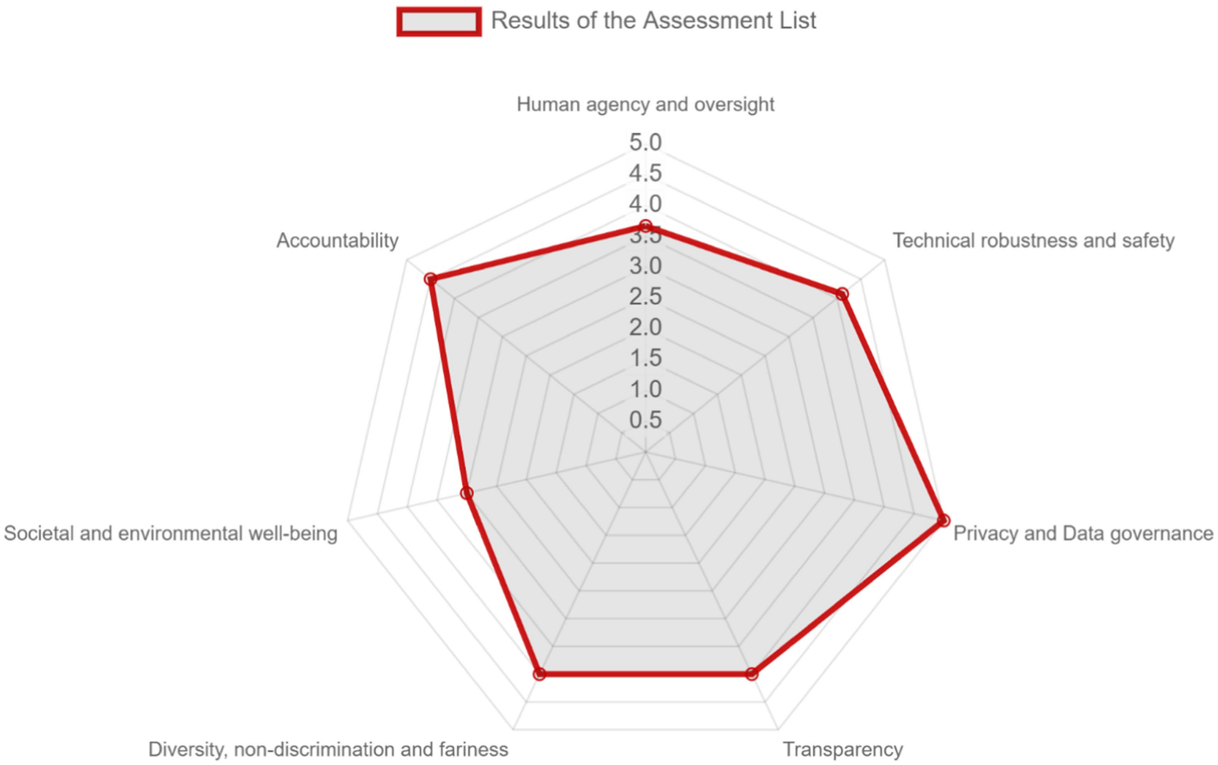


FIGURE 9
The results of the 7 requirements score as derived from ALTAI.

TABLE 4 SUS items.

Code	SUS Items
SUS_1	I think that I would like to use this system frequently.
SUS_2	I found the system unnecessarily complex.
SUS_3	I thought the system was easy to use.
SUS_4	I think that I would need the support of a technical person to be able to use this system.
SUS_5	I found the various functions in this system were well integrated.
SUS_6	I thought there was too much inconsistency in this system.
SUS_7	I would imagine that most people would learn to use this system very quickly.
SUS_8	I found the system very cumbersome to use.
SUS_9	I felt very confident using the system.
SUS_10	I needed to learn a lot of things before I could get going with this system.

a 'Good' rating (68–80.3) from another two, while the remainder assigned it a 'Poor' rating (51–68). A notable variation is evident among the Expert Operators, with scores ranging from 65 to 92.5, whereas the scores of the two Supervisors are more closely aligned.

In Figure 11, the average scores (ranging from 1 to 5) for each SUS question are illustrated. Notably, there is unanimity in the scores for questions 4, 7, and 10, which pertain, respectively, to the perceived need for technical support to use the system, the general consensus that most people would learn to use it quickly, and the anticipated necessity for users to learn a lot of things to operate the system. This uniformity in responses might stem from the users' limited experience and interaction with the system. Moreover, it is evident that the lowest score was garnered by question 2, which probes into the perceived complexity of the system (Figure 12).

The same types of graphs are presented for the VUS scores. Table 5 cites the VUS questions. In Figure 13, the plot illustrates the average VUS score per user. Initially, it is observed that the VUS scores are generally higher compared to the SUS scores, with all users awarding an 'Excellent' rating (>80.3), except for one user whose score is marginally below, at 80. This suggests a generally favorable user response to voice usability as compared to overall system usability.

In Figure 14, the average scores (1–5) for each VUS question are depicted. Within this scale, four questions received unanimous ratings with the maximum score. These questions, numbered 1, 6, 8, and 10, address, respectively, the ease of understanding the voice, the level of frustration experienced using the assistant in a noisy environment, the difficulty in customizing the voice assistant, and the difficulty in using the voice assistant. It's important to note that the questions with even numbering (6, 8, 10) are phrased with a negative sentiment, but the scores are inverted, meaning high values indicate a positive user experience. Additionally, all the scores are above the midpoint value of 3, suggesting a generally positive user response to these aspects of voice usability.

In the final plot for VUS, depicted in Figure 15, the average scores per user role are illustrated. The initial observation drawn from the bar values is that all user roles have, on average, rated the VUS above 80. Additionally, there is a slight variation in the scores among the Expert users, and an even smaller variation among the

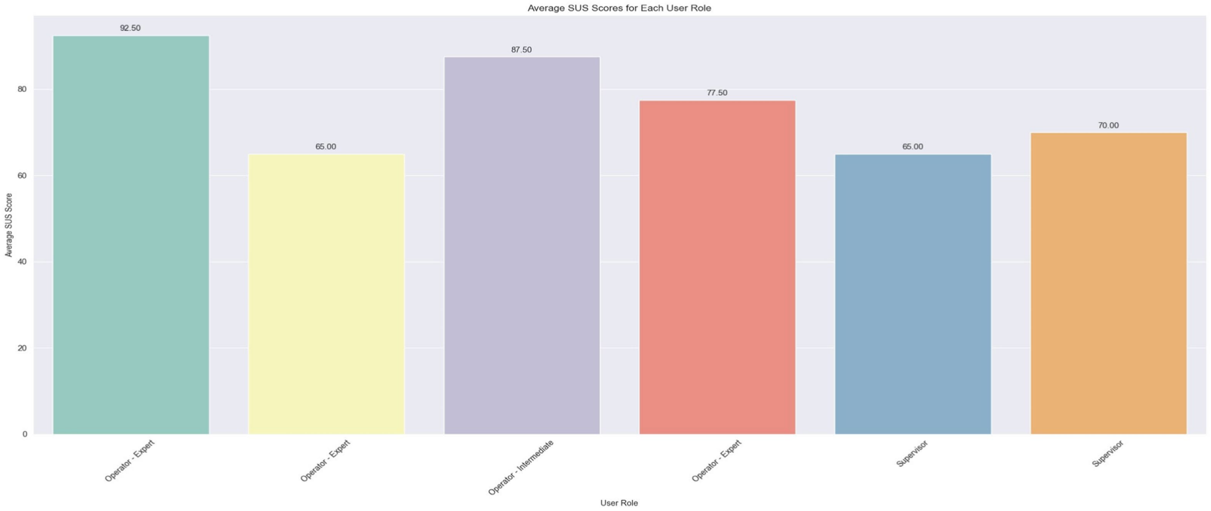


FIGURE 10
Average SUS score per user.

