



OPEN ACCESS

EDITED BY

Seamus Simpson,
University of Salford, United Kingdom

REVIEWED BY

Minna Horowitz,
University of Helsinki, Finland
Manuel Hernandez Perez,
University of Salford, United Kingdom

*CORRESPONDENCE

Emine Ozge Yildirim-Vranckaert
✉ emineozge.yildirim@kuleuven.be
Aaron Hyzen
✉ aaron.hyzen@asc.upenn.edu

RECEIVED 17 January 2025

ACCEPTED 16 June 2025

PUBLISHED 23 July 2025

CITATION

Yildirim-Vranckaert EO and Hyzen A (2025)
Untouched minds in a tangled web:
navigating mental autonomy and epistemic
welfare amidst digital propaganda.
Front. Commun. 10:1562368.
doi: 10.3389/fcomm.2025.1562368

COPYRIGHT

© 2025 Yildirim-Vranckaert and Hyzen. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Untouched minds in a tangled web: navigating mental autonomy and epistemic welfare amidst digital propaganda

Emine Ozge Yildirim-Vranckaert^{1*} and Aaron Hyzen^{2*}

¹Centre for IT & IP Law, Faculty of Law and Criminology, KU Leuven, Leuven, Belgium, ²Media, Inequality & Change Center, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, United States

In this article, we propose a new theoretical account of mental autonomy through which policymakers can develop new legal instruments to mitigate the harms caused by propaganda. We argue for a renewed understanding of mental autonomy, informed by relational autonomy, highlighting its interdependent nature, shaped by technological mediation and social structures in the digital public sphere. We begin by defining propaganda and discussing its potential to inflict harm through the transformational forces of datafication, algorithmization, and platformization. A historical review of legal approaches to propaganda reveals critical gaps in existing frameworks, which continue to rely on outdated perceptions of autonomy that assume the mind is largely immune to external influences. To address these inadequacies, we build upon the novel concept of epistemic welfare—societal structures and conditions to ensure epistemic agency—and extend it to mental autonomy, which we argue is a necessary precursor to such welfare. Finally, while recognizing the challenges of implementing legal protections against propaganda, we advocate for a governance approach that balances protection and freedom within the broader notions of free thought and expression.

KEYWORDS

law, propaganda, mental autonomy, computational propaganda, epistemic welfare, public sphere, freedom of thought, mental harm

Introduction

This paper redefines mental autonomy through a relational lens, specifically focusing on technological mediation, and situates it within the framework of epistemic welfare by examining the social context and harms of propaganda. To this end, we lay the theoretical foundation for developing more effective and comprehensive legal safeguards from propaganda. Recent years have seen the role of the propagandist revitalized and democratized (Wanless and Berk, 2017). In a world where political and ideological polarization have collided with the digital transformation of the public sphere (Latzer, 2022; Splichal, 2022; Habermas, 2022) and the so-called Web 3.0, characterized by the toolkit of computational propaganda, e.g., automated bots, algorithms, deepfakes, and generative AI (Ghosh and Scott, 2018; DiResta, 2018; Hyzen, 2023), we argue it is necessary to rethink notions of mental autonomy. These conditions allow for unprecedented production and distribution of disinformation and, particularly, propaganda, conceived of here as “a sustained campaign of communication to enforce ideological goals, manage opinion and codify loyalties” (Hyzen, 2021 p. 3482). Efforts to enact legal protections against propaganda have been tempered by the fear that such measures would amount to censorship or give even more power to digital platforms (Helberger, 2020). Beyond direct calls

to war or violence, the current legal and policy landscape presents significant challenges and shortcomings in combating the harms of disinformation, conspiracy theories, and highly personalized, targeted content spread by propaganda, as exemplified during the COVID-19 pandemic, QAnon movement, and the US Jan 6th riots. These instances demonstrate the potential of propaganda to inflict both individual and societal harm—not only to physical capacities, but also, critically, to mental autonomy—defined by us as the capacity of individuals to govern their cognitive processes free from illegitimate external influence.

Consequently, apart from the challenges of restricting propaganda due to concerns of freedom of expression, existing legal frameworks also present a challenge for protecting mental autonomy, which finds its basis in their implied view of the mind and its relation to external influences. These conceptions, influenced by traditional autonomy theories such as Cartesian Dualism and Kantian Mind Mediation, view the mind as a sacred, inherently autonomous, and largely untouchable entity (Descartes, 1641; Kant, 1781/1787; Yildirim-Vranckaert, 2023; Boire, 2000; De Jong, 2000). This outdated perspective limits the effectiveness of current laws in addressing the nuanced and pervasive nature of propaganda practices that systematically undermines mental autonomy through technological means. International human rights instruments, including Article 18 of the Universal Declaration of Human Rights (UDHR), its enforceable counterpart Article 18 of the International Covenant on Civil and Political Rights (ICCPR), and Article 9 of the European Convention on Human Rights (ECHR), explicitly protect these mental processes under the right to freedom of thought. However, current interpretations of these instruments, while illustrative of shared normative commitments across international, regional, and domestic contexts, fail to adequately account for the subtle manipulations enabled by modern technologies, leaving individuals' minds vulnerable to such manipulation without adequate legal recourse.

To bridge this gap, this contribution combines insights from legal and communication studies to adopt the notion of *epistemic welfare* (Hyzen et al., 2025) as a framework centered on fostering the necessary conditions and capabilities for individuals to exercise their *epistemic agency*—defined as individuals' control over how their knowledge, opinions and beliefs are formed, revised, and discussed with others (Coeckelbergh, 2023 p. 1342). The concept of epistemic welfare was first introduced from a legal perspective (Majcher, 2023), conceived of as directly corresponding “to individuals' ‘right to know’, and receive trustworthy, independent and varied information” (p. 3). Continuing, “digital technologies [can] empower citizens... co-existing harmful phenomena,” such as disinformation, have increased in “form and magnitude” (p. 3). Expanding on Majcher's legal conception and incorporating research from communication studies, welfare studies, social epistemology and governance, we follow a more expansive conception of epistemic welfare as “creating and maintaining the conditions and capabilities for individuals' epistemic agency in the public sphere” (Hyzen et al., 2025). Importantly, social epistemology (Goldman, 1987; Godler et al., 2019) stresses and illuminates the social organization of knowledge pursuit and dissemination, i.e., how knowledge is communicated in society, and maintains a normative goal to optimize organizations, institutions and practices that lead individuals to attain epistemically valuable states. As such, we see epistemic welfare as a comprehensive framework to evaluate knowledge producing institutions, media systems and to make legal and governance recommendations. By promoting a

well-functioning media and information ecosystem, epistemic welfare aims to prevent and protect individuals from mental harms inflicted by manipulative practices, and importantly propaganda.

