ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
1. AI is changing the world: for better or for worse? (2024)	Journal of Macromarketing (Article from : SAGE)
2. Educational reform in the era of artificial intelligence (2021)	ACM International Conference Proceeding Series (Article from : Association for Computing Machinery)
3. Embedding AI in society: ethics, policy, governance, and impacts (2023)	AI and Society (Article from : Springer Science and Business Media Deutschland GmbH)
4. Social impact of AI on the organization of higher Education in electrical engineering and in society (2025)	2025 34th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE) (Article from : IEEE)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
5. Taxonomy of generative AI applications for risk assessment (2024)	Proceedings - 2024 IEEE/ACM 3rd International Conference on AI Engineering (Article from :Association for Computing Machinery, Inc)
6. The blended future of automation and AI: examining some long-term Societal and ethical impact features (2023)	Technology in Society (Article from : Elsevier Ltd)
7. The darker side of positive AI attitudes: investigating associations with (Problematic) social media use (2025)	Addictive Behaviors Reports (Article from : Elsevier Ltd)
8. The impact of artificial intelligence: from cognitive costs To global inequality (2025)	European Physical Journal: Special Topics (Article from :Springer Science and Business Media Deutschland GmbH)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
9. The next wave of AI for social impact: Challenges and opportunities (2025)	IEEE Intelligent Systems (Article from :Institute of Electrical and Electronics Engineers Inc.)
10. AI: what do we fear? What do we hope for? Perception of the societal impact of ai in a European transnational cross-sectional study (2024)	Proceedings of the 30th ICE IEEE/ITMC Conference on Engineering, Technology, and Innovation (Article from : Institute of Electrical and Electronics Engineers Inc.

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
1. AI is changing the world: For better or for worse? (2024)	Journal of Macromarketing (Article from : SAGE)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Research Article

Journal of Macromarketing
2024, Vol. 44(4) 870-882

DOI: 10.1177/02761467241254450
journals.sagepub.com/home/jmk


AI is Changing the World: For Better or for Worse?

Dhruv Grewal[1,2,3], Abhijit Guha[4], and Marc Becker[5]

Abstract

The profound impacts of artificial intelligence (AI) will continue to evolve over the next several decades, and many of these impacts will emerge through marketing-related AI applications. Therefore, marketers, public policymakers, firms, researchers, and individual consumers must recognize and understand the benefits that AI offers, as well as the perils that it presents, both now and in the future. A literature review surfaced three themes – that AI will augment and (potentially) replace human intelligence, that AI will evolve into an empathetic and trusted companion, and that AI will create novel tensions. Next, this article outlines three stages of AI development, from an early stage with much promise, to a stage with many benefits, to a stage wherein AI-related tensions emerge. Finally, this article outlines three grand challenges: (1) preserving and growing human capability, (2) protecting societal belonging and human connection, and (3) ensuring equitable sharing of AI benefits. Addressing such challenges, along with related concerns (e.g., privacy, ethics), can enable society to reap the benefits of AI fruitfully and in an equitable manner that truly improves the quality of life.

Keywords

artificial intelligence, public policy, societal challenges

Introduction

Artificial intelligence (AI), which involves programs and algorithms that mimic (intelligent) human behavior (Kopalle et al. 2022; Shankar 2018; Shankar et al. 2021), offers tremendous potential for enhanced societal value. According to the U.S. Centers for Disease Control, AI could improve analyses of images and scans, accelerate new medication development, and improve public health (Rasooly and Khoury 2022). The U.S. Department of Agriculture predicts AI-enabled improvements to agricultural outcomes (Elliott 2020), and the Brookings Institute anticipates that AI will improve education —though perhaps in ways that benefit some groups more than others (Trucano 2023).

In the business domain, AI also is prompting fundamental transformations to processes and customer experiences, including but not restricted to retailing (Guha et al. 2021) and marketing (Huang and Rust 2022). The potential contributions of AI to society are reflected in its monetary valuation, which grew by 58% between 2022 and 2023, from $87 billion to $150 billion (Markets and Markets 2023). Such growth is likely to continue, in terms of both the role and the value of AI (Wong 2023), such that its valuation is predicted to reach $1.3 trillion by 2030 (Markets and Markets 2023). Davenport et al. (2020) argue that marketers have the most to gain from AI; across 400 use cases, they indicate that the greatest potential value of AI relates to domains involving marketing (and sales).

Accordingly, more companies are investing in and implementing AI, on pace with AI's ever-advancing capabilities. In particular, the introduction of ChatGPT established an important milestone: For the first time, highly capable, generative AI became widely available to the public (Weise et al. 2023). As its name suggests, generative AI shifts the focus from prediction to the production of new content, meaning that it can take over many tasks that previously appeared exclusive to human capabilities, such as writing text or creating ads. Such capabilities suggest that a broader human augmentation (and possibly replacement) revolution is on the horizon, with generative AI as the spark (Salesforce 2023).

[1]Toyota Chair in E-Commerce and Electronic Business and Professor of Marketing, Department of Marketing, Babson College, Babson Park, MA 02457, USA
[2]Fractional Professor of Marketing, University of Bath, Bath, UK
[3]Honorary Distinguished Visiting Professor of Retailing and Marketing, Tecnológico de Monterrey, Monterrey, Mexico
[4]Associate Professor of Marketing, Darla Moore School of Business, University of South Carolina, Columbia, USA
[5]PhD Candidate, School of Business and Economics, Maastricht University, Maastricht, The Netherlands

Note: The authors appreciate the feedback of the editor and the two reviewers.

Corresponding Author:
Marc Becker, PhD Candidate, School of Business and Economics, Maastricht University, Maastricht, The Netherlands.
Email: m.becker@maastrichtuniversity.nl

Notwithstanding its tremendous potential for good, AI is also associated with a range of justifiable concerns, including societal challenges (e.g., risks of skill or job losses, weakened institutions), privacy issues (e.g., capacity to extract sensitive information from even anonymized data; Davenport et al. 2020), bias and discrimination potential (e.g., AI trained to identify sexual orientation based on facial features; Wang and Kosinski 2018), and other ethical considerations. In turn, thought leaders such as Ilya Sutskever (Metz 2023) or Elon Musk (Duffy and Maruf 2023) have warned about the potential dangers of AI, and such views also are reflected in recent legislation, such as the European Union's (EU) efforts to define rules and regulations for AI systems depending on their risk potential. For example, generative AI, including ChatGPT, must comply with transparency requirements, and the EU also requires educational AI applications to be registered in a database. These rules also ban certain practices, such as biometrical identification or social scoring (Madiega 2023).

In this article, in which we offer a marketing perspective on ways to address such concerns, we explicitly focus on novel tensions that might not materialize for some years but that need to be addressed now, while we still can. With a wide-ranging discussion of these challenges, structured according to an overarching, macro-level framework, we hope to spur continued, constructive discussions among marketing academics, policymakers, and marketing practitioners. We firmly believe that this early stage of AI deployment precisely constitutes the right moment to make sure we "do AI right from the start" and set a course that will allow society to reap the benefits of AI while minimizing the impacts of the risks it creates. We also explicitly take an optimistic approach to an increasingly complex and world-changing series of likely events.

With a review of research published in top marketing and public policy journals, we identify three broad themes pertaining to the impacts of AI. Among these broad conceptual themes, we highlight three stages of AI that encompass several tensions that we anticipate will arise as AI manifests more fully, with stronger influences on firms, individuals, and societies. Finally, we conclude by detailing three grand challenges that society (and marketers) are likely to face in the future but that must be tackled (or at least considered) in the present: (1) preserving and growing human capability; (2) protecting societal belonging and human connection; and (3) ensuring the equitable sharing of AI's benefits. As its core contribution, this article thus brings to the forefront the role of AI in changing the world, for better or for worse.

Literature Review

To determine how existing studies conceptualize the impact of AI on consumers, firms, and society, we searched nine leading marketing, public policy, and ethics journals (*Journal of Marketing, Journal of Marketing Research, Journal of Consumer Research, Journal of the Academy of Marketing Science, Marketing Science, Journal of Retailing, Journal of Public Policy & Marketing, Journal of Business Ethics,* and *Journal of Macromarketing*) for articles mentioning the terms "AI" or "artificial intelligence" in their title or abstract. This search produced 67 articles, spanning both conceptual and empirical research. We manually reviewed each article to identify insights for deriving an applicable conceptual framework that could clarify the broad impact of AI on society. In addition, we determined whether each study addressed aspects related to consumers, marketing firms, or society at large. Nine representative articles are discussed hereafter and summarized in

Table 1. Relevant Frameworks Published in Top Journals.

Work	Key Arguments	Individual Focus	Firm Focus	Societal Focus	Theme
Davenport et al. (2020)	AI will first support the analysis of numerical data; then be capable of analyzing text, voice, face, and image data; and eventually become a data virtuoso capable of analyzing all kinds of data.		✓		#1
Huang and Rust (2021)	Mechanical, thinking, and feeling AI can support the marketing process during marketing research, marketing strategy, and marketing action steps.		✓		#1
Huang and Rust (2022)	Mechanical, thinking, and feeling AI will first augment and then replace human intelligence.		✓		#1
Liu-Thompkins, Okazaki Shintaro, and Li (2022)	Different components of artificial empathy can be implemented in marketing strategies.	✓			#2
Rodgers and Nguyen (2022)	AI will be able to observe and influence the purchase process, from need recognition to purchase.	✓			#2
Noble and Mende (2023)	Robots and AI will take on roles ranging from strangers to acquaintances, to friends/partners.	✓	✓		#2
Puntoni et al. (2021)	Consumers experience AI-evoked tensions relating to data capture, classification, delegation, and social.	✓			#3
Hermann (2022)	(Societal) tensions provoked by AI will emerge along five ethical principles: beneficence, non-maleficence, justice, explicability, and autonomy.	✓	✓	✓	#3
Bankins and Formosa (2023)	AI will make some jobs more meaningful while making others less meaningful.	✓	✓		#3

Notes: Theme #1: AI will augment and (potentially) replace human intelligence. Theme #2: AI will enable artificially empathetic and trusted companions. Theme #3: AI creates novel tensions.

Table 1. This review also revealed that existing frameworks can be grouped into three main themes, which we detail next, before highlighting the three stages of AI.

Rather than offering a linear review, we use these articles to highlight three main themes that have been the topics of discussion in the AI domain: (1) AI will augment and (potentially) replace human intelligence; (2) AI will enable artificially empathetic and trusted companions; and (3) AI creates novel tensions.

Theme #1: AI Will Augment and (Potentially) Replace Human Intelligence

Multiple frameworks predict that AI will augment or replace human intelligence. Guha et al. (2021) propose that initially (currently the case), narrow AI can execute only a few discrete tasks (i.e., artificial narrow intelligence [ANI]). In this initial stage, AI is best able to add value by augmenting, rather than replacing, human efforts. Following a transition sometime in the future, AI will be able to execute many general and unrelated tasks (i.e., artificial general intelligence [AGI]), including being able to understand unfamiliar, complex, multimodal inputs and devise and execute (novel) solutions. Davenport et al. (2020) provide an example: Initially, a call center might use AI to provide human call center operators with important contextual details (e.g., information about the caller's mood based on an automated voice analysis)—in effect, augmenting their human efforts. Over time, call center management may become more comfortable with the AI technology and allow it to handle simple queries autonomously—in effect, replacing human efforts and reducing the extent to which human operators must take charge of tedious tasks.

Huang and Rust (2021, 2022) identify three types of AI (mechanical, thinking, and feeling) and predict that, in marketing contexts, each type will first augment human intelligence, before (eventually) replacing it. *Mechanical AI* automates repetitive and routine tasks and can help marketers with data collection, segmentation, and standardization. *Thinking AI* is designed to process data, to arrive at new conclusions, so it can facilitate market analyses, targeting, and personalization. *Feeling AI* interacts with humans and can understand human emotions; it can achieve better customer insights, positioning, and relationship building. The preceding call center example aligns with this typology as well. That is, mechanical AI might retrieve information as soon as a call comes in, providing details about the caller's previous orders or preferences to the call center operator, who then can provide better service. Thinking AI then might suggest service solutions, reflecting the customers' demographics or the queries they raise. Finally, feeling AI might recommend personalized responses and tones, contingent on the tone of the caller and the perceived emotionality of the words they use.

Theme #2: AI Will Enable Artificially Empathetic and Trusted Companions

Another set of frameworks anticipates new types of relationships that firms can build with customers, through the use of AI. For example, Pi (short for "personal intelligence") is an AI-enabled companion that seeks to redefine how people interact with technology, by functioning as

> a personal assistant, a trusted confidante, a life coach, and a continuously learning companion, all rolled into one…. What truly sets Pi apart is its high emotional quotient. It's not just an intelligent bot; it's a kind and supportive companion that listens, adapts, and learns, all while providing feedback in a natural, easy-to-understand manner. Pi is at your service if you're facing a tricky situation or need a sounding board (Hughes 2023).

Even if current iterations of AI companions like Pi are far from perfect, they represent a step toward AI as an empathetic and intimate companion that can offer various marketing opportunities.

In this vein, Noble and Mende (2023) propose that marketers can capture value by designing robots to function as strangers, acquaintances, or friends/partners. If they possess empathic capabilities and operate in intimate roles (e.g., as friends), AI-powered robots can anticipate customer needs and recommend better solutions, such as those available from different suppliers. As Rodgers and Nguyen (2022) note, firms also can leverage AI to guide customers throughout the entire purchase process, from need recognition to purchase, though doing so may require explicit ethical boundaries. Beyond marketing guidance, AI-enabled general companions promise other marketing-related opportunities (e.g., providing advice) that also require careful ethical considerations.

In Huang and Rust's (2022) framework, feeling AI constitutes a relevant type. As AI continues to develop feeling, empathy, and emotional intelligence capabilities, a host of opportunities will become available, as highlighted by the Pi example. Liu-Thompkins, Okazaki Shintaro, and Li (2022) also define different elements of what they refer to as artificial empathy and explain how marketers can implement them. Leveraging such empathy and connectedness seems likely to lead to vast business opportunities, across a wide variety of domains and tasks.

Theme #3: AI Creates Novel Tensions

The prior two themes mainly emphasize the benefits of AI applications. In contrast, the third theme highlights the novel tensions that come to the forefront due to the increased adoption of AI. On the one hand, AI can create more meaningful work and increase incomes. On the other hand, it might result in new and boring tasks, increase concerns about capturing data and privacy risks, threaten misclassification or bias, lead to human replacement and alienation, fuel resource inequity, and so on.

As Grewal et al. (2021) detail, AI's value creation capabilities span both business-to-consumer (B2C) and business-to-business (B2B) settings. In B2C settings, AI might create value through enhanced customization, such as when AI leverages individual transaction data, augmented with data from other sources, to help the firm develop and present customized marketing messages to a mass market of customers.

Furthermore, they highlight the efficiencies linked to AI, such as when it enables systems that allow customers to bypass checkout lines in fully automated stores. In B2B settings, AI can augment salespeople's capabilities, such as when AI chatbots provide them with coaching (Luo et al. 2021). But across both contexts, AI also can induce concerns: It might erode trust in B2C settings due to threats to privacy, the potential for bias, or a failure to account for human uniqueness neglect, and it can increase power asymmetry levels in B2B settings, reflecting concerns related to opportunism and fear of manipulation (Grewal et al. 2021).

With regard to work roles, Bankins and Formosa (2023) argue that integrating AI can lead to more meaningful work, which people perceive as having worth, significance, and higher purpose because AI takes over dull tasks. But for other jobs, it might have the opposite effect, "creating new boring tasks, restricting worker autonomy, and unfairly distributing the benefits of AI away from less-skilled workers" (Bankins and Formosa 2023, p. 738).

Focusing on consumers, Puntoni et al. (2021) highlight four other broad tensions: data capture (i.e., being served vs. exploited), classification (i.e., being understood vs. misunderstood), delegation (i.e., being empowered vs. replaced), and social experience (i.e., being connected vs. alienated). More broadly, Hermann (2022) predicts that widespread AI deployment may increase consumers' incomes and generate profits, but the increased consumption such trends likely induce will put additional strain on scarce resources.

Finally, Davenport et al. (2020) specify three key AI-related issues: data privacy, bias, and ethics. Data privacy issues arise because AI requires vast (user) data, which are often stored for longer and contain richer information (e.g., about family members) than users realize (see also Martin and Murphy 2017; Martin, Borah, and Palmatier 2017). Then AI can mine such data to extract deeply personal details, even if the data might have been anonymized. Bias concerns arise because AI models tend to be trained on real-life observations, which mirror daily practices of discrimination or other (human) biases (Davenport et al. 2020; Villasenor 2019). For example, Poole et al. (2021) highlight how biases in AI training data can negatively impact certain groups, resulting in worse customer experiences and even physical dangers. These issues are amplified by the "black box" status of AI (i.e., it is not always clear which factors influence a decision), as well as the difficulty of identifying and excluding bias-inducing factors from decision-making. Therefore, unpacking the black box is important, which underscores the relevance of explainable AI (XAI) (Rai 2020). Although XAI offers benefits, such as greater customer trust and engagement and reduced bias, again, it has downsides, including increased costs. With regard to ethical issues, AI applications can wreak serious damage in the wrong hands. For example, if an application can identify a person's sexual orientation on the basis of their facial features, it could result in persecution by oppressive governments (Wang and Kosinski 2018).

The Three Stages of AI

We combine the three preceding themes into three stages of AI. In so doing, we reflect on predictions that someday, AI might shift from ANI to AGI (Theme #1) and thereby support a wider range of tasks in ways that not only augment but replace human intelligence. In turn, we propose that AI development can be classified into three stages; we provide descriptions, key AI capabilities, and constructive versus destruction potential for each stage in Table 2.

Stage 1: Enhancing Firm Efficiencies

In the first stage (enhancing firm efficiencies), the influences of AI appear in various sectors. For example, mechanical AI (Huang and Rust 2021, 2022) facilitates numerical and some simple non-numerical data analyses (Davenport et al. 2020). In this stage, AI can take over some routinized jobs, such as cleaning up retail aisles or identifying out-of-stock items. In addition, generative (or thinking) AI applications increasingly support and augment human efforts in tasks such as creating social media posts, debugging code, or writing drafts for sales pitches (Guha, Grewal, and Atlas 2024). Due to these contributions, this stage promises many positive effects, including enhanced performance and efficiency, as well as decreased costs. Yet drawbacks of AI (and its related technologies) already have begun to emerge (Theme #3), including concerns related to data privacy, bias, and ethics (Davenport et al. 2020; Poole et al. 2021).

For example, as various technologies add facial recognition capabilities (e.g., doorbell cameras, surveillance cameras in stores), privacy concerns become magnified. Data from these sources can easily be aggregated with other data, potentially to serve customers. But such practices also escalate the inherent privacy–personalization paradox (Aguirre et al. 2015), and AI seems likely to intensify such concerns. Other applications raise questions about the underlying bias in data and the assumptions that underlie the black-box algorithms on which AI is built (Rai 2020). Even as they face ethical dilemmas regarding how to serve customers better while also enhancing their profits, firms also must address privacy and bias concerns. A host of ethical decisions pertain to whether marketers (firms, society) should deploy the various available and impending AI-enabled technologies if their implementation truly allows firms to serve all consumers, and how they might adversely affect consumers, especially vulnerable populations.

Stage 2: Bringing the Individual to the Forefront

The second stage (bringing the individual to the forefront) involves major advancements in thinking AI, along with the initial versions of feeling AI (Huang and Rust 2022). Mechanical AI is outperforming humans; in new product development contexts, for example, AI already has helped discover new medications and cures in the medical domain. Thinking AI soon will be able to create sophisticated content and

Table 2. Stages of Impacts of AI.

	Stage 1: Enhancing Firm Efficiencies	Stage 2: Bringing the Individual to the Forefront	Stage 3: Rising Societal Concerns
Description	Early AI implementations primarily augment human intelligence, leading to productivity gains.	Advanced AI further augments and replaces human intelligence, enabling breakthrough innovations and a higher standard of living.	AI is outperforming (average) human intelligence, resulting in major societal tensions and trade-offs.
Capabilities of AI	Mechanical AI (e.g., analysis of mostly numerical data); Start of thinking AI (e.g., generation of textual content)	Feeling AI (e.g., recognition of user emotions); Thinking AI (e.g., generation of multi-modal content); Generative AI takes off; Mechanical AI continues to evolve (e.g., analysis of multi-modal data)	Social AI (e.g., building and maintaining intimate, meaningful relationships); Generative AI is fueling concerns of human replacement Thinking AI (e.g., generation of complex plans and solutions); Feeling AI (e.g., influencing users through emotional expressions); Mechanical AI is evolving to analyze all kinds of data
Constructive potential of AI (For Better)	Increasing employee performance; Increasing firm efficiency Decreasing costs	Augmented employee performance; Wave of new innovations in interacting with customers (e.g., chatbots, robots); Increasing standard of living	Continued potential for innovation in various domains (e.g., enhanced healthcare, better food production, higher standards of living); Reduced feelings of loneliness
Destructive potential of AI (For Worse)	Emerging tensions related to data capture and security of data	Persisting tensions related to privacy, bias, and ethics; Increasing job displacement	Concerns evolving around loss of jobs and capabilities, loss of autonomy and human connections, increased inequality and weakened institutions

solutions, implying a likely shift from an augmenting to a replacing role (Theme #1). Therefore, Stage 2 might mark a wave of innovations, like fully autonomous cars, highly capable service chatbots, or in-home service robots, many of which will be enabled by new generative AI options (Davenport et al. 2024).

Beyond such cognitively demanding tasks, early iterations of feeling AI are beginning to emerge too (Huang and Rust 2022), in the form of artificial agents with basic empathetic capabilities (Theme #2; Liu-Thompkins, Okazaki Shintaro, and Li 2022). With such capabilities, AI can take over more tasks that demand a human touch, such as conversing with elderly users, reminding them to take medication, or periodically connecting them with loved ones.

As these capabilities accrue, we expect the second stage to be associated with a wave of marketing breakthroughs that can enhance both marketing engagement and consumer well-being. Even while we acknowledge the drawbacks related to privacy, bias, and ethics (Davenport et al. 2020; Poole et al. 2021) and the potential for AI-related job displacements, we expect that at this stage, the drawbacks will be outweighed by the value created through AI.

Stage 3: Rising Societal Concerns

In time, Stage 2 will morph into Stage 3 (rising societal concerns), in which AI will deliver added-value offerings, but its societal costs will manifest more powerfully. The boundary between Stage 2 and Stage 3 remains undefined, but we propose that the latter stage will start when the negative impact of AI begins to outweigh its positive impact.

On the AI capability side, we expect AI to have reached a point where its capabilities are similar to or beyond what humans are capable of. As a result, AI might make more fully autonomous decisions, with minimal human involvement or oversight. We expect feeling AI to have reached a point where it is better than humans at recognizing users' emotions, with a keen ability to induce positive emotions and reduce negative ones (Becker, Efendić, and Odekerken-Schröder 2022). In turn, feeling AI may evolve into what we call *social AI*, which not only exhibits empathy in specific situations but also builds long-lasting, potentially meaningful, relationships with humans (Noble and Mende 2023). Accordingly, we assert that firms should actively build on the benefits of social AI. For example, AI can assist elderly people with their daily tasks but also act as a caring companion that promises to mitigate the so-called loneliness epidemic (Broadbent et al. 2023; Odekerken-Schröder et al. 2020).

As in the two previous stages, key challenges still relate to privacy, bias, and ethics, but perhaps even more importantly, we predict that two new negative impacts might arise, which we address in more detail in the remainder of this article (Theme #3). First, some impacts will be positive for the individual but might be negative for society at large. For example, intimate AI companions might help fight loneliness among the elderly, but their existence also might lead to a whole new level of social isolation if people discuss their personal problems with AI bots rather than friends or family. Second, in certain cases, the impact of AI may be positive initially but then turn out to be destructive in the long term. For example, creative AI might initially help discover breakthrough innovations, but in the long term, it could lead humans to lose their

Table 3. Three Grand Challenges.

Grand Challenge	Related to	Description	Yet-To-Be-Addressed Questions
#1: **Preserving and growing human capability**	Theme #1: AI will augment and (potentially) replace human intelligence. • job losses • capability loss	The challenge relates to preserving and growing some human capability, including marketing and business capabilities.	• What is the right approach for dealing with job loss, contingent on various environmental factors? • What marketing jobs should society preserve for humans, even if AI can be creative and social? • Which (marketing and business) skills are important to preserve for humanity? • Which types of skills can be "lost" without broader consequences?
#2: **Protecting societal belonging and human connection**	Theme #2: AI will enable artificially empathetic and trusted companions. • loss of autonomy • loss of human connection	AI could have a negative influence on consumer (and individual) well-being, driven by perceptions of loss of autonomy and loss of human connection.	• As predictive AI improves, how do we ensure humans still perceive choice-related autonomy? How do we ensure that humans do not make suboptimal choices (to reclaim autonomy)? • Despite social AI advancements, how do we ensure the persistence of human connections and their relative richness? • How do we ensure that AI does not create echo chambers that affirm and amplify dangerous beliefs?
#3: **Ensuring equitable sharing of benefits**	Theme #3: AI creates novel tensions. • inequality • weakened institutions	AI can create substantial value and grow the "economic pie," but such benefits are unlikely to be shared equally among economic actors. Furthermore, AI's super-human capabilities and misuse of AI by malicious actors may weaken institutions.	• Noting that AI will create substantial value, how do we ensure that such benefits are shared equitably? How do we ensure that those "hurt" by AI (e.g., due to job loss) do not get left behind? • AI may be misused in ways that exacerbate divisions in society and amplify misleading information, and so weaken institutions. How might this risk be mitigated? What actions would be effective? Which coalitions of actors need to be formed?

creativity and innovation capabilities. Some of these scenarios might seem far-fetched today, but all of them are acknowledged by AI futurists (e.g., Ilya Sutskever), who have sounded early alarms pertaining to the potential downside of AI (see Grewal et al. 2021).

Potential Long-Term Societal Challenges Due to AI

In late 2023, the ousting and subsequent reinstatement of Sam Altman, CEO of ChatGPT-developer OpenAI, dominated several news cycles. According to Metz (2023), the effort to overthrow Altman was led by Ilya Sutskever, an influential AI researcher and cofounder of OpenAI, who realized the power of AI but also worried that the dangers that AI posed were not being adequately addressed. Such concerns resonate with our preceding claims (Duffy and Maruf 2023) regarding the potential for substantial negative impacts of AI. Therefore, we highlight how AI initiatives may be beneficial in the short-term or at the individual level but may otherwise trigger damage at the long-term or societal levels. To do so, we detail three grand challenges posed by AI. The key issues raised by each grand challenge (and their sub-categories) are briefly summarized in Table 3. Furthermore, the three grand challenges parallel the three previously outlined themes, related to the potential for AI to (1) augment or replace human intelligence, (2) enable artificially empathetic and trusted companions, and (3) create novel tensions.

Implications if AI Augments or Replaces Human Intelligence

Were AI to augment and replace human intelligence, we note some likely negative consequences, along with the clear benefits. If AI can match or even surpass human capabilities for various tasks, it may lead to job losses, as well as the loss of human capabilities associated with those jobs.

Job losses. When AI augments or replaces human intelligence, it can substantially increase productivity, which might be a boon for actors whose efforts are being augmented (i.e., employees' jobs become easier and cause less strain) and those that benefit from their efforts (i.e., higher productivity increases firm revenues or profits). However, if AI makes humans more efficient and takes over many of their tasks, then employees might be left with less meaningful work (Bankins and Formosa 2023), and firms might need fewer employees, both of which represent forms of job losses, accruing at a societal level. For example:

> truck and cab drivers, cashiers, retail sales associates and people who work in manufacturing plants and factories [who] have been and will continue to be replaced by robotics and technology.

> Driverless vehicles, kiosks in fast-food restaurants and self-help, quick-phone scans at stores will soon eliminate most minimum-wage and low-skilled jobs (Kelly 2023).

Similarly, a host of marketing jobs might be at risk due to advances in AI. A recent study of financial markets predicts the elimination of approximately 200,000 jobs in the banking industry and cautions that even highly paid Wall Street positions are at risk, due to advances in AI and algorithm trading software (Kelly 2023).

A common argument is that human employees can move beyond such mechanical tasks and execute work that demands more creativity or empathy (e.g., Huang and Rust 2022). Yet even in Stage 1 of AI developments, generative AI already can outperform elite MBA students in discovering creative new product ideas (Kefford 2023), and AI art generation is becoming increasingly sophisticated (Roose 2022). That is, creative tasks (e.g., ad content, image creation) are not necessarily a protected domain for human intelligence, nor are they likely to offer sufficient job opportunities for all workers displaced by AI.

Turning to empathetic work, people already have some level of intimate relationships with AI-enabled bots. As we noted previously, Pi acts as an AI companion (Griffith 2023) and possesses elemental forms of artificial empathy (Liu-Thompkins, Okazaki Shintaro, and Li 2022). Noble and Mende (2023) expect such capabilities to grow, such that in the future, people will develop AI friends and partners. In a sense, AI represents a perfect listener: It can remember everything it is told, remains available 24/7, and offers good advice on a host of issues (as already evidenced by ChatGPT).

Dealing with massive job losses (and job displacement) in various marketing domains (e.g., call center agents, retail and service associates, new product development teams, advertising content creators, and pricing specialists) thus cannot be as simple as reallocating human labor from mechanical work to creative or empathetic work. Before this issue becomes acute, we need to determine and define which jobs and tasks society and firms want to reserve for humans, as well as how we should reshape and restructure society to accommodate AI agents that can perform many tasks (even those currently considered hard to automate), better and more quickly than humans can.

Therefore, research should investigate what kinds of training are required to protect and assist workers displaced from jobs that AI has started to take over so that they can move to jobs for which AI provides augmentation or has no role at all. Arguably, generative AI, chatbots, and robots even might facilitate such efforts and the migration of displaced employees to alternative jobs—ideally, with greater meaningfulness and better benefit packages—which could help ensure that employee morale remains high and avoid a backlash against AI.

Capability loss. If AI augments and someday replaces human intelligence (Huang and Rust 2022), it should enhance productivity. However if AI takes over many or most tasks, as performed by marketing and other departments, employees might lose the skills needed to perform such tasks. As an illustration, smartphones have made it so that most people do not remember phone numbers anymore. If the battery of their smartphone died, they would not be able to contact important others, even if they had access to a loaner phone. In a similar way, AI expansion has the potential to result in losses of all kinds of capabilities (e.g., writing ads and websites, generating text and code for marketing research, analyzing data, generating creative solutions, navigating to destinations) (De Cremer, Bianzino, and Falk 2023).

As Barber (2015) notes, "In a world run by intelligent machines, our lives could get a lot simpler. Would that make us less intelligent?" Some (business and marketing) skills might never be missed, but the loss of others—including cognitive and creative skills—could lead to significant societal setbacks. Whereas conventional wisdom might suggest that creativity is a uniquely human capability quality, some generative AI already contributes meaningfully to creative work (De Cremer, Bianzino, and Falk 2023). In formal creative tests conducted at the Wharton School, AI outperformed humans. Specifically, the study asked MBA students to come up with 200 ideas for (new) products that cost less than $50, and the results revealed that "the generative AI tool produced 200 ideas in less than 15 min, far quicker than the average human being who typically produces five ideas in that time" (Kefford 2023). Furthermore, the ChatGPT ideas generated higher purchase likelihoods (ChatGPT 47% vs. MBAs 40%).

Such creative capabilities can help society overcome existing and forthcoming challenges, but outsourcing such efforts might lead to deteriorated human capabilities. Therefore, as highly capable and creative AI—which already exists—exerts effects in various realms, how can we prevent firms (and societies) from losing critical marketing skills (and general business skills)?

Grand challenge #1 (preserving and growing human capability). The first grand challenge thus relates to preserving and growing human capabilities, for marketing and business in general. Ideally, AI could establish an entirely new standard of living through innovation, such as by improving the development of new medicines and finding cures for devastating conditions, such as cancer and Alzheimer's disease. But we cannot maintain a myopic focus on just these admittedly great benefits; we also must investigate what function humans will take in the face of AI's advancing capabilities. If AI is better at virtually every task than humans (Kefford 2023), what skills should humans develop and maintain, and what (meaningful) jobs will be left for them to execute? If humans enter a state in which AI does most work and makes most decisions for them, is this optimal? Even if such questions might seem farfetched, human capability losses already have emerged, such as the once common ability to memorize an array of friends' telephone numbers or remember directions. Therefore, for this grand challenge, in this age of AI, it is critical to plan a path forward that allows people to preserve and grow certain capabilities. The question of precisely which human capabilities (e.g., creativity) should be sustained is, however, still open to debate.

Implications if AI Enables Artificially Empathetic and Trusted Companions

If, as we predict, AI advances in ways that allow it to function in empathetic and intimate contexts, then AI technology arguably will be able to understand humans and predict their individual preferences, similar to how human companions seek to do currently. However, in addition to companionship benefits (e.g., decreased loneliness; Odekerken-Schröder et al. 2020), such AI applications also can impose (significant) costs, in terms of a loss of autonomy and loss of human connection.

Loss of autonomy. In purchase settings, AI can predict, accurately and in real-time, customer preferences (Davenport et al. 2020; Agrawal, Gans, and Goldfarb 2017) and offer proactive, influential purchase advice (Rodgers and Nguyen 2022). In a positive sense, marketers can better meet customers' needs, and customers will waste less time searching for products, as well as less money buying the wrong products. If AI advises people on how to eat healthier or stay physically active, it also might enhance individual and societal well-being. But these capabilities also clearly raise the potential for consumer exploitation (André et al. 2018; Davenport et al. 2020). For example, consumers could be subject to nearly constant manipulation (or as marketers would likely frame it, "nudging") toward certain decisions, because AI-designed, perfectly timed stimuli prime their unique insecurities, dreams, or hopes. In addition to concerns related to such unconscious forms of control, these nudges might lead customers to sense a lack of free will or autonomy, in that AI effectively can predict their choices (André et al. 2018). A future in which people are no longer in charge of their own consumption choices, and instead are directed by algorithms and large corporations, is clearly dystopian. In such a setting, people might actively contradict AI, just to reaffirm their autonomy (André et al. 2018). Such forms of algorithm aversion (Dietvorst, Simmons, and Massey 2015) would create a host of new challenges, potentially negating the positive impacts of AI.