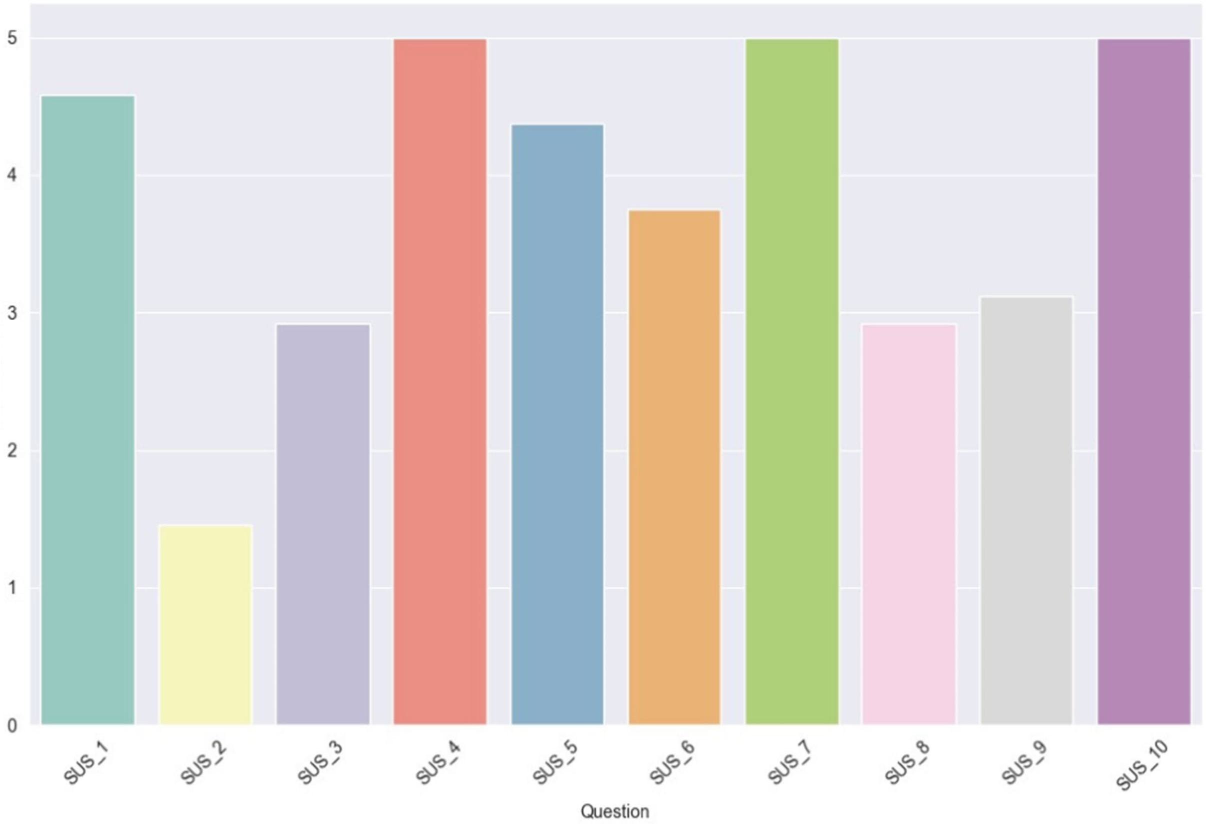


FIGURE 11
Average scores for SUS questions.

Supervisors, as indicated by the vertical black lines atop the bar plots.

Finally, Figure 16 illustrates a comparison of each user's average scores between the SUS and VUS. As previously observed, variations in scores among the Expert Operators are evident, with two of them also recording the highest VUS scores. Moreover, the two Supervisors, along with one of the Expert Operators, appear to occupy the lower spectrum of scores for both scales, indicating a less favorable assessment of usability. Intriguingly, the Intermediate Operator showcased consistent ratings across the two scales,

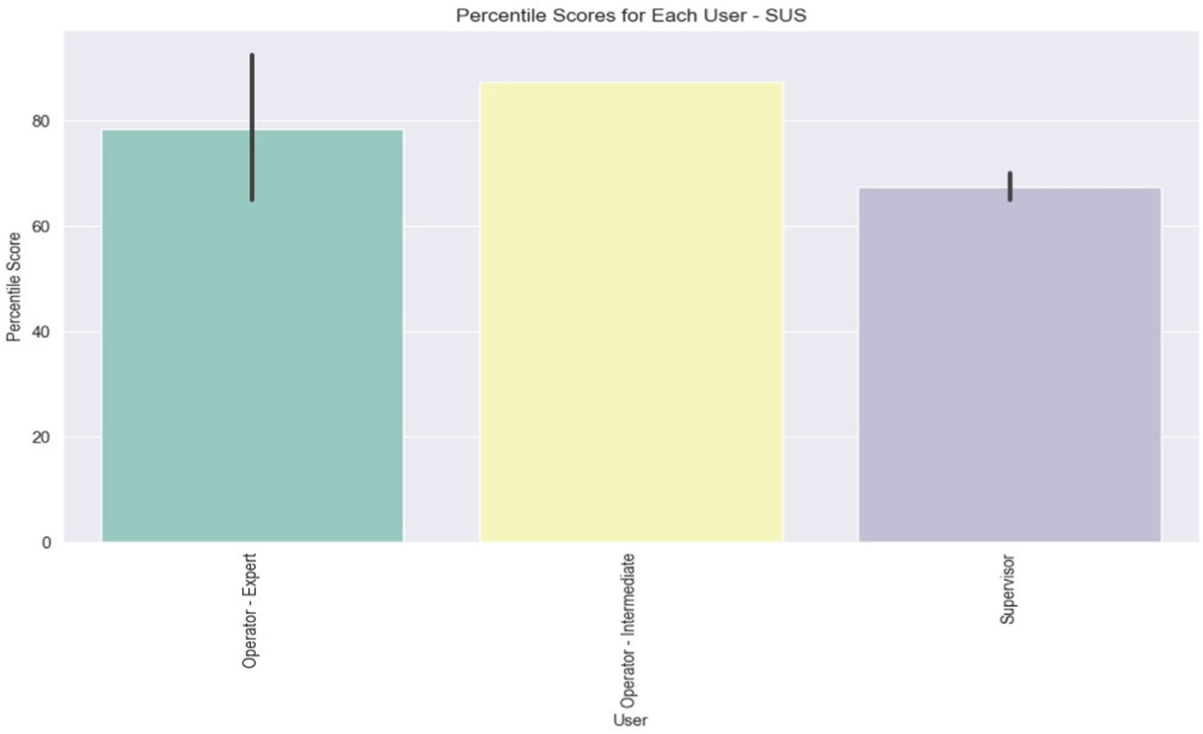


FIGURE 12
Average percentile SUS score for each user role.

TABLE 5 VUS items.

Code	VUS Items
VUS_1	I thought the response from the voice assistant was easy to understand.
VUS_2	I thought the information provided by the voice assistant was not relevant to what I asked.
VUS_3	I felt the response from the voice assistant was sufficient.
VUS_4	I thought the voice assistant had difficulty in understanding what I asked it to do.
VUS_5	I felt the voice assistant enabled me to successfully complete my tasks when I required help.
VUS_6	I found it frustrating to use the voice assistant in a noisy and loud environment.
VUS_7	The voice assistant had all the functions and capabilities that I expected it to have.
VUS_8	I found it difficult to customize the voice assistant according to my needs and preferences.
VUS_9	Overall, I am satisfied with using the voice assistant.
VUS_10	I found the voice assistant difficult to use.

reflecting a uniform perception of both system and voice usability. This Figure also serves as a comparison between SUS and VUS, since the variation between them demonstrates the complexity while evaluating voice interface-based systems compared to GUI-based systems.

5.3.3 Cognitive workload

In Figure 17, the average scores for each question of the NASA-TLX questionnaire are presented. The most noticeable observation initially is the difference in score between the question regarding perceived performance and all other questions. The former has unanimously the highest score, reaching 21. In contrast, the questions about the perceived mental, physical, and temporal demands, along with the one regarding effort, all exhibit low scores under 2.5. This suggests that the system was not perceived as particularly demanding by the users. A slightly higher value is observed for perceived frustration, with a score of 4.33. This could potentially be derived from the users' lack of familiarity with voice assistants.