However, meaningful mental autonomy is a foundational prerequisite for achieving epistemic welfare in the public sphere. Considering the outdated perception of the mind within current legal frameworks, in order to foster epistemic welfare, the complexities of mental harm in the digital age should be addressed better by a redefinition of mental autonomy fit for the digital age. Through a relational autonomy perspective, we propose a nuanced redefinition of mental autonomy that recognizes the reciprocal relationship between individuals and technological systems. This approach recognizes the dynamic and interactive nature of mental autonomy, viewing it as a capacity cultivated through technological mediation and socio-cultural contexts, rather than as an isolated or inherent trait.

The epistemic welfare framework, with the redefined concept of mental autonomy, offers a synergistic methodology to address the limitations of existing legal frameworks, particularly concerning fundamental rights. Epistemic welfare establishes the conditions necessary for individuals to exercise their epistemic agency, while the redefined mental autonomy ensures that these conditions are meaningfully realized within a relational and dynamic context. This interdisciplinary approach is not only a step towards bridging existing gaps in legal protections but also provides a more nuanced understanding of how mental harm can manifest in the digital information age. We begin by defining propaganda and examine how the digital transformation of the public sphere has amplified its capacity to inflict harm. Next, we explore the inadequacies of existing legal frameworks in addressing these challenges, review traditional perspective on mental autonomy, and propose a mediation approach grounded in a relational view of technology. Finally, we discuss the implications of this framework for developing more comprehensive legal safeguards, both on redress and remediation and prevention and proactive layers, that promote epistemic welfare in a society marked by increasing interconnectedness and pervasive manipulation.

Propaganda and the digital public sphere

Propaganda defined

Propaganda is an elusive concept. Is it an unpopular or controversial idea? An absurd expression? Lies, falsity, disinformation, or misinformation? Propaganda studies have a long history of theorizations that debate a firm definition, however an aggregated consensus is that propaganda is disseminated information designed and spread to influence public opinion or behavior (Lasswell, 1927; Hyzen, 2021). Propaganda is a term used in a variety of ways and contexts. Propaganda is often conflated with mis- and disinformation, defined as mistakenly false information and intentionally false information, respectively, (Wardle and Derakhshan, 2018; Baines and Elliott, 2020), or used alongside terms like fake news and alternate facts which lack refined definition. Mis/disinformation, by definition, undermines individuals' epistemic agency and is often used as a form of propaganda. However, propaganda can also selectively incorporate true information for intentional manipulation (Hyzen, 2023). Following our definition above and Lukes (2005) conception of power, we “situate propaganda as a

tangible expression of ideology: a mode of communication to spread ideas, achieve ideological goals, and exercise power” (Hyzen, 2023, p.52). Propaganda campaigns are a technique of social control to intentionally mold belief and opinion by spreading, repeating and elaborating ideological views through communication for a desired outcome. As such, we focus on propaganda whose defining characteristics are manipulative, persuasive expressions of power (Bakir et al., 2019; Benkler et al., 2018; Hyzen, 2023), that, when permeating the public sphere unchecked, undermine agency and autonomy, potentially leading to harms.

The digital transformation and computational propaganda

The digital transformation of the public sphere has increasingly captured the attention of scholars (Seeliger and Sevnigani, 2022; Staab and Thiel, 2022), coupled with the fact that the “vast majority of political speech acts [occur] over digital platforms governed by terms-of-service agreements” (Woolley and Howard, 2016, p. 4882), raises serious concerns for democracy and individuals’ ability to exercise agency and mental autonomy without protections. Latzer (2022) argues this transformation of the public sphere is driven by the trinity of “datafication,” or big data, a new asset class and revenue stream, “algorithmization,” or the automated selection processes that “assigns relevance to this data in order to extract economic, social, and political capital from it,” and “platformization” which “restructures markets and business models” towards commercialization and neoliberal logics, creating “organizational forms” that entrench and enhance datafication and algorithmization (p. 4). Splichal (2022) argues this transformation of communication, particularly datafication, enables networked platforms “to systematically... influence users’ online communication and even offline behavior” (p. 2), including “influencing and manipulating opinions at a highly personalized level” (p. 5). Computational propaganda, defined as “the assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion” (Woolley and Howard, 2016, p. 4883), allows propagandists to operate in the space created by the aforementioned trinity. The affordances of the emerging Web 3.0, machine learning, and generative AI, deepfakes have proliferated sophisticated dissemination capabilities for targeting individuals, and near infinite repetition of message, to any organization or individual, what Ghosh and Scott (2018) call “precision propaganda.” This democratizes the role of the propagandist, alongside traditional practitioners, in the digital public sphere (Wanless and Berk, 2017; Hyzen, 2023). Driven by neoliberal logic and a lack of any oversight beyond self-governance, platforms remain unmotivated to curtail potential harms. Meta, for example, rather than strengthening its content moderation, has removed its fact-checkers entirely and announced it will favor even less interference in content across its platforms (Thompson, 2025).

Unbridled propaganda in the digital public sphere is a key threat to epistemic agency. As an expression of power to mold and manage public opinion, propaganda is explicitly produced to distort justified beliefs, undermine the accuracy and integrity of information in favor of the propagandist’s desired outcomes. Computational and precision propaganda, characterized by automated tools, near infinite digital repetition, and widespread dissemination (Woolley and Howard, 2016; Hyzen, 2023), combined with the participatory nature of social media

(Wanless and Berk, 2017), have been shown to pollute and disrupt the communication of verified information and knowledge in digital spaces (Benkler et al., 2018). Propaganda can reproduce itself on social media and digital platforms through audience participation, including interactions between authentic users or programmed chatbots continuously (re)posting content or responding to comment threads (Wanless and Berk, 2017; Hyzen, 2023). Here, propagandists increasingly leverage new media and digital popular culture, e.g., meme and remix culture, to act as “ideological intermediaries” to achieve their goals (Hyzen and Van den Bulck, 2021, p. 180). Datafication, in particular, allows propagandists to identify vulnerable groups (Splichal, 2022) and precisely target said individuals to proliferate content in the digital public sphere cheaply and easily (Ghosh and Scott, 2018). In an ecosystem where propaganda, often in the form of disinformation, thrives, citizens’ ability to exercise epistemic agency and mental autonomy is inevitably compromised, we maintain epistemic welfare is a well-positioned framework to inform solutions.

Propaganda and legacy of traditional mental autonomy in law

When to regulate propaganda

Building on Hyzen (2021) definition, it remains crucial to delineate when propaganda becomes problematic. Regardless of its intentional nature, whether its content is based on false or accurate information, or technological architecture, in a legal sense, propaganda is still a dissemination of an idea or expression. Under many legal systems, both such dissemination and expression are protected under the right to freedom of expression and access to information. People cannot be denied their right to “offend, shock or disturb” (*Handyside v. 1976*, The United Kingdom, no. 5493/72, ECtHR, 1976 §49). This protection exists because what one holds sacred, either in thought or expression, might sound “absurd or anathema to another” (*Skugar and Others v. Russia (dec.)*, no. 40010/04, ECtHR, 2009). Consequently, we safeguard the “freedom for the thought we hate” (*United States v. Schwimmer*, 279 U.S. 644, 1929, 655).