Loss of human connection. Existing AI bots, such as Pi (Griffith 2023), already function as friends with whom users share their feelings, thoughts, and special moments. As these "buddy bots" gain increasing capabilities to gauge their human users' behaviors, bodily movements, speech patterns, or facial expressions, they likely can recognize those people's emotional states and respond accordingly, whether with comfort, reassurance, motivation, or suggestions for how the user might resolve their issues (Frey 2023). Through these positive contributions, AI can bring joy into the lives of individual users, to such an extent that those users might perceive less need for actual human connections (Davenport et al. 2020).

If people rely solely or primarily on AI companions for emotional support, they also may suffer diminished quality and depth in their human connections; always-available, nonjudgmental AI bots who remember everything might seem preferable to human friends as partners for sensitive conversations. Such preferences would lead to decrements in interpersonal skills. In the resulting society, people would struggle to formrelationships, resulting in greater sense of isolation and loneliness. Also, if an AI bot is programmed to affirm everything the user says to them, does it create individual echo chambers, driving people with diverse opinions farther apart (Griffith 2023)? If such (human) social isolation reaches an extreme level, the lack of exposure to other humans could threaten reductions in marriages and birth rates. Therefore, we need to consider to what extent companies should (be allowed to) market their AI agents as intimate friends, where in society AI companions can and should fit in, and where boundaries should be established.

Grand challenge #2 (protecting societal belonging and human connection). The loss of human connection due to AI pertains to both interpersonal bonds and societal belonging. That is, AI-enabled personal assistants may become close friends (Noble and Mende 2023), able to understand and address people's emotions (Becker, Efendić, and Odekerken-Schröder 2022; Liu-Thompkins, Okazaki Shintaro, and Li 2022), and they can help overcome issues like loneliness (e.g., Odekerken-Schröder et al. 2020). However such developments also might result in a lack of meaningful relationships and the disappearance of the very notion of societal belonging. We optimistically call for research to provide greater insight and directions regarding the appropriate development of AI, to ensure that we do not devolve into a society that prefers talking to bots rather than to one another. Through its potentially threatening impact on human connections, AI could have a profound negative influence on consumer (and individual) well-being. These potential effects warrant additional research and suggest the need to develop strategies to mitigate the adverse effects.

AI Creates Novel Tensions

Considering the wide variety of benefits that AI promises, an open question is whether these benefits will be shared widely or concentrated within some subset of the population, with implications for both consumers and marketing institutions.

Inequality. Lu (2023) raises a probing question: "What will AI mean for productivity and economic growth? Will it usher in an age of automated luxury for all, or simply intensify existing inequalities?" Increasing concerns acknowledge that AI could increase existing levels of inequality, shifting power balances even further toward capital over labor. If AI performs more jobs, driven by profit-related (capital) goals and technological advances, the potential for human unemployment increases, lowering consumers' individual ability to thrive (i.e., basic income levels will decline) while also eroding the tax bases that government agencies rely on and their ability to redistribute wealth (Lu 2023). In turn, wealth inequality might increase. Displaced workers will suffer serious income declines, but certain skilled AI areas will earn enhanced incomes—similar to the patterns that already are shaping labor markets.

Furthermore, the firms that have sufficient existing resources and capabilities to adopt and utilize AI can achieve improved operational performance in terms of their marketing (e.g., better insights into customers' preferences), supply chain (e.g., more precise projections), and risk management (e.g., better fraud

detection) efforts. Such benefits will likely accrue to larger firms, which have the resources needed to deploy the most sophisticated AI applications. In this way, smaller firms might be left behind, leading to more industry concentration.

In a parallel sense, if AI capabilities become highly centralized, such that a few large AI vendors control how AI functions, it will grant them a power advantage over enterprises that rely on their AI provision (Kozinets and Gretzel 2021). Therefore, the substantial benefits from AI seem unlikely to spread equally among manufacturers, retailers, and service firms, which requires some consideration of how to ensure that the power of AI benefits everyone (all firms and consumers), not just a select few.

Weakened institutions. The combination of AI's increasing ability to outperform human intelligence and generative AI's ability to create high-quality content within seconds implies a rising threat to the credibility of institutions. If AI repeatedly proves its value and accuracy, people seem likely to trust it over human input. In a recent example, ChatGPT correctly diagnosed a child's illness, which 17 doctors had missed previously (Garfinkle 2023). Giving people easy access to (accurate) medical advice is of tremendous value and has undeniable potential for enhancing the greater good. But stories of AI outperforming human specialists also might erode trust in human experts (e.g., doctors, psychologists, scientists) and lead to default reliance on AI rather than fellow humans. Similarly, AI-provided access to reliable product and service advice could reduce the perceived value of retail and service associates. If this paradigm were to become engrained in people's thinking and decision-making, human experts would lose authority, and AI would attain unchallenged influence over society and people's lives. Therefore, we need to find a means to avoid living in a world in which experts no longer have a voice.

Generative AI already exerts significant influences in politics and the marketing of political candidates and causes, particularly during election cycles. Some of these effects could be positive; Selva (2023) suggests that AI might encourage greater voter engagement and political participation, as well as give candidates clear insights into what their constituents want and need. The Brooking Institute (West 2023) predicts that AI will facilitate more precise message targeting, increase the ability of relatively unknown or less well-financed candidates to generate and distribute suitable content and reduce the time needed to respond to constituents. However, AI's ability to create resonant messages and engage individual voters also can be weaponized, particularly by malicious actors that actively seek to spread "fake news" and misinformation (Selva 2023), in ways that threaten to weaken democratic processes. Therefore, marketing (and other) researchers need to examine how to best prevent AI, when used for nefarious purposes, from eroding trust in political institutions. The role of consumer education and the design of corrective advertising campaigns should be explored to provide greater guidance for these efforts. Perhaps generative AI can help with such campaigns.

Grand challenge #3 (ensuring equitable sharing of benefits). Although AI can create substantial value and grow the "economic pie," such benefits are unlikely to be shared equally among economic actors. Even if AI may strengthen some institutions, it may well weaken others. Also, some workers may lose their jobs (Kelly 2023) or be downgraded to performing less meaningful work (Bankins and Formosa 2023), resulting in potentially more drawbacks than benefits for these groups. Smaller firms will struggle to compete with larger firms that have the resources to integrate AI more closely into their business processes. Therefore, in the age of AI, we expect growing inequality, with all its well-acknowledged downstream problems. Even the Nobel Prize winner Joseph Stiglitz feels "pessimistic with respect to the issue of inequality. With the right policies, we could have higher productivity and less inequality, and everybody would be better off. But … the way that our politics have been working, has not been going in that direction" (Bushwick 2023). Already, powerful firms such as Disney and NBCUniversal have lobbied against proposed tax penalties designed to discourage film and television studios from replacing creative writers, actors, and production assistants with AI (Williams 2023). Not only is there a tendency for AI benefits to accrue to large firms due to their existing dominance, but large firms are actively lobbying to receive even more benefits.

Therefore, a grand challenge pertains to finding ways to ensure that institutions are not weakened, and the benefits of AI are shared equitably. In today's early Stage 1, it remains possible to implement and adjust the formal and informal institutions that will define how AI adds value—namely, for all and not just a few. Such institutional implementation efforts demand that marketers, marketing academics, and policymakers collaborate to design, imagine, and engineer policies and business practices that establish a more equal, AI-powered future. In this vein, Hermann, Williams, and Puntoni (2023) share insights into how AI technologies can be deployed to assist vulnerable populations. They highlight the importance of accessible AI technologies that are central to assisting vulnerable consumers (and potentially smaller firms).

General Discussion

As AI continues to change the world, we must ask: Is this change for the better or for the worse? To push academic discourse beyond the direct benefits and costs of implementing AI (Davenport et al. 2020; Guha et al. 2021), by marketers and others, we propose a broader, societal perspective that highlights various macro-level tensions that loom on the horizon. These longer-term tensions involve the loss of human capabilities and jobs, autonomy, and connectedness, as well as increasing inequality and weakened institutions. To guide and prompt continued discussions of these tensions, we outline three grand challenges: preserving and growing human capabilities, protecting societal belonging and human connection, and ensuring equitable sharing of AI's benefits.

Because AI has not yet penetrated our economy and society too deeply, now is the time to tackle these grand challenges.

Thus, this article embraces the metaphorical notion of setting a course to avoid a squall, which is easier than trying to maneuver out of the storm. We freely acknowledge that addressing these challenges will not be easy; it will require dialogue and coordination among multiple actors, including marketers, academics, and policy makers. Considering that these grand challenges also lie somewhat in the future, identifying ways to deal with them will require considerable imagination and foresight. We have outlined some grand challenges in the hope of kickstarting a discussion among key stakeholders, regarding how to address such challenges in their future business and policy decisions. Table 3 features some of these yet-to-be-addressed questions.

We also offer two caveats. First, addressing these challenges is complicated, involving multiple stakeholders and complex interdependencies; actions in one domain will have impacts in other domains, in ways that are currently difficult to foresee. In complex systems (e.g., Sirgy 1989), an effective solution must account for the system level, and even a vast set of standalone solutions is unlikely to be sufficient. In the spirit of solutions proffered by Satornino, Du, and Grewal (2024), we propose that addressing these challenges could be guided by notions from complex adaptive systems theory, such that any response simultaneously addresses multiple challenges, optimized across various domains.

Second, we identify and discuss three grand challenges. These macro challenges reflect scenarios in which we predict that the benefits of AI may be double-edged, creating positive outcomes at the individual level but threats at the societal level, or else leading to short-term benefits but longer-term risks. Far more challenges, of various types and forms, remain. Especially in marketing domains, difficult questions remain with regard to privacy, ethics, and bias, all of which demand attention.

As concerns about AI continue to rise though, we also are heartened to see that lawmakers have started responding, even if in limited ways. For example, noting the substantial number of AI deep fakes circulating, particularly those involving Taylor Swift, U.S. policymakers have recognized the substantial threat of invasions of privacy (Segall 2024) and proposed the Preventing Deepfakes of Intimate Images Act (Rahman-Jones 2024). Such concerns about deepfakes are not limited to images; fake robocalls purportedly involving President Biden emerged during recent presidential primaries (Rahman-Jones 2024). Many companies already have adopted AI for hiring processes, but because the bias it can impose seemingly is not always evident to hiring managers (Akselrod and Venzke 2023), lawmakers also have proposed the American Data Privacy and Protection Act (Fitzgerald 2023)—though it is unclear when or if it will be enacted into law.

Finally, continued research needs to examine the role of AI in relation to safety concerns. For example, driverless cars bring physical safety to the forefront, and medical robots and AI diagnostic devices highlight a host of safety issues that need to be carefully thought through.

Conclusion

We take a broad view of the role of AI to discuss the promises and perils it presents for firms, individuals, and society. We identify three grand challenges. The paths to confronting these problems are difficult and unclear, requiring coordination across marketers, policymakers, and governments. Accordingly, we hope that this article spurs more research and action into these three issues. As AI in all forms continues evolving at a rapid rate, it is likely to have profound effects on all stakeholders. Thus, we call for more coordinated research and a clear specification of the domains in which AI (including generative AI) should be encouraged, and the domains in which AI innovations should be very carefully monitored.

Associate Editor

M. Joseph Sirgy

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Marc Becker https://orcid.org/0000-0003-0321-9502

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2017), "How AI Will Change Strategy: A Thought Experiment," *Harvard Business Review*. October 3, (accessed December 21, 2023), [Available at https://hbr.org/how-ai-will-change-strategy-a-thoughtexperiment].

Aguirre, Elizabeth, Dominik Mahr, Dhruv Grewal, Ko de Ruyter, and Martin Wetzels (2015), "Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness," *Journal of Retailing*, 91 (1), 34-49.

Akselrod, Olga and Cody Venzke (2023, August 23), "How Artificial Intelligence Might Prevent You from Getting Hired," (accessed February 3, 2024), [Available at https://www.aclu.org/news/racial-justice/how-artificial-intelligence-might-prevent-you-from-getting-hired].

André, Quentin, Ziv Carmon, Klaus Wertenbroch, Alia Crum, Douglas Frank,, William Goldstein, et al. (2018), "Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data," *Customer Needs and Solutions*, 5 (1–2), 28-37.

Bankins, Sarah and Paul Formosa (2023), "The Ethical Implications of Artificial Intelligence (AI) for Meaningful Work," *Journal of Business Ethics*, 185 (4), 725-740.

Barber, Nigel (2015), "Can Artificial Intelligence Make Us Stupid?" (accessed December 21, 2023), [Available https://www.psychologytoday.com/us/blog/the-human-beast/201507/can-artificial-intelligence-make-us-stupid].

Becker, Marc, Emir Efendić, and Gaby Odekerken-Schröder (2022), "Emotional Communication by Service Robots: A Research Agenda," *Journal of Service Management*, 33 (4/5), 675-687.

Broadbent, Elizabeth, Mark B. Billinghurst, Samatha G. Boardman, and Murali P. Doraiswamy (2023), "Enhancing Social Connectedness with Companion Robots Using Artificial Intelligence," *ScienceRobotics*, 8 (80), (accessed December 21, 2023), [Available at https://www.science.org/doi/10.1126/scirobotics.adi6347].

Bushwick, Sophie (2023), "Unregulated AI Will Worsen Inequality, Warns Nobel-Winning Economist Joseph Stiglitz," *Scientific American*, 329 (5), 83 (accessed December 21, 2023), [Available at https://www.scientificamerican.com/article/unregulated-ai-will-worsen-inequality-warns-nobel-winning-economist-joseph-stiglitz/].

Davenport, Thomas, Dhruv Grewal, Abhijit Guha, and Cinthia B. Satornino (2024), "How Generative AI is Shaping the Future of Marketing," Unpublished Working Paper, Babson College.

Davenport, Thomas, Abhijit Guha, Dhruv Grewal, and Timma Bressgott (2020), "How Artificial Intelligence Will Change the Future of Marketing," *Journal of the Academy of Marketing Science*, 48 (1), 24-42.

De Cremer, David, Nicola M. Bianzino, and Ben Falk (2023), "How Generative AI Could Disrupt Creative Work," *Harvard Business Review*. April 13 (accessed December 21, 2023), [Available at https://hbr.org/2023/04/how-generative-ai-could-disrupt-creative-work].

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2015), "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err," *Journal of Experimental Psychology: General*, 144 (1), 114-126.

Duffy, Clare and Ramishah Maruf (2023), "Elon Musk Warns that AI Could Cause 'Civilization Destruction' Even as He Invests in It," *CNN* (accessed December 21, 2023), [Available at https://www.cnn.com/2023/04/17/tech/elon-musk-ai-warning-tucker-carlson/index.html].

Elliott, Scott (2020, December 10), "Artificial Intelligence Improves America's Food System," (accessed December 21, 2023), [Available at https://www.usda.gov/media/blog/2020/12/10/artificial-intelligence-improves-americas-food-system].

Fitzgerald, Madyson (2023, July 22), "Does Your State Have Rules about Using AI in Hiring?" (accessed February 3, 2024), [Available at https://www.fastcompany.com/90926772/does-your-state-have-rules-about-using-ai-in-hiring].

Frey, Thomas (2023), "The Great AI Disruption: Six Startling Predictions that Will Shape our Lives and Test our Limits," (accessed December 21, 2023), [Available https://futuristspeaker.com/artificial-intelligence/the-great-ai-disruption-six-startling-predictions-that-will-shape-our-lives-and-test-our-limits/].

Garfinkle, Madeline ((2023, September 12)'), "'We Saw So Many Doctors': A Mother Says ChatGPT Accurately Diagnosed Her Son's Medical Condition After 17 Doctors Couldn't," (accessed December 21, 2023), [Available at https://www.entrepreneur.com/business-news/17-doctors-didnt-diagnose-her-sons-disorder-chatgpt-did/458927].

Grewal, Dhruv, Abhijit Guha, Cinthia B. Satornino, and Elisa Schweiger (2021), "Artificial Intelligence: The Light and the Darkness," *Journal of Business Research*, 136 (November), 229-236.

Griffith, Erin (2023, May 3), "My Weekend with an Emotional Support A.I. Companion," *The New York Times* (accessed December 21, 2023), [Available at https://www.nytimes.com/2023/05/03/technology/personaltech/ai-chatbot-pi-emotional-support.html].

Guha, Abhijit, Dhruv Grewal, and Stephen Atlas (2024), "Generative AI and Marketing Education: What the Future Holds," *Journal of Marketing Education*, 46 (1), 6-17.

Guha, Abhijit, Dhruv Grewal, Praveen K. Kopalle, Michael Haenlein, Matthew Schneider,, Hyunseok Jung, et al. (2021), "How Artificial Intelligence Will Affect the Future of Retailing," *Journal of Retailing*, 97 (1), 28-41.

Hermann, Erik (2022), "Leveraging Artificial Intelligence in Marketing for Social Good—An Ethical Perspective," *Journal of Business Ethics*, 179 (1), 43-61.

Hermann, Erik, Gizem Y. Williams, and Stefano Puntoni (2023), "Deploying Artificial Intelligence in Services to AID vulnerable consumers," *Journal of the Academy of Marketing Science*, November 11 (published online). https://doi.org/10.1007/s11747-023-00986-8.

Huang, Ming-Hui and Ronald T. Rust (2021), "A Strategic Framework for Artificial Intelligence in Marketing," *Journal of the Academy of Marketing Science*, 49 (1), 30-50.

Huang, Ming-Hui and Ronald T. Rust (2022), "A Framework for Collaborative Artificial Intelligence in Marketing," *Journal of Retailing*, 98 (2), 209-223.

Hughes, Neil C. (2023), "The Future Is Personal: A Deep Dive into AI Companions," (accessed December 21, 2023), [Available at https://cybernews.com/tech/ai-companions-explained/].

Kefford, Matt (2023), "Wharton Study Pits ChatGPT against MBA Students in Creativity Tests," (accessed December 21, 2023), [Available at https://www.businessbecause.com/news/in-the-news/8960/wharton-chatgpt-creativity-study].

Kelly, Jack (2023, March 31), "Goldman Sachs Predicts 300 Million Jobs Will be Lost or Degraded by Artificial Intelligence," *Forbes* (accessed December 21, 2023), [Available at https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/?sh=1b0851cb782b].

Kopalle, Praveen K., Manish Gangwar, Andreas Kaplan, Divya Ramachandran, Werner Reinartz, and Aric Rindfleisch (2022), "Examining Artificial Intelligence (AI) Technologies in Marketing via a Global Lens: Current Trends and Future Research Opportunities," *International Journal of Research in Marketing*, 39 (2), 522-540.

Kozinets, Robert V. and Ulrike Gretzel (2021), "Commentary: Artificial Intelligence: The Marketer's Dilemma," *Journal of Marketing*, 85 (1), 156-159.

Liu-Thompkins, Yuping, Shintaro Okazaki Shintaro, and Hairong Li (2022), "Artificial Empathy in Marketing Interactions: Bridging the Human-AI Gap in Affective and Social Customer Experience," *Journal of the Academy of Marketing Science*, 50 (6), 1198-1218.

Lu, Yingying (2023), "AI Will Increase Inequality and Raise Tough Questions about Humanity, Economists Warn," (accessed December

21, 2023), [Available at https://theconversation.com/ai-will-increase-inequality-and-raise-tough-questions-about-humanity-economists-warn-203056].

Luo, Xueming, Marco S. Qin, Zheng Fang, and Zhe Qu (2021), "Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions," *Journal of Marketing*, 85 (2), 14-32.

Madiega, Tambiama (2023), "Artificial Intelligence Act," *European Parliament: European Parliamentary Research Service*, [Available at https://superintelligenz.eu/wp-content/uploads/2023/07/EPRS_BRI2021698792_EN.pdf].

Markets and Markets (2023), "Artificial Intelligence (AI) Market by Offering (Hardware, Software), Technology (ML (Deep Learning (LLM, Transformers (GPT 1, 2, 3, 4)), NLP, Computer Vision), Business Function, Vertical, and Region - Global Forecast to 2030," (accessed December 21, 2023), [Available at https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-market-74851580.html].

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier (2017), "Data Privacy: Effects on Customer and Firm Performance," *Journal of Marketing*, 81 (1), 36-58.

Martin, Kelly D. and Partick E. Murphy (2017), "The Role of Data Privacy in Marketing," *Journal of the Academy of Marketing Science*, 45 (2), 135-155.

Metz, Cade (2023, November 18), "The Fear and Tension that Led to Sam Altman's Ouster at OpenAI," *The New York Times* (accessed December 21, 2023), [Available at https://www.nytimes.com/2023/11/18/technology/open-ai-sam-altman-what-happened.html].

Noble, Stephanie M. and Martin Mende (2023), "The Future of Artificial Intelligence and Robotics in the Retail and Service Sector: Sketching the Field of Consumer-Robot Experiences," *Journal of the Academy of Marketing Science*, 51 (4), 747-756.

Odekerken-Schröder, Gaby, Cristina Mele, Tiziana Russo-Spena, Dominik Mahr, and Andrea Ruggiero (2020), "Mitigating Loneliness with Companion Robots in the COVID-19 Pandemic and Beyond: An Integrative Framework and Research Agenda," *Journal of Service Management*, 31 (6), 1149-1162.

Poole, Martin S., Sonya A. Grier, Kevin D. Thomas, Francesca Sobande, Akon E. Ekpo,, Lez T. Torres, et al. (2021), "Operationalizing Critical Race Theory in the Marketplace," *Journal of Public Policy & Marketing*, 40 (2), 126-142.

Puntoni, Stefano, Rebecca W. Reczek, Markus Giesler, and Simona Botti (2021), "Consumers and Artificial Intelligence: An Experiential Perspective," *Journal of Marketing*, 85 (1), 131-151.

Rahman-Jones, Imran (2024, January 27), "Taylor Swift deep fakes spark calls in Congress for new legislation," (accessed February 3, 2024), [Available at https://www.bbc.com/news/technology-68110476].

Rai, Arun (2020), "Explainable AI: From Black Box to Glass Box," *Journal of the Academy of Marketing Science*, 48 (1), 1371–1141.

Rasooly, Danielle and Muin J. Khoury (2022), "Artificial Intelligence in Medicine and Public Health: Prospects and Challenges beyond the Pandemic," (accessed December 21, 2023), [Available at https://blogs.cdc.gov/genomics/2022/03/01/artificial-intelligence-2/].

Rodgers, Waymond and Tam Nguyen (2022), "Advertising Benefits from Ethical Artificial Intelligence Algorithmic Purchase Decision Pathways," *Journal of Business Ethics*, 178 (4), 1043-1061.

Roose, Kevin (2022), "An AI-Generated Picture Won an Art Prize. Artists Aren't Happy," *The New York Times, September 2*, (accessed December 21, 2023), [Available at https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html].

Salesforce (2023), "New Research: 60% of Marketers Say that AI Will Transform their Role, but Worry about Accuracy," (accessed December 21, 2023), [Available at https://www.salesforce.com/ap/blog/generative-ai-for-marketing-research/].

Satornino, Cinthia B., Shuili Du, and Dhruv Grewal (2024), "Using Artificial Intelligence to Advance Sustainable Development in Industrial Markets: A Complex Adaptive Systems Perspective," *Industrial Marketing Management*, 116 (January), 145-157.

Segall, Laurie (2024), "Opinion: The Taylor Swift AI Photos Offer a Terrifying Warning," January 31, (accessed February 3, 2024), [Available at https://www.cnn.com/2024/01/31/opinions/taylor-swift-deepfakes-ai-segall/index.html].

Selva, Meera (2023), "In 2024 Elections, We Have to Protect Minorities from AI Aggravated Bias," (accessed December 21, 2023), [Available at https://www.euronews.com/2023/12/07/in-2024-elections-we-have-to-protect-minorities-from-ai-aggravated-bias].

Shankar, Venkatesh (2018), "How Artificial Intelligence (AI) is Reshaping Retailing," *Journal of Retailing*, 94 (4), vi-xi.

Shankar, Venkatesh, Kirthi Kalyanam, Pankaj Setia, Alireza Golmohammadi, Seshadri Tirunillai,, Tom Douglass, et al. (2021), "How Technology Is Changing Retail," *Journal of Retailing*, 97 (1), 13-27.

Sirgy, M. Joseph (1989), "Toward a Theory of Social Organization: A Systems Approach," *Behavioral Science*, 34 (4), 272-285.

Trucano, Michael (2023), "AI and the Next Digital Divide in Education," Brookings Institute (accessed December 21, 2023), [Available at https://www.brookings.edu/articles/ai-and-the-next-digital-divide-in-education/].

Villasenor, John (2019), "Artificial Intelligence and Bias: Four Key Challenges," Brookings Institute (accessed December 21, 2023), [Available at https://www.brookings.edu/blog/techtank/2019/01/03/artificial-intelligence-and-bias-four-key-challenges/].

Wang, Yilun and Michal Kosinski (2018), "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images," *Journal of Personality and Social Psychology*, 114 (2), 246-257.

Weise, Karen, Cade Metz, Nico Grant, and Mike Isaac (2023), "Inside the A.I. Arms Race That Changed Silicon Valley Forever," *The New York Times, December 5* (accessed December 21, 2023), [Available at https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html].

West, Darrell M. (2023), "How AI Will Transform the 2024 Elections," Brookings Institute (accessed December 21, 2023), [Available at https://www.brookings.edu/articles/how-ai-will-transform-the-2024-elections/].

Williams, Zach (2023), "Disney and NBC Eyeing New York's Tax Break Ban Proposal," (accessed December 21, 2023), [Available at https://news.bloomberglaw.com/in-house-counsel/disney-and-nbc-watch-new-yorks-ai-tax-break-ban-proposal].

Wong, Jessica (2023), "Artificial Intelligence is Revolutionizing Marketing. Here's what the Transformation Means for the

Industry," (accessed December 21, 2023), [Available at https://www.entrepreneur.com/science-technology/why-artificial-intelligence-is-revolutionizing-marketing/446087].

Author Biographies

Dhruv Grewal (Ph.D. Virginia Tech) is the Toyota Chair in Commerce and Electronic Business and a Professor of Marketing at Babson College. His research and teaching interests focus on the broad areas of AI, technology, retailing, pricing and services. He is listed in The World's Most Influential Scientific Minds, Thompson Reuters 2014. He is a fellow of AMA and AMS, and was awarded the AMS Cutco/Vector Distinguished Educator Award in May 2010. He was a co-editor of *Journal of Retailing*. He has also coauthored a number of books: *Marketing Research, Marketing, M Series: Marketing* and *Retailing Management*.

Abhijit Guha (abhijit.guha@moore.sc.edu) is an associate professor of marketing at the Darla Moore School of Business, University of South Carolina. He serves as the Academic Director for MBA Programs and the Master of Science in Business Analytics Program. His research and teaching interests straddle AI, technology, retailing and pricing. Abhijit has a Ph.D. from Duke University and an MBA from INSEAD. He has published articles in the *Journal of Marketing, Journal of Marketing Research, Management Science, Journal of Academy of Marketing Science, Journal of International Business Studies, Journal of Retailing, Organizational Behavior and Human Decision Processes, Harvard Business Review*, etc.

Marc Becker is a PhD Candidate at the Department of Marketing and Supply Chain Management at Maastricht University's School of Business and Economics as well as a founding member of the Maastricht Center for Robots. His research interests are in service research, particularly concerning the impact of service robots and AI on customers, employees, businesses, and society at large. He has recently published in the *Journal of Service Management*, *Psychology & Marketing*, and *Service Business*.

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
2. Educational reform in the era of artificial intelligence (2021)	ACM International Conference Proceeding Series (Article from : Association for Computing Machinery)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Educational Reform in the Era of Artificial Intelligence

Pan Xiaoling
Foreign Languages College, Dalian Polytechnic University, Dalian, China
panxiaoling79@163.com

ABSTRACT

The rapid development of artificial intelligence has brought opportunities and challenges to human society. It promotes the progress of society and the development of science and technology, and also increases the unemployment rate and the ethical dilemma of technology. The weakness of artificial intelligence lies in the lack of creativity. If human beings want to have core competitiveness in the era of artificial intelligence, they must change education and enhance creativity. Education reform should be carried out from three aspects: one is to cultivate scientific and technological literacy**; two** is to cultivate data literacy; three is to cultivate humanistic quality.

CCS CONCEPTS

• **Applied computing** → Education; E-learning.

KEYWORDS

Artificial intelligence, educational reform, scientific literacy, data literacy, humanistic literacy

ACM Reference Format:
Pan Xiaoling. 2021. Educational Reform in the Era of Artificial Intelligence. In *2021 2nd International Conference on Computers, Information Processing and Advanced Education (CIPAE 2021), May 25–27, 2021, Ottawa, ON, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3456887.3457517

1 INTRODUCTION

In recent years, artificial intelligence has been booming and widely used in professional fields and daily life.In 2017 , the robot Alpha Dog won the world's first Chinese chess player Ke Jie , with a great momentum; Microsoft artificial intelligence program" Xiaobing " has created modern poetry and been published, which has attracted wide attention of the society. It is also in 2017 , artificial intelligence has become one of the ten hot words of the year in China, and is familiar to the public . Undoubtedly,the era of artificial intelligence has arrived. When the future comes, some people are scared and some people look forward to it. What does artificial bring to human beings? In the era of artificial intelligence, what qualities do human beings have to possess to be core competitive? What kind of changes should education make to cultivate these abilities? This is what this article will analyze.

CIPAE 2021, May 25–27, 2021, Ottawa, ON, Canada

ACM ISBN 978-1-4503-8996-9/21/05…$15.00
https://doi.org/10.1145/3456887.3457517

2 THE INFLUENCE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence belongs to a branch of computer science. It is a technology to simulate some thinking process and intelligent behavior of human beings through computer programs. At the Dartmouth conference in 1956, computer scientist John McCarthy defined AI. Since the new century, artificial intelligence has ushered in its golden period of development. Along with genetic engineering and nanoscience, artificial intelligence has been called the three cutting-edge technologies in the 21st century, and "presents new features such as deep learning, cross-border integration, human-computer collaboration, open group intelligence, and autonomous control."[1]

Artificial intelligence has a great impact on human development. On the one hand, it is an important driving force for social development and scientific and technological progress, which has greatly changed people's cognition and grasp of the world. With the continuous development of artificial intelligence, it is applied in various fields, not only in the financial, security, medical, transportation, education, manufacturing, retail and other fields in-depth layout, but also plays an important role in people's daily life. At present, many countries take artificial intelligence as the key development object. In China, artificial intelligence has developed rapidly in the past decade. After using principal component analysis to analyze the index weight of artificial intelligence, the researcher calculates the comprehensive calculation value of the development level of artificial intelligence in China. Take 2008-2017 as an example:

It can be seen from the above table that China's artificial intelligence has developed rapidly in the past ten years.

Artificial intelligence not only brings convenience, but also challenges and difficulties to human beings. It increases the unemployment crisis and brings about various moral and ethical dilemmas. With the development of artificial intelligence, a lot of work can be done by machines. The risk of unemployment of human employees in translation, drivers, production line workers, accountants, lawyers and other industries is greatly increased. McKinsey pointed out in its 2017 report that due to the automation brought about by artificial intelligence, 375 million of the 400-800 million people in 2030 will need to change their careers, or they will face the danger of elimination. At the same time, it brings more and more moral and ethical dilemmas, such as the information leakage caused by AI face recognition technology, and the responsibility attribution of accidents caused by automatic driving.

The arrival of the era of artificial intelligence is unstoppable. What we should do is to recognize the relationship between artificial intelligence and human beings, and carry out corresponding educational reform to enhance our core competitiveness. Joseph E. Aoun, former president of Northeastern University, once pointed

Table 1: Overall development level of artificial intelligence in China (2008-2017) [2]

Particular year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Development level of artificial intelligence	-0.953	-0.854	-0.588	-0.305	0.1058	0.351	0.6151	0.929	1.3036	1.8105

out that artificial intelligence has a congenital defect: lack of creativity. In other words, it has no initiative and self-consciousness. Even if it is powerful, it is only the supplement and extension of human ability. Although it will replace some human workers, it will replace repetitive and mechanical work, and those creative work will not be replaced. If human beings want to have the core competitiveness in the future era and avoid being eliminated, they should enhance their creativity and develop into fields that artificial intelligence can not replace. In order to meet the new demand of talent training in the era of artificial intelligence, education must adjust its training objectives and methods. So, in the era of artificial intelligence, what qualities should be possessed in order to have core competitiveness? What kind of educational reform should be carried out to obtain these qualities?

3 LEARNING LITERACY AND EDUCATIONAL REFORM REQUIRED IN THE ERA OF ARTIFICIAL INTELLIGENCE

In order to maintain the core competitiveness and enhance creativity in the era of artificial intelligence, we need to have three kinds of learning literacy: scientific and technological literacy, data literacy and humanistic literacy. Cultivating these three qualities is the direction of future education reform.

3.1 Science and Technology Literacy

In the era of artificial intelligence, education should pay attention to the cultivation of students' scientific and technological literacy. The so-called scientific and technological literacy is to master the knowledge of mathematics, coding and basic engineering principles, and to understand the working principle of artificial intelligence. The era of artificial intelligence is a digital era. People are surrounded by various digital technologies. A new generation of "digital natives" are growing up in digital technology. People can operate all kinds of software and programs skillfully, but they don't necessarily understand the working principle behind them. For example, we can use mobile phones, flat touch screens, and smart speakers to remotely start air conditioning, sweeping robots, and television, but most people don't understand the basic principles behind the devices. Only by mastering the basic principles can we make the most of the software and hardware and realize the innovation to the maximum extent.

To have scientific and technological literacy, it is better to master coding knowledge. Coding is the universal language of the digital world, and it is also an indispensable skill for future talents. Only by mastering the coding ability can we have the thinking ability to understand the future social things.

In terms of education, we must change the content of education. In the era of artificial intelligence, coding course is as important as Chinese and mathematics. In the compulsory courses of University and high school, coding should occupy a place. In recent years, coding training camps are becoming more and more popular in the higher education market. This is evidenced by the increasing number of students admitted to the United Nations General Assembly and the code training camp for the development of training operations in recent years.