5.3.4 Technical robustness

In this sub-section, we present the results related to the technical validation of the solution's main components. As far as the Analytics Service is concerned, we present metrics related to the accuracy of the AutoML models. The datasets used in these experiments were tabular containing datetime, numeric, categorical and text values. During preprocessing, text values were manually discarded due to inconsistencies and entries with missing values were removed. In this setting, we incorporated the Python libraries AutoKeras and FEDOT. The AutoKeras library generates Neural Networks (NN) and performs a Neural Architecture Search (NAS) guided by Bayesian Optimization and Gaussian processes. In our experiments, we employed the *StructuredDataClassifier* and the *StructuredDataRegressor* for the classification and regression tasks, respectively. They were also configured with the respective metrics

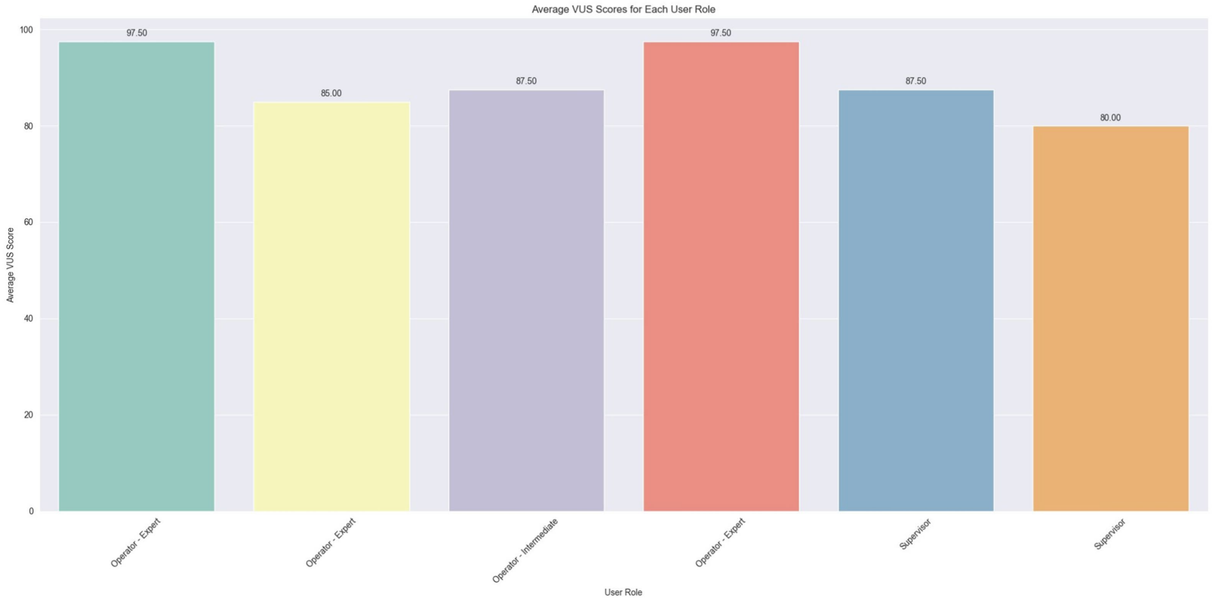


FIGURE 13
Average VUS score per user.

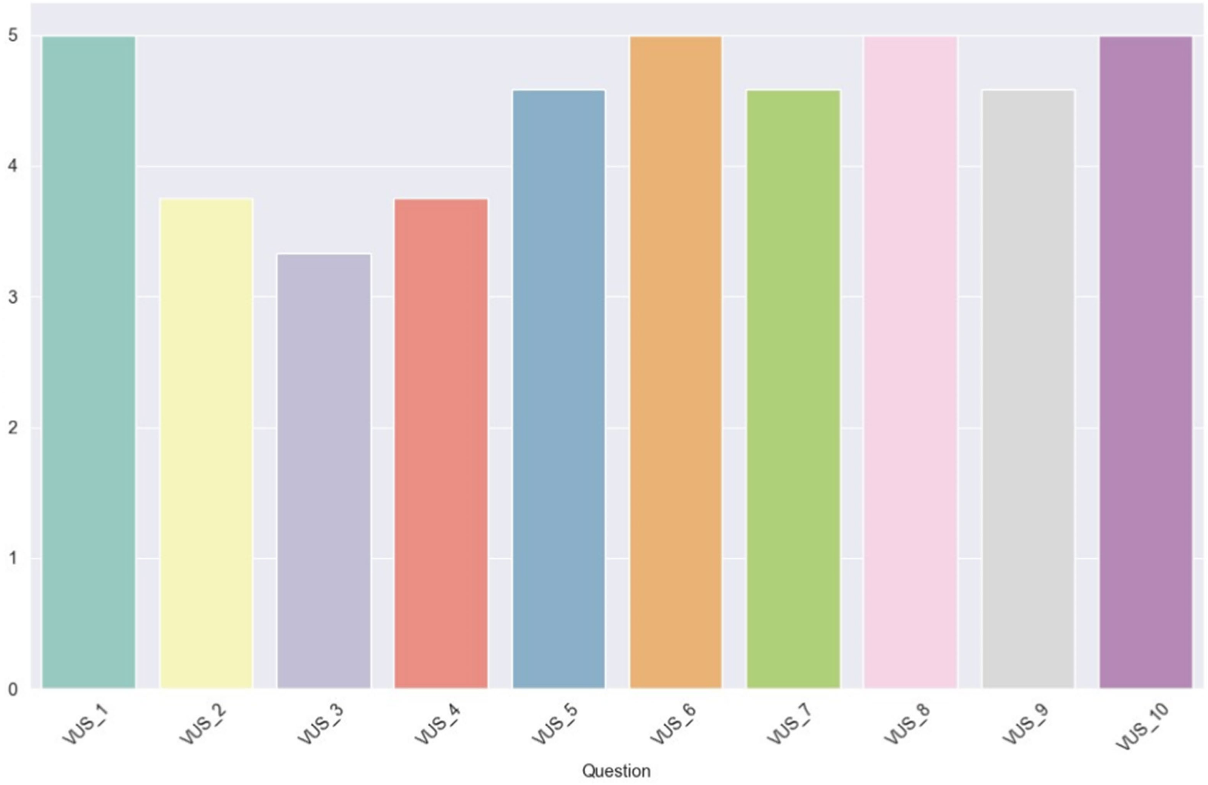


FIGURE 14
Average scores for VUS questions.

and the maximum number of trails to perform was set to 20. For the FEDOT library, the *Fedot* pipeline was employed with task-specific parameters in each task. For the classification task, the task and problem types were set to *classification* while for the regression task, it was set to time series forecasting. In both tasks, the optimization was conducted simultaneously with the maximum number of tuning