Such protection, however, is not absolute. Propaganda, including its content and means of dissemination, cannot be subjected to a blanket ban; but it can be regulated when such regulation is “provided by law” and “necessary” to protect a legitimate aim. A legitimate aim might arise, for instance, when propaganda interferes with the “rights of others” or poses a threat to “public order.”¹ Thus, there are two possibilities for restriction without creating an undue chilling effect or suppressing access to information: when propaganda harms or controls another’s rights or interests, causing individual harm, and when “one or more interests of society are wrongfully thwarted,”

1 See the three-part test under Article 19 of the International Covenant on Civil and Political Rights, which only allows for the right to freedom of expression to be legitimately subject to certain restrictions if such a restriction is justified: (i) the restriction must be provided for in law; (ii) it must pursue a legitimate aim; and (iii) it must be necessary to protect a legitimate aim. Such legitimate aims are: (a) for respect of the rights or reputation of others; (b) for the protection of national security or of public order, or of public health or morals.

causing societal harm (Trenchard and Gordon, *Cato's Letters*, No. 15, 1721; Smuha, 2021, p. 5). Accordingly, while we believe that propaganda can have destructive effects on a societal level when left unchecked, this contribution focuses on assessing its impact at the individual level, which we recognize ultimately contributes to collective and societal harm.

Using a bottom-up approach, one avenue through which problematic propaganda can affect individuals is by undermining their mental autonomy, thereby causing mental harm² (Yildirim-Vranckaert, 2023). Given mental autonomy is such an “important facet of individual's existence” (Evans v. The United Kingdom [GC], no. 6339/05, ECtHR, 2007, §77; Hämäläinen v. Finland [GC], no. 37359/09, ECtHR, 2014, §§42, 67; O'Callaghan and Shiner, 2021; Yildirim-Vranckaert, 2025, forthcoming), its interference raises profound concerns regarding the regulation of problematic propaganda practices. Consequently, it is crucial to examine existing legal frameworks to determine whether they adequately address—or fail to address—such harm.

Outdated view of the mind in legal frameworks

When examining how legal frameworks address mental harm, two critical observations emerge. First, these frameworks are insufficient in meeting the challenges posed by the mediated digital public sphere, specifically the computational and precision propaganda disseminated within it. Second, this inadequacy stems not from the laws themselves but from their foundational design and interpretation, which are grounded in outdated conceptions of the mind.

To elucidate this point, it is essential to define these outdated conceptions. Two influential philosophical perspectives are particularly relevant: Cartesian Dualism and Kantian Mind Mediation. René Descartes posited that the mind is a distinct, self-aware entity—a “thinking thing”—completely separate from the body and the material world (Descartes, 1641). Cartesian Dualism suggests that while the mind interacts with external stimuli, its core essence remains autonomous and impervious to external manipulations, as illustrated by the “evil genius” thought experiment (Meditation I-II). In contrast, Immanuel Kant emphasized the mind's active role in shaping reality by organizing sensory input through innate (*a priori*) categories such as space and time (Kant, 1781/1787). Kantian Mind Mediation asserts that although the mind structures our experiences, it maintains autonomy by limiting knowledge to *phenomena* (the world as it appears to us) rather than *noumena* (the world as it is in itself) (Kant, 1781/1787), and extends this autonomy to moral reasoning, where *rational agents* adhere to self-imposed moral laws, preserving the

mind's independence from external pressures or influences (Kant, 1785).

These philosophies collectively conceptualize the mind as an autonomous mediator, self-contained and resistant to external influences. This conceptualization has profoundly influenced legal and philosophical understandings of mental autonomy, grounding legal instruments such as human rights laws and domestic constitutional provisions in these outdated views.

For instance, international human rights frameworks—including Article 18 UDHR, Article 18 ICCPR, and Article 9 ECHR—explicitly protect freedom of thought to safeguard the *forum internum*—“person's inner sanctum (mind)” (U.N. Doc. A/76/380–Special Rap. Shaheed, 2021, ¶2), both from state and private actors. These provisions uphold freedom of thought as an “inviolable” space that is “above the law,” deserving “absolute” protection (U.N. Doc. E/CN.4/SR.14, 1947, §3; U.N. Doc. E/CN.4/AC.1/SR.8, 1947, §12–13; U.N. Doc. E/CN.4/SR.60, 1948, §7; Bossuyt, 1987, p. 355; Yildirim-Vranckaert, 2023; Bublitz, 2025, forthcoming).

However, the scarcity of legal precedents from bodies such as the United Nations Human Rights Committee (UNHRC) and the European Court of Human Rights (ECtHR) indicates that this principle is rarely invoked. This rarity is attributable to the narrow interpretation of the right to freedom of thought, which typically recognizes only extreme practices—such as brainwashing, coercion, indoctrination, ideology conversion systems, the use of force or violence—as violations (Nowak, 2005; Vermeulen and Van Roosmalen, 2018; Kokkinakis v. Greece, no. 14307/88, ECtHR, 1993; Masaev v. Moldova, no. 6303/05, ECtHR, 2009; United Nations Human Rights Committee, 2003). To this day, the impact of more subtle and pervasive manipulative practices capable of causing mental harm has not been *prima facie* adjudicated under the freedom of thought framework.

Similarly, as a domestic example, the United States Constitution's First Amendment indirectly protects freedom of thought (Whitney v. California, 274 U.S. 357, 1927; Griswold, 1965), describing it as “the matrix, the indispensable condition, of nearly every other form of freedom” (Palko v. Connecticut 1937, 327). This protection manifests itself as a negative right, against state action, by treating the mind as an autonomous sphere beyond “all official control” (West Virginia State Board of Education v. Barnette, 1943, 642; Reynolds v. United States, 98 U.S. 145, 1878; Stanley v. Georgia, 394 U.S. 557, 1969; Ashcroft v. Free Speech Coalition 535 U.S. 234, 2002). However, such constitutional protection does not extend to private actors, unlike the positive obligations imposed on states under the ICCPR or ECHR, operating under the assumption that government non-interference suffices to safeguard the mind. Remedies for private interference, such as the tort of Intentional Infliction of Emotional Distress (IIED), require stringent proof that conduct was “extreme” or “outrageous” and that the distress suffered was “severe” (Hyatt v. Trans World Airlines Inc., 1997, 297). These high thresholds reflect a persistent reliance on the outdated view of the mind as largely impervious to subtle external influences and a historical bias towards addressing physical rather than mental harm (Grey, 2011). Consequently, the abovementioned approaches reinforce the notion of the mind as a sacred, untouchable entity, protected by impenetrable barriers against external forces (Yildirim-Vranckaert, 2023; Yildirim-Vranckaert, 2025; De Jong, 2000; Boire, 2000; Bublitz, 2013). The presumption that individuals possess inherent autonomy

² It is important to note that while much scholarly discussion focuses on how manipulative practices can undermine mental autonomy or mental integrity, however it remains a concept without a concrete definition in legal or philosophical terms. Only in recent decades has it begun to receive attention, largely due to growing awareness of the impact of emerging and increasingly intrusive technologies. See, for instance, among others, Blitz and Bublitz (2021), Ligthart and Van de Pol, 2025, Lavazza (2018), and McCarthy-Jones (2019).

capable of resisting external influences—even highly intrusive yet less “extreme” ones—reflects a partially Cartesian dualist and predominantly Kantian Mind Mediation perspective.