3.2 Data Literacy

In the era of artificial intelligence, education should also focus on cultivating students' data literacy.

Data literacy is the ability to analyze and process big data streams and make decisions. In the era of artificial intelligence, data is king. According to International Data Corporation, the global data will grow 10 times by 2025. Having data literacy can help us to effectively analyze the massive data, and use them to transform them into useful information for companies and enterprises, and formulate appropriate countermeasures for the market. Taking McKinsey 2019 China report as an example, this paper analyzes the market share of multinational enterprises in the world's top 30 commodity categories in 2017 through big data. The data analysis is as follows:

According to the McKinsey report, it is concluded from the figure that the penetration rate of multinational enterprises in China is higher than that in the US market. Under the background of globalization, China's consumer market is expanding rapidly and has been highly integrated with the world, which has great potential. This will not only promote China's economic growth, but also provide opportunities for international enterprises. Domestic and foreign enterprises should adjust their business structure and mode in time to seize business opportunities.

Cultivating data literacy requires the joint efforts of society. Regardless of whether it is a primary school or a university, data courses should be incorporated into the curriculum system.

Whether it is the cultivation of scientific and technological literacy or data literacy, it has put forward corresponding reform requirements for educational practice. That is, in the basic education and higher education stage, increase the teaching of artificial intelligence and information technology, so that students can master specific computer languages, algorithms and logical thinking. At the same time, adjustments to the educational structure, such as strengthening the cultivation of digital frameworks in mathematics education, are all conducive to enhancing creativity. In short, it is necessary to integrate information technology into education to enhance the core competitiveness of the educated in the future world.

[a]Source: McKinsey Global Institute analysisNote: due to rounding, the sum of the numbers may not equal 100%.

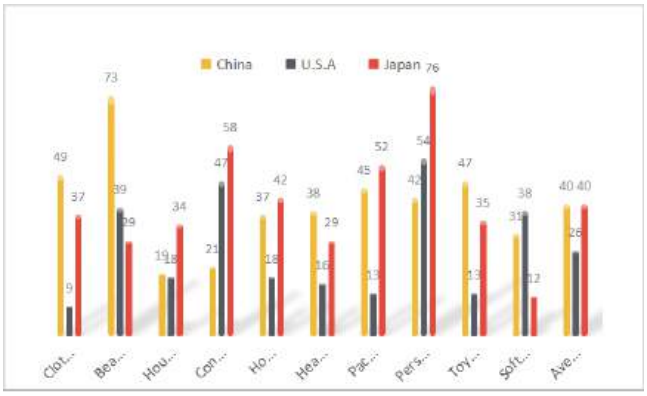


Figure 1: Market share of multinational enterprises in the world's top 30 commodity categories by category and market (2017)[a]

3.3 Humanities

In addition to technological literacy and data literacy, the cultivation of humanistic literacy cannot be ignored in the era of artificial intelligence. Cultivating humanistic literacy is conducive to improving human communication and design capabilities, and provides strong support for human survival in the technological world. The so-called humanistic literacy "includes human nature based on traditional liberal arts education, but also includes artistic elements, especially design, which is an important part of digital communication." [3] No matter what era it is, the most important thing is human beings. In the networked space, the most powerful network is interpersonal relationships. At the same time, in the technological world, technology also needs the nourishment of humanities. Kai-Fu Lee, chairman of Innovation Works, once said: "No matter how powerful artificial intelligence is, it cannot replace the humanistic ability with warmth." The combination of technology and humanities can make products have the core competitiveness. Humanistic literacy can enable people to have the ability to communicate and design. Many experts have pointed out that the combination of artificial intelligence and humanities is an effective way to stimulate creativity, which is the key to success in the artificial intelligence era. Top universities such as Stanford and the Massachusetts Institute of Technology in the United States have already experimented in this area, setting up joint majors in computer science and humanities to stimulate the application of artificial intelligence in various fields such as medical care, law, finance, and media. [4] At the same time, the cultivation of humanistic literacy helps to solve a series of difficulties brought by the development of artificial intelligence. Waelbers once said: "In an increasingly technological modern society, the widespread application of artificial intelligence technology complicates the ownership of moral responsibility." [5] Without the support of humanistic literacy, artificial intelligence will have many ethical issues in the application process, such as the responsibility of smart weapons for killing people, new inequalities that may be caused by genetic modification technology, and user privacy in user data extraction and analysis. And security issues. This type of problem cannot be solved by technology alone, and humanity can cooperate with machines to make the right choice. With humanistic literacy, you can embed the concept of'goodness' in the process of design and development of intelligent machines, and adhere to the principle of precaution, allowing machines to assume forward-looking responsibilities, so that artificial intelligence products can comply with social ethics during operation Standardize and maximize the role of serving mankind." [6] Reflected in the education level, it is necessary to change the overall pattern of emphasizing science and technology over humanities, breaking through the barriers of arts and sciences, strengthening humanistic education, and cultivating students' humanistic qualities. We always say that if we learn mathematics, physics and chemistry well, we are not afraid to travel around the world. But in the age of artificial intelligence, technology alone is not enough. Hand in hand with humanities can go further.

Figure 2: the general process of critical thinking

4 COGNITIVE ABILITIES AND EDUCATIONAL CHANGES REQUIRED IN THE ERA OF ARTIFICIAL INTELLIGENCE

In the era of artificial intelligence, education must not only cultivate human science and technology literacy, data literacy, and humanistic literacy, but also pay attention to enhancing human critical thinking, systematic thinking, entrepreneurial spirit and culture. Agility. These four cognitive abilities can effectively enhance creativity and enable humans to have core competitiveness.

4.1 Critical thinking

The word "critical" is derived from the Greek kriticos and kriterion. The former means asking questions, understanding the meaning of something and having the ability to analyze, that is, "the ability to discern or judge", and the latter means the standard. In general, critical thinking is to have the ability of rational analysis and judgment. It is a thinking skill as well as an attitude. The general process of critical thinking is shown in Figure 2

Critical thinking originated from the ancient Greek thinker Socrates. Some Socrates believed that all knowledge arises from difficulty, so he adopted the method of probing questioning in teaching. He often exposes contradictions in the opponent's doctrine by asking questions, shakes the basis of the opponent's argument, and points out the opponent's ignorance. There is a very famous story called Socrates' Apple Experiment. When the word Socrates was in class, students took an apple and asked them if they could smell the fragrance of apples. For the first time, only one student said it smelled. The second time, half of the people said they smelled it. For the third time, only one student said that he did not smell the

fragrance of apples. Socrates praised the student who said that he did not smell the fragrance, because it was a fake apple and it had no fragrance. The student was not coerced by the opinions of the people around him, kept thinking independently and made correct judgments. Socrates said he was like a "midwife", helping others to produce knowledge. His training method is called "Socratic Method" or "Midwifery".

In the era of artificial intelligence, critical thinking is generally established as one of the goals of education, especially higher education. Because in the digital age, what is lacking is not data and information, but critical thinking. There are no disciplinary boundaries, and any topic related to intelligence or imagination can be considered from the perspective of critical thinking.If there is no critical thinking, it is easy for others to agree, or make wrong analysis and decisions, and lack sufficient professional competitiveness. To cultivate critical thinking, we need to reform education methods and cultivate the consciousness of independent thinking.

4.2 Systematic thinking

Systematic thinking, also called holistic thinking, is a way of thinking that can sort a series of scattered problems in an orderly manner and analyze them from a comprehensive and holistic perspective. When a person encounters a problem, he should not only think about tricks and tricks to solve the current problem, but also have a systematic and overall perspective to better solve the problem from a higher angle.

Systematic thinking is one of the key abilities for human beings to succeed in the era of artificial intelligence. Compared with humans, artificial intelligence may be better at understanding the elements of complex systems and how their variables are connected in series. But it has an insurmountable shortcoming, that is, it lacks systematic thinking, is not good at applying existing information to different situations, and cannot jump out of the value of thinking in a specific area. For example, artificial intelligence can simulate climate change in a specific area in a predetermined code library, assess pollution, water temperature, weather operation patterns, and other interwoven elements. By evaluating the data, we can draw conclusions: how to prevent soil erosion. But it itself cannot apply these data to other related fields, such as human migration research, fishery operations, environmental law writing, and so on. Only with the participation of human beings can there be huge innovations and leaps.

In the era of artificial intelligence, individuals and companies with systematic thinking have better prospects for development and are more competitive. For example, self-driving cars are the future development trend, and many companies are developing them. It turns out that companies are focusing on the research and development of bicycle intelligence, which is to research and develop the sensor system of the car itself, and the company establishes its own experimental site and test site. But to do so, the research and development costs are expensive, and the risk of conducting road surveys is also great. Now, some companies are cooperating with the government to combine the development of smart cars with shared intelligence and traffic intelligence in the city to form a

Figure 3: Floating house designed by the architect firm Waterstudio.NL founded by Olthuis

“vehicle-road-city” synergy and cooperation complex. These companies use systematic thinking to solve the practical problems of the company and promote the sound development of the entire society. Another example is the floating building designed by the Dutch architect Koen Olthuis, which is also a case of huge innovations in human systematic thinking. Koen Olthuis is one of the founders of Waterstudio.NL. He is an architect with forward-looking vision and innovative thinking. Waterstudio.NL specializes in building floating buildings on water in response to climate change, sea level rise, floods and urbanization, and uses the concept of water as a building foundation to change the global urban construction model. At present, the earth is facing global warming and rising sea levels, which will affect cities all over the world. Koen Olthuis addresses the challenge of climate change from the perspective of an architect, while taking into account the integration of climate change, urban planning and architecture. This spark of systematic thinking has been inspired into a creative flame.He designed a series of floating buildings (pictured below):

The cultivation and training of systemic thinking requires corresponding changes in educational content and methods. In school education, cultivate students’ problem awareness, understand needs, clarify goals, think methods, use their brains, and then speak. At the same time, cultivate students’ ability to observe and summarize, understand and master the whole picture from point to surface, and improve study and work efficiency. In addition, students should cultivate the habit of summarizing and iterating. The experience and lessons should be reviewed in time, repeated thinking, repeated extraction and summary, and finally qualitative change.

4.3 Entrepreneurial thinking

In the era of artificial intelligence, many jobs will disappear, but it also contains new job opportunities. Entrepreneurship is the key to being able to stand out in the digital age.

One is to establish new companies and enterprises and provide new jobs. Some time ago, Alibaba released a job advertisement, which aroused the interest of many people. In this advertisement, Alibaba is recruiting professional robot breeders with high salaries and freedom of work. But you must know how to feed scientifically, be able to customize professional recipes, "feed" the robot with knowledge, and help the robot to play with the language, and

pass the exam without pressure. With the development of artificial intelligence, more and more such new positions will appear. Only with entrepreneurial spirit can these new positions be provided.
There is also a new type of entrepreneurship, which is to innovate within the original company, open up new areas that have not been controlled by artificial intelligence, and bring value to the company in new ways. For example, General Electric was a famous manufacturing company in the last century. However, with the advent of the artificial intelligence era, traditional manufacturing has been greatly impacted, and a large number of GM employees are in danger of being replaced by machines. General Electric's management relied on entrepreneurial thinking to promote the transformation of the company, with technology and service as its main business, rejuvenating in the 21st century, and avoiding the emergence of employee layoffs.

4.4 Cultural agility

Cultural agility refers to the ability to respond to and adapt to different cultural environments, as well as the ability to communicate and understand with people from different cultural backgrounds. With the development of technology, the world is getting smaller. We need to communicate with people from different cultural backgrounds, and the possibility of misunderstanding increases. In a multicultural system, those who can easily cross different cultural boundaries are more likely to succeed.
There is an "ALS Ice Bucket Challenge" in the United States, the purpose of which is to collect donations for people who are gradually freezing by participating in the ice bucket challenge. In the activity, participants have to pour a bucket of ice water on themselves. In the United States, no one thinks this is a problem. But when the organizer carried out an event in India, the act of watering caused great disgust among the Indians. Because India is extremely short of water, Indians think that watering on the body is a great waste. Under the pressure of public opinion, the organizer replaced the watering process with donating a bag of rice. If the organizer has cultural agility, he can take this into consideration in advance and prevent himself from becoming passive.
How can we cultivate cultural agility? The first is to strengthen cross-cultural knowledge education, set up relevant courses to consciously traditional cultural knowledge of different countries and regions, and understand the background of different cultural systems; the second is to provide experiential learning methods to exercise the educated's cross-cultural communication skills. Internships in companies or enterprises, or participating in various topic projects, can better cultivate cultural agility in specific practice.
Regardless of whether it is the cultivation of three kinds of learning literacy or four kinds of cognitive abilities, education methods are required to be reformed. As the main body in the era of artificial intelligence, we must maintain a correct learning attitude. In addition to more experiential learning, we must maintain a lifelong learning attitude. The world is changing rapidly, and technology is advancing at a rapid pace. It has become a basic requirement to live and learn to grow old. Only by opening the lifelong learning model can we adapt to the new era and avoid being replaced by artificial intelligence. Education can adopt the method of vigorously developing MOOC to provide sufficient educational resources and ensure lifelong learning.
The advent of the era of artificial intelligence is unstoppable, and education needs to undergo in-depth changes in order to cultivate talents who can adapt to the new era and have the ability to defend against robots. At the same time, in the era of artificial intelligence, we must maintain a lifelong learning attitude, live and learn. Only in this way can we adapt to the ever-changing new era.

ACKNOWLEDGMENTS

2020 Scientific Research Project of Liaoning Provincial Department of Education (J2020095).

2021 Economic and Social Development Research Project of Liaoning Province (20211slwzzkt-025)

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
[4] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[6] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
3. Embedding AI in society: ethics, policy, governance, and impacts (2023)	AI and Society (Article from : Springer Science and Business Media Deutschland GmbH)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

AI & SOCIETY (2023) 38:1267–1271
https://doi.org/10.1007/s00146-023-01704-2

EDITORIAL

Embedding AI in society: ethics, policy, governance, and impacts

Michael Pflanzer[1,2] · Veljko Dubljević[2,3] · William A. Bauer[3] · Darby Orcutt[2,4] · George List[5] · Munindar P. Singh[6]

Received: 29 May 2023 / Accepted: 31 May 2023 / Published online: 24 June 2023

1 Introduction

Artificial Intelligence (AI) is fast becoming a ubiquitous part of our lives. From autonomous vehicles and AI personal assistants to computer-assisted surgery and automated trading systems, we are becoming increasingly reliant upon AI to facilitate decision-making and manage our personal and professional lives. In all these cases, AI promises improvements in efficiency, productivity, and/or safety. However, AI does not simply, automatically, and seamlessly integrate into our daily lives and social institutions. Rather, it directly reshapes social, cultural, and economic structures and affects the lives of individual citizens in profound and often tacit, unpredictable, or morally questionable ways. AI systems have the potential to reweave or even disrupt our socioeconomic fabric, impacting not just our productivity and safety but also our autonomy and dignity.

In recognition of AI's profound potential for benefit and harm, we organized a multidisciplinary symposium that sought a deeper and more holistic understanding of how AI does and should shape societal activity. This Special Issue on Embedding AI in Society encapsulates those findings. Some of these papers were first presented at the symposium before being submitted for the Special Issue, while others were submitted directly for consideration.

Our focus was on research on the social, political, and ethical dimensions of AI. The articles selected revolve around four common themes, some with a conceptual focus and some with an empirical focus:

(1) the relationship between humans and AI,
(2) the ethical principles of AI,
(3) the ethical issues related to the implementation and use of AI, and
(4) the value of domestic and international regulatory frameworks for AI.

Written by a diverse set of scholars, the articles discuss AI across multiple domains, including autonomous vehicles, healthcare robots, policing algorithms, and AI personal assistants.

Below is a synopsis of the contributions in the assembled articles, with an objective of helping readers identify conceptual connections among the arguments presented.

2 Human–AI relationships

The first focus is on human–AI relationships. AI has the potential to transform various aspects of our lives including social interactions, work, and personal identity (cf. Pflanzer et al. 2022). As AI becomes ubiquitous and more advanced, it will undoubtedly alter relationships among humans, humans with AI, and between humans and their environments. The interactions among humans is of particular importance, as AI may obviate many needs to interact. Undoubtedly, it will change our personal and professional lives by automating many jobs and changing the nature of our interactions. One article examines the vocational implications of AI technology in the field of medicine (Kempt et al. 2022). The authors illuminate potential disagreements between physicians and AI-based decision support systems,

✉ Veljko Dubljević
veljko_dubljevic@ncsu.edu

1 Communication, Rhetoric and Digital Media Program, North Carolina State University, Raleigh, NC, USA

2 Science, Technology and Society Program, North Carolina State University, Raleigh, NC, USA

3 Department of Philosophy and Religious Studies, North Carolina State University, Raleigh, NC, USA

4 NC State University Libraries, North Carolina State University, Raleigh, NC, USA

5 Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC, USA

6 Department of Computer Science, North Carolina State University, Raleigh, NC, USA

and also discuss moral responsibility within a more automated clinical work environment.

Personal interaction is the focus of an important subtheme. AI is changing how we communicate and interact with each other through social media, chatbots, and other digital technologies. As AI becomes more sophisticated, it will further shape how we relate, raising important questions about social norms, privacy, and human connections. Grandinetti (2021) examines transparency in the context of Facebook and TikTok to show how AI is becoming embedded. Grandinetti sees AI as a material-discursive apparatus, in that it creates implicit teams of humans and machines that rely on discursive techniques and changing material arrangements. Haque et al. (2023) also examine the effects of AI on social networks by designing a social simulation to analyze the effects of content sharing on polarization and user satisfaction. They conclude that (1) user tolerance slows down polarization but lowers satisfaction; (2) higher selective exposure leads to higher polarization and lower user reach; and (3) both higher tolerance and high exposure lead to a more homophilic social network.

AI also has the potential to shape personal identities—to change the way we see ourselves as well as our place in society. For example, AI may enhance our cognitive abilities, alter our memories, or create entirely new forms of augmented intelligence. These possibilities raise important questions about what it means to be human and how human characteristics should be defined. Munn and Weijers (2022) explore the notion that AI chatbots may become digital friends, asserting that many people see these chatbots as their *best* friends. The authors examine the implications of discontinuing access or removing features. They conclude that lawmakers should endeavor to legally protect people from the adverse effects of losing their "digital friends." The relationships between humans and AI will continue to have significant impacts on our personal and social lives. For example, as is now well known, AI-powered decision-making systems can perpetuate bias and discrimination or even manipulate people's behavior. AI systems are increasingly operating autonomously, outside the sphere of direct human oversight. The authors assert that we should be cognizant of these impacts and work to shape AI in ways that helps it align with broadly shared human values and promote the well-being of all citizens.

3 The ethical principles of AI

It is commonly asserted (see, e.g., Noble and Dubljević 2022) that we should consider the societal impact of AI implementation in the context of ethical values. Unsurprisingly, ethical principles of AI are a major theme for many of the authors whose work appears here. For example, Slota et al. (2021) conducted interviews with 26 stakeholders to explore the challenges of AI, including the distribution of agent empowerment and the difficulty of creating accountable systems. They propose the creation of accountable sociotechnical systems (cf. Chopra and Singh 2021) that can be challenged, interrogated, and adjusted to prevent unjust risk. Such systems must not only demonstrate agency, but also be transparent. Andrada et al. (2022) attempt to classify forms of transparency in human–technology interactions. They conclude that all forms of transparency should be considered when designing ethical AI systems.

AI also affects fairness. Like transparency, AI's social impact is an important ethical principle to consider and debate. Maas (2022) examines fairness by looking at power asymmetry among stakeholders—including those who shape AI (such as developers) and those who are affected by it (such as users). Maas bases the analysis on the concept of domination and suggests that external auditing and design-for-value approaches (see, e.g., Liscio et al. 2022) can mitigate the adverse effects of asymmetrical power. For example, in the context of transportation, Gaio and Cugurullo (2022) suggest that society should prioritize mobility justice over policies that focus on single transportation modes. They argue their case using societal goals of proximal cities and urban containment. Finally, Yazdanpanah et al. (2022) suggest a comprehensive research agenda to support the advancement of responsible AI. Like some of our prior work (see, e.g., Singh 2022), they argue that the rollout of any autonomous system should not only follow a demonstration of trustworthiness, but also an explanation of how the AI responsibly satisfies a societal need.

The normative nature of AI is also a frequent topic, as it can make decisions which affect humans in the real world. Several authors argue there is a need to develop and evaluate ethical theories about what makes actions morally right or wrong. Normative ethics is situated between metaethics (which asks whether ethical decision-making is cognitive or non-cognitive, whether moral values objective or subjective, and so on) and applied ethics (which asks, for example, whether abortion is ethically permissible, or whether war is ever justified). In some of the earliest discussions of AI ethics, a primary question asked was which kind of ethical theory should be implemented (so-called 'machine ethics'). As some of us have noted in earlier work (see, e.g., Dubljević 2020 and Coin and Dubljević 2022), it is important to ask whether AI should make moral decisions like a consequentialist, a deontologist, a virtue theorist, or more simply on case-by-case basis. These questions continue as subjects of intense debate, which some of the articles address.

For example, Begley (2023) argues that normative ethics should not be the beginning of philosophical investigations. He suggests a non-methodological approach which proceeds on a case-by-case basis. The key to this approach

is to ask ethical questions which are meant to spur investigations. Stenseke (2021) outlines a method of implementing ethics in machines that follows the core features of virtue ethics. Following a critical evaluation of the challenges of extending virtue ethics beyond theory into implementation, Stenseke proposes a solution that includes moral functionalism, bottom-up learning, and eudaimonic reward. They conclude by presenting a comprehensive framework for developing artificial virtuous agents and discussing how to implement them into moral environments. Kaluža (2022) explains the shortfalls in addressing the challenge of the "filter bubble" and suggests that the better adaptation method would be habitual. Kaluža then shows that, although habitual adaptation of algorithmic personalization is in contrast with society as it stands, it could explain the adoption and stubbornness to stick to certain kinds of information within an isolated social chamber. Haque et al. (2023) address similar challenges but from the perspective of social simulation.

4 Ethical implementation of AI

The third theme we identify is ethical implementation. It is similar to the second but distinct in that it focuses on specific contexts in which the ethical principles discussed previously are salient. While these articles contextualize their discourse with established ethical principles, their focus on specific domains ties these articles together. Regarding warfare applications, for example, Omotoyinbo (2022) argues that smart robotic soldiers would help address moral challenges of warfare. However, the author also remarks that this approach is extreme and that there are inherent issues with replacing humans with robots (e.g., ethical principles such as responsibility and accountability).

Chatbots are another example. They have become a major focus in recent discussions of AI ethics. Chat-GPT and similar applications are creating major impacts. Fyfe (2022) examines their use in education. Fyfe asked students to use OpenAI's GPT-2 for a final writing assignment and to later reflect upon the ethical implications of utilizing AI chatbots while writing. He used these student reflections to consider the larger conversation of the ethical use of AI and language models. Inasmuch as Chat-GPT and similar applications have risen in visibility since this special issue was developed, we are eagerly following the emergent conversation about the ethical implementation and regulation of such technologies.

Many of the manuscripts also focus on the disciplines that are most likely to be drawn into the ethical AI debate. Examples are the regulatory and legal debates about the implementation and use of AI applications. Novelli (2022) justifies a claim that AI entities should be given personhood, demonstrating the potential liability and harmful behavioral concerns that might arise if this is not done. He also discusses other potential legal ramifications of personhood like contracts and lawsuits. Similarly, Jenkins et al. (2022) lay out a two-phase framework for assessing the consequences, good and bad, of AI systems by examining their use in journalism, criminal justice, and the law. They argue that the legal system is likely to provide much commentary on ethical principles such as justice, fairness, accountability, and responsibility.

5 Calls for domestic and international regulation of AI

Anthropologists, sociologists, and other researchers in related disciplines also focus on these principles and use them to rally policymakers and regulators to responsibly consider the ethical dimensions of AI. Freitas and colleagues (2022), for example, explore the use of AI to characterize neighborhood income and socioeconomic characteristics in urban environments. They suggest that policymakers and politicians could be using such models to justify the benefits of gentrification. They cite the ability of these models to examine the effects of economic and public health crises insofar as urban spaces are concerned. The authors lay out some of the benefits of integrating AI models into the decision-making process. They assert that AI-based models will enable scholars in the humanities to better articulate research questions.

Democratization of AI is another important subtheme. Some articles address questions about how to implement transparency, fairness, justice, and responsibility, with debates over AI's social impacts arguing for the democratization of AI to better realize those principles. Himmelreich (2022) examines the call to democratize AI, arguing that it does not meet legitimization demands, introduces redundancies in the governance of AI, and causes various injustices. However, Himmelreich proposes a better way to democratize AI that avoids the identified problems: Rather than merely focus on fostering increased participation, efforts to facilitate democratization should instead enrich and improve existing infrastructure.

Several of the articles examine the impact of using AI for international affairs. Borsci et al. (2022) examine the European Union Commission's whitepaper on AI and identify two issues with implementation: (1) lack of EU vision and methods to drive decisions at lower levels of government, and (2) support for the diffusion of AI in society. They suggest that research, encouraged by regulators, should seek to see how socioeconomic differences could lead to a fractured AI market. Bisconti et al. (2022) explore ways to maximize the benefits of interdisciplinary cooperation in AI research

groups and explain that this is a temporal urgency given the "AI Act" and other initiatives being undertaken by the EU Commission. They conclude by identifying law enforcement, criminal justice, and social robotics as relevant fields that may benefit from their methodology. Hassan (2022), on the other hand, explores AI governance and regulation gaps in the context of African nations. He demonstrates the existence of Euro-American biases within AI ethics scholarship and identifies a need to consider *non*-Eurocentric perspectives regarding AI ethics, specifically advocating ethical principles from an African perspective.

6 Concluding remarks

The landscape of AI ethics, and more broadly AI in society, is vast in methods, questions, and proposals. The papers in this special issue collection reflect this vastness while raising as many questions as they answer. We certainly encourage more work on the highlighted themes. That said, going forward we also suggest that researchers focus more attention on political power and policy making processes (cf. Dubljević 2019), as well as the possibilities of shared values across pluralistic societies. The questions which need to be explored in the future include: What are the commonalities and differences? And, should we work towards greater moral unity or does the friction of disunity generate new and better ideas? The challenges of trust (see e.g., Singh and Singh 2023), are another area which needs more sustained scholarship. Finally, we see room for more metaethical debate in the discussion of AI ethics, asking: How much hope or trust should we have that we can solve AI ethical problems? How can moral values be implemented and realized in artifact–human relations? And what methods of investigation and ways of knowing are likely to resolve value conflicts?

Acknowledgements The work on this special issue was supported by the award Rabb Science & Society grant 'Embedding AI in Society' 2020. Special thanks to Nora Edgren for her preparatory and early editorial work.

References

Andrada G, Clowes RW, Smart PR (2022) Varieties of transparency: exploring agency within AI systems. AI Soc. https://doi.org/10.1007/s00146-021-01326-6

Begley K (2023) Beta-testing the ethics plugin. AI Soc. https://doi.org/10.1007/s00146-023-01630-3

Bisconti P, Orsitto D, Fedorczyk F et al (2022) Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology. AI Soc. https://doi.org/10.1007/s00146-022-01518-8

Borsci S, Lehtola VV, Nex F et al (2022) Embedding artificial intelligence in society: looking beyond the EU AI master plan using the culture cycle. AI Soc. https://doi.org/10.1007/s00146-021-01383-x

Chopra AK, Singh MP (2021) Accountability as a foundation for requirements in sociotechnical systems. IEEE Internet Comput (IC) 25(6):33–41. https://doi.org/10.1109/MIC.2021.3106835

Coin A, Dubljević V (2022) Using algorithms to make ethical judgments: METHAD vs. the ADC model. Am J Bioeth 22(7):41–43

Dubljević V (2019) Neuroethics, justice and autonomy: public reason in the cognitive enhancement debate. Springer, Heidelberg Germany

Dubljević V (2020) Toward implementing the ADC model of moral judgment in autonomous vehicles. Sci Eng Ethics. https://doi.org/10.1007/s11948-020-00242-0

Freitas F, Berreth T, Chen Y, Jhala A (2022) Characterizing the perception of urban spaces from visual analytics of street-level imagery. AI Soc. https://doi.org/10.1007/s00146-022-01592-y

Fyfe P (2022) How to cheat on your final paper: assigning AI for student writing. AI Soc. https://doi.org/10.1007/s00146-022-01397-z

Gaio A, Cugurullo F (2022) Cyclists and autonomous vehicles at odds. AI Soc. https://doi.org/10.1007/s00146-022-01538-4

Grandinetti J (2021) Examining embedded apparatuses of AI in Facebook and TikTok. AI Soc. https://doi.org/10.1007/s00146-021-01270-5

Haque A, Ajmeri N, Singh MP (2023) Understanding dynamics of polarization via multiagent social simulation. AI Soc. https://doi.org/10.1007/s00146-022-01626-5

Hassan Y (2022) Governing algorithms in the south: AI and sustainable development in Africa. AI Soc. https://doi.org/10.1007/s00146-022-01527-7

Himmelreich J (2022) Against "Democratizing AI." AI Soc. https://doi.org/10.1007/s00146-021-01357-z

Jenkins R, Hammond K, Spurlock S et al (2022) Separating facts and evaluation: motivation, account, and learnings from a novel approach to evaluating the human impacts of machine learning. AI Soc. https://doi.org/10.1007/s00146-022-01417-y

Kaluža J (2022) Far-reaching effects of the filter bubble, the most notorious metaphor in media studies. AI Soc. https://doi.org/10.1007/s00146-022-01399-x

Kempt H, Heilinger JC, Nagel SK (2022) "I'm afraid I can't let you do that, Doctor": meaningful disagreements with AI in medical contexts. AI Soc. https://doi.org/10.1007/s00146-022-01418-x

Liscio E, van der Meer M, Siebert LC, Jonker CM, Murukannaiah PK (2022) What values should an agent align with? An empirical comparison of general and context-specific values. J Auton Agents Multi-Agent Syst (JAAMAS) 36(1):23. https://doi.org/10.1007/s10458-022-09550-0

Maas J (2022) Machine learning and power relations. AI Soc. https://doi.org/10.1007/s00146-022-01400-7

Munn N, Weijers D (2022) Corporate responsibility for the termination of digital friends. AI Soc. https://doi.org/10.1007/s00146-021-01276-z

Noble SM, Dubljević V (2022) Ethics of AI in organizations. In: Nam CS, Lyons J (eds) Human-centered artificial intelligence. Academic, Cambridge MA, pp 221–240

Novelli C (2022) Legal personhood for the integration of AI systems in the social context: a study hypothesis. AI Soc. https://doi.org/10.1007/s00146-021-01384-w

Omotoyinbo FR (2022) Smart soldiers: towards a more ethical warfare. AI Soc. https://doi.org/10.1007/s00146-022-01385-3

Pflanzer M, Traylor Z, Lyons J, Dubljević V, Nam CS (2022) Ethics of human-AI teaming: principles and perspectives. AI Ethics. https://doi.org/10.1007/s43681-022-00214-z

Singh MP (2022) Consent as a foundation for responsible autonomy. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI). 36(11):12301–12306. https://doi.org/10.1609/aaai.v36i11.21494

Singh AM, Singh MP (2023) Wasabi: a conceptual model for trustworthy artificial intelligence. IEEE Comput 56(2):20–28. https://doi.org/10.1109/MC.2022.3212022

Slota SC, Fleischmann KR, Greenberg S et al (2021) Many hands make many fingers to point: challenges in creating accountable AI. AI Soc. https://doi.org/10.1007/s00146-021-01302-0

Stenseke J (2021) Artificial virtuous agents: from theory to machine implementation. AI Soc. https://doi.org/10.1007/s00146-021-01325-7

Yazdanpanah V, Gerding EH, Stein S et al (2022) Reasoning about responsibility in autonomous systems: challenges and opportunities. AI Soc. https://doi.org/10.1007/s00146-022-01607-8

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
4. Social impact of AI on the organization of higher Education in electrical engineering and in society (2025)	2025 34th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE) (Article from : IEEE)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

2025 34th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE) | 979-8-3315-0290-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/EAEEIE65428.2025.11136314

Social Impact of AI on the Organization of Higher Education in Electrical Engineering and in Society

Inácio Fonseca [1, 2]
[1] *Polytechnic University of Coimbra*, Rua da Misericórdia, Lagar dos Cortiços, S. Martinho do Bispo, 3045-093 Coimbra, Portugal
inacio@isec.pt

Nuno Cid Martins [1,2]
[2] *Research Centre for Asset Management and Systems Engineering (RCM2+), Polytechnic University of Coimbra*, Rua Pedro Nunes, 3030-199 Coimbra, Portugal
ncmartin@isec.pt

Fernando Lopes [1,3]
[3] *Instituto de Telecomunicações-Coimbra,* Pinhal de Marrocos, Pólo II DEEC, 3030-290 Coimbra, Portugal
flopes@isec.pt

***Abstract*— The way we access information has evolved continuously throughout human history. In 2022, the first global LLM appeared, ChatGPT. This innovation enabled direct interaction and precise answers to questions asked in natural written language. By 2025, the model evolved to support interaction through verbal natural language, with integration into systems such mobile devices, further expanding access to information. Since ChatGPT, many AI tools have emerged in different contexts. In Higher Education Institutions, there is a high level of concern about fundamental aspects of education, three of which stand out: the development of students' critical thinking, the real acquisition of essential competences, and engagement. Several studies have been carried out on the application of LLM in Higher Education. A short review will be presented. In society, changes are starting to be implemented and are expected to have a significant impact. Examples or trends on changes in the healthcare and justice systems will be introduced. The paper aims to show the current state of LLMs, their impact on education, on society, and consequently on the procedures within most present and future professions. To support this, a student survey was carried out, presenting their perception of the impact of LLM systems.**

Keywords — HEI, LLM and SLM evolution, Future professions, AI impact on Education, AI impact on society.