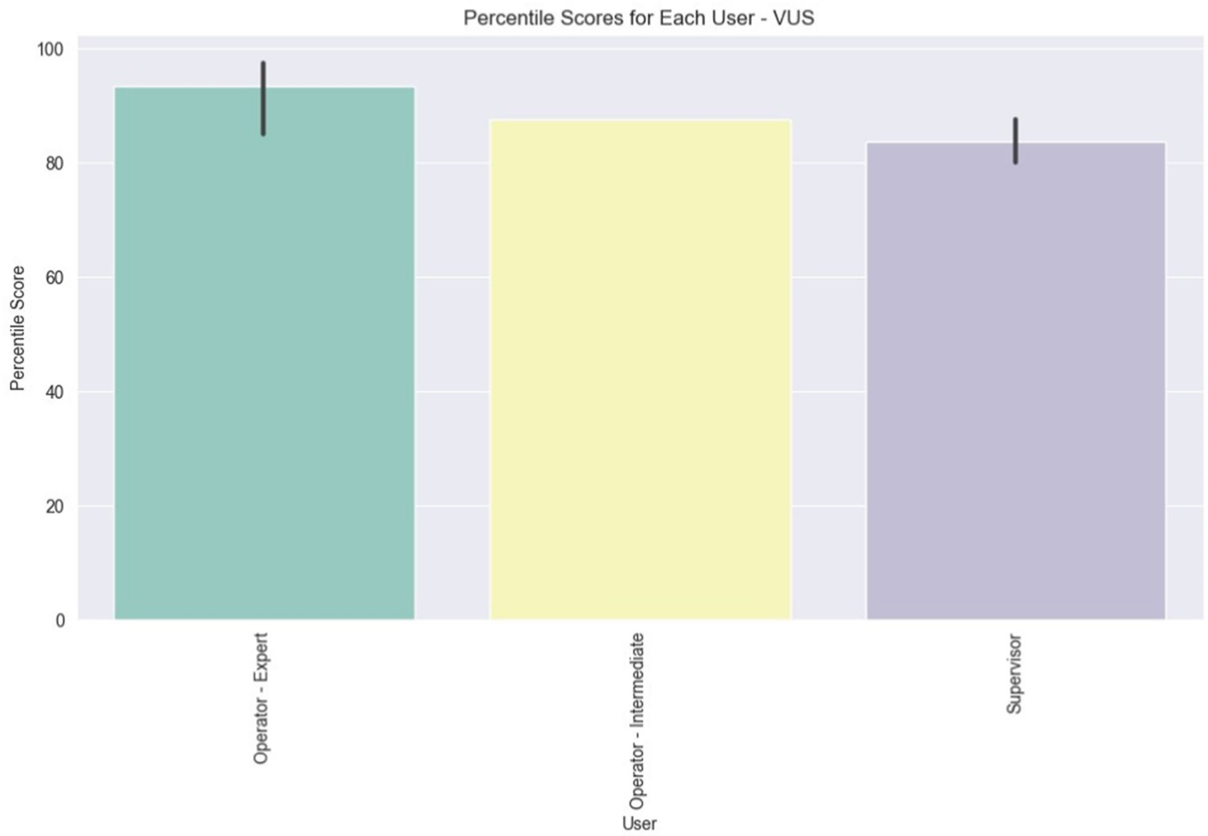


FIGURE 15
Average percentile VUS score for each user role.

iterations set to 100. We implemented three models per algorithm: (i) an initial model trained with the 80% of the dataset; (ii) a retrained model, with the whole dataset; and (iii) a new model which executes the AutoML process on the whole dataset from the beginning. Moreover, it should be noted that the execution time depends on the pre-configured stopping conditions.

Table 6 presents the results of the Defect Group Prediction case. The data processing algorithms selected 6 features (i.e., Date Created, Product Type (SKU), Defect Source, Station ID, Part Family) and the performance of the AutoML models was evaluated by measuring F1-macro, F1-micro, Receiver Operating Characteristic Area Under Curve (ROC-AUC), and execution time. Table 7 presents the results of the Defective Orders Prediction case. The data processing algorithms summed the Defect Instances were on the attribute Date Created to produce the necessary timeseries, and the performance of the AutoML models was evaluated by measuring the Mean Square Error (MSE), the Mean Absolute Error (MAE), and the execution time. The FEDOT AutoML framework selected the Extreme Gradient Boosting (XGBoost) ML algorithm, while the AutoKeras AutoML framework selected the Neural Network ML algorithm.

Throughout the evaluation period, as it was presented in Section 5.2, we were continuously improving the technical developments activities. In order to increase the intent recognition accuracy, we performed elimination of unusual words, streamline of sentences, usage of keywords, as well as usage of selection button on the tablet. These improvements drove the high accuracy in the user's intention understanding. The intent recognition accuracy for the conversations in September 2023, with 402 conversation turns, was 95.30%, as depicted in Figure 18.

5.3.5 Discussion on lessons learned and managerial implications

We performed workshops among the involved users and bilateral interviews in order to acquire qualitative feedback and to draw the main lessons learned. Below, we summarize the main conclusions derived from this procedure.

The solution provided the general validation of the cognitive worker assistance technology as one of the key enablers of a relevant change in the execution of industrial operations: the possibility of being assisted by an intelligent system, with a voice interface, which may drive the workers through the execution of complex operation, has been very appreciated by the users who confirmed the potential application not only in "off-line" working places, like laboratories or indirect activities (quality control, product repair, maintenance, logistics, material management) but also in "on-line" tasks execution like the quality control on the assembly where the pace has to be respected. A great advantage of the solution is the possibility to leave the hands free to execute tasks while the cognitive support understands the request, collects the right information, and provides it in a user-friendly format.

The possibility of having a unique interface to get access to all the information is very important as it boosts the efficiency in the

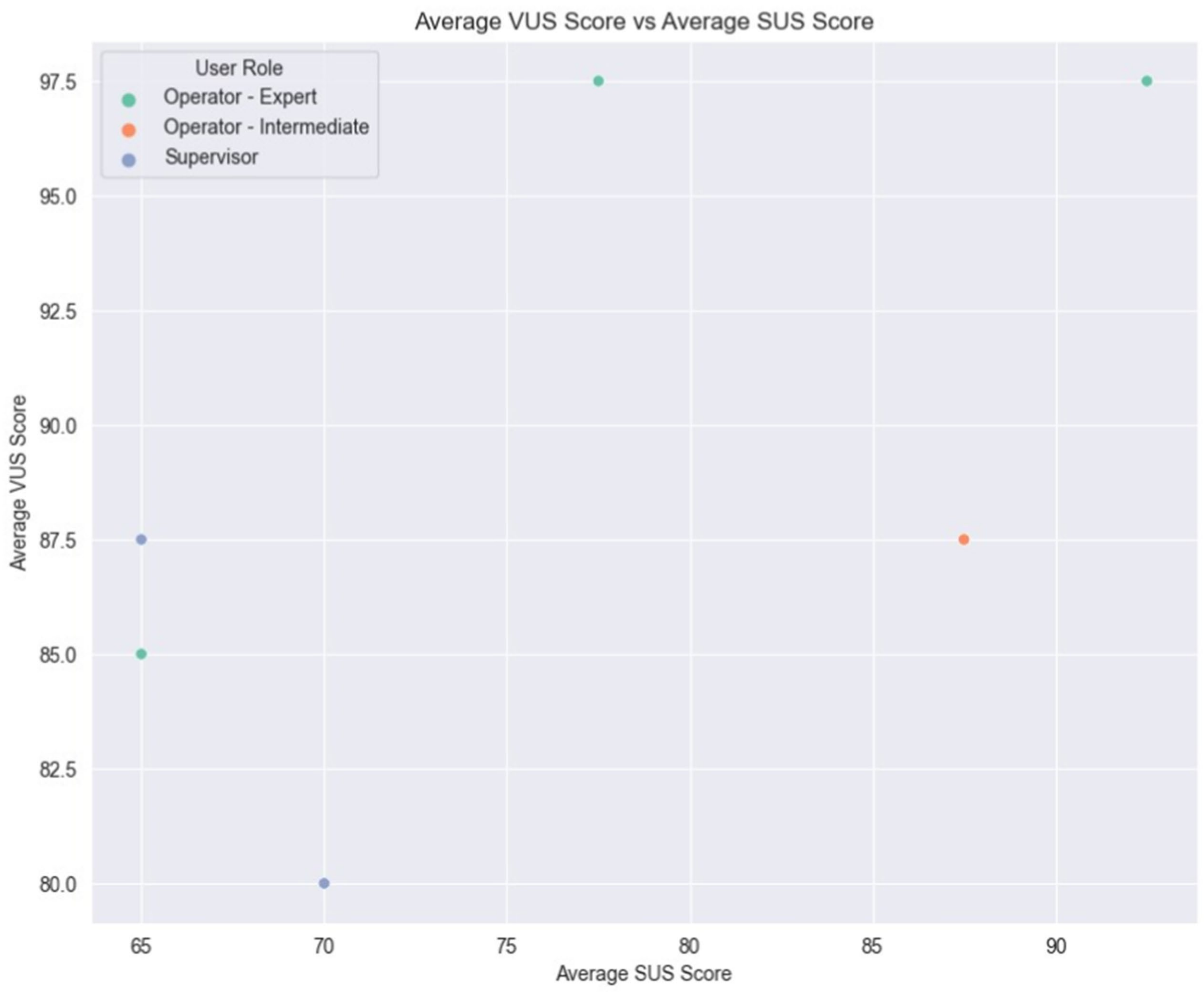


FIGURE 16
Average SUS vs. VUS scores.

navigation among the different systems of the IT landscape, ensuring the right information to the right person at the right time.