From the mind to technology as a mediator

Emerging research demonstrates that computational propaganda, as discussed above, and similar intrusive practices grounded in technological systems, leverage digital footprints to predict behavior—sometimes with greater accuracy than individuals themselves or their loved ones (Kosinski et al., 2013; Youyou et al., 2015; Kosinski, 2021; Ramon et al., 2021). These systems pose a threat of significantly influencing cognitive processes, manipulating behavior, belief formation, and individual decision-making (Tufekci, 2015; Panagopoulos, 2016; Woolley and Howard, 2016; Matz et al., 2017; Ribeiro et al., 2019; Zarouali et al., 2022; Acemoglu et al., 2023; Simchon et al., 2024; Yildirim-Vranckaert, 2023). DiResta (2018) notes that “curatorial algorithms are designed to process simple social signals,” yet they have no underlying ethics or human editors that recognize the harms or consequences of spreading extremist propaganda (p. 20–21). Consequently, it is no longer viable to view the mind as an isolated mediator; external technological forces profoundly shape cognitive processes and influence mental autonomy.

Given this shift, it becomes imperative to move beyond the assumption of human cognition as the primary driver of understanding and interacting with the world. Instead, we must recognize technology as a significant mediator in cognitive processes. This perspective draws on the work of post-phenomenologists Don Ihde and Peter-Paul Verbeek, who have extensively explored how technologies mediate human experience.

Ihde's concept of *multistability* illustrates how technologies are not fixed in function or meaning but take on multiple forms through interaction with users (Ihde, 1990) (e.g., a smartphone can serve as a communication device, a navigation tool, or a source of entertainment, depending on the user's context and intention). By shaping perception and enabling or constraining experiences, technologies become active components within relational contexts (Ihde, 1990, 1995, 2009). Similarly, Verbeek, building on Ihde, argues that technologies “actively mediate” (Verbeek, 2005, p.114) human experience by influencing “the way in which humans have access to their world” (p.119), thereby *transforming* perception, interaction, values, and moral decision-making (Verbeek, 2005, 2011). Importantly, this mediation can both support and undermine mental autonomy. For example, while technologies facilitating “speech abundance” and fostering knowledge sharing can enhance civic participation (Volkov, 1995; Hasen, 2017), they can also be exploited in ways that deceive, manipulate, coerce, and harass, distorting individuals' cognitive processes (Norton, 2018; Cohen, 2016).

Therefore, while recognizing the dual nature of this mediation, encompassing both its beneficial and detrimental effects, these frameworks underscore that technology is not merely a passive tool employed by the mind but an active participant in co-constructing cognitive processes. This underscores the urgent need to redefine mental autonomy, and mental harm, in the digital era in line with this updated understanding; otherwise, the traditional conception of the mind risks yielding narrow interpretations in current legal frameworks, thus hindering the development of robust protections

against the intricate and pervasive methods of cognitive manipulation.

Restructured understanding of mental autonomy and harm to Foster epistemic welfare

As previously identified, while epistemic welfare constitutes the conditions that ensure a healthy and equitable information sphere to enhance epistemic agency, achieving this goal is contingent upon ensuring meaningful mental autonomy for individuals. Mental autonomy serves as a foundational prerequisite to acquire knowledge, beliefs and therefore formulate opinions for the exercise of epistemic agency. Consequently, in Part II, we highlighted the necessity for redefining mental autonomy. This section aims to address this gap by first elaborating on the concept of epistemic welfare, then redefining mental autonomy through the lens of technological mediation and finally demonstrating how this redefined mental autonomy, can inform the interpretation and application of existing legal frameworks—across diverse jurisdictions and particularly within human rights law—better promote epistemic welfare in the era of digital propaganda.

The goal of epistemic welfare is to allow individuals to exercise their epistemic agency, enabling them to, ideally, reach epistemically valuable states and freely from beliefs and opinions. Below, we offer a brief explication of the epistemic welfare framework and then how it can be useful to envisage rights, laws and governance to curtail harmful aspects of certain propaganda. To discuss the usefulness of epistemic welfare as concept, we must first explain the definition we follow. The framework combines conceptions of the *epistemic* and of *welfare*, and is grounded in the fields of epistemology and social epistemology. Epistemology encompasses the study of what constitutes knowledge, what justifies or warrants “true belief,” and to what extent humans can obtain knowledge (Zimmer et al., 2019). The core of social epistemology studies the testimony of knowledge between persons and how these social interactions bare overwhelming on an individual's conception of the world and their knowledge. As such, how beliefs are formed and held under certain conditions and the social “influences exerted by other knowers” (Goldman, 1987, p. 109). While social epistemology has several schools, epistemic welfare is rooted in the contemporary reformist school that follows a veristic, truth-seeking conception, which postulates “standards for valid knowledge claims” (Godler et al., 2019, p. 217), but also critically acknowledges the influence of societal forces on knowledge production and testimony. Importantly, the veristic approach evaluates the institutions and practices with “truth-linked standards, any or all of which can be used” for the appraisal of their procedures and processes (Goldman, 1987, p. 187), in our case the digital platforms and automated tools distributing targeted propaganda. Veristic social epistemology postulates that institutions and knowledge practices that meet such standards will produce and promote epistemically value states for individuals: (i) having true beliefs, (ii) avoiding errors, (iii) having justified beliefs, (iv) having rational beliefs (or partial beliefs), and (v) having knowledge (Goldman, 2011, p. 14). In turn, institutions or social practices which fail to meet standards will undermine individuals' epistemic agency and the goals of epistemic welfare. The notion of “welfare” in

epistemic welfare refers to the conditions and capabilities of a society, including the structural forces that enhance or undermine epistemic agency. Including what “should be organized by the state” in how societies establish circulate knowledge and disseminate information (Hyzen et al., 2025), and what legal protections should be implemented against harm.

Likewise, exercising mental autonomy is a necessary condition and precursor to exercising epistemic agency. As such, epistemic welfare's precepts can equally serve mental autonomy, e.g., to enhance agency in forming and holding justified beliefs or acquiring knowledge, and to minimize power relations in the digital sphere. We argue propaganda, especially in its targeted computational forms, represents one such disruptor of epistemic agency and mental autonomy. As such, novel imaginaries are required to curtail proliferation, including protections for mental autonomy. In the next subsection, we will discuss mental autonomy and its reformulation in detail.