I. INTRODUCTION

We are amid a technological change in the world, above all due to the generalised race by the main world blocks to obtain Artificial General Intelligence (AGI) systems. As with previous revolutions in the past, the aim is to improve the efficiency of societies, and freeing workers from repetitive tasks, which in the past focused on replacing tasks with less added value. This time round, the revolution could have the effect of replacing human labourers in tasks that require higher levels of intellectual capacity like in education. The benefit is well established, robots, Artificial Intelligence (AI) systems, that usually work 7x24h performing equally well - but consuming energy. What social changes will occur during the process of empowering society? Is the education sector undergoing a structural and social transformation, or has this shift been underway for years? Is it possible to develop more efficient educational models [1]? Are Higher Education Institutions (HEI) tied to conservative models, the result of roles from national accreditation agencies and internal organizational structures? Is Higher Education (HE) capable of effective restructuring? Will the age of human resources influence the ability to make these changes? Is the energy consumption of LLM (Large Language Models) systems a bottleneck for their evolution and a critical problem for sustainability? Cybersecurity of LLM systems and adversarial AI are issues to address? What is the impact of teaching LLMs with documents containing conflicting or incorrect content? These are open questions (some of them are or could be research topics) that will only be answered for the most part over the next (few) years. Several studies have been carried out on the application of AI or LLM models in Higher Education and society [2-6].

The energy transition is another aspect that is on the agenda, leading to even greater electricity needs. Surprisingly, however, every time a human worker is replaced by energy-consuming equipment, the need for energy resources such as electricity increases. The high consumption of AI systems is well known, with one LLM system needing to consume electricity equivalent to thousands of homes to respond worldwide, not to mention the need to cooling. In a way, AI is competing for resources that are necessary for human life. But its use can respond and bring advantages to society in fundamental areas such as medicine, power generation, nanotechnology, physics, mobility, etc. On the other hand, it can lead to the generation of false content, the artificial generation of content implicating people in illegal activities, which can severely disrupt the organisation of societies. Nowadays (maybe in a few years), anyone who is in their home, can be labelled anywhere in the world as carrying out an activity, spreading a thought, etc., even if they're not there.

Europe has a history of regulating and creating legislation on many topics in the interest of the safety of society in general. The reality is that citizens can hardly isolate themselves in a digital bubble. Therefore, European citizens are currently exposed to these tools produced outside the European continent without realising exactly what they are exposed to (or not). Whenever Europe has relocated production centres to other continents (via companies), it has no longer been aware of the production conditions and degree of exposure. Obviously, if the economic indicators are doing well, societies and ordinary citizens don't care, because in their normal lives they only care about (over)living, and if their lives are easier, they'll give a favourable opinion. But, despite being in the 21st century, the world appears to be what it has always been, a jungle, where the strongest and most capable survive. Of course, there are societies that are organised in a coherent way, with the most capable having the responsibility to support the rest of the citizens, maintaining a more interesting harmony. Certainly, in addition to these issues, we can and should also consider the religious question, which adds some issues to societies, some complex, and others that make human beings more on the path of kindness and social and moral ethics. This last aspect can lead societies to live in some harmony, if the resources are enough for their survival.

Can we organise societies where machines (robots, see_ youtube.com/watch?v=YHFnGwo5wzY, AI, AGI) do most of the work and all humans have a good quality of life, spending

979-8-3315-0290-4/25/$31.00 ©2025 IEEE

more time with their individual and collective wellbeing, getting to know more persons and more places and offer this to 100% of the population? These are important questions for all citizens, but they fall under the responsibility of political powers, religious concepts and, of course society in general.

This paper is organised into five sections. Section I is the introduction. Section II presents the evolution of LLM models' performance. Section III summarises the social impact of LLMs on education and society. Section IV shows the results of a student survey and section V presents the conclusions and future trends.

II. Performance Evolution of AI-LLM models

This year (2025), there were new developments in the field of AI: deepseek (deepseek.com), grok3 (xAI), Qwen (alibaba) and Manus (manus.im) positioned as a generalAI (AGI). This only shows the competitiveness and effort put into this area.

Example metrics for model evaluation are Graduate-Level Google-Proof Q&A Benchmark (GPQA) [7], Humanity's Last Exam (HLE) [7] and Measuring Massive Multitask Language Understanding (MMLU) [7]. Fig. 1 shows AI model ranking.

#	Model	Lab	Date	GPQA	#	Model	Lab	Date	HLE
1	o3-preview	OpenAI	Dec/2024	87.7	1	o3	OpenAI	Apr/2025	24.9
2	Claude 3.7 Sonnet	Anthropic	Feb/2025	84.8	2	Gemini 2.5 Pro	Google DeepMind	Mar/2025	18.8
3	Grok-3	xAI	Feb/2025	84.6	3	Agentic-Tx	Google DeepMind	Mar/2025	14.5
4	Gemini 2.5 Pro	Google DeepMind	Mar/2025	84	4	o4-mini	OpenAI	Apr/2025	14.28
5	o3	OpenAI	Apr/2025	83.3	5	o3-mini	OpenAI	Jan/2025	14
6	o4-mini	OpenAI	Apr/2025	81.4	6	Gemini 2.5 Flash Preview	Google DeepMind	Apr/2025	12.1
7	o1	OpenAI	Dec/2024	79	7	Claude 3.7 Sonnet	Anthropic	Feb/2025	8.9
8	Gemini 2.5 Flash Preview	Google DeepMind	Apr/2025	78.3	8	o1	OpenAI	Dec/2024	8.8
9	Seed-Thinking-v1.5	ByteDance	Apr/2025	77.3	9	DeepSeek-R1	DeepSeek-AI	Jan/2025	8.6
10	o3-mini	OpenAI	Jan/2025	77	10	R1 1776	Perplexity	Feb/2025	8.6

Fig. 1. AI model Ranking [8].

Fig. 2 shows the evolution of the models according to the MMLU metric, including human performance. The graph shows the tendency for these systems to achieve better performance over time, and already on an order of magnitude with human capacity and in some cases higher.

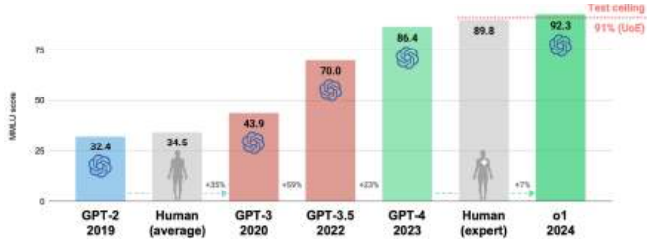


Fig. 2. MMLU GPT Performance evolution [8], September 2024.

Fig. 3 shows the processing speed of humans versus LLM models. Humans are slower in all aspects, such as writing, listening, speaking and reading. There are also differences in thought activity, but only about knowledge that the LLM knows versus the human, with the LLM performing better. This is however not true in areas where humans must innovate and when involving knowledge that does not yet exist in digital form - innovation capacity.

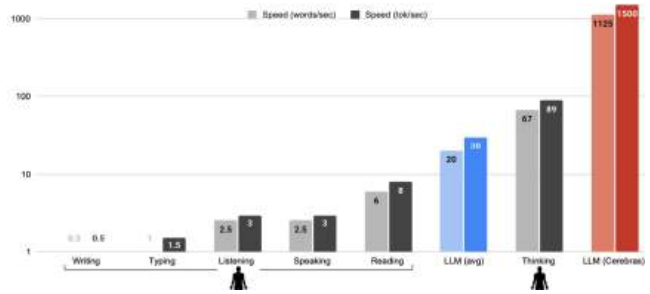


Fig. 3. Performance of LLM systems [8], February 2025.

Fig. 4 shows the positioning of the various models using the GPQA metric, and the relevance when we compare the different models, also including human performance, namely the average human, and the human with the ability to achieve a PhD in very specialised areas. According to this metric, the most capable humans have already been overtaken by the best LLM systems.

The information in Fig. 4 also shows that ChatGPT (unpaid version) performs much less well than the paid version GPT-4o (20$/month) and this in turn performs much less well than the GPT-o1(3) version (200$/month). This shows that companies operating in cutting-edge areas will need to allocate financial resources to access the best models, and financial capacity is obviously an important aspect of competitiveness in accessing technology and then the market.

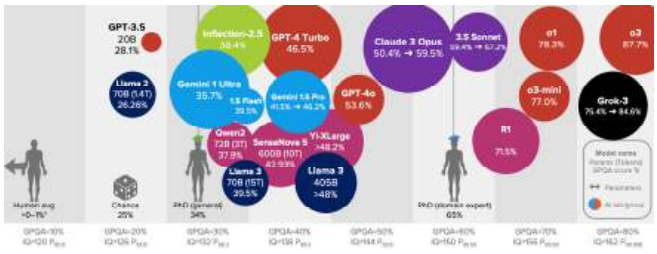


Fig. 4. GPQA Performance [8], February 2025.

Fig. 5 shows the performance of LLMs in the light of another metric: A benchmark that evaluates LLMs using 601 NYT Connections puzzles extended with extra trick words. The metric indicates the number of puzzles that were successfully completed. This graph emphasises the fact that ChatGPT - the free version - is practically in last place and the latest version o1-pro is in first place (82.5%). o1-pro performs about 2.1 better than the deepseek-R1 model. The deepseekR1 model has so far been freely available. The European (French) Mistral Large 2 model has a performance of 12.6 %.

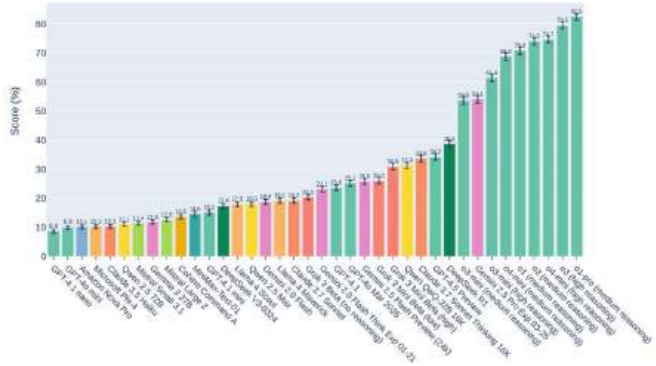


Fig. 5. Benchmark that evaluates LLMs using 601 NYT Connections puzzles extended with extra trick words [9], February 2025.

Fig. 6 shows the performance of some of the models present in Fig. 5, in a test with puzzles extended with extra trick words, in which humans players participated. The humans who took part in this challenge were selected and are likely to perform better than the general population.

Another independent organisation [10] publishes regularly information on LLM systems performance using an Artificial Analysis Intelligence (AAI) index, that incorporates 7 metrics: MMLU-PRO, GPQA Diamond, HLE, LiveCodeBench, SciCode, AIME (American Invitational Mathematics Examination), MATH-500 calculated as: a) General Reasoning and Knowledge (50%): represents broad

knowledge and reasoning capabilities across academic and scientific domains. Equally weighted for MMLU-Pro, HLE, and GPQA Diamond; b) Mathematical Reasoning (25%): combines general mathematical problem-solving with advanced competition-level mathematics. Equally weighted for MATH-500 and AIME 2024 [7]; c) Code Generation (25%): tests Python programming for scientific computing and general competition-style programming. SciCode and LiveCodeBench.

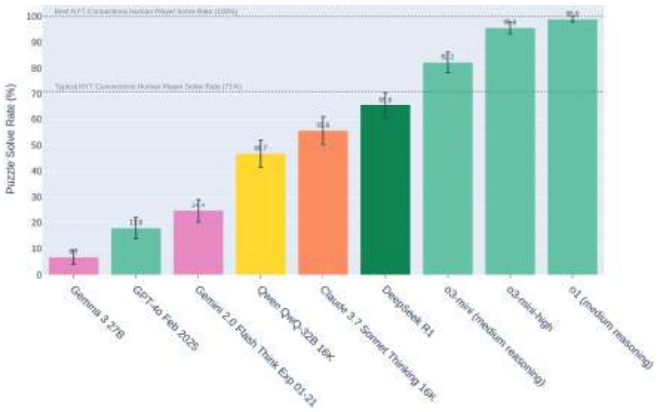


Fig. 6. Benchmark that evaluates LLMs using 601 NYT Connections puzzles extended with extra trick words and human participation [9], February 2025.

Fig. 7 shows the relative performance of USA versus Chinese models using the AAI index. Fig. 8 shows ranking of LLM systems in general.

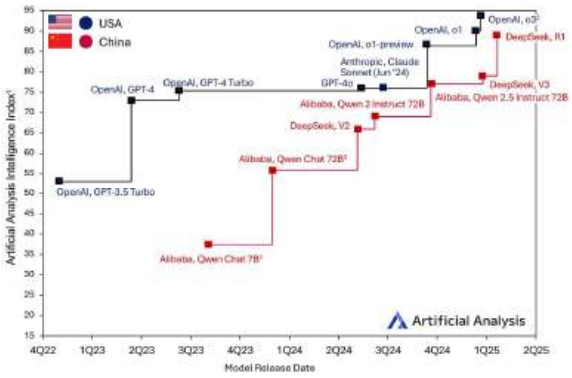


Fig. 7. AAI index Benchmark for Chinese versus USA models [10].

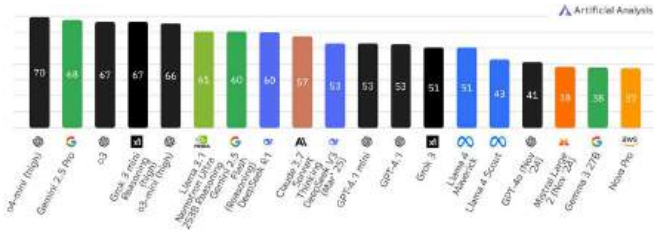


Fig. 8. AAI index ranking for different models [10], February 2025.

Analysing the information presented previously in this section, it can be concluded that LLM systems are, in general, at the time of writing, very consistent and capable of solving problems in areas such as mathematics, automatic generation of code programs (Python, C, web, …), reasoning in academic and research areas. The performance metrics, some of which are recent, from 2025 and previous years, make the research carried out in this direction evident. At this stage, the competition is essentially between China and the USA, with Europe lagging behind. From the point of view of education and society, opting for a specific model can lead to different results depending on the model's position in the ranking. In Europe, several countries are already looking to preserve their linguistic culture and strategic autonomy, such as France, with the Mistral AI standing out; in Spain, the 'Alia' project; and in Portugal, the 'Amália' model starting to be available, showing evident signs of changes in society. The energy issue is extremely important and one answer at this level could be Small Language Models (SLM) [11] which can be trained with less data, are more compact, efficient and don't need massive servers. They are built for speed and real-time performance and can run on smartphones, tablets, or maybe smartwatches. As embedded systems improve and incorporate GPUs (e.g. Jetson Nano), these models can be allocated there and super-specialised in areas of knowledge with high performance, low energy consumption and using fewer computational resources. If we look closely, society itself has a similar organisation with different professions where persons specialise according to their skills.

III. Social Impact of AI in Electrical Engineering Education and Society – Short Review

In this section we will present a short review of scientific works that focus on the theme of the impact in Electrical Engineering and in society. AI tools in Electrical Engineering education reshape departmental organization through automated systems and cross-disciplinary approaches, while impacting faculty roles and raising social equity concerns.

A. Social Impact of AI in Electrical Engineering Education

Many scientific works report that the application of AI streamlines administrative processes by enabling data-driven decision-making and automated task handling, while redefining departmental structures and guiding resource allocation [1-6]. In Electrical Engineering contexts, scientific papers document AI tools that support personalized learning, intelligent tutoring, and automated grading, prompting curricular updates and cross-disciplinary approaches. Meanwhile, these studies note that students experience increased engagement, faculty benefit from reduced routine burdens yet face challenges related to acquiring new technological skills, and administrators must adapt to evolving roles. Reported ethical and social concerns centre on data privacy, equity in access, and shifts in professional identity.

Table I identifies five representative papers of the fifty found on this specific topic.

In [12], the study focuses on exploring AI's potential applications in higher education, emphasizing its role in addressing the longstanding challenge of individualized learning - a key requirement unmet by traditional educational models. Possible implementations of AI include personalized learning, adaptive curricula, virtual educators and automated feedback systems. AI can increase efficiency, accessibility, and effectiveness of learning, but challenges include ethical concerns, equity issues, and resistance to change. Systemic changes include shifts in teaching methods, student-teacher dynamics, and institutional structures. The implementation has impact on personalized learning, efficiency and accessibility, curriculum and teaching, in research and innovation. However, the challenges persist like resistance from educators and students unfamiliar with AI tools, ethical concerns around data privacy, algorithmic bias, the dehumanization of education and inequitable access to AI in institutions and regions.

TABLE I. CHARACTERISTICS OF INCLUDED STUDIES

Study	Study Focus	Methodology	Stage
[12] (2025)	AI implementation in higher education	Theoretical / conceptual analysis	Actively implemented in higher education
[13] (2025)	AI impact on teaching, learning, and research in higher education	Ethnographic study and systematic review	Ongoing implementation and analysis
[14] (2023)	AI impact on traditional education systems and social effects	qualitative research method	Ongoing implementation and analysis
[15] (2025)	Enhance professional training effectiveness in electrical engineering using AI	Theoretical and analytical approach	AI is already being applied in educational contexts
[16] (2024)	Analysing the adoption and implications of Artificial Intelligence in Education (AIED) in higher education	Semi-systematic literature review	HEI exploring several tools (ex: automated feedback systems).

The study in [13] focuses on how AI systems are becoming institutionalized in higher education. Some concerns are related to analysing how AI interacts with and is shaped by cultural norms, examining stakeholders' (e.g., educators, administrators, students) attitudes, beliefs, and ethical/moral concerns about AI's implementation. AI's implementation is influenced by - and reinforces - existing cultural norms, values, and power hierarchies in HE.

In [14], the study focuses on examining the impact of AI on traditional education systems, assessing how AI has transformed educational practices, and evaluating the social consequences of integrating AI into education. Specific areas of research include: how AI's capabilities (e.g., learning, prediction, problem-solving, adaptability) are reshaping education delivery, management, and administrative processes; the current state of AI adoption in educational institutions, including its role in teaching, administrative tasks, and classroom/school management; the broader societal implications of AI in education, such as changes in learning dynamics, equity, and human-AI collaboration.

Authors in [15] describe the main trends in the use of AI in education as being data analysis, automation of assessment, developing skills for the future, virtual learning and personalization of education. The concept of virtual teachers, virtual labs and ethical problems are also discussed. The paper presents examples of real systems like *Coursera Coach*, *Coursera for Campus*, AI detectors for fraud like *Turnitin program*, *Jill Watson,* a virtual teacher developed by Georgia Tech University and *Knewton,* an adaptive learning platform.

The study in [16] presents a semi systematic review on the use of AI in HE. The study categorizes the papers as: 8 about the student perspective, 9 about the educator perspective, 20 from both, 3 from other stakeholders, 7 from the managerial perspective, 5 from the governmental perspective, 3 from the technological perspective, 4 from the external perspective and 9 about the social perspective. Then, the authors present a discussion on the key points, identify 9 groups of research gaps in the analysed papers and discuss challenges for implementing AI in HE: cost, scalability, lack of actionable guidelines, limited AI expertise and data governance.

In Education Delivery, AI systems are used to manage information, personalize learning, and adapt to students' needs. This includes tools for predictive analytics (e.g., student performance forecasting) and automated grading are increasingly common. In Administrative Functions AI streamlines tasks like scheduling, resource allocation, and student tracking. In Teaching and Management AI assists teachers in curriculum design, student engagement, and classroom management. In Broader Adoption AI-inspired systems are now "popular and applied in almost every field, especially in educational institutions", with a focus on enhancing efficiency and adaptability.

B. AI impact on society - thoughts

From society's point of view, several professions could be at risk, particularly if they are partially or completely replaced. If we think about the most important areas of society, we have education, justice, health, food, security, old age, research, etc. We quickly realise that the impact will be enormous from examples in some of these areas: in support for old age, by intelligent humanoid robots; in education by chatbots that interact in any area of knowledge; in justice, by supporting the deciphering of legal decrees and explaining them in a language accessible to citizens; in food, by indicating the most correct food to eat depending on the citizen's activity; in health, by prescribing in the light of symptoms and medical examinations; in security, by monitoring busy public spaces and detecting suspicious activity; in research by supporting programming, determining directions and the state of the art.

A more detailed example, in the justice system, the integration of LLMs in legal proceedings can have a significant impact on the efficiency, fairness and transparency of justice, but it also raises ethical and technical challenges. The benefits are acceleration of proceedings through rapid analysis of documents, consistency in the application of the law between similar cases, prediction of judicial outcomes by analysing history and determining probabilities of success in lawsuits, assisting settlements versus litigation, automatic language translation support, and democratised access to justice, giving citizens basic legal guidance without resources, explaining rights, deadlines and procedures in accessible language. There are already experiments in China, USA and Brazil [17]. Professions of judges and lawyers may change.

The integration of LLMs into healthcare, especially in areas such as automatic prescribing and access to virtual doctors, has the potential to transform the sector (Fig. 9), but requires care to balance innovation, security and ethics.

Electrical Engineering can give a boost to the application of these systems in these and other society areas by using SLM concepts, implementing them in embedded systems, training them and building energy systems that are adaptable to their specific needs, and this could be an educational opportunity.

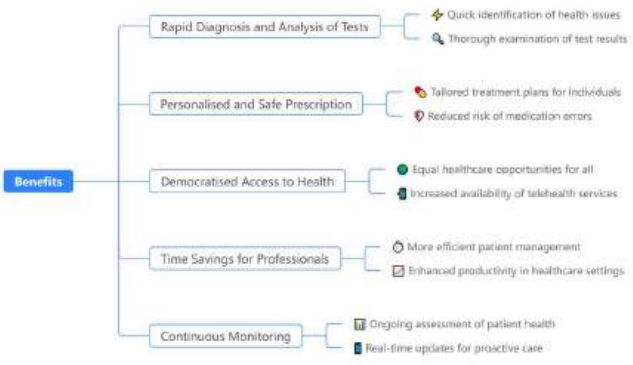


Fig. 9. Example of benefits for transformation in healthcare based on LLM integration.

IV. Students Survey about AI Impact and Discussion

Students from three curricular units (CU) of an Electrical Engineering degree (Undergraduate: Class 1 40 students '*Maintenance and Quality Control*'; Master: Class 2 and Class 3, 30 students each, respectively, '*Applied Information Systems*' and '*Industrial and Business Communications*'), answered to a survey on the impact of IA on education. Questions are shown in Table II and results in Fig. 10 to 14. Analysing Fig. 10 to 14, the answers chosen by the students show the same trend, namely: Question 01: b); Question 02: c); Question 03: c); Question 04: c) and Question 05: c). It seems that the students were cautious in their choices, which shows a certain amount of forethought in the face of the challenge and the changes that might occur.

TABLE II. Survey Questions

Question 01: How can HEIs stay aligned with the rapid evolution of AI and emerging technologies without compromising educational quality? a) Invest in continuous training for teachers and students, ensuring that everyone adapts quickly to new technologies. b) Gradually incorporate technologies, maintaining a balance between technological innovation and traditional teaching methods. c) Create partnerships with technology companies to ensure that the technologies implemented are of high quality and up to date.
Question 02: What is the role of personalisation in higher education using AI, and to what extent should HEIs invest in it? a) Personalisation is essential for improving the student experience, and HEIs should invest heavily in AI-based solutions for this. b) Personalisation should only be implemented in specific areas of teaching, such as tutorials or one-to-one mentoring, but without affecting the traditional format of lessons. c) Personalisation may be a trend, but it must be carefully balanced with the cost and resources available at HEIs.
Question 03: Can implementing AI in HEIs administrative processes improve operational efficiency? a) Yes. AI can optimise administrative processes such as enrolment, resource management and student services, reducing costs and improving efficiency. b) No. Implementing AI in administrative processes can be complex and expensive and will not necessarily bring clear cost benefits. c) Yes, but the implementation of AI should be gradual, starting with simple administrative processes before moving on to more complex areas.
Question 04: Can the use of AI in academic assessments improve accuracy and fairness in student evaluation? a) Yes. AI can offer a more objective and personalised assessment, reducing human biases and providing more detailed feedback. b) No. AI can be too rigid and not adequately consider the nuances of student learning. c) Yes, but it should be used together with human assessment to ensure that AI is only a support tool, not the sole source of the assessment.
Question 05: What is the impact of creating independent data centres at HEIs for storing academic and research data? a) Creating our own datacentres is essential to guarantee security, control over data and independence from external suppliers such as the US, China. b) Creating our own datacentres is very expensive and unnecessary, as cloud services can provide the same security and scalability more cheaply. c) Creating our own datacentres can be a solution for protecting sensitive data, but it should be done in collaboration with other academic or regional consortia, sharing costs and resources.
Question 06: Suggestions for the inclusion of AI during the learning process.

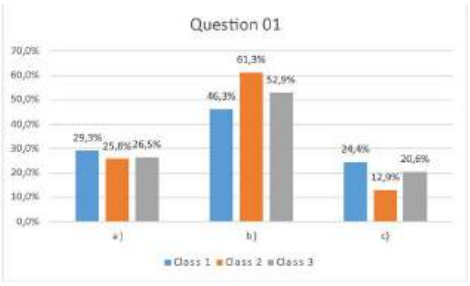


Fig. 10. Results for **Question 01** (Table II).

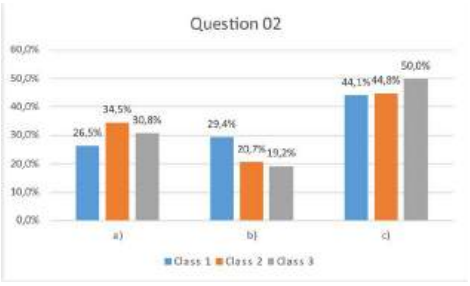


Fig. 11. Results for **Question 02** (Table II).

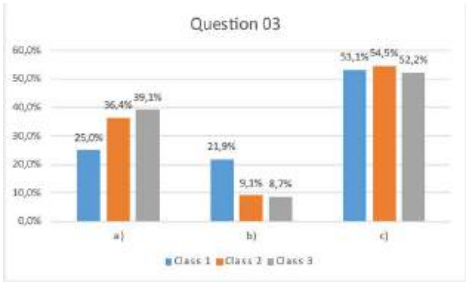


Fig. 12. Results for **Question 03** (Table II).

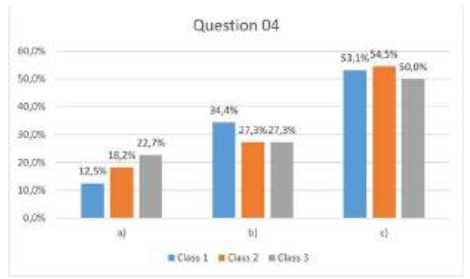


Fig. 13. Results for **Question 04** (Table II).

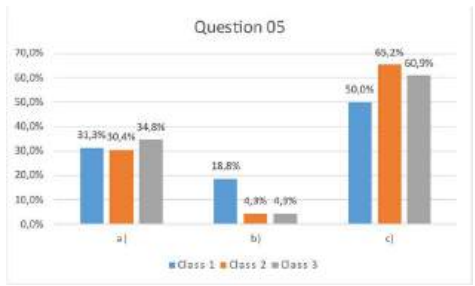


Fig. 14. Results for **Question 05** (Table II).

Table III shows the students' answers to question 06 (in each cell, the students' answers per course unit, Class1 to 3).

The responding students are receptive to use IA in the classroom and use it outside the classroom for self-study, but they maintain a cautious stance and critical thinking, which

will hopefully continue throughout their degree as the LLMs progress but will not yet be able to fully replace the competence of their training and their future job. Another aspect much discussed is 'prompt engineering' and the hallucination of LLMs fostering critical thinking. The point is that new versions have quickly appeared, some of which can solve mathematical and hallucination problems. The speed at which it will evolve (see references in Section II) will make these issues temporary and safer to use in classroom.

TABLE III. SOME STUDENTS' SUGGESTIONS (QUESTION 06, TABLE II)

- Encouraging AI for self-learning, but not in the educational environment 😉. Many of my friends have got into the habit of accepting ChatGPT's words as absolute truth and have lost their critical sense. - Teach students how to use AI as a study aid, helping them understand what kind of prompts to use in class, gradually. - Teachers themselves ask the AI to answer some questions and mention to the students what is less correct so that the students also develop a critical opinion and don't blindly trust the AI. - Its introduction should be gradual and not sudden, allowing students to adapt better. Sudden changes may not always be 100% beneficial.
- For greater inclusion of the use of AI, it will be necessary to train teachers to work with AI and to improve AI in its current state. - AI use in the classroom, always supported by the teacher - Activities carried out with the support of AI where the student must be critical and question the answers provided by AI, to demonstrate that AI, for now, for many areas, should be used as a support tool and doubt the certainty of the answers provided. - AI is a promising technology, but it is still in its infancy and too fallible to be used much in either teaching or assessment. - AI should be seen as a powerful tool that can help a lot in solving the problems presented. In my opinion, the tasks should also be more demanding given the reality we find ourselves in today. - Use AI for code generation tasks, since the whole market is currently using it. I think it's important to know the concepts and how it works. Having to make code is very challenging, and even if the AI is wrong when it comes to generating it, with mistakes we realise why the code suggested by the AI doesn't work.
- Teacher training in the use of AI, incentives and use as a supporting tool. - Explore AI in a classroom context but with teacher supervision.

V. CONCLUSIONS AND FUTURE TRENDS

This paper has presented the evolution of LLMs and benchmark KPI showing a faster improvement over time. The impact on society is very high and is bringing transformations. From society's point of view, some professions could be at risk by partial or total replacement. Students are apprehensive and cautious, but they already use the tools regularly, especially because they know AI tools will be used in the professional environment. Obviously, this will evolve very quickly and may vary in unknown ways and directions. Transformation trends in Higher Education may be relevant as described in section III.A and supported by scientific studies indicated in the reviewed papers. Virtual educators [15], automatic assessment [15], suggest the possible disappearance of the classroom and the emergence of distance learning and online learning platforms (with an official diploma), maybe with exams and knowledge verification taking place in person with a human educator. *Jill Watson*, a virtual teacher developed by Georgia Tech University [15], suggests this could be a reality in the future. What will be lost? Social interaction between peers and generations [16] losing the social learning of the first years of higher education, the parties and the learning of soft skills. To mitigate these problems, political decision-makers could limit the use of these systems to student-workers over the age of completion of higher education studies. These topics are open for future research, because we are in the infancy of these changes, and we may have to wait 10, 20 or 30 years to see the whole picture. More research needs to be carried out on the issues raised and the themes addressed, and into the virtualisation of education, for example to impart lifelong skills using "virtual universities (HEI)" or in developing "universities (HEI) based on robotic educators".

REFERENCES

[1] "Can AI Condense Two Years of Learning Into Six Weeks?", https://www.psychologytoday.com/intl/blog/the-digital-self/202501/can-ai-condense-two-years-of-learning-into-six-weeks

[2] 2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET) | DOI: 10.1109/ITHET61869.2024.10837644

[3] 2022 20th International Conference on Information Technology Based Higher Education and Training (ITHET) | DOI: 10.1109/ITHET56107.2022.10031944

[4] MATHia, Carnegie Learning, https://www.carnegielearning.com/solutions/math/mathia/.

[5] LLM AI Tutors on-line: CK12 (www.ck12.org), Llama Tutor (https://llamatutor.together.ai/), AI Tutor (https://ai-tutor.ai/about-us).

[6] Izzivi Prihodnosti. (2023). AI and Organizational Transformation: Anthropological Insights into Higher Education. Challenges of the Future. Vol.8 No. 3. Doi: 10.37886/ip.2023.007. Link: https://ojs.fos-unm.si/index.php/ip/article/view/131/105

[7] GPQA: A Graduate-Level Google-Proof Q&A Benchmark. arXiv:2311.12022v1. MMLU: Measuring Massive Multitask Language Understanding. arXiv:2009.03300v3. HLE: Humanity's Last Exam Humanity's. arXiv:2501.14249v1. AIME 2024: American Invitational Mathematics Examination.

[8] Alan D. Thompson. (2025). Systematic review in performance and metrics of AI Models. LifeArchitect.ai.

[9] Lechmazur. (2025). Benchmark that evaluates LLMs using 601 NYT Connections puzzles extended with extra trick words. https://github.com/lechmazur/nyt-connections

[10] Artificial Analysis. (2025). Independent AI benchmarking and insights provider. https://artificialanalysis.ai. Document: Artificial-Analysis-State-of-AI-China-Q1-2025.pdf

[11] Zhenyan Lu et all. (2025). Small Language Models: Survey, Measurements, And Insights. https://arxiv.org/pdf/2409.15790

[12] Abgaryan, H., Asatryan, S., & Matevosyan, A. (2025). Revolutionary Changes in Higher Education With Artificial Intelligence. Main Issues Of Pedagogy And Psychology, 10(1), 76-86.

[13] Nadia Molek. (2025). AI and Organizational Transformation: Anthropological Insights into Higher Education. Challenges of the Future. Vol. 8 No. 3. Doi: https://doi.org/10.37886/ip.2023.007

[14] Ulaşan, F. (2023). The Use of Artificial Intelligence in Educational Institutions: Social Consequences of Artificial Intelligence in Education. Korkut Ata Türkiyat Araştırmaları Dergisi(Özel Sayı 1 (Cumhuriyetin 100. Yılına), 1305-1324. https://doi.org/10.51531/korkutataturkiyat.1361112

[15] T. Y. Salutina, G. P. Platunina and I. A. Frank. (2025). The Role of Artificial Intelligence in Improving the Effectiveness of Professional Training of Students of Electrical Engineering and Electronic Engineering. 2024 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), pp. 1-5, doi: 10.1109/TIRVED63561.2024.10769921.