Mobile devices, such as tablets, are not the best equipment solution: tablets have to be managed by hand, interrupting task execution, and introducing inefficiency. The best solution has been identified in the usage of a headset and touch big screen, fixed in front of the operator to be easily seen and touched for the manual commands input, connected to a simple barcode scanner for SKU barcode reading. However, the adoption of wearable devices such as headsets introduces some constraints from a personnel management point of view: any wearable has to be strictly individual for safety and health reasons, and this may inflate the deployment cost of the solution in the factory.

The intensity of the information support and level of detail provided to the different people has been appreciated. However, the text referring to domain knowledge (e.g., quality checklists) includes long sentences, repetition, and wrong or incorrect words. This element highlighted the need to completely review the information content that has to be designed to be used in these types of digital systems and cannot replicate the structure that currently is used for the description on paper. The key reason behind this is that there was the need for a long knowledge acquisition period in order to simplify and create more robust relations for the syntax to be used.

The functionality to provide different services according to the user's profile has been appreciated, as well as the deployment of the learning path through the different users' skills profile (from novice to intermediate to expert). One remark has arisen by users related to this topic: the risk of having operators who can pass from novice to expert level faster than today could be penalized by the fact that these operators can be more "passive" towards the task execution as it is always suggested by the system. In this way, the risk of a passive approach will drive to the lack of "real expert" operators who are not able to execute without the system. The risk has to be mitigated by the real shift of the operator's attention from the pure mechanical execution of tasks to the interpretation of the information provided. To achieve this objective, an intensive, focused training action has to be put in place, combined with the collection of ideas and rewarding management.

The quality risk assessment functionality has been much appreciated, and, for the first time in the factory, there is clear visibility

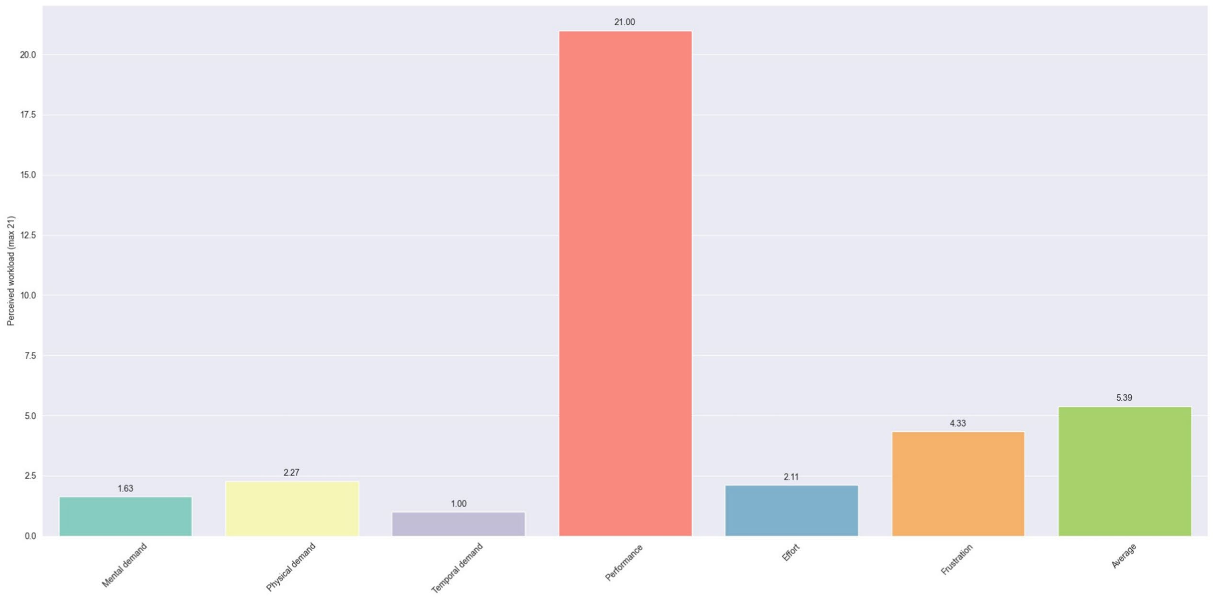


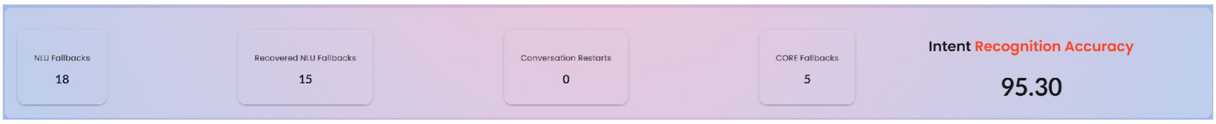
FIGURE 17
Perceived workload using NASA-TLX scores for the white goods production.

TABLE 6 Results for the defect group prediction.

	AutoML models	Metrics			
		F1-macro	F1-micro	ROC-AUC	Execution time (s)
Initial models	FEDOT	0.5055	0.8363	0.9023	1212.50
	AutoKeras	0.4509	0.7813	0.7141	1019.81
Retrained models	FEDOT	0.4969	0.8368	0.9019	9.35
	AutoKeras	0.4510	0.7813	0.7141	0.58
New models	FEDOT	0.4909	0.8277	0.8722	91.96
	AutoKeras	0.4631	0.7681	0.7177	953.36

TABLE 7 Results for the defective orders prediction.

	AutoML models	Metrics		
		MSE	MAE	Execution time (s)
Initial models	FEDOT	0.2338	0.2017	101.88
	AutoKeras	0.0402	0.1624	187.39
Retrained models	FEDOT	0.1361	0.1002	0.32
	AutoKeras	0.0201	0.1082	2.57
New models	FEDOT	0.1391	0.0988	127.48
	AutoKeras	0.0191	0.1036	11.17



NLU Fallbacks
18

Recovered NLU Fallbacks
15

Conversation Restarts
0

CORE Fallbacks
5

Intent Recognition Accuracy
95.30

FIGURE 18
Intent recognition accuracy and fallback summary.

of predictive quality analytics results. The good quality of the prediction has also been confirmed by users who got access to predictive results and confirmed them with real experience on the various defective parts and products in the actual production process.

The huge amount of data with poor quality is one of the most common cases in the IT landscape of manufacturing companies, which, for years, have been collecting data from the shop floor without real and effective usage of them. This element forces the IT organization to create a stronger link with the shop floor to deploy a more effective solution: for Whirlpool this will pass through a redesign of the data architecture in the cloud in order to more effectively and efficiently support the analytics functionalities deployment. Therefore, the need for a deep review of the actual data management strategy has arisen: the need to count on a high-quality database, real-time updated, and designed to be efficiently integrated with digital functionalities proved to be one of the higher priorities in digital transformation. Whirlpool identified that they need to be based upon a completely different quality control data model.

A focused change management strategy has been defined, and it has been used as the backbone for the creation and deployment of communication and training actions not only toward the involved users but also engaging the overall factory organization at different levels. These activities put evidence on the need for great attention on how these types of technology are presented and deployed to the impacted population in order to create the right level of awareness on potential and risks and enable the real adoption of the solution.

The adoption of the solution seriously and structurally faces privacy and GDPR compliance management in the deployment of voice assistance applications. The execution of the change management actions and the "privacy by design" effort spent for system development and consensus form finalization, put evidence on the real poor level of awareness of people on the potential impact of this technology in the working environment, and, in parallel, also at home and in their personal life.

5.4 Discussion on generalizability criteria

The evaluation results showed that the integration of voice assistance technology with AutoML has the potential to significantly contribute to the increase of operational efficiency. In order to apply the proposed approach to different manufacturing use cases, the challenges of AutoML when it uses a voice interface for user interaction should be taken into account. Therefore, the following criteria need to be considered by manufacturers, software developers, data scientists, and practitioners:

Accuracy: It should achieve high prediction accuracy. It should be taken into account that higher accuracy may need more computational resources.