Redefining mental autonomy

Building on the concept of *technological mediation*, we propose a framework for mental autonomy that evaluates the reciprocity between individuals and technologies.³ *Reciprocity*, characterized by mutual dependence and interactive engagement, is operationalized through three dimensions: *power relations*, *openness versus closedness*, and *reversibility*. These dimensions assess whether individuals can engage with, influence, or reshape the technologies that simultaneously shape their cognitive processes.

- *Power relations* examine how technologies structure control dynamics, fostering either balanced interactions or asymmetries that undermine individual autonomy.
- *Openness versus closedness* evaluates whether systems enable meaningful engagement and adaptability (openness) or constrain agency through rigidity and opacity (*closedness*).
- *Reversibility* emphasizes the bidirectional nature of interactions, where technologies influence individuals but also remain subject to their feedback and modification.

This framework offers a conceptual tool for evaluating mental autonomy by identifying the conditions under which it is fostered or undermined. While not prescribing specific implementation measures, it provides a lens for analyzing risks such as manipulation, cognitive distortion, and opacity that compromise individuals' ability to exercise epistemic agency.

By emphasizing balanced power dynamics, openness, and reversibility, the framework establishes the conditions necessary for

epistemic welfare. It facilitates a critical examination of how digital systems shape autonomy and highlights pathways for addressing manipulative practices. For instance, power relations can be assessed by examining whether users have meaningful control over their data. Openness involves evaluating whether platforms enable users to adapt targeting mechanisms or operate with transparency, rather than opacity, that restricts agency. Reversibility considers whether users can challenge and modify algorithmic categorizations or if these processes remain unilateral.

However, when mental autonomy is illegitimately disrupted—particularly when individuals cannot engage openly or reversibly with external forces, including technological systems—*mental harm* arises. Such harm, as conceptualized here, stems from the manipulation of these interactions, resulting in diminished autonomy and a compromised ability to critically engage with information.

From concept to governance: a legal framework for mental harm to promote epistemic welfare

While the conceptual framework above advances our understanding of the mind's relationship with technology, translating it into actionable legal and policy measures is crucial. Without such translation, the framework risks remaining abstract and intangible, failing to address practical challenges or fulfill the conditions necessary for fostering epistemic welfare in the contemporary context of propaganda.

First, it is essential to recognize that mental harm is neither static nor universally measurable. However, this does not preclude its tangibilization within a legal framework. Instead of reinventing the wheel, drawing on established harm frameworks, such as those under the ECHR, provides valuable guidance. Yildirim-Vranckaert (2025) parallels the *threshold of application* found in Article 3 ECHR, which prohibits torture and inhuman or degrading treatment. Like Article 9 ECHR, which protects freedom of thought, Article 3 ECHR is an *absolute* right (Ramirez Sanchez v. France [GC], no. 59450/00, ECtHR, 2006), meaning no explicit limitations or restrictions are permitted. However, this does not imply that every form of ill-treatment automatically qualifies as a violation (Savran v. Denmark [GC], no. 57467/15, ECtHR, 2021). Instead, the ill-treatment must meet a *minimum severity threshold*,⁴ requiring a detailed examination of the facts and circumstances of each case (Ireland v. The United Kingdom, no. 5310/71, ECtHR, 1978; Khlaifia and others v. Italy [GC], no. 16483/12, ECtHR, 2016; Bouyid v. Belgium [GC], no. 23380/09, ECtHR, 2015).

Considering that both Articles 3 and 9 ECHR aim to protect individuals' “fundamental” *dignity* and *integrity* (Soering v. The United Kingdom, no. 14038/88, ECtHR, 1989, §88; Kokkinakis v.

³ This is a novel framework developed by us, inspired by the elements of Maurice Merleau-Ponty's principle of reciprocity or reversibility, which emphasizes the mutual and dynamic relationship between subjects and objects. Additionally, it synthesizes Don Ihde and Peter-Paul Verbeek's technological mediation theories, which build upon phenomenological insights such as Merleau-Ponty's. For a comprehensive understanding of Merleau-Ponty's concept of reciprocity between objects and subjects, see Merleau-Ponty (1962).

⁴ The ECtHR has established that not all forms of ill-treatment automatically constitute a violation of Article 3 ECHR. Instead, the treatment must reach a minimum level of severity, assessed through a process of contextualization that considers factors such as its duration, physical or psychological impact, and the victim's circumstances (e.g., age, health, vulnerability). Once this threshold is met, Article 3's absolute prohibition applies, and the ill-treatment is unequivocally considered a violation.

Greece, no. 14307/88, ECtHR, 1993, §33), whether physical or mental, legislation addressing mental harm should adopt a similarly refined and robust approach (Yildirim-Vranckaert, 2025, forthcoming).

Establishing a threshold of application facilitates the adoption of a flexible, case-specific methodology that accounts for all relevant circumstances. This approach allows for a nuanced examination of power relations, openness versus closedness, and reversibility to determine the degree of reciprocity in human-technology interactions. If reciprocity is significantly diminished, such that mental autonomy is undermined to a degree meeting the minimum severity threshold, interference can be presumed.

Furthermore, any comprehensive legal framework must address two key layers of protection against mental harm: (1) redress and remediation mechanisms, and (2) prevention and proactive measures. These layers ensure both a response to existing harm and safeguards against future violations, fostering an environment where meaningful mental autonomy is ensured to promote epistemic welfare.

(1) Redress and remediation mechanisms

Addressing mental harm through redress and remediation presents significant challenges. For instance, take illegitimate manipulation via computational propaganda as an example, one immediate challenge lies in the difficulty of producing concrete evidence or proof of harm and establishing causation (Bublitz, 2014), particularly given the covert and subtle nature of emerging manipulative propaganda practices. These policies must therefore account for this difficulty; they should rather consider all the circumstances (and context) surrounding claims of mental harm. Thus, includes assessing the interference itself, the characteristics of the individual affected, and the environment in which the harm occurred (Yildirim-Vranckaert, 2025, forthcoming). A rigid, one-size-fits-all approach would fail to address the complexity of mental harm and its often-invisible effects.

(2) Prevention and proactive measures

In an optimal ecosystem, prevention should take precedence over ex-post remediation, averting harm before it occurs. A prevention-oriented framework—grounded in the redefined understanding of mental autonomy and mental harm—prioritizes the establishment of supportive environments that enhance meaningful autonomy. This approach does not eliminate the nuanced approach discussed above and the need for contextual analysis; rather, it broadens the scope to anticipate potential harms through systemic safeguards and proactive measures.

Given that mental autonomy is deeply intertwined with one's environment, the environment itself must facilitate individuals' full and meaningful exercise of autonomy. The responsibility for creating such an environment primarily rests with policymakers, whose duties extend beyond merely granting rights and expecting individuals to exercise them in a vacuum. Instead, policymakers must provide the necessary safeguards, guardrails, and foundational support systems to ensure these rights can be effectively realized.