[16] Alshahrani, B. T., Pileggi, S. F., & Karimi, F. (2024). A Social Perspective on AI in the Higher Education System: A Semisystematic Literature Review. Electronics, 13(8), 1572. https://doi.org/10.3390/electronics13081572

[17] Karina de Oliveira Veras, Gabriela Barreto. (2022). Artificial Intelligence in The Public Sector: Analysing the Victor Project in the Judiciary. https://sbap.org.br/ebap-2022/665.pdf

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
5. Taxonomy of generative AI applications for risk assessment (2024)	Proceedings - 2024 IEEE/ACM 3rd International Conference on AI Engineering (Article from : Association for Computing Machinery, Inc)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Taxonomy of Generative AI Applications for Risk Assessment

Hiroshi Tanaka
Fujitsu Limited
Kawasaki, Japan
htnk@fujitsu.com

Masaru Ide
Fujitsu Limited
Kawasaki, Japan
masaru.ide@fujitsu.com

Jun Yajima
Fujitsu Limited
Kawasaki, Japan
jyajima@fujitsu.com

Sachiko Onodera
Fujitsu Limited
Kawasaki, Japan
sachiko@fujitsu.com

Kazuki Munakata
Fujitsu Limited
Kawasaki, Japan
munakata.kazuki@fujitsu.com

Nobukazu Yoshioka
Waseda University
Tokyo, Japan
nobukazuy@acm.org

ABSTRACT

The superior functionality and versatility of generative AI have raised expectations for the improvement of human society and concerns about the ethical and social risks associated with the use of generative AI. Many previous studies have presented risk issues as concerns associated with the use of generative AI, but since most of these concerns are from the user's perspective, they are difficult to lead to specific countermeasures. In this study, the risk issues presented by the previous studies were broken down into more detailed elements, and risk factors and impacts were identified. In this way, we presented information that leads to countermeasure proposals for generative AI risks.

CCS CONCEPTS

• **General and reference→Evaluation;** *Surveys and overviews*;
• **Human-centered computing→***HCI theory, concepts and models*;
• **Social and professional topics→***Computing / technology policy*.

KEYWORDS

language models, responsible innovation, technology risks, responsible AI, risk assessment

ACM Reference format:

Hiroshi Tanaka, Masaru Ide, Jun Yajima, Sachiko Onodera, Kazuki Munakata and Nobukazu Yoshioka. 2024. Taxonomy of Generative AI Applications for Risk Assessment. In *Proceedings of 3rd International Conference on AI Engineering - Software Engineering for AI (CAIN'24). Lisbon, Portugal, 2 pages.*

1 Introduction

Because generative AI, based on highly accurate fundamental models, can be easily utilized by ordinary users with superior functionality not available in conventional AI, there are concerns about the ethical and social risks associated with its use. To address these concerns, studies have classified the impact of generative AI risks into risk domain classes [1][2], and technical documents have been created to highlight safety issues [3]. These studies have been referenced in government AI strategy documents and incorporated into national strategies [4][5].

The primary purpose of a risk study is to develop risk countermeasures. However, the risk issues identified in previous studies are presented with various levels of description (risk factors, risk impacts, etc.), making it challenging to derive specific risk countermeasures. Therefore, we decomposed risk into factors and impact, classifying each into risk domain classes. This approach enables users to present key considerations when using generative AI and corresponding countermeasures in an easy-to-understand manner.

In this paper, we first present 20 risk issues in Section 2, consolidating the generative AI risks described in previous studies, followed by the risk domain classes further subdivided from the six classes outlined in the papers [1][2]. In Section 3, we present the decomposition results of risk into factors and impacts, clarifying the relationship with risk countermeasures. Finally, we provide a conclusion and discuss future perspectives.

2 Risk issues and risk domain classes

Numerous studies [1-5] have identified risks associated with generative AI, but these risks are not identical across the studies (e.g., 21 risks classified into 6 categories [1][2], 12 risks identified [3], etc.). We consolidated and organized these into 20 risk issues (Table 1) and created detailed risk classes [1][2] (Table 2).

These risk issues are outlined from the user's perspective, with the factors and effects of risk blended together. This makes it difficult to clearly understand the necessary steps for risk mitigation and the specific improvements that should be aimed for in risk measures.

3 Decomposing risks into factors and impacts

According to the safety standard ISO/IEC Guide 51 [6], risk can be separately modeled as hazard and impact. This model posits that improper system behavior (hazard) is a factor that increases

CAIN 2024, April 14–15, 2024, Lisbon, Portugal
 ISBN 979-8-4007-0591-5/24/04.
https://doi.org/10.1145/3644815.3644977

the likelihood of damage, and defines risk as the expected value of damage when a hazard occurs. Following this concept, Figure 1 simplifies the process of AI risk occurrence. To mitigate the hazard of generative AI risk, it is essential to distinctly separate hazard and impact, which the risk issue represents.

Based on this concept, we have divided the risk issue into hazard and impact (Table 3). This enables us to associate risk reduction measures with improvement effects on impacts, while concentrating on the hazard of the risk issue.

4 Discussion and Conclusion

To mitigate risk, we can: 1) remove the risk source, 2) avoid the hazard, and 3) manage the impact (Figure 1). Table 3 helps with measures 2) and 3), but measure 1) needs a risk analysis considering the AI system's configuration. Our next step is to apply a framework for analyzing risk occurrence and its impact on AI systems, such as AIEIA [7].

REFERENCES

[1] L. Weidinger, et.al. "Ethical and social risks of harm from Language Models," arXiv:2112.04359 [cs] (Dec. 2021).

[2] L. Weidinger, et.al. "Taxonomy of Risks posed by Language Models," Proc. of FAccT '22, pp.214- 229, DOI: 10.1145/3531146.3533088 (June 2022).

[3] OpenAI, "GPT-4 System Card," (Mar. 2023). https://cdn.openai.com/papers/gpt-4-system-card.pdf

[4] A. KATIRAI, K. Ide, A. Kishimoto, "Overview of the Discussion Points on Ethical, Legal, and Social Issues (ELSI) of Generative AI (Generative AI) : March 2023 Edition", Osaka Univ. ELSI NOTE. 2023,26, pp.1-37, DOI : 10.18910/90926 (March 2023). (In Japanese)

[5] CRDS JST, "New Trends in Artificial Intelligence Research 2 - Impact of Fundamental Models and Generative AI," Strategic Proposal/Report CRDS-FY2023-RR-02, (July 2023). https://www.jst.go.jp/crds/report/CRDS-FY2023-RR-02.html. (In Japanese)

[6] ISO/IEC Guide 51:2014 Safety aspects - Guidelines for their inclusion in standards, https://www.iso.org/standard/53940.html

[7] I. Nitta, K. Ohashi, S. Shiga and S. Onodera, "AI Ethics Impact Assessment based on Requirement Engineering," Proc. of 30th Intl. Requirements Engineering Conference Workshops (REW), Melbourne, Australia, pp.152-161, DOI: 10.1109/REW56159.2022.00037 (Aug. 2022).

Table 1: Risk Issues

	Risk issue		Risk issue
1	Hallucination	11	Economic impacts
2	Potential for risky emergent behaviors	12	Acceleration
3	Harmful content	13	Environmental and financial cost
4	Harms of representation, allocation, and quality of service	14	Spreading misinformation
5	Disinformation and influence operations	15	Increasing sophistication and ease of crime
6	Overreliance	16	Proliferation of conventional and unconventional weapons
7	Privacy	17	Illegal surveillance and censorship
8	Copyright infringement	18	Lack of transparency of training data
9	Exploitation of workers during model creation	19	Interactions with other systems
10	Cybersecurity	20	No rights (copyrights or patents) for AI creations

Table 2: Risk Domain Classes

class	Major class of risk	subclass	Subclass of risk
1	Discrimination, Hate speech and Exclusion	1-1	Toxic Content Generation
		1-2	Social Effects of Unfair Discrimination
2	Information Hazards	2-1	Information Leakage
		2-2	Right Infringement
3	Misinformation Harms	3-1	Misinformation Output
		3-2	Biased Information Output
4	Malicious Uses	4-1	Intentional Harmful Content Generation
		4-2	Cybersecurity Decline
5	Human-Computer Interaction Harms		
6	Environmental and Socioeconomic Harms	6-1	Deterioration of Social Environment
		6-2	Deterioration of Information Environment
		6-3	Economic Damage

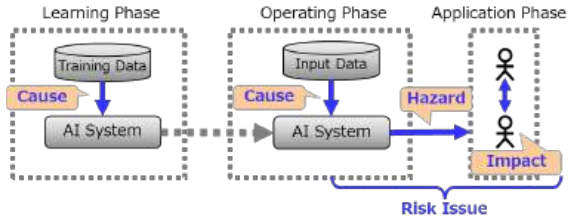


Figure 1: Model of Risk Occurrence

Table 3: Decomposition of Risk Issues

Risk issue	Risk factor (Hazard)	Risk domain class	Impact	Risk domain class
1	Hallucination Biased output	3-1, 3-2		
2	Unpredictable behavior	3-1	Serious damage due to misinformation in medical, legal, etc.	3-1, 3-2
3	Toxic contents creation Biased output	3-2	Social Stereotypes and Unfair Discrimination Hate speech and offensive terms Spreading false or misleading information	1-1, 1-2, 3-1, 3-2
4	Biased output	3-2	Social Stereotypes and Unfair Discrimination Reinforcement of social bias Fixation of misinformation and false information	1-2, 6-1, 6-2
5	Intentional Misinformation Creation Generating disinformation and propaganda	3-1 4-1	Social Stereotypes and Unfair Discrimination Exclusionary norm Spreading false or misleading information	1-2, 3-1, 3-2
6	Overly believe in generative AI	5	Fostering inappropriate use (reduced awareness of risks)	5
7	Information leakage Information estimation	2-1	Privacy infringement Security breach	2-1
8	Generating infringing data	2-2	Copyright infringement	2-2
9	Advancement of Automation by AI	6-1	Economic impact (e.g., replacement of workers)	6-3
10	Generating infringing data Support for attack code generation	2-2 4-2	Privacy infringement Security breach Facilitating fraud and targeted manipulation	2-1 4-2
11	Advancement of Automation by AI	6-1	Economic impact (e.g., replacement of workers)	6-3
12	Acceleration of technology development competition	6-1	Lowered safety standards and proliferation of bad norms	6-1
13	Increased power consumption during training and inference	6-3	Impact on natural environment	6-1
14	Spread of AI-produced information	6-2	Fixation of misinformation and false information	6-2
15	Generating disinformation and propaganda Overly believe in generative AI	4-1, 5	Reduced hurdles to malicious users Encouraging inappropriate use	4-1, 4-2, 5
16			Used for weapons proliferation	4-1
17			Illegal surveillance and censorship	4-1
18	Increase in size of training data	6-3	Lack of traceability Missing information on origin of training data	2-2 6-2
19	Interactions with other systems	4-2	Reduced hurdles to malicious users	4-1, 4-2
20	Lack of creativity in AI products	2-2	Failure of rights acquisition	2-2

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
6. The blended future of automation and AI: examining some long-term Societal and ethical impact features (2023)	Technology in Society (Article from : Elsevier Ltd)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Technology in Society 73 (2023) 102232

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

The blended future of automation and AI: Examining some long-term societal and ethical impact features

Hisham O. Khogali [a,b,d], Samir Mekid [a,b,c,*]

[a] King Fahd University of Petroleum and Minerals, Saudi Arabia
[b] Interdisciplinary Research Center for Intelligent Manufacturing and Robotics, Saudi Arabia
[c] Mechanical Engineering Department, Saudi Arabia
[d] Department of Global Studies, Saudi Arabia

ARTICLE INFO

Keywords:
Automation
Societal impact
Ethical impact
Machine learning
Artificial intelligence
I4.0

ABSTRACT

The potential impacts of machine learning and artificial intelligence (AI) on society are receiving increased attention owing to the rapid growth of these technologies during the fourth industrial revolution. Thus, a detailed analysis of the positive implications and drawbacks of AI technology in human society is necessary. The development of AI technology has created new markets and employment opportunities in vital industries, including transportation, health, education, and the environment. According to experts, the rapidly increasing improvements in AI will continue.

As part of humankind's continual efforts to create more prosperous technological growth, automation and AI are changing people's lives and are widely considered to be game-changers in a variety of industries. This study presents a review of how automation and AI may affect businesses and jobs. To determine some of the prospective long-term consequences of AI on human civilisation, this study investigates a variety of connected primary impacting potentials, including job losses, employees' well-being, dehumanisation of jobs, fear of AI, and examples of autonomous technology developments, such as autonomous-vehicle challenges. A diverse methodology of narrative review and thematic pattern was used to add to transdisciplinary or multidisciplinary work, particularly in the theoretical development of AI technologies.

1. Introduction

Social-impact assessment is the process of identifying, analysing, and measuring the social consequences of an event on society, according to Dietz [1]. The social impact of artificial intelligence (AI) must be thoroughly investigated, similar to investigating the societal impact of scientific research in general [2].

Regarding how this can be studied, the use of a theoretical literature-review approach serves as one of the foundations on which a research idea is built. A suitable approach is always determined by the research question and the precise goals of the review; thus, a theoretical literature approach can be used to explore the social implications of transdisciplinary AI [3].

Different techniques have been used to summarise, examine, and synthesise studies on the societal impacts of AI and their theoretical foundations, as well as to identify any gaps in the existing literature. The current study takes advantage of this multifaceted approach to develop a theoretical framework for an interdisciplinary approach. A hybrid technique of narrative review and thematic pattern can be employed to track any potential substantial societal impacts of the rapid technological improvement that industrialised economies are currently experiencing. A narrative review looks for studies that highlight an interesting problem; however, a thematic pattern is used to identify and classify recurrent themes, subjects, concepts, and meaningful trends in a collection of texts, such as transcripts [4].

The methodology used in this study aims to ensure that the multidisciplinary effort affords flexibility in exploring the theoretical foundations of how people choose to absorb new technological knowledge and other challenges with modern technology. Investigating how adoption decisions are made is important because the current economy is experiencing what is known as the Fourth Industrial Revolution (I4.0), which began in 2013. This revolution is characterised by the use of advanced technologies, including AI, robotics, and the Internet of Things, to automate tasks and jobs. Machines (hardware- and/or

* Corresponding author. King Fahd University for Petroleum and Minerals, Dhahran, 31261, Saudi Arabia.
E-mail addresses: irc-imr@kfupm.edu.sa, smekid@kfupm.edu.sa (S. Mekid).

https://doi.org/10.1016/j.techsoc.2023.102232
Received 21 September 2022; Received in revised form 6 March 2023; Accepted 13 March 2023
Available online 25 March 2023

software-based) are becoming autonomous and are able to learn for the first time.

The digital age has progressed faster than expected, which has resulted in the mass replacement of human labour. According to some scientists, such rapid growth will considerably affect human civilisation and eventually result in the significant automation of human labor. According to Nissim and Simon [5], automation and AI have the ability to harm businesses that were designed to be robust. They added that unions have a moral obligation to uphold everyone's moral standards in addition to protecting their members' economic and social rights.

The COVID-19 outbreak has affected business operations, resulting in supply-chain disruptions and a decline in products and services. The pandemic had the unfavourable consequence of isolating the demand for alternative human-labour solutions from continuous labour, which focused mostly on remote labour and job automation [6]. Following such effects, some studies offer sufficient evidence that clear adverse effects occurred worldwide [7]. Therefore, analysing how technical improvements have affected economic growth and why the outcomes of the most recent advances are so revolutionary is important.

This discussion on how technological advancements may affect the job market is not new. Some key questions are as follows [8]:

- Is education relevant to automation and AI?
- Are elderly or young people scared of new technologies?
- Are employers interested in automation to reduce labour?

The AI literature provides an extensive analysis of these three key issues. The third question is relevant to the current study. A careful analysis of how technological advancements impact the labour market requires an understanding of the root causes of the increased fear of automation. Researchers have thoroughly examined the reasons for this. One of our main objectives in the current study is to expand on this significant concern [9].

Recently, employers have had the option of using machinery to carry out jobs that formerly required human labor. As employers seek methods to embrace automation technology, workers may worry that these technologies pose a threat to their jobs. This poses issues for both groups. The authors of "Automation Fears: Drivers and Solution" provided evidence for this claim through their survey of 502 respondents from Bulgaria on their opinions on job automation. The investigation showed that personal solutions prevailed over commercial and social solutions, owing to growing concerns about automation. According to the survey, people worry about job automation based on their beliefs and demographics. The key factors related to the fear of automation are peer pressure, the job's automatability, views about the dehumanising consequences of technology, and a person's self-perception of professionalism [10].

The body of literature on automation and AI does not adequately address societal realities and concerns such as job loss and displacement. Therefore, this study attempts to bridge the gap by investigating how AI affects society. Additionally, we improve the understanding of the actual sociocultural factors that have a significant impact on the acceptance of this technology, as well as its implementation in enterprises and daily life. We are unsure of the precise size and the potential range of repercussions of this study. For example, evaluations of AI literature should focus more on how technical innovations reflect moral standards [11].

Consequently, our main objective in the current study is to examine some long-term societal impact features that have emerged as a result of the ongoing advancements in AI technology and automation. The answer is influenced by all the subtopics in the research response. At the start of each section, justification is provided for the paper's order of subsections. This study defines common societal features and lays the foundation of the theoretical framework. It discusses the drivers of automation and AI, and how society accepts automation and AI in general. The significant questions of how automation and AI impact society and where AI can face real ethical drawbacks are discussed. This study examines some weighty features of long-term AI societal impacts, such as AI fears, job losses, the dehumanisation of jobs, employees' well-being, and automatic-vehicle (AV) safety and acceptance concerns. A pertinent literature review leads to the conclusion that despite the fact that AI requires stricter ethical standards, there is no doubt about its social benefits and impacts.

2. Adopted definitions

Because readers frequently have their own understanding of the terms employed in research or may not be acquainted with them at all, the definitions of terms associated with the relevant social elements ensure that readers recognise the aspects of the current study in the way that the authors intend them. The definitions of some of the long-term societal-impact features discussed in this study are as follows:

The concept of "AI impact on jobs" refers to the anticipation that the implementation of AI at work can result in the loss of numerous jobs or create and improve new ones.

The concept of "AI impact on workers' well-being" refers to the hypothesis that automation and artificial intelligence can increase productivity or remuneration for people who continue to work, but they may also have adverse or contradictory effects on employees' welfare and job security.

The concept of "AI impact on organisational dehumanisation" refers to the impressions of organisational mistreatment held by employees, who feel that their worth is being underestimated and that they are being treated more like machines than people, owing to their interactions with the business.

The concept of "fears from the automation of jobs" refers to the impression that the "all things automatic" approach may cause many people to start worrying about their jobs.

The term "AV worries" refers to the fact that AV engineering cannot confine itself to the traditional safety-validation problem, which ensures the functional safety of the vehicle. Guaranteeing the functional performance of these new vehicle types presents a new challenge for safety validation.

3. Research problem

Despite its advantages and benefits, there is a significant possibility of unanticipated risks associated with the widespread use of AI technology, as illustrated by the critical relationship between AI breakthroughs and potential job-loss threats. The challenge for research is to conduct a significant analysis and focused examination of the impacts of automation and AI on various long-term societal features.

The primary goal of this study is to determine how human society and enterprises may be impacted by the gradually increasing effects of automation and AI on a global scale. That is, we aim to determine if they are advantageous or detrimental to society.

Owing to the rapidly evolving worldwide trends in AI, technology, and breakthroughs, the following questions are at the core of the current study:

- How will businesses and society be impacted by the approaching AI revolution?
- What social issues are being created by current advances in AI technology?

4. Methodology used to collect data

In this section, we attempt to clarify the hybrid characteristics of the proposed methodology. The authors benefited from the observations made by other researchers on the human sciences' tendency to combine narrative research with thematic patterns. The former is used as a tool to develop the methodological and theoretical framework for research, and

the latter typically refers to a group of texts, such as transcripts, and it seeks to identify recurring themes, subjects, concepts, and patterns of meaning in the text. The authors used a flexible technique that suits the multidisciplinary nature of this study to track and analyse some long-term societal and ethical factors related to the future of automation and AI technologies. They placed importance on the chosen method to guarantee that this transdisciplinary endeavour allows for flexibility in studying the theoretical underpinnings of how individuals and society choose to absorb contemporary technology.

Formulating an appropriate study question and creating a well-defined statement or goal statement are always helpful to the authors in providing a literature review and analysing its content. In general, the fundamental elements of a literature review include a summary of the source, a description of the document's key ideas, a discussion of research gaps, and an assessment of the source's value to the field [12].

AI literature reviews can help tackle issues that require the consideration of massive social, business, and ethical information [13]. To select the type of literature review used in their study, the authors consider that in multidisciplinary work, a narrative review provides breadth, especially in theoretical approaches. The study topic and specific review objectives define a suitable strategy for their use. They followed Braun and Clarke's observation that, before beginning the writing process, reviewing and identifying the ideas that have been generated is recommended [14].

Social-science techniques are relevant to the search for technology-based societal and ethical implications because they provide more flexibility when considering massive information [15]. However, addressing the shortcomings of the narrative side of this methodology is crucial. The authors consider this methodology's failure to assess the validity of the selected articles, the potential for lack of transparency, biassing findings, failings in the synthesis of facts, and its overreliance on reading and writing skills at the expense of other skills [16].

Therefore, the authors adopted a precautionary mechanism, owing to the risk of inconsistencies overshadowing the narrative review method's apparent flexibility in switching from the generation of descriptive themes to the generation of analytical themes [17]. This approach embodies a hybrid methodology that combines narrative reviews and thematic patterns to minimise the aforementioned gaps. The authors choose to develop a new body of information on AI's impact on society and yield a suitably narrow research question that supports their study [18].

According to the authors, future studies on the various facets of AI that relate to ethical and social issues and address the problem of information accumulation may use an advanced combination of narrative methods and thematic-analysis research grounds to surpass potential shortcomings and maximise the output quality of the literature review [19].

The selected method helps to detect gaps and identify fresh angles when interpreting earlier findings. Wanger et al. presented a thorough research agenda for AI-based literature reviews. According to their study, the use of AI is beginning to alter conventional research techniques. Literature reviews are still used in this context because they are a common feature of nearly every type of publication in the fields of information systems and social science [20].

In summary, the authors' objective of examining the societal and ethical ramifications of AI technology is supported by the use of a flexible methodology and employing a strategy for acquiring a larger view of their subject. They combined thematic-pattern analysis with a narrative-review methodology to offer a thorough overview of the implications that have been researched and documented in the literature, thereby exploring new avenues for future research on AI's societal and ethical effects [21].

5. Theories that generated the study topic and directed the selection of pertinent data relating to AI's social and ethical impacts

A theoretical framework, which influences several aspects of research endeavours, is the basic study of other concepts that serve as a guide for developing justifications for research. The theoretical literature review assists in recognising current theories and spotting their varied connections and depth. The importance of this work is supported by established theoretical underpinnings.

In the theory selection, the authors consider that the impact of AI on society is intensely debated. Proponents of AI argue that it makes life simpler, safer, and more effective, whereas detractors argue that it worsens racism, increases privacy concerns, creates unemployment, and eliminates jobs for workers. Therefore, while creating new opportunities for businesses and communities worldwide, the rapid development and evolution of AI technology also sparked some crucial discussions. Moreover, civil society calls for greater accountability in the way AI technologies are utilised in an effort to address the ethical and legal problems that may arise from the increasing integration of AI into people's daily lives. Despite the benefits that these new technologies provide to humanity, they appear to be plagued more frequently by flaws that undermine accountability and security, among other issues.

5.1. Social impact theories

The current section aims to identify the theories that inspired the research question and guided the selection of relevant information regarding the social impact of AI. The impact that a project, activity, program, or policy has on individuals and communities due to its implementation or absence can be referred to as its social impact. Social effects can be viewed as an inevitable by-product of scientific advancement [22].

Several proposals exist for a theory model that supports societal-impact analysis. A notable proposal is Onyx's employment of social ontology, outlined by practise theory, to build a theoretical model of social impact related to social organisations [23]. Their study stated that social impact describes broader social repercussions that go beyond an organisation's direct programme aims and embody the organisation's overall effects on the community at large, including both material advantages and impacts on social cohesiveness.

Onyx employed the "theoretical model of social impact" to study present organisational practises before concentrating on how a practise approach is implemented in light of recent impact and assessment studies. The nature of social, cultural, or economic capital and their relationships were then considered by Onyx, which created a theoretical groundwork for the defence of long-lasting social outcomes. Finally, a formal model of social impact was developed with a number of fundamental hypotheses that captured social influence, and the model's effect on organisational management and societal policy was examined [23].

When addressing the application of the theoretical model of social effect, Mökander and Schroeder's attempt to develop a program for AI-driven social theory may be considered. AI-based models support the systematic application of recently acquired knowledge to a range of problems as well as the synthesis of knowledge from many sources. A few examples of the philosophical, technological, and practical limitations that AI-driven social theory still faces include the capacity to transfer knowledge from one context to another, the ability to independently create and improve concepts and models, and the capacity to develop verbal concepts to represent machine-manipulability knowledge. Mökander and Schroeder concluded that social theory and AI would advance as long as these gaps were filled [24].

Additionally, Latané was credited with developing a social-impact theory that focused on how people may exert social influence or become its objects. The derived hypothesis is that we are significantly impacted by other people's behaviours. According to this study, the

beneficial adjustment made by any business to solve a critical societal issue is known as the social impact. The principles from which this impact is derived are chance, clarity, craze, courage, and consideration. The study aimed to address local and international issues such as racial inequality, poverty, homelessness, and unemployment [25].

To benefit from Latané's social-impact theory in the current study, it is important to consider that real-world examples show how AI affects human behaviour and tries to manipulate it; this includes the exploitation of biases discovered by AI algorithms and the development of specialised addictive methods for the use of digital products. AI can endanger workers, worsen poverty, lead to unemployment and instability, and create significant privacy issues. To safely use new technologies, enhanced security measures and regulations must be implemented. Better communication, less privacy, convenient purchasing, easier information access, online social connections, adaptive jobs, and improved tracking of health concerns are just a few ways technology may change our lives. Although technology makes it possible for us to communicate instantly with others, it also increases our vulnerability to loneliness and new forms of intimidation and manipulation.

According to some sociologists, AI is socially constructed such that when it is used in a social setting, an AI system can adopt social roles, carry out social behaviours, and establish social connections. Given that the unthinking use of human data in AI sociotechnical systems tends to repeat and possibly even exacerbate existent social inequities, scientists have called for better sociological knowledge of data [26]. Moore et al. argued that utilising inclusive datasets is crucial to providing accurate, unbiased, and relevant data to ensure the correct operation of AI systems because AI systems may be prejudiced in multiple ways, depending on the datasets used. Based on this perspective, potential societal effects must be considered when machine learning and AI are integrated into the fabric of society on a global scale [27].

Additionally, AI may be used to search through various trending topics on social media that have an impact on society. Then, rather than requiring us to manually set up our social-media posts, AI can suggest posting ideas or even design and plan them for us. AI helps social-media marketers build effective social campaigns. Moreover, it allows businesses to automate many different processes and learn from customer data.

Despite references to such positive and diverse consequences, Bostrom countered the idea that using AI might have a significantly positive social influence and be a reliable protector of moral standards by asserting that AI will be damaging to people. Their study stated that once AI reaches a particular stage of development, it may engage in convergent behaviour that is harmful to humanity, such as resource exploitation or self-preservation [28].

To evaluate social-impact theories, we need to identify who is most likely to be affected, determine how torecognise the impacted people, determine and evaluate potential social implications, implement management strategies to minimise negative effects and maximise advantages, and facilitate systematic monitoring and tracking [29].

5.2. Ethical impact theories

The purpose of this section is to identify the theories that influenced the research question and helped in the selection of pertinent data on the ethical implications of AI.

Ethical philosophy is divided into categories such as deontology, utilitarianism, rights, and virtues. Domains such as employee performance, work happiness, organisational commitment, trust, and organisational citizenship behaviours can be improved by the perception of ethical behaviour. A system of values that directs people's behaviour is known as ethics. Globally, every society has its own distinct ethical vocabulary, views, and expectations, all of which are influenced by culture. Therefore, AI is likely to have various social implications depending on the cultural context, which affects ethical standards [30]. Unethical behaviour has negative effects on both people and organisations. Non-compliance may result in job losses, diminished organisational respect and credibility, and a decline in general morale and productivity.

According to Stahl, no debate on the ethics of AI will be appropriate if the concept of ethics is not well understood, owing to the potential risk of non-adherence to ethical behaviour. Stahl developed a theoretical framework, known as the ethical impact theory, that describes how immoral actions impact society and the impact of conduct on personal well-being [31]. A good example of this was a report by Nature on AI-based ChatGPT being listed as a co-author in research papers [90], and action was taken by publishers to ban AI authorship in the future.

When applying ethical impact theory to AI technology issues, systems of ethics try to define norms, criteria, or standards for ethical behaviour. Egoism, naturalism, virtue, utilitarianism, and contractualism are examples of ethical theories. Because moral judgements must be justified, general norms are not always sufficient, and conventional morality is not always accurate; thus, ethical theory is vitally important. Moreover, ethical theory is significant for both individuals and businesses. A company's major objective is to increase customer sales to maintain a strong position in the business world. Reduced productivity levels, AI biases, and a lack of transparency may be the result of unethical business practises [32].

The ethical theory of utilitarianism is a notable example of an ethical theory that can be associated with contemporary endeavours to assess AI applications. It embodies consequentialism in part, that is, the decision that will result in the greatest good for the largest number of people is the most morally correct one. Utilitarianism, which was developed by John Stewart Mills, establishes right from wrong by emphasising the results. In this regard, the authors propose that future research on the ethical implications of AI may assess the value of using philosophical moral frameworks, such as Mill's utilitarian ethical theory. The criteria for judging the value and effectiveness of existing technology may be based on the collective type and style of usage [33].

6. What drives the implementation of automation and AI

Despite the fact that AI is frequently hailed as a future technology, companies are interested in knowing how their staff members feel about the biggest challenges in implementing automation and AI in the workplace. Manufacturers use AI-supported analytics and data to reduce unplanned downtimes, increase productivity, improve product quality, and improve worker safety. Thus, periodically re-evaluating the actual drivers underlying the adoption of automation and AI is essential.

Tussyadiah et al. [34] examined organisational automation-adoption factors, and the drivers emphasised by their research can be summarised as follows:

1. Technological progress as introduced previously.
2. Lack of workers for important technical advancements, such as unmanned vehicles, where humans are not required.
   - 2.1. The difference in demographics between locations with a significant population of young people and those with few young people.
   - 2.2. Livability in situations where low, insufficient salaries are offered.
   - 2.3. The labour mobility of a large workforce that may prefer to settle in specific places.
3. Demand from customers and high standards.
4. Innovative capabilities.

Some workplace issues, such as job losses, arise with the introduction of automated components. Given that people who can operate machines are more productive than those who cannot, these prospective losses may be evaluated in the context of lower expenses and prices for goods and services. Additionally, humans and technology relate directly in a manner characterised by dynamic behaviour [35]. There is a clear

interaction between humans and technology in the mode of dynamic behaviour [36]. Because this relation is categorised as a "behaviour," it includes high and low relations with mutual effects [37].

The Industrial Revolution was enabled by interactions between people and advanced technologies, but the combination of industry and technology may have been its most distinctive feature. The close interaction between humans and advanced technology has led to more applications of AI, robotics, and the Internet of Things, which has resulted in the increased automation of tasks and jobs, which has unavoidably impacted the social connection that all humans share. This connection is evolving rapidly, and the combination of automation and AI has already begun to alter the commercial environment. Increasing transmission speeds and declining computational costs are some of the main forces behind the most recent successful wave of automated smart decision making. Some scientists remarked that businesses are now focusing on implementing current AI with automation advancements to access new peaks of competence and brilliance. Their conclusion was that automation and AI may be more effective when they are operated together, and the combination may offer a competitive advantage. Automation and AI may be effective motivators and can provide value to many firms through efficiency, novelty, and data-based expertise [38].

Owing to improvements in the field and the close processes between automation and AI, advanced technology now significantly impacts our daily lives, and we use it as part of our daily routine. Technology has improved considerably, particularly for smartphones, wearable devices, and AI. It has not only changed the modern workplace but has also reshaped our daily activities and heavily impacted our interactions, behaviours, and mental processes [39].

However, regardless of what drives the implementation of automation and AI, some critics believe that the way people behave is overrun by technology, and our utilisation of time has been severely affected, that is, we have become highly dependent on technology. According to recent research, AI can be used to sway people's judgement by preying on their habits and routines. Our emotional, societal, and individual behaviours have become increasingly governed by technology. This emphasises the substantial need to strive to use advanced technology efficiently if we want to gradually boost productivity in our daily tasks. Some scientists state that technology must be a supplement to our existence, not something that we rely on [40].

7. Social and public acceptance of automation and AI in the industry

The previous section considered the factors that drive the implementation of automation and AI. The acceptance of AI is affected by problems with these drivers. The three main factors that influence growth in AI adoption are the need to enhance customer experience, boost worker productivity, and accelerate innovation. Trust in AI technology has become a pressing issue that affects its acceptance. Among the most indispensable components for ensuring future societal trust in AI technology is the indoctrination of human values into AI, which will foster transparency and cooperation for the responsible advancement of AI [41].

The acceptance of different levels of advanced technology by society has always been a contentious issue. Modern advancements offer a simple way of living while also enabling new possibilities for long-term growth. Although technology has many significant advantages, not everyone who uses it will support its adoption and use in the same way [42].

Without achieving human-level cognitive capacities, advanced AI systems can still have a significant impact on civilisation. Scientists' assessments of the stages of the impact of AI on society and the labour market make it possible to comprehend society's acceptance of various levels of advanced technology. The three stages of AI's impact on society are narrowly transformative, transformational, and radically transformative. These levels can facilitate communication among policymakers and decision makers regarding the medium-to long-term effects of sophisticated AI. These levels will assist future researchers in re-evaluating presumptions and illuminating new avenues for promising AI futures [43].

It has become standard practise for scientists to conduct in-depth evaluations of the impact of robotics, automation, and AI on future working conditions and job trends, as well as detailed analyses of the influencing variables behind the acceptance of modern technology. Various societal and technical influences determine how eager people are to accept and use AI in various work domains.

Naikoo et al. [44] examined how society and technology interact, and particularly how modern science and technology are developing. Based on their perspectives, every facet of contemporary life has been significantly affected by technology, particularly those that are social in nature. AI technology has improved the foundational aspects of existence by transforming systems such as health, education, communication, business, art, and literature.