Efficiency: It should be efficient in terms of time and computational resources. The automated processes should be designed to optimize the use of resources and reduce the overall time required to train models and make predictions. This particularly applies to applications that are critical in terms of time (high sampling time). In these cases, prediction accuracy may negatively be affected.

Scalability: It should be capable of scaling up to handle large datasets and complex ML tasks and pipelines. They should be able to handle increasing amounts of data and computational demands without compromising performance or efficiency.

Flexibility: It should provide flexibility by supporting several ML algorithms. It should allow users to experiment with different approaches and customize the automated processes according to their specific requirements and preferences.

Explainability: It should be accompanied with explainability mechanisms, suitable for voice interfaces, in order to enable the human to understand the inner working and the outcomes of the "black-box" AutoML pipelines.

Transparency: It should be able to provide evidence about the ML models and pipelines that were used for generating some specific insights in order to contribute to the increase of user trust.

Robustness: It should be robust against unexpected inputs and real-world industrial data challenges, such as noise, missing values, and imbalances. This is particularly important in safety-critical tasks.

Adaptability: It should be capable of adapting across various domains and use cases covering diverse business requirements and manufacturing operations.

Integration with existing IT systems: It should incorporate interfaces, based on related standards, for facilitating the integration with existing IT infrastructures and production systems, i.e., legacy systems, Enterprise Resource Planning (ERP), Manufacturing Execution Systems (MES), quality management systems, etc., according to the application domain and the manufacturing operation at hand. It should also support both a cloud-based integration and an on-premise integration, providing sufficient documentation.

Cost-effectiveness: It should minimize the additional effort required for installing and configuring the solution.

Compliance with ethical and regulatory standards: It should be compliant with trustworthy and ethical Artificial Intelligence (AI) principles, taking into account related regulations, such as ALTAI and EU AI Act.

6 Conclusion and future work

Augmented intelligence puts together human and artificial agents to create a socio-technological system so that they co-evolve by learning and optimizing decisions through intuitive interfaces, such as conversational, voice-enabled interfaces. However, existing research works on voice assistants rely on knowledge management and simulation methods instead of data-driven algorithms in order to take advantage of the large amounts of data existing in modern manufacturing environments. In addition, practical application and evaluation in real-life scenarios are scarce and limited in scope due to the aforementioned challenges.

In this paper, we proposed the integration of voice assistance technology with AutoML in order to enable the realization of the augmented intelligence paradigm in the context of Industry 5.0. AutoML automates the building and deployment of ML pipelines without requiring ML knowledge, while the voice interface exposes the data analytics outcomes to the user in an intuitive way. On the other hand, the user is able to interact with the assistant, and consequently with the ML models, through voice in order to receive immediate insights while performing their task. The proposed approach was evaluated in a real manufacturing environment. We followed a structured evaluation methodology and analyzed the results, which demonstrate the effectiveness of our proposed approach.

Our future work will move towards the following directions: (i) We will extend the proposed solution by utilizing LLM for user interaction in order to enhance the scalability and adaptability of the DIA to various industrial settings with less human effort at the design time of the solution; (ii) We will evaluate the proposed solution in additional real-life manufacturing scenarios from various sectors taking into account the generalizability criteria that were derived from the current research work.; and, (iii) We will focus on how transparency and explainability approaches for AutoML can be incorporated when there is a voice user interface instead of a GUI.

Data availability statement

The datasets for this article are not publicly available due to legal and privacy-related restrictions in relation to confidential human data. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Professor Nicholas Asher - Centre National de Recherche Scientifique (CNRS), Enrica Bosani - Whirlpool EMEA, Professor Gregoris Mentzas - National Technical University of Athens, Dr. Alexandros Bousdekis - Institute of Communication and Computer Systems (ICCS), Dr. Stefan Wellsandt - BIBA (Bremer Institut für Produktion und Logistik GmbH), Karl Hribernik - BIBA (Bremer Institut für Produktion und Logistik GmbH). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in these studies.

Author contributions

AB: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. MFo: Conceptualization, Data curation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. MFi: Data curation, Software, Validation, Writing – original draft, Writing – review & editing. SW: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. KL: Methodology, Software, Writing – original draft, Writing – review & editing. EB: Investigation, Resources, Validation, Writing – original draft, Writing – review & editing. GM: Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. K-DT: Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the European Union's Horizon 2020 project COALA "Cognitive Assisted agile manufacturing for a Labor force supported by trustworthy Artificial Intelligence" (Grant agreement No 957296). The work presented here reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2025.1538840/full#supplementary-material

SUPPLEMENTARY TABLE S1
Resulting ALTAI recommendations for each out of the 7 requirements.

References

Abner, B., Rabelo, R. J., Zambiasi, S. P., and Romero, D. (2020). Production management as-a-service: a softbot approach. In B. Lalic, V. Majstorovic, U. Marjanovic, CieminskiG. von and D. Romero Advances in production management systems towards smart and digital manufacturing: Part II (pp. 19–30). Cham: Springer International Publishing.

Afanasev, M. Y., Fedosov, Y. V., Andreev, Y. S., Krylova, A. A., Shorokhov, S. A., Zimenko, K. V., et al. (2019) A concept for integration of voice assistant and modular cyber-physical production system. In 2019 IEEE 17th international conference on industrial informatics (INDIN), IEEE.

Amirian, M., Tuggener, L., Chavarriaga, R., Satyawan, Y. P., Schilling, F. P., Schwenker, F., et al. (2021). Two to trust: automl for safe modelling and interpretable deep learning for robustness. In Trustworthy AI-Integrating Learning, Optimization and Reasoning: First International Workshop, TAILOR 2020, Virtual Event, September 4–5, 2020, Revised Selected Papers 1 (pp. 268–275). Springer International Publishing.

Artificial Intelligence *High-level independent group on artificial intelligence (AI HLEG). Ethics Guidelines for Trustworthy AI*. (2019) European Commission.

Azevedo, K., Quaranta, L., Calefato, F., and Kalinowski, M. (2024). A multivocal literature review on the benefits and limitations of automated machine learning tools. *arXiv preprint arXiv*:2401.11366.

Babel, M., McGuire, G., and King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS One* 9:e88616. doi: 10.1371/journal.pone.0088616

Bangaru, S. S., Wang, C., Hassan, M., Jeon, H. W., and Ayiluri, T. (2019). Estimation of the degree of hydration of concrete through automated machine learning based microstructure analysis–a study on effect of image magnification. *Adv. Eng. Inform.* 42:100975. doi: 10.1016/j.aei.2019.100975

Barbudo, R., Ventura, S., and Romero, J. R. (2023). Eight years of AutoML: categorisation, review and trends. *Knowl. Inf. Syst.* 65, 5097–5149. doi: 10.1007/s10115-023-01935-1

Bernard, D., and Arnold, A. (2019). Cognitive interaction with virtual assistants: from philosophical foundations to illustrative examples in aeronautics. *Comput. Ind.* 107, 33–49. doi: 10.1016/j.compind.2019.01.010

Bousdekis, A., Apostolou, D., and Mentzas, G. (2020). A human cyber physical system framework for operator 4.0–artificial intelligence symbiosis. *Manuf. Lett.* 25, 10–15. doi: 10.1016/j.mfglet.2020.06.001

Bousdekis, A., Mentzas, G., Apostolou, D., and Wellsandt, S. (2022). Evaluation of AI-based digital assistants in smart manufacturing. In IFIP international conference on advances in production management systems (pp. 503–510). Cham: Springer Nature Switzerland.

Bousdekis, A., Mentzas, G., Apostolou, D., and Wellsandt, S. (2024). Assessing trustworthy artificial intelligence of voice-enabled intelligent assistants for the operator 5.0. In IFIP international conference on advances in production management systems (pp. 220–234). Cham: Springer Nature Switzerland.

Bousdekis, A., Wellsandt, S., Bosani, E., Lepenioti, K., Apostolou, D., Hribernik, K., et al. (2021). Human-AI collaboration in quality control with augmented manufacturing analytics. In Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems: IFIP WG 5.7 International Conference, APMS 2021, Nantes, France, September 5–9, 2021, Proceedings, Part IV (pp. 303–310). Springer International Publishing.