Aligned with this responsibility, we propose a two-dimensional conception of liberty, encompassing the most basic negative liberty (freedom from interference) and the crucial positive liberty (freedom to self-determination and self-realization) (Berlin, 1969; Heyman, 1992). While protecting individuals' minds from governmental

interference fulfills a baseline level of providing conditions, a more comprehensive framework demands proactive steps—positive obligations—to enable meaningful engagement.

This necessity is accentuated by the fact that most digital power asymmetries today are not solely the result of governmental action but are primarily created by private actors or hybrid public-private entities, thereby complicating the assessment of interference (Cohen, 2019; De Gregorio, 2022; Balkin, 2018). This poses unique challenges in jurisdictions emphasizing negative rights, such as the US, where proactive obligations may be limited or non-existent. Without such obligations, achieving the necessary prerequisites for exercising and realizing mental autonomy remains incomplete, particularly in contexts where individual agency is subtly but systematically undermined by manipulative practices.

Balancing freedom and protection

It is, however, crucial to avoid overreach in the name of protection. Preventive and proactive measures must not cross into overbearing paternalism or excessive control over individuals' cognitive freedom. Once a state has fulfilled its obligation to protect, prevent, and provide; individuals retain the freedom to make their own choices—even if those choices may result in (mental) harm. For instance, individuals may choose to engage with propaganda, weaponizing manipulative technologies. Any further state interference beyond these foundational duties would violate negative rights (and obligations) and infringe on personal freedoms.

Striking this balance requires careful calibration. The goal is to create an enabling environment where mental autonomy is protected and nurtured without undermining individual freedom under the guise of protection. A robust mental harm framework must acknowledge this dynamic interaction between mental autonomy, external forces, and the preventive and remedial roles of governance, ensuring that protection and freedom coexist to foster epistemic welfare.

Conclusion

This contribution has argued to redefine mental autonomy and frame it within the concept of epistemic welfare to address the challenges of propaganda in the digital public sphere. Though we follow a broader, more robust concept of epistemic welfare, nevertheless we return and connect it to Majcher's (2023) original legal conception and the active role the law can play to empower citizens' use of digital technologies in the public sphere. We maintain that the transformations of the digital public sphere, largely through datafication, algorithmization and platformization, has profoundly changed the dynamics and management of public opinion (Splichal, 2022) and has given novel avenues for propagandists to precisely targeted vulnerable individuals. By critically examining the outdated conceptions of the mind—as self-contained and inherently autonomous—embedded in current legal frameworks, we have demonstrated the insufficiency of existing safeguards in tackling the subtle and pervasive harms enabled by contemporary propaganda, particularly its computational and precision forms.

The redefined concept of mental autonomy emphasizes its relational and mediated nature, shaped by dynamic interactions

between individuals, technology, and socio-cultural environments. This concept, paired with the epistemic welfare framework, offers a dual approach: mental autonomy provides a lens to identify and address cognitive interferences, while epistemic welfare establishes the broader societal conditions needed for individuals to exercise their epistemic agency. Together, these concepts offer a holistic model for evaluating and mitigating the impacts of propaganda while fostering environments that support the unfettered exercise of cognitive processes.

While this research lays the theoretical groundwork for a reinterpreted understanding of mental autonomy within the context of the right to freedom of thought, further work is required to operationalize this framework at regulatory and policy levels. Advancing this interpretation allows for the adjudication of subtle interferences with cognitive processes, creating both the impetus and legitimacy for regulatory measures that acknowledge the relational nature of the mind in a technologically mediated world. Such measures, while universally normative in scope, can provide guidance across diverse jurisdictional contexts.

Ultimately, this contribution aspires to promote a more equitable and resilient digital public sphere—one where individuals and communities can exercise their mental autonomy and epistemic agency free from undue interference.

Data availability statement

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

Author contributions

EY: Conceptualization, Formal Analysis, Methodology, Resources, Writing – original draft, Writing – review & editing. AH: Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing.

References

- Acemoglu, D., Makhdomi, A., Malekian, A., and Ozdaglar, A. E.. (2023) A model of behavioral manipulation, NBER Working Paper No. w31872.
- Ashcroft v. Free Speech Coalition (2002). Supreme Court of the United States, 535 U.S. 234.
- Baines, D., and Elliott, R. J. (2020). Defining misinformation, disinformation and malinformation: an urgent need for clarity during the COVID-19 infodemic. *Discuss. Pap.* 20, 20–06.
- Balkin, J. M. (2018). Free speech in the algorithmic society: big data, private governance, and new school speech regulation. *U.C. Davis Law Rev.* 51, 1149–1187. doi: 10.2139/ssrn.3038939
- Bakir, V., Herring, E., Miller, D., and Robinson, P., (2019). Organized Persuasive Communication: A new conceptual framework for research on public relations, propaganda and promotional culture. *Critical Sociology*, 45, pp.311–328.
- Benkler, Y., Faris, R., and Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford: Oxford University Press.
- Berlin, I. (1969). *Four essays on liberty*. Oxford: Oxford University Press.
- Blitz, M. J., and Bublit, J. C. (2021). *The law and ethics of freedom of thought*, vol. 1: Neuroscience, autonomy, and individual rights. Cham, Switzerland: Palgrave Macmillan/Springer Nature.
- Boire, R. G. (2000). On cognitive liberty. *J. Cogn. Liberties*. 1, 7–13.
- Bossuyt, M. J. (1987). Guide to the "Travaux Préparatoires" of the international covenant on civil and political rights. Dordrecht: Martinus Nijhoff.
- Bouyid v. Belgium [GC], no. 23380/09, ECtHR, (2015)
- Bublit, J. C. (2013) 'My mind is mine!? Cognitive liberty as a legal concept', in E. Hildt and A. Franke (eds.) *Cognitive enhancement: Trends in augmentation of human performance*, vol. 1. Dordrecht: Springer, pp. 241–259
- Bublit, J. C. (2014). "Cognitive liberty or the international human right to freedom of thought" in *Handbook of neuroethics*. eds. J. Clausen and N. Levy (Dordrecht: Springer), 1309–1313.
- Bublit, J. C. (2025). "The mind and conscience are the person's most sacred possessions": the origins of freedom of thought in the universal declaration of human rights and the international covenant on civil and political rights" in P. O'Callaghan and B. Shiner (eds.) *The Cambridge handbook of the right to freedom of thought*. Cambridge: Cambridge University Press. 15–30. doi: 10.1017/9781009539616.004
- Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. *AI Ethics* 3, 1341–1350. doi: 10.1007/s43681-022-00239-4

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the Research Foundation -Flanders (FWO) under the Weave interdisciplinary research project ALGEPI (Understanding Algorithmic Gatekeepers to Promote Epistemic Welfare), Grant Agreement No. G098223N. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

The authors would like to thank Hilde Van den Bulck and Laurens Naudts for their careful reading of the manuscript, along with helpful comments and suggestions.

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.

Generative AI statement

The authors 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.