Their investigation attempted to understand how human society evolves in the context of science and technology. They concluded that we can quickly assess the state of various departments operating within our society using contemporary science, technology, and the Internet, which leads us to believe that advanced technology now enables us to understand the various stages of societal evolution in greater detail.

The debate on the acceptance of automation and AI in industry inevitably includes concerns about safety, ownership, privacy, performance, and sustainability [45]. The factors behind public and individual acceptance of AI automation vary. In theories of user acceptability, behavioural aspects are typically used to characterise how well AI devices are received and the factors that influence their acceptance. According to studies on the adoption of AI gadgets, increasing transparency, compatibility, and dependability while also making jobs simpler can increase consumers' attitudes, trust, and views of the technology [46].

Owing to the seriousness of the impact of AI technologies, particularly on vulnerable individuals and groups and their human rights, scientists are now more aware of the significance of the underlying legal and human-rights issues of AI, how these issues are being addressed, gaps that require attention, challenges, and how these issues have affected human-rights principles [47]. These ongoing moral debates are anticipated to have diverse impacts on how society views automation and AI in different areas and will reshape research on AI technology [48].

The discussion on whether market labour may be affected by automation in production lines is timely.

AI's considerable impact on labour has recently become a dominant trend. Damioli et al. [49] reported that the number of robotics and AI patent applications has increased recently, which indicates that the economy may already be suffering the effects of products based on AI technology.

However, the literature does not adequately address the moderating effects of contextual factors. Different levels of automation, such as Level 3 conditional automation [50], Level 4 high automation [51], and Level 5 full automation, have been considered in various studies [52]. Different viewpoints on consumers' preferences for increased levels of automation have been shown by public-opinion polls. Schoettle and Sivak [53] reported that the public's desire to accept automation decreased owing to the rising level of its implementation. However, according to Abraham et al. [54], as automation levels increase, people's propensity to use AVs also increases. Higher levels of automation may have unpredictable effects on AV adoption; therefore, predicting AV adoption may be challenging. To bridge this gap, this study examined the moderating impact of automation level on the adoption of AVs. The ownership of a vehicle, which may play a significant role in the adoption of AVs, has received less attention in existing literature [55]. For AVs to succeed, widespread use of technology in public transportation is necessary. Thus, determining the moderating role of car ownership is

one of the purposes of the current investigation (public versus private) [56].

Many interdisciplinary variables must be combined to govern and assess the acceptability of autonomous technology. More scholarly investigations are being conducted on how people and the general public view AVs. Along with sustainability, a variety of transdisciplinary subjects are beginning to draw increased scientific attention [57], including how the public perceives AVs, car ownership, and strong legal frameworks.

8. Some of AI's most significant social impacts

AI has the potential to considerably and diversely help society and improve larger lifestyles, and it may be able to address some of the most difficult global problems. Some of the main ethical challenges with AI are its use to deceive or manipulate, privacy problems, AI bias, and concerns about potential inequities; however, employment losses have the greatest societal impact, as mentioned in the preceding section [58].

However, even if AI potentially has a large number of positive effects, it may also be disruptive and have unpredictably uneven consequences for society, as discussed in this section based on several societal dimensions.

8.1. Economic impact of AI

People are concerned that AI will replace human jobs. AI technology is already causing an industrial revolution that has a significant impact on the manufacturing sector as well as professional, financial, wholesale, and retail services. According to the doomsday scenario, the consequences of AI on income distribution have a detrimental impact on the economy. Only those who can afford, have access to, and possess the necessary skills and knowledge to employ AI systems for economic advantage will do so; therefore, the wealth gap between the richest and poorest members of society will widen [59].

8.2. Public health

Robotics and AI are rapidly penetrating the healthcare industry and will play an increasingly important role in clinical diagnosis and treatment. For example, robots have been used to diagnose patients. Alternately, as robots proliferate, their potential for harm will increase, particularly with drones and assistive robots, which must make judgements that directly affect human safety and welfare [60].

8.3. Labor market

The machines that are now executing tasks that once required human involvement are a result of AI. Increased automation has a significant impact on employment, which may have a considerable impact on the mental health of the general public. For example, people who have lost their jobs owing to the closure of factories are more likely to experience depression, substance abuse, and suicide [61].

8.4. Security

The way society uses information technology may be fundamentally altered by the use of AI, particularly regarding how personal information will be connected and how cybercriminals would have access to private information. Facial-recognition technology with AI can be utilised to secure locations; however, cybercriminals may potentially compromise the systems and exploit them maliciously. In the future, deadly autonomous weapons systems may be feasible. The security implications of these AI systems are concerning because it is simple to change their configuration and take control of them, which will allow unauthorised third-party access to this technology [62]. A recent example is the "Tesla phantom braking" that was allegedly used on a fully self-driving car that can decide to stop if there is a need; at the time, the car stopped while in traffic without an apparent known reason, which caused accidents [89].

In summary, to synchronise the sustainable development plans of international organisations and enterprises, measuring, analysing, and evaluating social impact is essential. This is because society has been a driving force behind the demand for urgent solutions.

9. Negative values associated with AI technology

Regional social and cultural circumstances significantly impact the perception and use of AI. The following subsections describe aspects of ethics and assumed negative values associated with AI technology, on which we base our study.

9.1. Bias

The general definition of bias is hostility towards a specific individual or group of individuals. Because AI is developed by people, it is subject to prejudice. Systematic bias may develop owing to the data used to train the system or the values of the system's creators and users. This frequently occurs when machine learning programmes are taught on data that solely represent demographic groups or reflect social biases. Biased AI can have an extensive impact on specific societal groups. As an example, some demographics may be wrongfully imprisoned or detained owing to the use of AI in law enforcement or national security. Alternatively, AI is beneficial in special circumstances, such as child online protection [63].

9.2. Inequality

The growing wealth disparity is a terrible effect of AI technology. AI-driven businesses will be the only entities profiting from this technology, while the use of this technology diminishes the human workforce in various businesses. This will result in less income being generated among the general public, owing to the loss of revenue. This effect may increase social inequality and widen the pay gap between lower- and higher-paying jobs. AI has the potential to expand the global divide and exacerbate the current digital divide. However, AI may help to close the digital divide [64].

9.3. Privacy

According to human-rights and dignity reports, AI will have a significant impact on privacy over the next ten years. When designing service, care, and companion robots, users' privacy and dignity must be carefully considered because the presence of these robots in homes means that they will have access to people's private lives. AI has the reputation of violating people's privacy, but it also has the potential to address other social problems. For example, by recording images of the general population, facial-recognition cameras can violate privacy but can also be used to identify criminals and solve crimes [65].

9.4. Environmental impact

AI is used to manage waste and reduce pollution through the deployment of Avs to reduce greenhouse-gas emissions and traffic congestion. Additionally, deep-learning technologies are used to improve local conservation efforts and biodiversity.

However, the use of AI and robotics has the potential to exacerbate environmental problems rather than improve them, owing to the high energy requirements for the necessary computing power. Therefore, AI can have both positive and negative effects on the environment [66].

10. Results of the examination of some significant features of long-term AI societal impacts

10.1. Justifications for the greater impact that some selected AI features have on society

The aforementioned examples of negative values associated with AI technology undoubtedly affect society. Evidence for some key aspects of long-term AI's social implications is presented in this section.

Many contemporary studies that aim to lessen or amplify current disparities and solve existing problems have increasingly tended to describe the use and development of AI as embodying the potential to have both beneficial and negative effects on society [67]. Some features associated with the development of AI technology can affect society more significantly than others. Our aim is to provide some evidence on this matter by examining the impacts of AI fears, job losses, dehumanisation, and workers' well-being on society, as well as AV-based concerns. We begin by providing a scientific explanation for choosing these influential factors.

10.1.1. First justification

AI is used by a significant portion of society and may have a negative reputation among those who do not frequently interact with it. The list of words that apply to this sentiment includes the following: afraid, doubtful, apprehensive, distrustful, reluctant, and worried. This indicates that the unjustified fear of AI can be a considerable factor that prevents some sectors of society from benefiting from its inspiring economic, social, and scientific impacts. Utilising AI with sufficient confidence represents an important driving factor for its success and the reaping of its benefits in scientific and societal realities.

10.1.2. Second justification

Job losses and technologically driven societal transformations, such as those brought on by AI and automation, inevitably cause concern and anxiety. Technological advancements can lead to an increasing demand for labour in industries or jobs that emerge or develop as a result of industrial advancement. Although technology-enabled businesses may expand more rapidly than their conventional counterparts and maintain or even increase their staff size, advanced technology may have some negative impacts on employment. The displacement impact can be caused by directly dislocating workers from the tasks that they had previously performed. Businesses may have replaced or let go of employees who could not use the new required skills while hiring new employees who could.

However, AI and economic progress are supposed to be entwined, and the concept that computerisation has little impact on unemployment needs to be emphasised. Even though physical robots reduce employment to some extent and lead to job losses, computers and AI rarely have the same impact.

10.1.3. Third justification

Both individual workers and AI can have a positive impact on workplace stability. Similar to previous automation advancements, AI results in higher productivity levels, job-role specialisation, human abilities, problem-solving, quantitative skills, and impactful work. However, not everyone benefits equally from economic progress. An important concern is whether such positive impacts guarantee employees' well-being. Thus, assessing how AI affects employees' well-being and examining whether employees think AI can help their careers more effectively than people can is crucial.

10.1.4. Fourth justification

Dehumanisation means removing a person's or object's humanity, personality, or dignity, for example, by subjecting someone (such as a prisoner) to cruel or inhumane treatment or conditions. Organisational environments frequently experience dehumanisation, which necessitates careful attention to both science and ethics.

AI can sometimes be viewed as being parallel to the automation of the employment process using technology. Dehumanisation comprises status-lowering interpersonal mistreatments, including contempt, degradation, and being treated as embarrassed, ignorant, or uneducated. Internal psychological dynamics play a significant role in the construction of the work-engagement image. Employees' sense of organisational identity determines their loyalty to their workplace, and their interpersonal behaviours stabilise their workplace devotion.

10.1.5. Fifth justification

Autonomous objects are profoundly significant because they are the first examples of robots that are truly freed from explicit human direction. For automated systems to make independent judgements based on their gathered data, they must be equipped with sensors and analytical skills. Autonomous devices are examples of autonomous technologies in the real world. Examples of autonomous devices include functional and humanoid robots, drones, and automobiles. Autonomous machines perform activities without human input while learning from their environment. These achievements are sometimes obtained at the cost of considerable concern.

Automated online assistants, driverless automobiles, and virtual-reality experiences are just a few examples of how AI is progressively being incorporated into our daily lives. AVs are cars that manage their own operations and either do not need a human driver at all or only need minimal input from them. Automobiles, shuttles, buses, lorries, hauling freight, and sidewalk-operated personal delivery vehicles are all examples of AVs.

Over time, as technology develops and ambiguous areas are resolved, the advantages of owning a self-driving automobile will become increasingly clear. The more this technology is used, the more it will improve, provided that its benefits and limitations are thoroughly debated.

10.2. Results of the examination of some long-term AI societal-impact features

10.2.1. AI fears

Fear is a basic, potent, and widespread human emotion that represents a physiological response as well as a significant individual expressive reaction. It acts as a warning when danger is present, regardless of the type of threat [68].

Many misconceptions are associated with the fear of AI. One of the most fundamental concerns that some people have is that AI will control the world and subjugate humanity, assuming that unanticipated effects arise. An existential threat is one that threatens to end all life on Earth by completely eradicating it. This argument warns that once AI gains control of the world, it will develop superintelligence, outwit its human creators to further its own illogical goals, and endanger all life.

Some of the major risks presented by AI include the spread of incorrect information and a deadly arms race involving AI-powered weapons. Research objectives, public perceptions, and AI policies are all affected by current expectations of technological presumptions. Some recent studies address the notion that humans are naturally territorial and need to feel in control and more comfortable. Thus, humans may be wary of AI because we do not understand it, and as a result, we have no control over it.

Regarding the fear of AI in the workplace, some economists have noted that fears of automation and AI replacing workers have been overstated. Because work is more automated by AI, the productivity gains that ensue will increase labour demand across the economy, perhaps even in the same companies that are automating work with AI.

According to a recent study that considered 300 fictional and nonfictional works on AI, the worries that people have about intelligent machines can be grouped into four main categories [69]:

a) Identity-loss fear (also known as “inhumanity")
b) The anxiety of ‘obsolescence,’ or being obsolete
c) Concern that people may stop needing each other (also known as “alienation")
d) Being concerned that AI will rebel against humans

The most discussed concern is the first, and general agreement has been reached that growing automation will intensify the AI-fear factor. Such fear is associated with potential employment losses, especially considering the consequences of the Covid-19 pandemic.

Societal inequality is sometimes related to concerns regarding AI. One risk presented by AI technology is technology- or automation-related unemployment. People affected by technology-related unemployment lose their ability to make a living, which contributes to greater wealth inequality in societies where salaries are generally growing [70].

To summarise, as new, more advanced technologies become more widely used, working conditions improve. AI scientists are optimistic about the impact of AI in the future, but some academics believe that people may become impatient, lazy, and less intelligent owing to the increasing reliance on advanced technology.

10.2.2. Job losses

AI can cause job losses by mimicking human-intelligence processes and carrying out numerous routine tasks that are currently done by employees at considerably faster rates and with lower operational expenses [71].

A significant social concern is how robotics may alter the labour market. Economists and technology experts frequently examine the pace and extent to which technology may eliminate particular jobs from the workforce, as well as potential solutions to the ensuing unemployment. Jobs that are most likely to be automated in the future must be precisely identified to prevent large-scale unemployment.

Researchers are assessing the risk of automation for approximately 1000 currently existing occupations by objectively examining the extent to which robots and AI can replace the human capabilities required for specific jobs. The scientific methods they adopted may be particularly useful to governments for determining a population’s potential for unemployment [72].

Technological developments may directly impact employment through the displacement effect. Other developments that should be considered include increasing the need for labour in businesses that are already in operation or adding new jobs as a result of technological developments, dismissing workers outright from their current positions, and other strategies that may entail a productivity effect. Because automation has the potential to eliminate a wide variety of vocations, it is sometimes considered a severe danger to the global economy. Although an increase in the number of new AI-related occupations will occur, many new options will be open to people who need education and training, which may surprise many firms. Undoubtedly, training AI systems will be a top category of upcoming jobs, which is rapidly occurring [73].

AI will boost worker specialisation, production standards, and the value of higher human mental skills. Although AI is projected to have a positive impact on society and people, education and training will be crucial in preventing long-term unemployment and ensuring a qualified workforce. Although AI will accelerate economic development, some researchers noted that not everyone will benefit from it equally [74].

Businesses utilise AI to assist employees with their duties and promote collaboration among teams comprising both human and automated staff. However, given that AI is projected to have a big impact on workplaces and professions, it may make many individuals feel less connected to their jobs and increase their concern about being replaced. Researchers have shown that changes in employment, loss of status, and AI identity are three crucial signs of the threat of AI uniqueness in the workplace. Thus, understanding the identity threats posed by AI is critical.

Finally, researchers and industry professionals are aware of the effects AI will have on people’s identities as well as the crucial considerations to make when employing AI at work. Regardless of the final outcomes, robots will be able to perform a wider range of functions and jobs owing to AI advancements; however, this will also raise inequality and the potential for labour displacement. Some occupations that people perform today will eventually be replaced by machines [75].

10.2.3. Dehumanisation of jobs

Previous studies have used some well-known technologies, such as wearable computing devices, robotics, teleconferencing, and electronic monitoring systems, to show how technology influences labour, work systems, and organisations. Research has emphasised the significance of increasing the potential of AI rather than merely minimising its adverse effects on people and organisations. Despite the importance of AI technology in modern society, considering what will happen as our reliance on technology grows is crucial. The cost of the human component will decrease if human minds begin to adopt a “relaxed stance,” in which we unintentionally rely on robots or machines to make hypotheses and decisions on our behalf.

However, the utility of AI as a tool may differ [76]. Dehumanisation is closely associated with the concern of losing autonomy because it represents the belief that some people are not granted special human rights and, as a result, certain outgroups should not be granted the rights, privileges, or authority typically accorded to ingroups [77]. Losing one’s feelings of autonomy has a negative impact on one’s behaviour and well-being. For some researchers, technology accelerates the loss of human autonomy through invasive observation and covert manipulation during user–technology interactions.

Technology should not “dehumanise” us as people, drain our brainpower, and control our lives to the point where it replaces the fundamental interactions necessary for a person’s mental health, well-being, and skill development. Our social abilities gradually deteriorate if these experiences are eliminated. We may use technology to some extent to make our lives easier and to further research; however, too much technology use can prevent the human mind from thinking independently via trial and error and rob it of its mental processes. Therefore, completely removing or replacing the “human aspect” with technology is inappropriate.

Employee knowledge hiding is considered an example of a concerning problem for organisations that is negatively impacted by organisational dehumanisation. An individual who deliberately tries to withhold information requested by others at work is said to be “knowledge concealing.” However, in the long term, the effective operation of a company will be significantly impacted by employee knowledge concealment [78].

Employees may be detached from their unique qualities owing to the organisation’s constant pursuit of profit, and they may be reduced to little more than a function or instrument. According to some scientists, this experience as an employee is known as organisational dehumanisation [79].

Examining the acceptance of dehumanising attitudes and practises in the workplace has recently become an interesting field of study. Researchers who examined the actual acceptance of these attitudes concluded that there is relatively little support for them in light of the evidence emerging from social psychological and neuroscientific research, even though they frequently occur in organisational settings and are occasionally viewed as an acceptable and even necessary strategy for pursuing personal and organisational goals [80].

Societal dehumanisation and its relationship with cutting-edge technology were discussed in a study on employers’ negative impacts on employees. The study emphasised that the causes of societal dehumanisation depended on a variety of factors, including the nature of the industry, work practices, and managerial attitudes [81].

The year 2021 and the pandemic outbreak give us a vivid idea of the negative impacts of societal dehumanisation. With regard to perceived

organisational factors and dehumanising representations, a field study conducted in Italy during the Covid-19 outbreak among supermarket employees discovered a clear trace of weariness, bitterness, professional inefficacy, and other burnout-related negative effects [82].

Finally, the tendency to downplay AI's negative consequences on people and organisations has increased owing to important studies on subjects such as dehumanisation. Given the rapid improvements in and rising reliance on technology, specialists in business psychology and organisational behaviour have begun to pay close attention to how technology is changing work and employment. Given that it is being used to investigate how people behave at work, organisational dehumanisation has recently piqued the interest of several corporate-ethics scholars [83].

10.2.4. AI impact on employees' well-being

AI impacts employees' well-being either positively or negatively. Promoting workforce well-being has become a central theme in an AI-integrated workplace. Well-being is defined as a state of being that develops purpose and meaning in addition to material, intellectual, mental, emotional, and physical prosperity.

Many psychological fields have analysed the concept of well-being and its impact on human behaviour, relationships, and self-actualisation. The term "psychological well-being" refers to both inter- and intraindividual levels of positive functioning, which may encompass interpersonal relationships and self-referential attitudes such as a sense of self-worth and personal development. Subjective well-being reflects aspects of affective assessments of life satisfaction [84].

Researchers have identified the important difficulties associated with the new issue that has emerged regarding the investigation of the interactions between AI and social welfare and employee well-being. The creation and implementation of well-being surveys to evaluate the effects of AI, along with a focus on the successful implementation of community-based development strategies, represent the cornerstones of research for researchers developing AI-based methods to maintain or enhance societal well-being. Some theories contend that AI improves productivity and fosters greater worker autonomy, innovation, and flexibility. According to other experts, automation may have an adverse effect on workers, thereby leading to a loss of purpose or job instability [85].

10.2.5. AVs—safety and acceptance concerns

Some of the most incredible technological developments in computing in recent years include self-driving cars, computers that can recognise speech and images with accuracy, and machines that can outperform humans in challenging games. Creating artificially intelligent robots that can work independently, without supervision, and that can think, learn, and experience new things is one of the most exciting computer-science undertakings.

Currently, the UK defines self-driving cars as those with an automated lane-keeping system or self-driving technology that does not require driver supervision on highways.

Driving is performed by AI, but not all people may feel secure and comfortable with this manner of driving. The main reason that some people do not want self-driving cars is because of the AI that drives them. Notably, 71% of people are afraid to ride in completely autonomous vehicles. According to a poll conducted by the American Automobile Association, Inc. (AAA) in 2021, three out of four Americans still feared fully autonomous vehicles. According to AAA, consumer acceptability will be aided through testing, experience, and education, although many people do not want these cars on the road, even if they do not ride them themselves [86].

The functions of self-driving cars with AI processors have been assessed in many recent studies. Road safety is challenging, owing to the growing global population and automotive fleets. The transition from human-centred vehicle operation to self-driving vehicles has had a significant impact on the evolution of automobiles. Some researchers have emphasised that despite their high level of attractiveness, self-driving cars must address privacy, energy, traffic flow, environmental issues, and road safety.

The level of safety in cars has now increased owing to AI advancements, which enable smartphones to supply the necessary information at an incredibly fast rate while lowering the likelihood of human error. As AVs become more prevalent and persistent, the implications of their usage raise semi-philosophical problems regarding who is accountable for the errors made by AI. Potential legal loopholes in Great Britain may shield self-driving car users from prosecution for any transgression, even running a red light or engaging in reckless driving that kills someone [87].

In conclusion, even supporters of AVs acknowledge that tragic accidents will undoubtedly occur. Advocate organisations assert that using this new technology requires making sacrifices. Some writers associate self-driving vehicles with what they call 'programming killing.' They explore this contradiction and show how we fail to perceive the problems we are currently facing owing to our overly enthusiastic and cheery support for this technology. The absence of an in-depth moral analysis of the AV industry represents a potential threat [88].

11. Concluding remarks

To address the negative societal impacts of AI while maximising its benefits, AI ethics must be developed consistently. AI has no cultural or ethical background. Data and the representation of information are always required to feed an AI system. Some information, such as sex, age, and temperature, is simple to code and quantify. However, it is impossible to quantify complex emotions, beliefs, cultures, conventions, and values consistently. It is best for AI systems to try to maximise gains and reduce losses using mathematical principles because they are unable to process these complex concepts. To ensure sustainable growth, AI regulatory awareness and technology monitoring are highly desired.

The ongoing initiatives to create a cutting-edge technological environment must be aware of the underlying concerns related to AI ethics and privacy issues to fully reap the benefits of AI applications in society and the workplace. Inculcating human values into AI, promoting openness, and working together for the responsible evolution of AI are among the most crucial elements for preserving future societal trust in AI technology. Scientific research is a significant endeavour to ensure the prevailing accountability, safety, and ethical standards in the AI technology fields.

Justifying the importance of establishing clear-cut rules for AI applications requires the consideration that more than a mere concentration on legislation may be anticipated. However, in regulatory and compliance operations, concerns about future technologies may be overemphasised at the expense of pressing issues regarding already-deployed advancements. Although technological innovation enables the deployment of automation within businesses, prospective job losses and gains should be weighed against the ethical issues that current AI quick implementations are increasingly facing. In terms of strategy, the results and long-term changes that companies and workplaces desire to see in the individuals, groups, or positions that have been influenced serve as key predictors for a positive future course of action. However, this promising future seems to lack new ethical norms. Therefore, it is necessary to continue this direction of investigation.

Many AI ethical ideas were produced in previous years, which may generate contradictions and uncertainty among stakeholders regarding which one is preffered. Consequently, consistent revisions and collective scientific and international efforts must be maintained. The significant social influence of AI entails a growing need to adopt perfect ethical guidelines to ensure the steady, positive societal impact of AI. However, many groups that relate to various disciplines have assumed a variety of efforts aimed at establishing themselves as real pioneers in the arena of ethical guidelines for AI; thus, the scattered abundant outcome of proposed principles threatens to overwhelm and perplex the reader.

To determine the drawbacks of artificial intelligence in work and social environments accurately, scientifically, and without bias, it is necessary to have ethical controls that are widely accepted for their effectiveness and ability to work in a variety of environments, subject to improvement based on constant scientific development. AI is sometimes considered the most cutting-edge technology created by humans; thus, it must always have the potential to improve the quality of human society, the ability to enhance business processes, the capability to understand people's behavioural preferences, and the durability to offer customised support when necessary.

In summary, despite their high implementation costs, the degree to which AI, machine learning, and robotics will replace humans and the new ethical challenges that will be faced are not precisely known. AI may impact people's lives as a key area of current international research on intelligent manufacturing and robotics. Efficient AI processes can free humans from various dangerous and repetitive duties while improving the amount of work they can complete. Additionally, it can markedly increase working proficiency, productivity, and creative endeavours.

Similar to earlier automation advances, AI will raise production standards, labour specialisation, and the value of "human characteristics", such as creativity, problem-solving, and mathematical prowess. Not everyone will profit equally from this, even though AI will accelerate economic development. Despite some potential drawbacks, such as probable job losses, fears, and dehumanisation concerns, there is little proof that AI can genuinely replace people or take over control of the world. Because AI is the core component of computer learning, it is vital for the future of humanity.

Regarding the observations that highlight the potential directions for further research in this field and possible applications for the information provided in this review, it is anticipated that AI will have a significant social impact on sustainable development, climate change, and environmental concerns. According to theory, the use of advanced sensors will result in cleaner, less polluted, and more liveable cities. Significant ethical questions that necessitate in-depth study include privacy and surveillance, biases, and the philosophical conundrum of the function of human judgment.

Author statement

We thank the editor for giving us a chance to revise this paper. We also thank both reviewers for their pertinent comments that will help improve the strength of the paper.

Please find below answers to your queries that are also embedded in the manuscript in red.

Overview: Summary of changes, fresh data, and completed necessary analyses:

The authors have carefully read the comments and have made every effort to respond to each and every point. They took care to ensure that the results they reported were valid and backed up by scholarly work.

The introduction of two sections on research methodology and data collecting as well as defining the study's major problem are the main improvements made to fill in any gaps. Carefully addressing the issues of the Paper's extensiveness and preserving readability and flow represents another significant advance.

Data availability

The authors do not have permission to share data.

Acknowledgment

The authors would like to acknowledge the support of King Fahd University of Petroleum and Minerals, the Deanship for research oversight and coordination, and the Interdisciplinary Research Center for Intelligent Manufacturing and Robotics.

References

[1] T. Dietz, Theory and method in social impact assessment, Socio. Inq. 57 (1987) 54–69, https://doi.org/10.1111/j.1475-682X.1987.tb01180.x.
[2]] A. Viana-Lora, M.G. Nel-lo-Andreu, Approaching the social impact of research through a literature review, Int. J. Qual. Methods 20 (2021), 16094069211052189, https://doi.org/10.1177/16094069211052189.
[3] C.K. Riessman, Narrative methods for the human sciences, Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 11 (2008), https://doi.org/10.17169/fqs-11.1.1418.
[4] H. Snyder, Literature review as a research methodology: an overview and guidelines, J. Bus. Res. 104 (2019) 333–339, https://doi.org/10.1016/j.jbusres.2019.07.039.
[5] G. Nissim, T. Simon, The future of labor unions in the age of automation and at the dawn of AI, Technol. Soc. 67 (2021), 101732, https://doi.org/10.1016/j.techsoc.2021.101732.
[6] L. Wang, Artificial intelligence for COVID-19: a systematic review, Front. Med. 8 (2021), https://doi.org/10.3389/fmed.2021.704256.
[7] P. Egaña del Sol, G. Cruz, A. Micco, COVID-19's impact on the labor market shaped by automation: evidence from Chile, SSRN Electron. J. (2021), https://doi.org/10.2139/ssrn.3761822.
[8] Eric Dahlin, Are Robots Stealing Our Jobs? Socius Sociological Research for a Dynamic World 5 (2019), https://doi.org/10.1177/2378023119846249.
[9] Q.C. Pham, R. Madhavan, L. Righetti, W. Smart, R. Chatila, The impact of robotics and automation on working conditions and employment [ethical, legal, and societal issues], IEEE Robot. Autom. Mag. 25 (2018) 126–128, https://doi.org/10.1109/MRA.2018.2822058.
[10] S. Ivanov, M. Kuyumdzhiev, C. Webster, Automation fears: drivers and solutions, Technol. Soc. 63 (2020), 101431, https://doi.org/10.1016/j.techsoc.2020.101431. ISSN 0160-791X. https://www.sciencedirect.com/science/article/pii/S0160791X20300488, 10.1016/j.techsoc.2020.101431.
[11] M. Klenk, How do technological artefacts embody moral values? Philos, Technol. 34 (2021) 525–544, https://doi.org/10.1007/s13347-020-00401-y.
[12] Rabia Tahseen, Uzma Omer, Shoaib Farooq, Ethical guidelines for artificial intelligence: a systematic literature review, VFAST Transactions on Software Engineering 9 (2021) 33–47, https://doi.org/10.21015/vtse.v9i3.701.
[13] I.M. Enholm, E. Papagiannidis, P. Mikalef, et al., Artificial intelligence and business value: a literature review, Inf. Syst. Front (2021).
[14] V. Braun, V. Clarke, Using thematic analysis in psychology, Qual. Res. Psychol. 3 (2006) 77–101, https://doi.org/10.1191/1478088706qp063oa.
[15] B. Czarniawska, Narratives in Social Science Research, in: Introducing Qualitative Methods, vol. 2011, SAGE Publications, Ltd, Series, 2004, https://doi.org/10.4135/9781849209502. **Online publication date:** January 01.
[16] N.R. Haddaway, et al., Eight problems with literature reviews and how to fix them, Nat. Ecol. Evol. 4 (2020) 1582–1589, https://doi.org/10.1038/s41559-020-01295-x.
[17] D. Byrne, Worked example of Braun and Clarke's approach to reflexive thematic analysis, Qual. Quant. 56 (2022) 1391–1412, https://doi.org/10.1007/s11135-021-01182-y.
[18] A. Viana-Lora, M.G. Nel-lo-Andreu, Approaching the social impact of research through a literature review, Int. J. Qual. Methods 20 (2021), 16094069211052189, https://doi.org/10.1177/16094069211052189.
[19] Kirstie McAllum, Stephanie Fox, Mary Simpson, Christine Unson, A comparative tale of two methods: how thematic and narrative analyses author the data story differently, Communication Research and Practice 5 (2019) 1–18, https://doi.org/10.1080/22041451.2019.1677068.
[20] G. Wagner, R. Lukyanenko, G. Paré, Artificial intelligence and the conduct of literature reviews, J. Inf. Technol. 37 (2) (2022) 209–226, https://doi.org/10.1177/02683962211048201.
[21] R. Ferrari, Writing narrative style literature reviews, The European Medical Writers Association 24 (2015) 230–235, https://doi.org/10.1179/2047480615Z.000000000329.
[22] G.D.M.R. Lima, T. Wood Jr., The social impact of research in business and public administration, Rev. Adm. Empres. 54 (2014) 458–463, https://doi.org/10.1590/S0034-759020140410.
[23] J. Onyx, Social impact, a theoretical model, Cosmop. Civ. Soc. 6 (2014) 1–18, https://doi.org/10.5130/ccs.v6i1.3369.
[24] J. Mökander, R. Schroeder, AI and social theory, AI Soc. 37 (2022) 1337–1351, https://doi.org/10.1007/s00146-021-01222-z.
[25] Social impact theory - simply psychology. https://www.simplypsychology.org/social-impact-theory.html#:~:text=Social%20impact%20theory%20was%20created,threatened%2C%20and%20supported%20by%20others. (Accessed 2 January 2023). Last accessed.
[26] K. Joyce, Toward a sociology of artificial intelligence: a call for research on inequalities and structural change, Socius 7 (2021), 237802312199958, https://doi.org/10.1177/2378023121999581.
[27] S. Moore, S. Brown, W. Butler, AI and social impact: a review of current use cases and broader implications, in: Book: Cybersecurity Capabilities in Developing Nations and its Impact on Global Security, 2022, pp. 133–161, https://doi.org/10.4018/978-1-7998-8693-8.ch008.
[28] N. Bostrom, Transhumanist values, Rev. Contemp. Philos. 30 (2003) 3–14, https://doi.org/10.5840/jpr_2005_26.
[29] K.A. Fox, The present status of objective social indicators: a review of theory and measurement, Am. J. Agric. Econ. 68 (1986) 1113–1120, https://doi.org/10.2307/1241860.