Brooke, J. (1996). "SUS: A "quick and dirty" usability scale" in Usability Evaluation in industry. eds. P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland (United Kingdom: Taylor and Francis).

Chaabi, M., Hamlich, M., and Garouani, M. (2022) Evaluation of AutoML tools for manufacturing applications. In International conference on integrated design and production (pp. 323–330). Cham: Springer International Publishing.

Colabianchi, S., Costantino, F., and Sabetta, N. (2024). Assessment of a large language model based digital intelligent assistant in assembly manufacturing. *Comput. Ind.* 162:104129. doi: 10.1016/j.compind.2024.104129

Conrad, F., Mälzer, M., Lange, F., Wiemer, H., and Ihlenfeldt, S. (2024). AutoML applied to time series analysis tasks in production engineering. *Proc. Comput. Sci.* 232, 849–860. doi: 10.1016/j.procs.2024.01.085

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., et al. (2017). "What can i help you with?" infrequent users' experiences of intelligent personal assistants. In Proceedings of the 19th international conference on human-computer interaction with mobile devices and services (pp. 1–12).

Crisan, A., and Fiore-Gartland, B. (2021). Fits and starts: Enterprise use of automl and the role of humans in the loop. In Proceedings of the 2021 CHI Conference on human factors in computing systems (pp. 1–15).

de Assis Dornelles, J., Ayala, N. F., and Frank, A. G. (2022). Smart working in industry 4.0: how digital technologies enhance manufacturing workers' activities. *Comput. Ind. Eng.* 163:107804. doi: 10.1016/j.cie.2021.107804

Denkena, B., Dittrich, M. A., Lindauer, M., Mainka, J., and Stürenburg, L. (2020). Using AutoML to optimize shape error prediction in milling processes. In Proceedings of the machining innovations conference (MIC).

Drozdal, J., Weisz, J., Wang, D., Dass, G., Yao, B., Zhao, C., et al. (2020). Trust in AutoML: exploring information needs for establishing trust in automated machine learning systems. In Proceedings of the 25th international conference on intelligent user interfaces (pp. 297–307).

Elshawi, R., Maher, M., and Sakr, S. (2019). Automated machine learning: state-of-the-art and open challenges. *arXiv preprint arXiv*:1906.02287.

Fikardos, M., Lepenioti, K., Bousdekis, A., Bosani, E., Apostolou, D., and Mentzas, G. (2022) An automated machine learning framework for predictive analytics in quality control. In IFIP international conference on advances in production management systems (pp. 19–26). Cham: Springer Nature Switzerland.

Freire, S. K., Panicker, S. S., Ruiz-Arenas, S., Rusák, Z., and Niforatos, E. (2022). A cognitive assistant for operators: Ai-powered knowledge sharing on complex systems. *IEEE Pervasive Comput* 22, 50–58. doi: 10.1109/MPRV.2022.3218600

Garouani, M., Ahmad, A., Bouneffa, M., Hamlich, M., Bourguin, G., and Lewandowski, A. (2022). Towards big industrial data mining through explainable automated machine learning. *Int. J. Adv. Manuf. Technol.* 120, 1169–1188. doi: 10.1007/s00170-022-08761-9

Gärtler, M., and Schmidt, B. (2021). *Proceedings of the 54th Hawaii International Conference on System Sciences*. (Hawali, USA), 4063–4072.

Gerling, A., Ziekow, H., Hess, A., Schreier, U., Seiffer, C., and Abdeslam, D. O. (2022). Comparison of algorithms for error prediction in manufacturing with AutoML and a cost-based metric. *J. Intell. Manuf.* 33, 555–573. doi: 10.1007/s10845-021-01890-0

Ghofrani, J., and Reichelt, D. (2019). Using voice assistants as HMI for robots in smart production systems. In CEUR Workshop Proceedings

Hart, S. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research" in Human mental workload. eds. P. A. Hancock and N. Meshkati (Amsterdam: Elsevier).

He, X., Zhao, K., and Chu, X. (2021). AutoML: a survey of the state-of-the-art. *Knowl. Based Syst.* 212:106622. doi: 10.1016/j.knosys.2020.106622

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., and McTear, M. (2019). Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces?. In Proceedings of the 31st European Conference on Cognitive Ergonomics (pp. 207–214).

Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). Automated machine learning: Methods, systems, challenges. Cham: Springer Nature.

Ionescu, T. B., and Schlund, S. (2021). Programming cobots by voice: a human-centered, web-based approach. *Proc. CIRP* 97, 123–129. doi: 10.1016/j.procir.2020.05.213

Jayasurya, B., Suguna, M., Saravanan, P., and Revathi, M. (2024) AutoML as a catalyst for predictive maintenance innovation: strategies and outcomes. In 2024 3rd international conference on artificial intelligence for internet of things (AIIoT) (pp. 1–6). IEEE.

Jwo, J. S., Lin, C. S., and Lee, C. H. (2021). An interactive dashboard using a virtual assistant for visualizing smart manufacturing. *Mob. Inf. Syst.* 2021, 1–9. doi: 10.1155/2021/5578239

Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., and Veeramachaneni, K. (2021). Automl to date and beyond: challenges and opportunities. *ACM Comput. Surv.* 54, 1–36. doi: 10.1145/3470918

Krauß, J., Pacheco, B. M., Zang, H. M., and Schmitt, R. H. (2020). Automated machine learning for predictive quality in production. *Proc. CIRP* 93, 443–448. doi: 10.1016/j.procir.2020.04.039

Leng, J., Sha, W., Wang, B., Zheng, P., Zhuang, C., Liu, Q., et al. (2022). Industry 5.0: Prospect and retrospect. *J. Manuf. Syst.* 65, 279–295. doi: 10.1016/j.jmsy.2022.09.017

Lepenioti, K., Bousdekis, A., Apostolou, D., and Mentzas, G. (2020). Prescriptive analytics: literature review and research challenges. *Int. J. Inf. Manag.* 50, 57–70. doi: 10.1016/j.ijinfomgt.2019.04.003

Li, C., Chrysostomou, D., and Yang, H. (2023). A speech-enabled virtual assistant for efficient human–robot interaction in industrial environments. *J. Syst. Softw.* 205:111818. doi: 10.1016/j.jss.2023.111818

Li, C., and Yang, H. J. (2021). Bot-x: an AI-based virtual assistant for intelligent manufacturing. *Multiag. Grid Syst.* 17, 1–14. doi: 10.3233/MGS-210340

Liang, D., and Xue, F. (2023). Integrating automated machine learning and interpretability analysis in architecture, engineering and construction industry: a case of identifying failure modes of reinforced concrete shear walls. *Comput. Ind.* 147:103883. doi: 10.1016/j.compind.2023.103883

Linares-Garcia, D. A., Roofigari-Esfahan, N., Pratt, K., and Jeon, M. (2022). Voice-based intelligent virtual agents (VIVA) to support construction worker productivity. *Autom. Constr.* 143:104554. doi: 10.1016/j.autcon.2022.104554

Longo, F., and Padovano, A. (2020). Voice-enabled assistants of the operator 4.0 in the social smart factory: prospective role and challenges for an advanced human–machine interaction. *Manuf. Lett.* 26, 12–16. doi: 10.1016/j.mfglet.2020.09.001

Ludwig, H., Schmidt, T., and Kühn, M. (2023). Voice user interfaces in manufacturing logistics: a literature review. *Int. J. Speech Technol.* 26, 627–639. doi: 10.1007/s10772-023-10036-x

Maddikunta, P. K. R., Pham, Q. V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T. R., et al. (2022). Industry 5.0: a survey on enabling technologies and potential applications. *J. Ind. Inf. Integr.* 26:100257. doi: 10.1016/j.jii.2021.100257

Mallouk, I., Sallez, Y., and El Majd, B. A. (2023) AutoML approach for decision making in a manufacturing context. In International workshop on service orientation in Holonic and multi-agent manufacturing (pp. 151–163). Cham: Springer Nature Switzerland.