- Cohen, J. E. (2016). The regulatory state in the information age. *Theor. Inq. Law* 17, 369–414. doi: 10.1515/til-2016-0015
- Cohen, J. E. (2019). *Between truth and power: The legal constructions of informational capitalism*. New York: Oxford University Press.
- De Gregorio, G. (2022). “The law of the platforms” in *Digital constitutionalism in Europe: Reframing rights and powers in the algorithmic society* (Cambridge: Cambridge University Press), 80–122.
- De Jong, C. D. (2000). *Freedom of thought, conscience and religion or belief in the United Nations (1946–1992)*. Antwerp: Intersentia.
- Descartes, R. (1641) *Meditations on first philosophy*. Translated by J. Cottingham. Cambridge University Press, Cambridge.
- DiResta, R. (2018). Computational propaganda: if you make it trend, you make it true. *Yale Rev.* 106, 12–29. doi: 10.1111/yrev.13402
- Evans v. The United Kingdom [GC], no. 6339/05, ECtHR, (2007)
- Ghosh, D., and Scott, B. (2018) Digital deceit: the technologies behind precision propaganda on the internet. Available online at: <https://scholar.harvard.edu/files/dipayan/files/digital-deceit-final-v3.pdf> (accessed July 20, 2024).
- Godler, Y., Reich, Z., and Miller, B. (2019). Social epistemology as a new paradigm for journalism and media studies. *New Media Soc.* 22, 213–229. doi: 10.1177/1461444819856922
- Goldman, A. I. (1987). Foundations of social epistemics. *Synthese* 73, 109–144. doi: 10.1007/BF00485444
- Grey, B. J. (2011). “Neuroscience and emotional harm in tort law: rethinking the American approach to free-standing emotional distress claims” in *Law and neuroscience: Volume 13: Current legal issues*. ed. M. Freeman (Oxford: Oxford University Press).
- Griswold, V. Connecticut, 381 U.S. 479 (1965).
- Goldman, A. I. (2011). A guide to social epistemology. In A. I. Goldman and D. Whitcomb (Eds.), *Social epistemology: Essential readings*. 11–37. Oxford: Oxford University Press.
- Habermas, J. (2022). Reflections and hypotheses on a further structural transformation of the political public sphere. *Theory Cult. Soc.* 39, 145–171. doi: 10.1177/0263276422112341
- Hämäläinen v. Finland [GC], no. 37359/09, ECtHR, (2014)
- Handyside v. (1976). The United Kingdom. European Court of Human Rights, Application No. 5493/72.
- Hasen, R. L. (2017). Cheap speech and what it has done (to American democracy). *First Amend. Law Rev.* 16, 200–231.
- Helberger, N. (2020). The political power of platforms: how current attempts to regulate misinformation amplify opinion power. *Dig. J.* 8, 842–854. doi: 10.1080/21670811.2020.1773888
- Heyman, S. J. (1992). Positive and negative liberty. *Chicago-Kent L. Rev.* 68, 81–117.
- Hyatt, V. Trans world airlines, inc., 943 s.w.2d 292 (mo. ct. app. 1997).
- Hyzen, A. (2021). Revisiting the theoretical foundations of propaganda. *Int. J. Commun.* 15, 3479–3496.
- Hyzen, A. (2023). Propaganda and the web 3.0: truth and ideology in the digital age. *Nord. J. media Stud.* 5, 49–67. doi: 10.2478/njms-2023-0004
- Hyzen, A., and Van den Bulck, H. (2021). Conspiracies, ideological entrepreneurs, and digital popular culture. *Media Commun.* 9, 179–188. doi: 10.17645/mac.v9i3.4092
- Hyzen, A., and Van den Bulck, H. (2024). “Putin’s war of choice”: US propaganda and the Russia–Ukraine invasion. *J. Media* 5, 233–254. doi: 10.3390/journalmedia5010016
- Hyzen, A., Van den Bulck, H., Puppis, M., Kulig, M., and Paulussen, S. (2025). Epistemic welfare, algorithmic recommender systems and the public sphere in the digital era: creating conditions and capabilities for epistemic agency and a way out of the epistemic crisis. *Commun. Theory*. doi: 10.1093/ct/ctaf018
- Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Bloomington: Indiana University Press.
- Ihde, D. (1995). *Postphenomenology: Essays in the postmodern context*. Evanston: Northwestern University Press.
- Ihde, D. (2009). *Postphenomenology and technoscience: The Peking University lectures*. Albany: SUNY Press.
- Ireland v. The United Kingdom, no. 5310/71, ECtHR, (1978)
- Kant, I. (1781/1787). *Critique of pure reason*. Translated by P. Guyer and A. W. Wood, 1998. Cambridge: Cambridge University Press.
- Kant, I. (1785) *Groundwork for the metaphysic of morals*. Translated by J. Bennett, Early Modern Texts, 2017. <https://www.earlymoderntexts.com/assets/pdfs/kant1785.pdf>
- Khlaifia and Others v. Italy [GC], no. 16483/12, ECtHR, (2016)
- Kokkinakis v. Greece, no. 14307/88, ECtHR, (1993)
- Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Sci. Rep.* 11:100. doi: 10.1038/s41598-020-79310-1
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. USA* 110, 5802–5805. doi: 10.1073/pnas.1218772110
- Lasswell, H. (1927). The theory of political propaganda. *Am. Polit. Sci. Rev.* 21, 627–631. doi: 10.2307/1945515
- Latzer, M. (2022). The digital trinity—controllable human evolution—implicit everyday religion: characteristics of the socio-technical transformation of digitalization. *KZfSS Koln. Z. Soziol. Sozialpsychol.* 74, 331–354. doi: 10.1007/s11577-022-00841-8
- Lavazza, A. (2018). Freedom of thought and mental integrity: the moral requirements for any neural prosthesis. *Front. Neurosci.* 12:82. doi: 10.3389/fnins.2018.00082
- Lighthart, S., and Van de Pol, N. (2025) ‘Freedom of thought: absolute protection of mental privacy and mental integrity? Considering the case of neurotechnology in criminal justice’, in P. O’Callaghan and B. Shiner (eds.) *The Cambridge handbook of the right to freedom of thought*. Cambridge: Cambridge University Press, 350–362.
- Lukes, S. (2005). *Power: A radical view*. London, UK: MacMillan/Red Globe.
- Majcher, K. (2023). *Coherence between data protection and competition law in digital markets*. Oxford, United Kingdom Oxford University Press.
- Masaev v. Moldova, no. 6303/05, ECtHR, (2009).
- Matz, S. C., Kosinski, M., Nave, G., and Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl. Acad. Sci. USA* 114, 12714–12719. doi: 10.1073/pnas.1710966114
- McCarthy-Jones, S. (2019). The autonomous mind: the right to freedom of thought in the twenty-first century. *Front. Artif. Intell.* 2:19. doi: 10.3389/frai.2019.00019
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. Translated by C. Smith. 1st Edn. London: Routledge.
- Norton, H. (2018) ‘(At least) thirteen ways of looking at election lies’, *Okla. Law Rev.* 71, 117–139.
- Nowak, M. (2005). U.N. Covenant on civil and political rights. 2nd rev. Edn. Kehl: Engel.
- O’Callaghan, P., and Shiner, B. (2021). The right to freedom of thought in the European convention on human rights. *Eur. J. Comp. Law Gov.* 8, 112–145. doi: 10.1163/22134514-bja10016
- Palko v. Connecticut, 302 U.S. 319 (1937).
- Panagopoulos, C. (2016). All about that base: changing campaign strategies in U.S. presidential elections. *Party Polit.* 22, 179–190. doi: 10.1177/1354068815605676
- Ramirez Sanchez v. France [GC], no. 59450/00, ECtHR, (2006)
- Ramon, Y., Farrokhnia, R. A., Matz, S. C., and Martens, D. (2021). Explainable AI for psychological profiling from behavioral data: an application to big five personality predictions from financial transaction records. *Information* 12:518. doi: 10.3390/info12120518
- Reynolds v. United States, 98 U.S. 145 (1878).
- Ribeiro, F. N., Saha, K., Babaei, M., Henrique, L., Messias, J., Benevenuto, F., et al. (2019). “On microtargeting socially divisive ads: a case study of Russia-linked ad campaigns on Facebook” in *Proceedings of the conference on fairness, accountability, and transparency (FAT ’19)* (New York: Association for Computing Machinery), 140–149.
- Savran v. Denmark [GC], no. 57467/15, ECtHR (2021)
- Simchon, A., Edwards, M., and Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus* 3:pgae035. doi: 10.1093/pnasnexus/pgae035
- Shaheed, A. (2021). (Special Rapporteur on Freedom of Religion or Belief), *Promotion and Protection of Human Rights: Human Rights Questions, Including Alternative Approaches for Improving the Effective Enjoyment of Human Rights and Fundamental Freedoms*, U.N. Doc. A/76/380.
- Skugar and Others v. Russia (dec.), no. 40010/04, ECtHR, (2009)
- Smuha, N. A. (2021). Beyond the individual: governing AI’s societal harm. *Internet Policy Rev.* 10, 1–32. doi: 10.14763/2021.3.1574
- Soering v. The United Kingdom, no. 14038/88, ECtHR, (1989)
- Splichal, S. (2022). *Datafication of public opinion and the public sphere: How extraction replaced expression of opinion*. London, United Kingdom Anthem Press.
- Staab, P., and Thiel, T. (2022). Social media and the digital structural transformation of the public sphere. *Theory Cult. Soc.* 39, 129–143. doi: 10.1177/02632764221103527
- Stanley v. Georgia, 394 U.S. 557 (1969).
- Seeliger, M., and Sevignani, S. (2022). A New Structural Transformation of the Public Sphere? An Introduction. *Theory, Culture and Society*, 39, 3–16. doi: 10.1177/02632764221109439
- Thompson, S. A. (2025). Meta says fact-checkers were the problem. Fact-checkers rule that false. *The New York Times* <https://www.nytimes.com/2025/01/07/business/mark-zuckerberg-meta-fact-check.html> (accessed January 15, 2025).
- Trenchard, J., and Gordon, T. (1721) *Of freedom of speech: that the same is inseparable from public liberty*. *Cato’s Letters*, No. 15, *The London Journal*, 4 February.

- Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency. *Colo. Tech. Law J.* 13, 203–218.
- United Nations Human Rights Committee (2003) Views of the Human Rights Committee under article 5, paragraph 4, of the Optional Protocol to the International Covenant on Civil and Political Rights: seventy-eighth session concerning communication no. 878/1999 (Yong-Joo Kang v. Republic of Korea). UN Doc. CCPR/C/78/D/878/1999, 15 July.
- United States v. Schwimmer, 279 U.S. 644 (1929).
- United Nations Commission on Human Rights (1948) *Third session, summary record of the sixtieth meeting*, UN Doc. E/CN.4/SR.60.
- United Nations Commission on Human Rights, Drafting Committee (1947) *First session, summary record of the eighth meeting*, UN Doc. E/CN.4/AC.1/SR.8.
- United Nations Commission on Human Rights. (1947) *Summary record of the fourteenth meeting*, UN Doc. E/CN.4/SR.14.
- Verbeek, P.-P. (2005). What things do: Philosophical reflections on technology, agency, and design. University Park: Penn State University Press.
- Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. Chicago: University of Chicago Press.
- Vermeulen, B., and Van Roosmalen, M. (2018). “Freedom of thought, conscience and religion” in Theory and practice of the European convention on human rights. eds. P. Van Dijk, F. Van Hoof, A. Van Rijn and L. Zwaak. 5th ed (Antwerp: Intersentia), 735–764.
- Volokh, E. (1995). Cheap speech and what it will do. *Yale Law J.* 104:1805.
- Wanless, A., and Berk, M. (2017). “Participatory propaganda: the engagement of audiences in the spread of persuasive communications” in Social media and social order. eds. D. Herbert and S. Fisher-Høyrem (Berlin: De Gruyter), 111–136.
- Wardle, C., and Derakhshan, H. (2018). Thinking about ‘information disorder’: formats of misinformation, disinformation, and mal-information. *J. Fake News Disinformat.* 43–54.
- West Virginia State Board of Education v. Barnette, 319 U.S. 624 (1943).
- Whitney v. California, 274 U.S. 357 (1927).
- Woolley, S., and Howard, P. (2016). Automation, algorithms, and politics: political communication, computational propaganda, and autonomous agents – introduction. *Int. J. Commun.* 10:9.
- Yildirim-Vranckaert, E. O. (2023). The right to construct yourself and your identity: the current human rights law framework falls short in practice in the face of illegitimate interference to the mind. *Am. J. Law Med.* 49, 267–285. doi: 10.1017/amj.2023.31
- Yildirim-Vranckaert, E. O. (forthcoming 2025) ‘The mind and the law: a contextualized framework for freedom of thought under the European Convention on Human Rights’, in M. Blitz and J. C. Bubltz (eds.) *The law and ethics of freedom of thought: Cognitive liberty and privacy*, Cham, Switzerland: Palgrave Macmillan. 2.
- Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. USA* 112, 1036–1040. doi: 10.1073/pnas.1418680112
- Zarouali, B., Dobber, T., De Pauw, G., and Vreese, C. (2022). Using a personality-profiling algorithm to investigate political microtargeting: assessing the persuasion effects of personality-tailored ads on social media. *Commun. Res.* 49, 1066–1091. doi: 10.1177/0093650220961965
- Zimmer, F., Scheibe, K., Stock, M., and Stock, W. G. (2019). Fake News in Social Media: Bad Algorithms or Biased Users? *Journal of Information Science Theory and Practice*, 7. doi: 10.1633/JISTaP.2019.7.2.4