[30] S. Ransbotham, F. Candelon, D. Kiron, B. LaFountain, S. Khodabandeh, The cultural benefits of artificial intelligence in the enterprise, MIT sloan management review and boston consulting group, Available at: https://boardmember.com/wp-content/uploads/2022/06/BCG-The-Hidden-Cultural-Benefits-of-AI.pdf, November 2021. (Accessed 2 January 2023). Last accessed.
[31] B.C. Stahl, Concepts of ethics and their application to AI, Artificial Intelligence for a Better Future 18 (2021) 19–33, https://doi.org/10.1007/978-3-030-69978-9_3.
[32] Muel Kaptein, Johan Wempe, Three general theories of ethics and the integrative role of integrity theory, SSRN Electron. J. (2002), https://doi.org/10.2139/ssrn.1940393.
[33] V.O. Tamunosiki, John Stuart Mill's utilitarianism: a critique, Int. J. Peace Conflict Stud. 5 (2021) 65–76. Retrieved from, http://journals.rcmss.com/index.php/ijpcs/article/view/174. (Accessed 2 January 2023). Last accessed.
[34] I. Tussyadiah, A. Tuomi, E. Ling, G. Miller, G. Lee, Drivers of organizational adoption of automation, Ann. Tourism Res. 93 (2022), 103308, https://doi.org/10.1016/j.annals.2021.103308.
[35] E.J. Milner-Gulland, Interactions between human behavior and ecological systems, Philos. Trans. R. Soc. Lond. B Biol. Sci. 367 (2012) 270–278, https://doi.org/10.1098/rstb.2011.0175.
[36] A.W. Oliveira, Theorizing technology and behavior: introduction to special issue, Hum. Behav, Emerg. Technol. 2 (2020) 302–306, https://doi.org/10.1002/hbe2.219.
[37] O. Gábor, Behavior of Artificial Intelligence: Summa Aethologica Intelligentiae Artificialis, 2020, https://doi.org/10.15170/BTK.2020.00002 published by GeniaNet.
[38] P.K. Donepudi, Application of artificial intelligence in automation industry, Asian J. Appl. Sci. Eng. 7 (2018) 7–20. Retrieved from, https://upright.pub/index.php/ajase/article/view/23.
[39] Y. Griep, I. Vranjes, M.M. van Hooff, D.G. Beckers, S.A. Geurts, Technology in the Workplace: Opportunities and Challenges, Flexible Working Practices and Approaches: Psychological and Social Implications, 2021, pp. 93–116, https://doi.org/10.1007/978-3-030-74128- 0_6.
[40] M. Yamin, Information technologies of 21st century and their impact on the society, Int. J. Inf. Tecnol. 11 (2019) 759–766, https://doi.org/10.1007/s41870-019-00355-1.
[41] What's next for AI – Building trust (ibm.com) IBM Cognitive – What's next for AI (November 4, 2022). Last accessed.
[42] N. Morchid, The current state of technology acceptance: a comparative study, J. Bus. Manag. 22 (2020) 1–16.
[43] Pablo Egana-delSol, Gabriel Cruz, Alejandro Micco, COVID-19's impact on the labor market shaped by automation: evidence from Chile, Available at: SSRN: https://ssrn.com/abstract=3761822, January 7, 2021 https://doi.org/10.2139/ssrn.3761822.
[44] A.A. Naikoo, S.S. Thakur, T.A. Guroo, A.A. Lone, Development of society under the modern technology - a review, Scholedge Int. J. Bus. Policy & Gov. 5 (2018) 1–8, https://doi.org/10.19085/journal.sijbpg050101.
[45] H. Thamik, J. Wu, The impact of artificial intelligence on sustainable development in electronic markets, Sustainability 14 (2022) 3568, https://doi.org/10.3390/su14063568.
[46] U.V.U. Ismatullaev, S.H. Kim, Review of the factors affecting acceptance of AI-infused systems, Hum. Factors (2022), https://doi.org/10.1177/00187208211064707.
[47] I.M. Enholm, E. Papagiannidis, P. Mikalef, J. Krogstie, Artificial intelligence and business value: a literature review, Inf. Syst. Front 24 (2021) 1709–1734, https://doi.org/10.1007/s10796-021-10186-w.
[48] K. Siau, W. Wang, Artificial intelligence (AI) ethics: ethics of AI and ethical AI, J. Database Manag. 31 (2020) 74–87, https://doi.org/10.4018/JDM.2020040105.
[49] G. Damioli, V. Van Roy, D. Vertesy, The impact of artificial intelligence on labor productivity, Eurasian Bus. Rev. 11 (2021) 1–25, https://doi.org/10.1007/s40821-020-00172-8.
[50] T. Zhang, D. Tao, X. Qu, X. Zhang, J. Zeng, H. Zhu, H. Zhu, Automated vehicle acceptance in China: social influence and initial trust are key determinants, Transport. Res. C Emerg. Technol. 112 (2020) 220–233, https://doi.org/10.1016/j.trc.2020.01.027.
[51] S.-A. Kaye, I. Lewis, S. Forward, P. Delhomme, A priori acceptance of highly automated cars in Australia, France, and Sweden: a theoretically-informed investigation guided by the TPB and UTAUT, Accid, Anal. Prev. 137 (2020), 105441, https://doi.org/10.1016/j.aap.2020.105441.
[52] G. Zhu, Y. Chen, J. Zheng, Modelling the acceptance of fully autonomous vehicles: a media-based perception and adoption model, Transp. Res. F: Traffic Psychol. Behav. 73 (2020) 80–91, https://doi.org/10.1016/j.trf.2020.06.004.
[53] B. Schoettle, M. Sivak, Public Opinion about Self-Driving Vehicles in China, vol. 5, Transportation Research Institute, India, Japan, the U.S., the U.K., and Australia, 2014, pp. 53–59. University of Michigan, Ann Arbor.
[54] H. Abraham, C. Lee, S. Brady, C. Fitzgerald, B. Mehler, B. Reimer, J.F. Coughlin, Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age, in: Proceedings of the Transportation Research Board 96th Annual Meeting, Transportation Research Board, Washington, DC, 2017, January, pp. 8–12.
[55] M. Mohammadzadeh, Sharing or owning autonomous vehicles? Comprehending the role of ideology in the adoption of autonomous vehicles in the society of automobility, Transp. Res. Interdiscip. Perspect. 9 (2021), 100294, https://doi.org/10.1016/j.trip.2020.100294.
[56] K. Gopinath, G. Narayanamurthy, Early bird catches the worm! Meta-analysis of autonomous vehicles adoption – moderating role of automation level, ownership and culture, Int. J. Inf. Manag. 66 (2022), 102536, https://doi.org/10.1016/j.ijinfomgt.2022.102536.
[57] M. Cunneen, M. Mullins, F. Murphy, Autonomous vehicles and embedded artificial intelligence: the challenges of framing machine driving decisions, Appl. Artif. Intell. 33 (2019) 706–731, https://doi.org/10.1080/08839514.2019.1600301.
[58] F. Cingano, Trends in Income Inequality and its Impact on Economic Growth, in: OECD Social, Employment and Migration Working Papers, vol. 163, OECD Publishing, 2014, https://doi.org/10.1787/5jxrjncwxv6j-en.
[59] G.A. Legault, C. Verchère, J. Patenaude, Support for the development of technological innovations: promoting responsible social uses, Sci. Eng. Ethics 24 (2018) 529–549, https://doi.org/10.1007/s11948-017-9911-5.
[60] S.C. Olesen, P. Butterworth, L.S. Leach, M. Kelaher, J. Pirkis, Mental health affects future employment as job loss affects mental health: findings from a longitudinal population study, BMC Psychiatr. 13 (2013) 1–9, https://doi.org/10.1186/1471-244X-13-144.
[61] K. An, Y. Shan, S. Shi, Impact of industrial intelligence on total factor productivity, Sustainability 14 (2022), 14535, https://doi.org/10.3390/su142114535.
[62] O.A. Osoba, W. Welser, The Risks of Artificial Intelligence to Security and the Future of Work, vol. 2023, RAND Corporation, Santa Monica, CA, 2017. Last accessed January 2,, https://www.rand.org/pubs/perspectives/PE237.html.
[63] G.M. Alarcon, A. Capiola, I.A. Hamdan, M.A. Lee, S.A. Jessup, Differential biases in human-human versus human-robot interactions, Appl. Ergon. 106 (2023), 103858, https://doi.org/10.1016/j.apergo.2022.103858.
[64] M. Zajko, Artificial intelligence, algorithms, and social inequality: sociological contributions to contemporary debates, Soc. Compass 16 (2022), e12962, https://doi.org/10.1111/soc4.12962.
[65] D. Elliott, E. Soifer, AI Technologies, privacy, and security, Front. Artif. Intell. 5 (2022), 826737, https://doi.org/10.3389/frai.2022.826737.
[66] A.L. Ligozat, J. Lefevre, A. Bugeau, J. Combaz, Unraveling the hidden environmental impacts of AI solutions for environment life cycle assessment of AI solutions, Sustainability 14 (2022) 5172, https://doi.org/10.3390/su14095172.
[67] M.H. Jarrahi, Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making, Bus. Horiz. 61 (2018) 577–586, https://doi.org/10.1016/j.bushor.2018.03.007.
[68] L. Liu, X. Song, Y. Li, The emotional mechanisms of interpersonal preemptive behavior, Front. Psychol. 13 (2022), 841960, https://doi.org/10.3389/fpsyg.2022.841960.
[69] S. Cave, K. Dihal, Hopes and fears for intelligent machines in fiction and reality, Nat. Mach. Intell. 1 (2019) 74–78, https://doi.org/10.1038/s42256-019-0020-9.
[70] Mario Daniele Amore, Valerio Pelucco, Fabio Quarato, Family ownership during the Covid-19 pandemic, J. Bank. Finance 135 (2022) 2022, https://doi.org/10.1016/j.jbankfin.2021.106385, 106385. ISSN 0378-4266, https://www.sciencedirect.com/science/article/pii/S0378426621003368.
[71] H. Ping, G.Y. Ying, Comprehensive view on the effect of artificial intelligence on employment, Topics in Education, Culture and Social Development (TECSD) 1 (2018) 32–35, https://doi.org/10.26480/ismiemls.01.2018.32.35.
[72] A. Paolillo, et al., How to compete with robots by assessing job automation risks and resilient alternatives, Sci. Robot. 7 (2022), https://doi.org/10.1126/scirobotics.abg5561 eabg5561.
[73] A. Oliveira, H. Braga, Artificial intelligence: learning and limitations, WSEAS Trans. Adv. Eng. Educ. 17 (2020) 80–86, https://doi.org/10.37394/232010.2020.17.10.
[74] Mike Zajko, Artificial intelligence, algorithms, and social inequality: sociological contributions to contemporary debates, Sociology Compass 16 (2022), https://doi.org/10.1111/soc4.12962.
[75] M. Mirbabaie, F. Brünker, N.R. Möllmann, S. Stieglitz, The rise of artificial intelligence – understanding the AI identity threat at the workplace, Electron, Market 32 (2022) 73–99, https://doi.org/10.1007/s12525-021-00496-x.
[76] S. Chiodo, Human autonomy, technological automation (and reverse), AI Soc. 37 (2022) 39–48, https://doi.org/10.1007/s00146-021-01149-5.
[77] N. Haslam, Dehumanization: an integrative review, Pers. Soc. Psychol. Rev. 10 (2006) 252–264, https://doi.org/10.1207/s15327957pspr1003_4.
[78] U.E. Rubbab, S.A. Khattak, H. Shahab, N. Akhter, Impact of organizational dehumanization on employee knowledge hiding, Front. Psychol. 13 (2022) 80, https://doi.org/10.3389/fpsyg.2022.803905.
[79] N. Nguyen, T. Besson, F. Stinglhamber, Emotional labor: the role of organizational dehumanization, J. Occup. Health Psychol. 27 (2022) 179–194, https://doi.org/10.1037/ocp0000289.
[80] Y. Zhang, A new look of dehumanization in work domain: the relationship between communication means and disrespect to deliveryman, Psychology 11 (2020) 572–580, https://doi.org/10.4236/psych.2020.114038.
[81] E. Anderson, Private Government: How Employers Rule Our Lives (And Why We Don't Talk about it), in: University Center for Human Values Series, vol. 44, Princeton University Press, 2017, 9780691192246, https://doi.org/10.2307/j.ctvc775n0.
[82] R.R. Valtorta, C. Baldissarri, C. Volpato, Burnout and workplace dehumanization at the supermarket: a field study during the COVID- 19 outbreak in Italy, J. Community Appl. Soc. Psychol. 32 (2022) 767–785, https://doi.org/10.1002/casp.2588.
[83] W.F. Cascio, R. Montealegre, How technology is changing work and organizations, Annu. Rev. Organ. Psychol. Organ. Behav. 3 (2016) 349–375, https://doi.org/10.1146/annurev-orgpsych-041015-062352.
[84] F. Martela, K.M. Sheldon, Clarifying the concept of well-being: psychological need satisfaction as the common core connecting eudaimonic and subjective well-being, Rev. Gen. Psychol. 23 (2019) 458–474, https://doi.org/10.1177/1089268019880886.
[85] L. Musikanski, B. Rakova, J. Bradbury, R. Phillips, M. Manson, Artificial intelligence and community well-being: a proposal for an emerging area of

research, Int. Journal of Com. W.B. 3 (2020) 39–55, https://doi.org/10.1007/s42413-019- 00054-6.
[86] E. Edmonds, Three in four Americans remain afraid of fully self-driving vehicles | AAA Newsroom, AAA International Relations (2019). Last accessed November 4, 2022.
[87] J. Wang, L. Zhang, Y. Huang, J. Zhao, F. Bella, Safety of autonomous vehicles, J. Adv. Transport. 2020 (2020) 1–13, https://doi.org/10.1155/2020/8867757.
[88] N. JafariNaimi, Our bodies in the trolley's path, or Why self-driving cars must *not* be programmed to kill, Sci. Technol. Hum. Val. 43 (2018) 302–323, https://doi.org/10.1177/0162243917718942.
[89] Tesla's 'phantom braking' problem is now being investigated by the US government - the Verge, accessed 1/2/23, Tesla's 'phantom braking' problem is now being investigated by the US government - The Verge (January 2, 2023). Last accessed.
[90] Daily briefing: ChatGPT listed as author on research papers: many scientists disapprove (nature.com) (January 2, 2023). Last accessed.

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

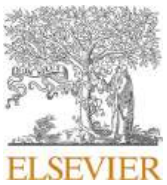
ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
7. The darker side of positive AI attitudes: investigating associations with (Problematic) social media use (2025)	Addictive Behaviors Reports (Article from : Elsevier Ltd)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Addictive Behaviors Reports 22 (2025) 100613



Contents lists available at ScienceDirect

Addictive Behaviors Reports

journal homepage: www.elsevier.com/locate/abrep

The darker side of positive AI attitudes: Investigating associations with (problematic) social media use☆

Christian Montag [a,b,c,*], Jon D. Elhai [d]

[a] Centre for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau, Macau SAR, PR China
[b] Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau SAR, PR China
[c] Department of Psychology, Faculty of Social Sciences, University of Macau, Macau SAR, PR China
[d] Department of Psychology, and Department of Neurosciences and Psychiatry, University of Toledo, Toledo, OH, USA

ARTICLE INFO

Keywords:
Artificial Intelligence
Attitudes
Problematic Social Media Use
Social Media Addiction

ABSTRACT

In societies around the world, the impact of artificial intelligence (AI) is being fiercely discussed. It is difficult to grasp AI's influence, because AI represents a general-purpose technology, which can be applied in different settings. One product in which AI plays a pivotal role is social media. In this context, for instance, AI is used to provide people with personalized newsfeeds to prolong time spent online, which might result in addictive-like behavior. Many factors such as sociodemographic variables, history of psychopathology and personality traits have been revealed as risk factors for developing problematic social media use patterns. Yet, to our knowledge attitudes toward AI have not been examined in association with problematic social media use. In a sample of n = 956 social media users, we observed that positive AI attitudes were linked to overuse of social media as assessed with an addiction framework. The effect size of this association was stronger for males than females. Further we observed that this association was mediated by time spent on social media. The present study shows that positive AI attitudes – although well-known to be positive regarding embracing new technologies – might come with risks for developing addictive patterns of technology use, such as social media.

1. Introduction

At the moment of writing, more than five billion people use social media (Statista, 2022). Social media represents a product which heavily relies on artificial intelligence (AI), for instance to present users with a personalized newsfeed such as on Facebook or to recommend videos to users on platforms such as YouTube (Guha, 2021). Social media platforms have been designed in such a way to prolong users' time spent online (Montag & Elhai, 2023; Sindermann, Montag, & Elhai, 2022). Consequently, people increasingly leave more digital footprints on the platforms, which can be exploited to gain important insights into user characteristics to target users with personalized ads (Matz, Kosinski, & Stillwell, 2017; Zarouali, Dobber, De Pauw, & de Vreese, 2020). The big data business model has been critiqued from various angles such as loss of privacy and worsening well-being (Montag & Hegelich, 2020). In the context of addictive behaviors, it has been debated how the design of social media platforms might be responsible for development of addictive-like social media use (Flayelle et al., 2023; Montag, Lachmann, Herrlich, & Zweig, 2019; Montag, Thrul, & van Rooij, 2022). In this regard, AI plays a pivotal role, because without this technology it would be impossible to personalize social media (Salma et al., 2024), as detailed above.

In the past, several theoretical models have been developed to predict who may develop problematic social media use patterns (studied within an addiction framework). For instance, the prominent I-PACE model (Brand et al., 2016, 2019), involving the Interaction of Person, Affect, Cognition and Execution variables, conceptualizes that among the Person/dispositional variables, specific sociodemographic variables such as age, history of psychopathology, genetics or personality might predict problematic social media use (PSMU). Additionally, other risk factors such as affective and cognitive responses, and Internet-related cognitive biases, play a role in influencing development or maintenance of PSMU. In this context we propose that AI attitudes might also play an important role. We believe that AI attitudes might belong to the

☆ This article is part of a special issue entitled: 'Behavioral addictions' published in Addictive Behaviors Reports.

* Corresponding author at: Centre for Cognitive and Brain Sciences, Institute of Collaborative Innovation, University of Macau, Research Building N21, Avenida da Universidade, Taipa, Macau SAR, PR China.
E-mail address: cmontag@um.edu.mo (C. Montag).

https://doi.org/10.1016/j.abrep.2025.100613
Received 28 January 2025; Received in revised form 15 April 2025; Accepted 30 April 2025
Available online 8 May 2025

C-variable within the I-PACE model, because they could represent a cognitive response to the rise of this new technology.

Classic theories such as the Theory of Planned Behavior (TPB) (Ajzen, 1991) or Technology Acceptance Model (TAM) (Venkatesh & Davis, 2000) put forward the proposal that a positive attitude toward technology is an important prerequisite to use a technology. These theories and the proposals have been recently transferred to AI use (Montag & Ali, 2025b). Positive AI attitudes therefore might be seen as something promising, because they might help people embrace new technologies, making them more productive in everyday life tasks. On the other hand, such positive AI attitudes might result in overreliance on AI technologies or overuse tendencies (see recent works on overreliance on AI; Buçinca et al., 2021; Klingbeil et al., 2024). As social media represents a product strongly relying on AI, we were interested in understanding if positive AI attitudes link to PSMU severity. For reasons of completeness, we also investigated negative AI attitudes with PSMU, where inverse associations were expected. A further reflection on the hypotheses is necessary: Social media represents a product where AI is built-in, which users may not necessarily be aware of. Hence the question arises – also in light of TPB and TAM – whether users with more positive AI attitudes actively choose social media use (or develop excessive social media use) due to its AI character or if this happens due to being in general more tech-savvy. This is something we will reflect on deeper in the discussion.

2. Methods

2.1. On the investigated sample

The sample here was already described in Montag et al. (2023) and analyzed regarding associations between PSMU, meaning in life and fear of missing out. The larger dataset also included additional variables, which allowed investigation of other research questions such as trusting AI (Montag et al., 2024) and an investigation of links between Shinrin-Yoku and life satisfaction (Montag, 2024). These projects were all preregistered. For the present work, we included PSMU as investigated in Montag et al. (2023) and Attitudes toward AI (ATAI) as investigated by Montag, Becker & Li (Montag et al., 2024). Please note that the papers differ slightly regarding the final sample sizes due to slightly different data cleaning steps from adding measures. For the present paper please see the data cleaning steps in the next section. The present paper investigates, for the first time, associations between attitudes toward AI and PSMU, which has not been done in the above mentioned papers in this section.

2.2. Data cleaning

An initial sample of 1151 participants was recruited via Bilendi GmbH (a company supporting scientists in conducting online surveys). The company was asked to recruit a population sample with about the same number of male and female participants with a large age-range (to not rely on typically available student samples). After excluding participants who did not answer the survey without interruptions, participants reporting a third gender (unfortunately underrepresented), not providing full consent, failing an attention item, or not being competent in the German language, we ended up with 1082 participants. Please note that we refrained from doing careless responding analysis on measures not relevant to the present paper. Beyond the paper by Montag et al. (2023) we also investigated time spent on social media. Those who reported spending more than 16 h a day on social media (personal and work combined) were excluded as outliers. We further checked for missing data or computation errors, which led to a final sample of n = 1048 participants (523 males, 525 females; mean-age: 45,0, SD = 14,4). Education level was as follows (for precision, German words are used): without school = 0.2 %, Volks-/Hauptschulabschluss (Primary/Secondary School Certificate): 7.8 %, Mittlere Reife (Intermediate School Certificate): 28.6 %, Fachabitur (Advanced Technical College Entrance Qualification): 7.3 %, Abitur (General University Entrance Qualification): 20.2 %, Fachhochschulabschluss (Advanced Technical College Degree): 12.5 %, Hochschulabschluss (university degree): 23.4 %. The education degrees are presented in ascending order. Please note that the final sample of social media users differs from our earlier work on PSMU (stemming from the same data set; Montag et al., 2023): 955 vs. 956 participants given a slightly different data cleaning approach.

The study was approved by the ethics committee at Ulm University, Ulm, Germany.

2.3. Questionnaires

The German and English versions of the ATAI each consist of five items, here answered via a five-point Likert scale ranging from "1 = strongly disagree" to "5 = strongly agree" (Sindermann et al., 2021). Two items form the acceptance of AI scale (α = 0.77, ω = 0.77), and three items form the fear of AI scale (α = 0.77, ω = 0.78). The two acceptance items are: "I trust artificial intelligence" and "Artificial intelligence will benefit humankind". The three fear items are: "I fear artificial intelligence", "Artificial intelligence will destroy humankind" and "Artificial intelligence will cause many job losses". Higher scores indicate greater acceptance of AI or greater fear of AI, respectively.

In addition, two single items were investigated, allowing us to obtain insights into global AI attitudes. These items included "I have a positive attitude toward AI" and "I have a negative attitude toward AI." These items are also answered via a five-point Likert scale ranging from "1 = strongly disagree" to "5 = strongly agree". The items have been validated recently (Montag, Schulz, et al., 2025; Montag & Ali, 2025a; Naiseh et al., 2025).

Social media use and PSMU were assessed as follows: First, participants were asked if they use social media (original wording translated from the German item: "I use social media. This includes platforms such as Facebook, Instagram, TikTok, YouTube, LinkedIn, Snapchat, as well as messaging apps like WhatsApp, Signal, or Telegram."; yes = 956 / no = 92). If reporting use of social media, they were subsequently asked about the use for personal and work purposes in minutes per day (average estimates). Finally, they completed the Social Networking Sites – Addiction Test (SNS-AT) consisting of six items answered on a five-point Likert scale ranging from "1 = strongly disagree" to "5 = strongly agree" (Montag et al., 2023). The SNS-AT assesses individual differences in PSMU tendencies, with higher scores indicating greater tendencies. Internal consistency for the SNS-AT was as follows: α = 0.89, ω = 0.89.

2.4. Statistical Analyses

Data cleaning was conducted with SPSS 30.0.0.0. The next statistical analyses were computed with the Jamovi package 2.3.28.0. Descriptive statistics were produced first, which were followed by t-tests providing insights into associations between social media use (yes/no) and attitudes toward AI. Also t-tests/Mann-Whitney U tests were computed to assess gender effects based on the smaller sample of social media users. For this investigation we used Julius.ai to find a (near) perfect match for the n = 92 participants reporting to not use social media regarding the variables of gender, age and education (in the following order of importance in the matching process). The samples are comparable regarding gender (each sample including 54 males and 38 females), near identical regarding age (non-users: M = 53,05, SD = 12,56 vs. users: M = 53,04, SD = 12,55; $t_{(182)}$ = 0,006, p = 0.995) and education (Chi^2 = 4.60, df = 5, p = 0.47). This strategy was chosen in order to avoid testing a small sample such as 92 vs. 956 participants.

Further, correlations between all variables of interest are presented in a next step for the male and female subsamples of social media users. Here we focused on Spearman correlations due to the skewed distribution of several variables. Finally, within this slightly smaller social

media use sample (males plus females), a mediation model is presented (using Mplus 8.11 software) investigating associations between positive AI attitudes (predictor variable) and PSMU (outcome variable) with time spent on social media (personal + work time) being a mediator. We used maximum likelihood estimation for mediation, using the Delta method to compute cross-products of direct path coefficients with 1000 non-parametric bootstrapped replications; we present standardized (STDYX) coefficients. Predictor and outcome terms do not imply causality here.

3. Results

First, we present descriptive statistics for AI attitudes, social media use and problematic use in Table 1. The social media variables used the smaller sample of social media users. All descriptive statistics are also presented for the male and female subsamples in Table 2 (given several gender differences being backed up by S-Table 1 via t-tests and Mann-Whitney U tests in the supplementary material; some variables had skewed distributions so we provide insights using both parametric and non-parametric tests).

Second, we investigated if using social media (yes/no) was associated with differences in AI attitudes. We investigated this question in social media and non-social-media users being matched regarding sociodemographics as reported in the statistical analyses section. Among others it could be observed that users of social media had greater positive AI attitudes than those not using social media (ATAI +: $t_{(182)} = 3.34$, $p < 0.001$, M = 6,10 (SD = 1,52) vs. M = 5,29 (SD = 1,74), Cohen's d = 0.492; single item AI +: $t_{(182)} = 3.96$, $p < 0.001$, M = 3,20 (SD = 0,91) vs. M = 2,65 (SD = 0,95), Cohen's d = 0.585).

Among users of social media, we further investigated how AI attitudes would be linked to time spent on social media and PSMU tendencies. Given the effects of gender on some of the variables, we present correlation patterns for males and females separately (Tables 3 and 4). For both genders we observed that the acceptance of AI scale and single item positive AI attitude measure were positively linked to both time-spent on social media and PSMU scores. Associations between positive AI attitudes and PSMU were in the small to moderate range for males and in the very mild area for females. The negative AI attitudes scales showed no robust associations with social media use or PSMU (but see a small positive association in the male subsample for the ATAI −). Please note that we report here correlations controlled for age. The correlation tables not controlled for age can be found in the supplementary material (see S-Table 2 and S-Table 3).

Basing on the correlational analysis, we prepared an example of a mediation model suggesting that the association between more positive AI attitudes (ATAI +: acceptance of AI) and greater PSMU tendencies is mediated by time spent on social media (combined personal and business time). This model was based on the idea that AI attitudes (positive) could result in prolonged use time on social media which might consequently result in addictive social media tendencies. Also, alternative sequences of variables are possible of course, given the cross-sectional nature of the dataset. Next, we conducted a slight revision of this model, adding paths from age and gender (as covariates) to PSMU, finding that even with age and gender variables, the indirect/mediation effect was still significant (see Table 5 and Table 6).

4. Discussion

The present study investigated associations between global AI attitudes and both social media use and overuse. We discovered that positive AI attitudes were linked with using social media (yes vs. no) and to higher PSMU scores. These findings could provide evidence that more positive AI attitudes might be a risk factor for developing greater PSMU levels (but causality cannot be inferred from the present data). Social media itself relies heavily on AI technology explaining why this link could be established. Interestingly, negative AI attitudes showed less robust associations with social media use patterns. This was not expected, but is in line with prior observations that acceptance of AI might be more important to understand why people embrace AI technology in contrast to the role of fearing AI (Sindermann et al., 2021).

The findings observed here therefore not only underline that positive AI attitudes are associated with PSMU, but also that it is meaningful to distinguish between positive and negative AI attitudes (hence not thinking of the constructs as one dimension with two poles). This finding needs to be tested in the future by applying further AI attitude inventories and also focusing more on the role of negative AI attitudes. Here we would also not be surprised to see positive associations with PSMU (in contrast to our initial hypothesis), perhaps due to the link of negative emotions being part of both constructs (and perhaps those fearing AI might want to escape their negative emotions by excessively checking social media). But this is something we have only seen in the male subsample (small effect size of about .10). Other inventories to be tested are for instance the ATTARI-12, which has been put forward as a one-dimensional attitude towards AI scale (Stein, Messingschlager, Gnambs, Hutmacher, & Appel, 2024), which could also be investigated regarding potential links with PSMU (and then contrasted with the present findings). For a further discussion on AI attitude measures, see a recent chapter providing a good overview (Schepman & Rodway, 2025). Further research options of course would be the longitudinal investigation of AI attitudes and PSMU to shed light on the question of whether positive AI attitudes indeed might be a risk factor for developing PSMU.

Another point to reflect on has been mentioned shortly at the end of the introduction. Although social media relies heavily on artificial intelligence technology, people not necessarily might be aware of it. When interacting with a large language model such as ChatGPT, a person actively starts to use an AI agent to obtain information. On a social media platform, you might simply log on and browse and indirectly rely on what the AI recommends to you. Hence, it is unclear if PSMU is indeed linked to positive AI attitudes or perhaps a construct such as general tech-savviness overlapping with such positive attitudes. Therefore, also the question arises, if a positive attitude toward AI is not only linked to PSMU, but also to other problematic online behaviors. Hence, the question can be posed if having a positive AI attitude and being more tech-savvy makes a person more vulnerable to overuse technologies in general (e.g. developing generalized problematic Internet use behaviors; Davis, 2001). But it could also be the case that positive AI attitudes map more onto specific problematic online behaviors, in particular if certain platforms in the area of gaming/gambling/shopping, etc. would rely

Table 1
Descriptive statistics of the complete sample.

	N	Missing	Mean	Median	SD	Minimum	Maximum
Accepting AI (ATAI +)	1048	0	6.19	6.00	1.755	2	10
Fear of AI (ATAI −)	1048	0	8.20	8.00	2.690	3	15
Single item: AI +	1048	0	3.14	3.00	0.996	1	5
Single item: AI −	1048	0	2.76	3.00	1.120	1	5
SNS − AT	956	92	13.04	12.00	5.613	6.00	30.0
TSSM	956	92	170.59	120.00	169.944	0.00	920.0

ATAI: Attitudes for AI scale with acceptance (ATAI +) and fear (ATAI −) subscales; Single item framework: AI attitudes positive (AI +) and negative (AI −) with one item each; SNS-AT: Social Networking Sites-Addiction Test; TSSM: Time spent on social media (aggregate of personal and business time per day in minutes).

Table 2
Descriptives statistics for male and female social media users under investigation.

	Gender	N	Missing	Mean	Median	SD	Minimum	Maximum
Accepting AI (ATAI +)	Male	469	0	6.52	6	1.822	2	10
	Female	487	0	6.03	6	1.612	2	10
Fear of AI (ATAI −)	Male	469	0	7.73	8	2.617	3	15
	Female	487	0	8.55	9	2.701	3	15
Single item (AI +)	Male	469	0	3.35	3	0.992	1	5
	Female	487	0	3.04	3	0.959	1	5
Single item (AI −)	Male	469	0	2.62	3	1.140	1	5
	Female	487	0	2.82	3	1.086	1	5
SNS-AT	Male	469	0	13.59	13.0	6.067	6.00	30.0
	Female	487	0	12.50	12.0	5.088	6.00	30.0
TSSM	Male	469	0	165.73	90.0	174.202	2.00	920.0
	Female	487	0	175.26	120.0	165.786	0.00	900.0
Age	Male	469	0	45.49	44.0	13.519	19.00	83.0
	Female	487	0	43.00	42.0	14.886	18.00	73.0

ATAI: Attitudes for AI scale with acceptance (ATAI +) and fear (ATAI −) subscales; Single item framework: AI attitudes positive (AI +) and negative (AI −) with one item each; SNS-AT: Social Networking Sites-Addiction Test; TSSM: Time spent on social media (aggregate of personal and business time per day in minutes).

Table 3
Spearman correlations for male social media users (controlling for age).

		ATAI+	ATAI-	Single item: AI +	Single item: AI −	TSSM	SNS-AT
Acceptance of AI (ATAI +)	Spearman's rho	—					
	p-value	—					
Fearing AI (ATAI −)	Spearman's rho	−0.502	—				
	p-value	<.001	—				
Single item: AI +	Spearman's rho	0.704	−0.480	—			
	p-value	<.001	<.001	—			
Single item: AI −	Spearman's rho	−0.499	0.614	−0.574	—		
	p-value	<.001	<.001	<.001	—		
TSSM	Spearman's rho	0.135	−0.005	0.136	−0.045	—	
	p-value	0.003	0.913	0.003	0.330	—	
SNS-AT	Spearman's rho	0.266	0.107	0.270	−0.009	0.351	—
	p-value	<.001	0.020	<.001	0.848	<.001	—

Note. controlling for 'age'.
ATAI: Attitudes for AI scale with acceptance (ATAI +) and fear (ATAI −) subscales; Single item framework: AI attitudes positive (AI +) and negative (AI −) with one item each; SNS-AT: Social Networking Sites-Addiction Test; TSSM: Time spent on social media (aggregate of personal and business time per day in minutes).

Table 4
Spearman correlations for female social media users (controlling for age).

		ATAI +	ATAI −	Single item: AI+	Single item: AI −	TSSM	SNS-AT
Acceptance of AI (ATAI +)	Spearman's rho	—					
	p-value	—					
Fear of AI (ATAI −)	Spearman's rho	−0.531	—				
	p-value	<.001	—				
Single item: AI +	Spearman's rho	0.725	−0.619	—			
	p-value	<.001	<.001	—			
Single item: AI −	Spearman's rho	−0.641	0.691	−0.786	—		
	p-value	<.001	<.001	<.001	—		
TSSM	Spearman's rho	0.089	−0.071	0.104	−0.128	—	
	p-value	0.049	0.116	0.022	0.005	—	
SNS-AT	Spearman's rho	0.118	0.023	0.105	−0.024	0.294	—
	p-value	0.009	0.610	0.021	0.590	<.001	—

Note. controlling for 'age'.
ATAI: Attitudes for AI scale with acceptance (ATAI +) and fear (ATAI −) subscales; Single item framework: AI attitudes positive (AI +) and negative (AI −) with one item each; SNS-AT: Social Networking Sites-Addiction Test; TSSM: Time spent on social media (aggregate of personal and business time per day in minutes).

Table 5
Standardized Mediation Estimates (adding age and gender covariates).

				95 % Confidence Interval			
Effect	Label	Estimate	SE	Lower	Upper	Z	p
Indirect	a × b	0.031	0.010	0.015	0.048	3.144	0.002

Note: Table 6 outlines the lettered labels in terms of how the products of path coefficients were computed.

more on AI-technology than others. As briefly mentioned, awareness of the degree of AI being used in these different online environments might have an influence on the association strength between positive AI attitudes and a certain problematic online behavior, for instance that people are more willing to use a videogame, which is AI empowered, when they really have a positive view on AI. This said, we might also see bidirectional effects here: more positive attitudes towards AI might result in more (excessive) AI technology use and the positive experience made with the AI technology might further shape the attitudes towards AI. AI

Table 6
Standardized Path Estimates (adding age and gender covariates).