Matt, C., Hess, T., and Benlian, A. (2015). Digital transformation strategies. *Bus. Inf. Syst. Eng.* 57, 339–343. doi: 10.1007/s12599-015-0401-5

Mentzas, G., Fikardos, M., Lepenioti, K., and Apostolou, D. (2024). Exploring the landscape of trustworthy artificial intelligence: status and challenges. *Intellig. Decis. Technol.* 18, 837–854. doi: 10.3233/IDT-240366

Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., and Frick, N. R. (2021). Understanding collaboration with virtual assistants–the role of social identity and the extended self. *Bus. Inf. Syst. Eng.* 63, 21–37. doi: 10.1007/s12599-020-00672-x

Mukherjee, A., Mertes, J., Glatt, M., and Aurich, J. C. (2024). Voice user Interface based control for industrial machine tools. *Proc. CIRP* 121, 121–126. doi: 10.1016/j.procir.2023.09.238

Murad, C., Munteanu, C., Cowan, B. R., and Clark, L. (2019). Revolution or evolution? Speech interaction and HCI design guidelines. *IEEE Pervasive Comput* 18, 33–45. doi: 10.1109/MPRV.2019.2906991

Nguyen Ngoc, H., Lasa, G., and Iriarte, I. (2022). Human-centred design in industry 4.0: case study review and opportunities for future research. *J. Intell. Manuf.* 33, 35–76. doi: 10.1007/s10845-021-01796-x

Nikitin, N. O., Vychuzhanin, P., Sarafanov, M., Polonskaia, I. S., Revin, I., Barabanova, I. V., et al. (2022). Automated evolutionary approach for the design of composite machine learning pipelines. *Futur. Gener. Comput. Syst.* 127, 109–125. doi: 10.1016/j.future.2021.08.022

Ninomiya, Y., Iwata, T., Terai, H., and Miwa, K. (2024). Effect of cognitive load and working memory capacity on the efficiency of discovering better alternatives: a survival analysis. *Mem. Cogn.* 52, 115–131. doi: 10.3758/s13421-023-01448-w

Norda, M., Engel, C., Rennies, J., Appell, J. E., Lange, S. C., and Hahn, A. (2023). Evaluating the efficiency of voice control as human machine interface in production. *IEEE Trans. Autom. Sci. Eng.* 99, 1–12. doi: 10.1109/TASE.2023.3302951

Rabelo, R. J., Romero, D., and Zambiasi, S. P. (2018). Advances in production management systems. smart manufacturing for industry 4.0: IFIP WG 5.7 international conference, APMS 2018, Seoul, Korea, august 26–30, 2018, proceedings, part IISoftbots supporting the operator 4.0 at smart factory environments. In (pp. 456–464). Springer International Publishing.

Radclyffe, C., Ribeiro, M., and Wortham, R. H. (2023). The assessment list for trustworthy artificial intelligence: a review and recommendations. *Front. Art. Intellig.* 6:1020592. doi: 10.3389/frai.2023.1020592

Rooney, S., Pitz, E., and Pochiraju, K. (2024). AutoML-driven diagnostics of the feeder motor in fused filament fabrication machines from direct current signals. *J. Intell. Manuf.*, 1–18. doi: 10.1007/s10845-024-02332-3

Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Appl. Psychol.* 53, 61–86. doi: 10.1111/j.1464-0597.2004.00161.x

Ruiz, E., Torres, M. I., and del Pozo, A. (2023). Question answering models for human–machine interaction in the manufacturing industry. *Comput. Ind.* 151:103988. doi: 10.1016/j.compind.2023.103988

Saka, A. B., Oyedele, L. O., Akanbi, L. A., Ganiyu, S. A., Chan, D. W., and Bello, S. A. (2023). Conversational artificial intelligence in the AEC industry: a review of present status, challenges and opportunities. *Adv. Eng. Inform.* 55:101869. doi: 10.1016/j.aei.2022.101869

Salehin, I., Islam, M. S., Saha, P., Noman, S. M., Tuni, A., Hasan, M. M., et al. (2024). AutoML: a systematic review on automated machine learning with neural architecture search. *J. Inform. Intellig.* 2, 52–81. doi: 10.1016/j.jiixd.2023.10.002

Schuh, G., Stroh, M. F., and Benning, J. (2022) Case-study-based requirements analysis of manufacturing companies for auto-ML solutions. In IFIP international conference on advances in production management systems (pp. 43–50). Cham: Springer Nature Switzerland.

Sousa, A., Ferreira, L., Ribeiro, R., Xavier, J., Pilastri, A., and Cortez, P. (2022) Production time prediction for contract manufacturing industries using automated machine learning. In IFIP international conference on artificial intelligence applications and innovations (pp. 262–273). Cham: Springer International Publishing.

Stahl, B. C., and Leach, T. (2023). Assessing the ethical and social concerns of artificial intelligence in neuroinformatics research: an empirical test of the European Union assessment list for trustworthy AI (ALTAI). *AI Ethics* 3, 745–767. doi: 10.1007/s43681-022-00201-4

Wang, D., Andres, J., Weisz, J. D., Oduor, E., and Dugan, C. (2021). Autods: towards human-centered automation of data science. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1–12).

Wang, B., Zheng, P., Yin, Y., Shih, A., and Wang, L. (2022). Toward human-centric smart manufacturing: a human-cyber-physical systems (HCPS) perspective. *J. Manuf. Syst.* 63, 471–490. doi: 10.1016/j.jmsy.2022.05.005

Wellsandt, S., Klein, K., Hribernik, K., Lewandowski, M., Bousdekis, A., Mentzas, G., et al. (2022). Hybrid-augmented intelligence in predictive maintenance with digital intelligent assistants. *Annu. Rev. Control.* 53, 382–390. doi: 10.1016/j.arcontrol.2022.04.001

Wellsandt, S., Rusak, Z., Ruiz Arenas, S., Aschenbrenner, D., Hribernik, K. A., and Thoben, K. D. (2020). Concept of a voice-enabled digital assistant for predictive maintenance in manufacturing. Cranfield, UK: Proceedings of the TESConf 2020 - 9th International Conference on Through-life Engineering Services.

Xu, X., Lu, Y., Vogel-Heuser, B., and Wang, L. (2021). Industry 4.0 and industry 5.0—inception, conception and perception. *J. Manuf. Syst.* 61, 530–535. doi: 10.1016/j.jmsy.2021.10.006

Yu, T., and Zhu, H. (2020). Hyper-parameter optimization: a review of algorithms and applications. *arXiv preprint arXiv*:2003.05689.

Zambiasi, L. P., Rabelo, R. J., Zambiasi, S. P., and Lizot, R. (2022) Supporting resilient operator 5.0: an augmented softbot approach. In IFIP international conference on advances in production management systems (pp. 494–502). Cham: Springer Nature Switzerland.

Zhai, W., Shi, X., Wong, Y. D., Han, Q., and Chen, L. (2024). Explainable AutoML (xAutoML) with adaptive modeling for yield enhancement in semiconductor smart manufacturing. *arXiv preprint arXiv*:2403.12381.

Zhai, W., Shi, X., and Zeng, Z. (2023) Adaptive modelling for anomaly detection and defect diagnosis in semiconductor smart manufacturing: a domain-specific AutoML. In 2023 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM) (pp. 198–203). IEEE.

Zheng, T., Grosse, E. H., Morana, S., and Glock, C. H. (2024). A review of digital assistants in production and logistics: applications, benefits, and challenges. *Int. J. Prod. Res.* 62, 8022–8048. doi: 10.1080/00207543.2024.2330631

Zöller, M. A., Titov, W., Schlegel, T., and Huber, M. F. (2023). Xautoml: a visual analytics tool for understanding and validating automated machine learning. *ACM Trans. Interact. Intellig. Syst.* 13, 1–39. doi: 10.1145/3625240

Zwakman, D. S., Pal, D., and Arpnikanondt, C. (2021). Usability evaluation of artificial intelligence-based voice assistants: the case of Amazon Alexa. *SN Comput. Sci.* 2:28. doi: 10.1007/s42979-020-00424-4