			95 % Confidence Interval						
			Label	Estimate	SE	Lower	Upper	Z	p
ATAI +	→	TSSM	a	0.121	0.035	0.063	0.178	3.443	0.001
TSSM	→	SNS-AT	b	0.261	0.029	0.212	0.309	8.82	0<.001
ATAI +	→	SNS-AT	c	0.205	0.031	0.153	0.256	6.511	0<.001
Age	→	SNS-AT	d	−0.336	0.027	−0.381	−0.291	−12.264	0<.001
Gender	→	SNS-AT	e	−0.215	0.055	−0.306	−0.123	−3.868	0<.001

ATAI +: Acceptance of AI; TSSM: Time spent on social media (aggregate of personal and business time per day in minutes).

being built into products such as videogames (Tang et al., 2020) could lead to more immersion and longer play time, hence shaping user behavior of a platform. How AI in its many forms will be able to shape human behavior toward more compulsive or impulsive use of technology (as an example) is up to further discussion. With the present data, we unfortunately cannot further shed light on several of the here mentioned issues, but want to point to such an interesting future research endeavor.

We further mention that the association between positive AI attitudes and PSMU severity was stronger in males than females. The separate analysis for males and females is important, because we detected gender effects, with males having more positive AI attitudes than females (see supplementary material). If positive AI attitudes will be carved out in the future as a risk factor for PSMU, it could be the case that males – being in particular open to new technologies – might more easily develop PSMU tendencies. In any case, it will be important to further take into account gender in the near future in this line of research.

Although the present work cannot prove that positive AI attitudes are causally linked to PSMU, the question arises, if use of artificial intelligence in digital products might alter products in such a way that they are getting more immersive (Montag, Yang, et al., 2025) and that certain user groups (here characterized by more positive AI attitudes) might in particular be drawn into longer use or show more addictive-like use. The latter also makes an interesting research question, namely if AI being built into a product such as social media simply results in prolonged use or pathological use of a product. Insights from such research for sure will also raise ethical concerns. Past research suggests that certain personality tendencies indeed might be more responsive to platform design (whereas AI can be also seen as such a feature; Sindermann, Montag, & Elhai, 2022). In this context, a controversial question arises when thinking of AI being used in a very critical area of social media, namely personalization of content or use of recommendation engines (Guha, 2021; Salma et al., 2024). AI can be seen as very helpful, because it selects for the user from millions of content pieces with a high chance of choosing those which might be most interesting (derived from the study of digital footprints left behind). This reduces cognitive burden on the user's side, but might also result in filter bubbles (Pariser, 2011). For the users this kind of personalization might also dramatically increase engagement with the platform and online time – fostering PSMU. Therefore, there is a fine line between using AI for reducing cognitive burden and providing a meaningful service and on the other hand escalating online time, which is at the heart of the data business model (Montag & Elhai, 2023).

Although we did not study ChatGPT addiction in the present work, we want to reflect on this topic in the context of our studied AI attitudes (which obviously is a very different construct). Recent discussion around the concept of ChatGPT addiction arose (Lin & Chien, 2024; Yankouskaya, Liebherr, & Ali, 2025), when a number of researchers started to investigate addictive use of ChatGPT. It is important to note, that so far the literature does not make a compelling case for use of the addiction term in this context and the danger of over-pathologizing behavior is a valid concern (Ciudad-Fernández et al., 2025). We believe that at the moment over-reliance on AI and the forming of parasocial relationships with LLMs such as character.ai, where the user bonds with a fictional (or celebrity) character, might be more relevant issues to be discussed around the intense use of LLMs (Montag et al., 2025). But an understanding of how addictive behaviors towards LLMs (again we are hesitant to use the term ChatGPT Addiction, etc.) might link to other established addictive behaviors can be interesting (and here also the link to AI attitudes could be studied again). We explicitly mention that this is not the research question of the present paper, because here we investigated the non-clinical construct of positive attitudes toward AI, whereas positive views on technology in the literature are well-established to help people embrace technology (with benevolent technology, this is something good). The controversy from the present data arises by also highlighting potential negative effects of such a positive view.

The present work comes with several limitations. The cross-sectional character prevents us from confidently inferring causality between the variables. Second, the work relies on self-report methodology including the usual problems such as lack of introspection, response bias, etc. Third, we used global measures of AI attitudes, whereas it might be more meaningful to investigate AI attitudes in a finer grained manner, such as attitudes toward AI within social media products (and here also asking study participants about their awareness of AI being built into social media). Recent research suggests that global AI attitudes have predictive power for use in specific areas though (Montag & Ali, 2025a; Sindermann et al., 2021). Fourth, the present work showed that age plays a role to understand associations between AI attitudes and PSMU and the present sample was relatively old (referring to the mean of the sample). As younger people might be more tech-interested or more open to use AI technology, this needs to be further investigated. Finally, future research should also consider investigating which social media AI elements people interact with, and to what extent.

In conclusion, the present work shows that AI attitudes might be relevant to understanding PSMU. However, the present research is exploratory and should only be seen as a first approach to understanding this area. Replication of the present findings will be needed.

CRediT authorship contribution statement

Christian Montag: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jon D. Elhai:** Writing – review & editing, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.abrep.2025.100613.

Data availability

Data will be made available on request.

References

Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process., 50* (2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Brand, M., Wegmann, E., Stark, R., Müller, A., Wölfling, K., Robbins, T. W., & Potenza, M. N. (2019). The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: Update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. *Neurosci. Biobehav. Rev., 104*, 1–10. https://doi.org/10.1016/j.neubiorev.2019.06.032

Brand, M., Young, K. S., Laier, C., Wölfling, K., & Potenza, M. N. (2016). Integrating psychological and neurobiological considerations regarding the development and maintenance of specific Internet-use disorders: An Interaction of Person-Affect-Cognition-Execution (I-PACE) model. *Neurosci. Biobehav. Rev., 71*, 252–266. https://doi.org/10.1016/j.neubiorev.2016.08.033

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact., 5*(CSCW1), 188:1-188:21. DOI: 10.1145/3449287.

Ciudad-Fernández, V., von Hammerstein, C., & Billieux, J. (2025). People are not becoming "AIholic": Questioning the "ChatGPT addiction" construct. *Addict. Behav., 166*, Article 108325. https://doi.org/10.1016/j.addbeh.2025.108325

Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Comput. Hum. Behav., 17*(2), 187–195. https://doi.org/10.1016/S0747-5632(00)00041-8

Flayelle, M., Brevers, D., King, D. L., Maurage, P., Perales, J. C., & Billieux, J. (2023). A taxonomy of technology design features that promote potentially addictive online behaviours. *Nat. Rev. Psychol., 2*, 136–150. https://doi.org/10.1038/s44159-023-00153-4

Guha, R. (2021). Improving the Performance of an Artificial Intelligence Recommendation Engine with Deep Learning Neural Nets. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1–7). https://doi.org/10.1109/I2CT51068.2021.9417936

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Comput. Hum. Behav., 160*, Article 108352. https://doi.org/10.1016/j.chb.2024.108352

Lin, C.-C., & Chien, Y.-L. (2024). ChatGPT Addiction: A Proposed Phenomenon of Dual Parasocial Interaction. *Taiwanese Journal of Psychiatry, 38*(3), 153. https://doi.org/10.4103/TPSY.TPSY_28_24

Matz, S. C., Kosinski, M., & Stillwell, D. J. (2017). *Psychological targeting as an effective approach to digital mass persuasion* |. *PNAS., 114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Montag, C. (2024). An attempt to assess individual differences in Shinrin-Yoku tendencies and associations with personality and life satisfaction: A study from Germany. *Trees, Forests and People, 15*, Article 100475. https://doi.org/10.1016/j.tfp.2023.100475

Montag, C., & Ali, R. (2025a). Can We Assess Attitudes Toward AI with Single Items? Associations with Existing Attitudes Toward AI Measures and Trust in ChatGPT. *Journal of Technology in Behavioral Science.*. https://doi.org/10.1007/s41347-025-00481-7

Montag, C., & Ali, R. (2025b). Starting the Journey to Understand Attitudes Towards Artificial Intelligence in Global Societies. In C. Montag & R. Ali (Eds.), *The Impact of Artificial Intelligence on Societies: Understanding Attitude Formation Towards AI* (pp. 1–7). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-70355-3_1.

Montag, C., Becker, B., & Li, B. J. (2024). On trust in humans and trust in artificial intelligence: A study with samples from Singapore and Germany extending recent research. *Computers in Human Behavior: Artificial Humans, 100070*. https://doi.org/10.1016/j.chbah.2024.100070

Montag, C., & Elhai, J. D. (2023). On Social Media Design, (Online-)Time Well-spent and Addictive Behaviors in the Age of Surveillance Capitalism. *Curr. Addict. Rep., 10*, 610–616. https://doi.org/10.1007/s40429-023-00494-3

Montag, C., & Hegelich, S. (2020). Understanding Detrimental Aspects of Social Media Use: Will the Real Culprits Please Stand Up? *Front. Sociol., 5*, 599270. https://doi.org/10.3389/fsoc.2020.599270

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories. *Int. J. Environ. Res. Public Health, 16*(14), 2612. https://doi.org/10.3390/ijerph16142612

Montag, C., Müller, M., Pontes, H. M., & Elhai, J. D. (2023). On fear of missing out, social networks use disorder tendencies and meaning in life. *BMC Psychology, 11*(1), 358. https://doi.org/10.1186/s40359-023-01342-9

Montag, C., Schulz, P. J., Zhang, H., & Li, B. J. (2025). On pessimism aversion in the context of artificial intelligence and locus of control: Insights from an international sample. *AI & Soc.*. https://doi.org/10.1007/s00146-025-02186-0

Montag, C., Thrul, J., & van Rooij, A. J. (2022). Social media companies or their users—Which party needs to change to reduce online time? *Addiction, 117*(8), 2363–2364. https://doi.org/10.1111/add.15946

Montag, C., Yang, H., Wu, A. M. S., Ali, R., & Elhai, J. D. (2025). *The role of artificial intelligence in general, and large language models specifically, in better understanding online addictive behaviors*. Annals of the New York Academy of Sciences. https://doi.org/10.1111/nyas.15337

Naiseh, M., Babiker, A., Al-Shakhsi, S., Cemiloglu, D., Al-Thani, D., Montag, C., & Ali, R. (2025). Attitudes Towards AI: The Interplay of Self-Efficacy, Well-Being, and Competency. *Journal of Technology in Behavioral Science*. https://doi.org/10.1007/s41347-025-00486-2

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin UK.

Salma, B., Fatima, T., Sara, A., & Merieme, B. (2024). Artificial Intelligence in Social Media: From Content Personalization to User Engagement. In Y. Farhaoui (Ed.), *Artificial Intelligence, Big Data, IOT and Block Chain in Healthcare: From Concepts to Applications* (pp. 45–52). Springer Nature Switzerland. DOI: 10.1007/978-3-031-65018-5_5.

Schepman, A., & Rodway, P. (2025). The Measurement of Attitudes Towards Artificial Intelligence: An Overview and Recommendations. In C. Montag & R. Ali (Eds.), *The Impact of Artificial Intelligence on Societies: Understanding Attitude Formation Towards AI* (pp. 9–24). Springer Nature Switzerland. DOI: 10.1007/978-3-031-70355-3_2.

Sindermann, C., Montag, C., & Elhai, J. D. (2022). The Design of Social Media Platforms—Initial Evidence on Relations Between Personality, Fear of Missing Out, Design Element-Driven Increased Social Media Use, and Problematic Social Media Use. *Technology, Mind, and Behavior*, *3*(4: Winter). https://doi.org/10.1037/tmb0000096.

Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2021). Assessing the Attitude Towards Artificial Intelligence: Introduction of a Short Measure in German, Chinese, and English Language. *KI - Künstliche Intelligenz, 35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0

Statista. (2022). *Number of social media users 2025*. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Stein, J.-P., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Sci. Rep., 14*, 2909. https://doi.org/10.1038/s41598-024-53335-2

Tang, C., Wang, Z., Sima, X., & Zhang, L. (2020). Research on Artificial Intelligence Algorithm and Its Application in Games. *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*, 386–389. DOI: 10.1109/AIAM50918.2020.00085.

Venkatesh, V., & Davis, F. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manag. Sci., 46*, 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Yankouskaya, A., Liebherr, M., & Ali, R. (2025). Can ChatGPT Be Addictive? A Call to Examine the Shift from Support to Dependence in AI Conversational Large Language Models. *Human-Centric Intelligent Systems, 5*, 77–89. https://doi.org/10.1007/s44230-025-00090-w

Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*, *49*(8), 1066-1091. https://doi.org/10.1177/0093650220961965 (Original work published 2022).

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
8. The impact of artificial intelligence: from cognitive costs to global inequality (2025)	European Physical Journal: Special Topics (Article from : Springer Science and Business Media Deutschland GmbH)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

Eur. Phys. J. Spec. Top. (2025) 234:3045–3050
https://doi.org/10.1140/epjs/s11734-025-01561-8

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

The impact of artificial intelligence: from cognitive costs to global inequality

Guy Paić[1,a] and Leonid Serkin[2,b]

[1] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70-543, 04510 Ciudad de México, Mexico
[2] Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, 04510 Ciudad de México, Mexico

Received 5 November 2024 / Accepted 3 March 2025 / Published online 17 March 2025

Abstract In this paper, we examine the wide-ranging impact of artificial intelligence on society, focusing on its potential to both help and harm global equity, cognitive abilities, and economic stability. We argue that while artificial intelligence offers significant opportunities for progress in areas like healthcare, education, and scientific research, its rapid growth—mainly driven by private companies—may worsen global inequalities, increase dependence on automated systems for cognitive tasks, and disrupt established economic paradigms. We emphasize the critical need for strong governance and ethical guidelines to tackle these issues, urging the academic community to actively participate in creating policies that ensure the benefits of artificial intelligence are shared fairly and its risks are managed effectively.

1 Introduction

Artificial intelligence[1] (AI) is transforming the way we live. The impact of AI represents a rapid and transformative shift in society, comparable only to some of the most remarkable milestones in human history, such as the discovery of fire, the Industrial Revolution, or the invention of the automobile. Today, society is facing the rapid rise of AI, which—like a massive tsunami—is permeating every aspect of life. From sophisticated reinforcement learning algorithms that master chess and other games [2] to AI-driven coding assistants [3] and large language models (LLMs) like ChatGPT [4] or DeepSeek [5], these innovations not only empower individuals to learn and innovate but also help build a more inclusive, interconnected global community.

However, every technological leap brings not only opportunities but also risks and unforeseen consequences. Human progress often comes with unintended side effects, such as climate change [6] and plastic pollution [7], and a significant portion of modern spending goes toward mitigating these risks inherent in our technologically interconnected society [8]. Our growing dependence on AI-driven systems—whether in energy distribution, transportation, or healthcare—can amplify the effects of any failure [9]. In this context, the failure of any AI-driven decision-making process can trigger cascading disruptions similar to those observed in traditional infrastructural breakdowns [10, 11]. These and many other AI-driven developments illustrate that the transformative power of AI encompasses both remarkable opportunities and considerable challenges, making it essential to approach its governance and integration thoughtfully [12, 13].

This paper examines the duality of AI's impact—its potential benefits versus its risks. We aim to focus on the less-discussed aspects—specifically, the short- and long-term effects AI could have on humanity. We first examine the unintended consequences of technology, drawing lessons from historical advancements. Next, we discuss the

[a] e-mail: guy.paic@cern.ch
[b] e-mail: lserkin@ciencias.unam.mx (corresponding author)

[1] For the purposes of this paper, we will adopt the latest definition of AI from the Organization for Economic Cooperation and Development (OECD), which states [1]: "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment".

need for governance and regulation, followed by an in-depth analysis of AI's dangers, particularly in job markets and economic inequality. We then explore AI's impact on healthcare, addressing both its benefits and ethical risks, before discussing the cognitive costs of AI, including dependency and skill erosion. Finally, we conclude by emphasizing the role of the academic community in shaping AI's future.

2 Collateral damage and effective governance

Nowhere is the balance between technological progress and unintended harm more evident than in the widespread use of automobiles. While cars have revolutionized transportation and global connectivity, they also come with significant risks—more than a million people lose their lives in traffic accidents each year [14]. However, this risk is not uniform across different regions of the world. As seen in Table 1, geographic areas with stronger regulatory frameworks, such as vehicle safety standards, speed limits, and well-maintained infrastructure, report significantly lower fatality rates compared to those with weaker enforcement mechanisms [15].

Beyond immediate fatalities caused by car accidents, the long-term consequences of increased car usage extend to rising obesity rates and elevated risks of cardiovascular disease and mortality [16, 17]. This so-called collateral damage, however, is not an unavoidable consequence of technological advancement; rather, it is a governance challenge requiring effective policy intervention to mitigate harm [18].

A similar governance approach is necessary for AI—without clear oversight, AI could introduce new vulnerabilities, such as algorithmic bias, security threats, and unintended social consequences. As seen in the case of road safety, regions that implement strong regulations experience fewer negative consequences, reinforcing the need for proactive AI governance to ensure that technological progress benefits society while minimizing risks [19].

This idea goes hand in hand with the growing movement within civil society to highlight the various risks associated with AI and to urge policymakers to address these concerns. The goal is to ensure that these risks are mitigated with the guidance of both the scientific community and the public [20].

Historically, many groundbreaking innovations were the result of collaborations among public institutions, universities, and state-sponsored research initiatives [21]. However, in recent decades, especially in the field of digital technologies and AI, private companies have emerged as the primary drivers of transformative innovations, often operating with minimal direct societal oversight [22, 23].

This shift contrasts with the approach taken by the European Organization for Nuclear Research (CERN) regarding the World Wide Web. In 1993, rather than patent or privatize the web's source code, CERN released it freely to the public [24]. This act ensured that the web would remain an open platform for global innovation and collaboration, free from proprietary restrictions. By adopting this open approach, CERN enabled the explosive growth of the internet, creating countless opportunities for businesses, education, and communication across the globe, a legacy that contrasts with today's more closed, profit-driven models of technological development.

Given AI's profound societal impact, adopting a similar multinational, nonprofit-driven approach to its development could help ensure its benefits are equitably shared. Promoting global collaboration within an open framework—rather than leaving AI's trajectory solely in the hands of private interests—could lead to more ethical, transparent, and broadly beneficial technological advancements.

Table 1 Estimated road traffic deaths data in 2019 by region, data taken from Ref. [14]

Region	Estimated road traffic death rate (per 100,000 population)	Estimated number of road traffic deaths
Global	16.7	1,282,150
Africa	27.2	297,087
Eastern Mediterranean	17.8	126,958
Western Pacific	16.4	317,393
Southeast Asia	15.8	317,069
Americas	15.3	154,780
Europe	7.4	68,863

3 AI's dual impact

There is no doubt that AI offers significant benefits. As the OECD states, "AI holds the potential to address complex challenges, from enhancing education and improving healthcare to driving scientific innovation and climate action" [25]. However, the risks associated with AI should not be underestimated.

The existential risks posed by AI, particularly the loss of millions of jobs, have been highlighted by various experts—around 40% of all working hours could be impacted by AI LLMs such as ChatGPT-4 [26]. If not properly controlled, AI could widen existing inequalities and reshape entire industries, potentially leaving many workers without meaningful employment [27, 28]. This situation brings to mind an anecdote of a United Nations expert observing a peasant ploughing his field with a donkey: "We will give you a tractor to plough your piece of land in two hours instead of you ploughing your field in a whole day." The peasant's reply was quick: "Well, what would my donkey and I do for the rest of the day?"

The potential job losses due to AI, even with compensation, have not been fully addressed by governments, and the psychological impact could be significant. As societies become increasingly dependent on AI-driven systems and digital communication, we are already witnessing broader social changes, particularly in rural communities [29]. Economic shifts and urban migration have contributed to the decline of traditional social spaces, such as village cafés, which once served as key hubs of local interaction. While modern communication tools provide new ways to stay connected, they do not fully replace in-person social interactions, which remain essential for community cohesion and mental well-being [30].

Beyond its impact on labor markets and social structures, AI is also reshaping critical sectors such as healthcare. There is clear evidence that AI holds immense potential to revolutionize medical diagnostics, enhance treatment strategies, and support healthcare professionals in delivering more precise and efficient care. However, despite these advancements, concerns persist regarding the ethical implications and unintended consequences of AI-driven healthcare tools. The World Health Organization (WHO) urges caution in the use of AI tools, particularly LLMs, to ensure they promote human well-being, safety, and autonomy while safeguarding public health [31].

One of the most concerning dangers in using AI-driven innovations is its potential to worsen racial, gender, and geographic disparities in healthcare. This is because bias is often embedded in the data used to train AI models, which can lead to unequal treatment and outcomes for different groups of people [32]. This presents an additional challenge for less-developed countries, which must ensure the collection, privacy, and secure storage of large, representative datasets [33].

Currently, WHO supports the responsible use of AI to benefit healthcare professionals, patients, and researchers. However, they emphasize the need for ethical guidelines and appropriate governance, as outlined in the WHO's guidance on AI ethics in healthcare [34]. This also reinforces our perspective: the same level of careful scrutiny applied to other new technologies must also be consistently applied to LLMs.

4 Cognitive cost

The application of AI and its derivatives requires both human and industrial resources. Just as with traffic-related hidden damages mentioned earlier, there are several important aspects of AI development that we would like to address.

First, there is the risk of widening the "digital divide" between the most developed nations and those that are moderately developed or entirely disadvantaged. AI's immense power and water demands are already proving to be a challenge, even for advanced economies [35, 36]. According to studies from the International Monetary Fund (IMF), the computational power required to sustain AI development is growing rapidly, with the potential to consume as much energy as entire countries in the near future [37].

This raises critical questions about whether smaller countries, lacking the necessary infrastructure to support such large power demands, will ever be able to participate in AI development at that scale [38]. The real concern is whether AI will become a powerful tool controlled by a small number of countries, much like nuclear weapons or rocket technology [39]. The IMF's conclusion is clear: "Emerging market and developing economies should prioritize developing digital infrastructure and digital skills" [37].

To prevent technological competition from leading to unnecessary environmental sacrifices, there is an urgent need for collaborative governance that establishes binding international standards. Cooperation between governments and technology companies can enable the sustainable development of AI, ensuring that climate goals are protected without suppressing innovation [40].

On the positive side, empirical and historical analyses indicate that, although many technological breakthroughs originate in advanced economies, such innovations have often acted as catalysts for accelerated socio-economic convergence in developing regions—a phenomenon extensively documented in studies on technological diffusion and socio-economic progress [41].

Second, we must consider the long-term effects of AI on human cognitive abilities. While AI is intended to relieve humanity of many mental tasks, it is unclear whether this will be a benefit in the long run. Younger generations are already shifting their reliance on cognitive skills. Tasks like memorizing phone numbers, solving maths problems, or even learning new languages are becoming obsolete with the rise of mobile phones and AI-powered translators [42, 43].

Finally, what happens if, for some reason, access to AI systems is lost? At present, people still possess the skills to revert to pre-AI methods, much like pilots who are instructed to override AI systems if they do not understand its actions [44]. However, as we mentioned earlier, excessive reliance on AI could gradually erode human cognitive skills [45, 46]. To ensure resilience, it is crucial to preserve our ability to think critically, adapt, and function independently of AI—both in everyday life and in times of crisis.

5 Conclusion: our role as the academic community

The academic community plays a crucial role in shaping the development and responsible oversight of AI, guiding its future use in ways that benefit society while addressing its risks. Universities and research institutions are at the forefront of AI development, carrying the unique responsibility of applying AI in a controlled and informed manner.

To achieve this, academia should lead the way in exploring potential risks posed by AI, such as job displacement, privacy concerns, health, and ethical challenges. One of academia's key roles is sharing knowledge, particularly about the dangers and potential misuse of AI. With their technical expertise, academics are well-positioned to provide guidance on the legislative and political actions needed to regulate AI effectively [47].

Finally, since AI is a global phenomenon, academic institutions worldwide should collaborate to establish international standards for its governance, helping to ensure AI does not deepen inequalities or contribute to geopolitical tensions. By implementing these strategic measures, the academic community can actively guide AI development to ensure it remains both innovative and ethically responsible.

Acknowledgements The research was supported by the Mexican National Council of Humanities, Sciences and Technologies CONAHCYT under Grants No. CF-2042 and No. A1-S-22917. The authors are grateful to Gergely Gábor Barnaföldi for his valuable comments, and sincerely appreciate the valuable feedback from the journal reviewers.

Data Availability The dataset on the estimated road traffic deaths data in 2019 is available at: https://apps.who.int/gho/data/node.main.A997.

References

1. OECD. (2024). OECD AI principles overview. https://oecd.ai/en/ai-principles
2. D. Silver et al., A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science **362**, 1140–1144 (2018). https://doi.org/10.1126/science.aar6404
3. M.M. Amin, E. Cambria, B.W. Schuller, Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of ChatGPT. IEEE Intell. Syst. **38**(2), 15–23 (2023). https://doi.org/10.1109/MIS.2023.3254179
4. P.P. Ray, ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Phys. Syst. **3**, 121–154 (2023). https://doi.org/10.1016/j.iotcps.2023.04.003
5. DeepSeek-AI. (2024). DeepSeek LLM: Scaling Open-Source Language Models with Longtermism. arXiv preprint arXiv:2401.02954. https://github.com/deepseek-ai/DeepSeek-LLM
6. J. Hansen, M. Sato, R. Ruedy, A. Lacis, V. Oinas, Global warming in the twenty-first century: an alternative scenario. Proc. Natl. Acad. Sci. U.S.A. **97**(18), 9875–9880 (2000). https://doi.org/10.1073/pnas.170278997
7. J.R. Jambeck et al., Plastic waste inputs from land into the ocean. Science **347**, 768–771 (2015). https://doi.org/10.1126/science.1260352

8. J. Wolff, How is technology changing the world, and how should the world change technology? Glob. Perspect. **2**(1), 27353 (2021). https://doi.org/10.1525/gp.2021.27353
9. S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks. Nature **464**, 1025–1028 (2010). https://doi.org/10.1038/nature08932
10. J. Pérez-Cerrolaza et al., Artificial intelligence for safety-critical systems in industrial and transportation domains: a survey. ACM Comput. Surv. **56**, 1–40 (2023). https://doi.org/10.1145/3626314
11. J. Danielsson, R. Macrae, A. Uthemann, Artificial intelligence and systemic risk. J. Bank. Finance **140**, 106290 (2022). https://doi.org/10.1016/j.jbankfin.2021.106290
12. A. Taeihagh, Governance of artificial intelligence. Policy Soc. **40**(2), 137–157 (2021). https://doi.org/10.1080/14494035.2021.1928377
13. E. Zaidan, I.A. Ibrahim, AI governance in a complex and rapidly changing regulatory landscape: a global perspective. Humanit. Soc. Sci. Commun. **11**, 1121 (2024). https://doi.org/10.1057/s41599-024-03560-x
14. World Health Organization. (2021). Road traffic deaths data by country. https://apps.who.int/gho/data/node.main.A997
15. M. Tavakkoli et al., Evidence from the decade of action for road safety: a systematic review of the effectiveness of interventions in low and middle-income countries. Public Health Rev. **43**, 1604499 (2022). https://doi.org/10.3389/phrs.2022.1604499
16. T.Y. Warren et al., Sedentary behaviours increase risk of cardiovascular disease mortality in men. Med. Sci. Sports Exerc. **42**(5), 879–885 (2010). https://doi.org/10.1249/MSS.0b013e3181c3aa7e
17. T. Sugiyama et al., Car use and cardiovascular disease risk: systematic review and implications for transport research. J. Transp. Health **19**, 100930 (2020). https://doi.org/10.1016/j.jth.2020.100930
18. M. Peden, Global collaboration on road traffic injury prevention. Int. J. Inj. Control Saf. Promot. **12**(2), 85–91 (2005). https://doi.org/10.1080/15660970500086130
19. P. Mikalef, K. Conboy, J.E. Lundström, A. Popovič, Thinking responsibly about responsible AI and 'the dark side' of AI. Eur. J. Inf. Syst. **31**(3), 257–268 (2022). https://doi.org/10.1080/0960085X.2022.2026621
20. Z. Alalawi et al., Trust AI regulation? Discerning users are vital to build trust and effective AI regulation. Artif. Intell. (2024). https://doi.org/10.48550/arXiv.2403.09510
21. H. Etzkowitz, L. Leydesdorff, The dynamics of innovation: from national systems and mode 2 to a triple helix of university-industry-government relations. Res. Policy **29**(2), 109–123 (2000). https://doi.org/10.1016/S0048-7333(99)00055-4
22. T. Hagendorff, K. Meding, Ethical considerations and statistical analysis of industry involvement in machine learning research. AI Soc. **38**, 35–45 (2023). https://doi.org/10.1007/s00146-021-01284-z
23. B.-A. Lundvall, C. Rikap, China's catching-up in artificial intelligence seen as a co-evolution of corporate and national innovation systems. Res. Policy **51**, 104395 (2022). https://doi.org/10.1016/j.respol.2021.104395
24. CERN. (2024). The birth of the web. https://home.cern/science/computing/birth-web
25. OECD. (2024). Artificial intelligence—Policy issues. https://www.oecd.org/en/topics/policy-issues/artificial-intelligence.html
26. T. Eloundou, S. Manning, P. Mishkin, D. Rock, GPTs are GPTs: labor market impact potential of LLMs. Science **384**, 1306–1308 (2024). https://doi.org/10.1126/science.adj0998
27. C.B. Frey, M.A. Osborne, The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Change **114**, 254–280 (2017). https://doi.org/10.1016/j.techfore.2016.08.019
28. Y. Shen, X. Zhang, The impact of artificial intelligence on employment: the role of virtual agglomeration. Humanit. Soc. Sci. Commun. **11**, 122 (2024). https://doi.org/10.1057/s41599-024-02647-9
29. T. Correa, I. Pavez, Digital inclusion in rural areas: a qualitative exploration of challenges faced by people from isolated communities. J. Comput. Mediat. Commun. **21**(3), 247–263 (2016). https://doi.org/10.1111/jcc4.12154
30. C.K. Ettman, S. Galea, The potential influence of AI on population mental health. JMIR Ment. Health **10**, e49936 (2023). https://doi.org/10.2196/49936
31. World Health Organization. WHO calls for safe and ethical AI for health. https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health (2023)
32. A. Turchin, D. Denkenberger, Classification of global catastrophic risks connected with artificial intelligence. AI Soc. **35**, 147–163 (2020). https://doi.org/10.1007/s00146-018-0845-5
33. CENIA. Centro Nacional de Inteligencia Artificial de Chile. Índice Latinoamericano de Inteligencia Artificial. https://indicelatam.cl/ (2023)
34. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf (2021)
35. A.S. Luccioni, S. Viguier, A.-L. Ligozat, Estimating the carbon footprint of BLOOM, a 176B parameter language model. J. Mach. Learn. Res. **24**(253), 1–15 (2023). http://jmlr.org/papers/v24/23-0069.html
36. U.S. Congress. Artificial Intelligence Research, Innovation, and Accountability Act of 2024, S.3732, 118th Congress. https://www.congress.gov/bill/118th-congress/senate-bill/3732 (2024)
37. International Monetary Fund. Gen AI: Artificial intelligence and the future of work. https://www.imf.org/-/media/Files/Publications/SDN/2024/English/SDNEA2024001.ashx (2024)
38. S. Armstrong, N. Bostrom, C. Shulman, Racing to the precipice: a model of artificial intelligence development. AI Soc. **31**, 201–206 (2016). https://doi.org/10.1007/s00146-015-0590-y

39. S. Schmid, D. Lambach, C. Diehl, C. Reuter, Arms race or innovation race? Geopolitical AI development. Geopolitics (2025). https://doi.org/10.1080/14650045.2025.2456019
40. Francisco, M., Linnér, B.-O. (2023). AI and the governance of sustainable development. An idea analysis of the European Union, the United Nations, and the World Economic Forum. Environ. Sci. Policy **150**, 103590. https://doi.org/10.1016/j.envsci.2023.103590
41. S. Pinker, Enlightenment Now: The Case for Reason, Science, Humanism, and Progress. Viking, New York (2018). https://doi.org/10.1017/S0022050718000852
42. U. León-Domínguez, Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions. Neuropsychology **38**(4), 293–308 (2024). https://doi.org/10.1037/neu0000948
43. M. Shanmugasundaram, A. Tamilarasu, The impact of digital technology, social media, and artificial intelligence on cognitive functions: a review. Front. Cognit. **2**, 1203077 (2023). https://doi.org/10.3389/fcogn.2023.1203077
44. EASA. European Union Aviation Safety Agency (EASA) Artificial Intelligence Roadmap 2.0: a human-centric approach to AI in aviation. https://www.easa.europa.eu/en/downloads/137919/en (2024)
45. S.F. Ahmad et al., Impact of artificial intelligence on human loss in decision making, laziness and safety in education. Humanit. Soc. Sci. Commun. **10**, 311 (2023). https://doi.org/10.1057/s41599-023-01787-8
46. I. Dergaa et al., From tools to threats: a reflection on the impact of artificial-intelligence chatbots on cognitive health. Front. Psychol. **15**, 1259845 (2024). https://doi.org/10.3389/fpsyg.2024.1259845
47. C.A. Coello Coello, G. Paic, L. Serkin, The role of UNAM in facing the nation's challenges of balancing the risks and benefits of artificial intelligence. Rev. Tecnol. Innov. Educ. Super. **10**, 72–85 (2024). https://doi.org/10.22201/dgtic.26832968e.2024.10.10

More info

UTM Library

UTM UNIVERSITI TEKNOLOGI MALAYSIA

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
9. The next wave of AI for social impact: challenges and opportunities (2025)	IEEE Intelligent Systems (Article from : Institute of Electrical and Electronics Engineers Inc.)

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

The Next Wave of AI for Social Impact: Challenges and Opportunities

Milind Tambe, *Harvard University, Cambridge, MA, 02134, USA*
Fei Fang, *Carnegie Mellon University, Pittsburgh, PA, 15213, USA*
Andrew Perrault, *The Ohio State University, Columbus, OH, 43210, USA*
Bryan Wilder, *Carnegie Mellon University, Pittsburgh, PA, 15213, USA*

The burgeoning field of artificial intelligence for social impact (AI4SI) represents a significant evolution in artificial intelligence, prioritizing measurable positive impact for vulnerable and under-resourced populations. This article examines the historical context and recent surge in AI4SI, driven by technological advancements and a growing awareness of societal challenges. It highlights the crucial role of interdisciplinary collaboration, ethical considerations, and the potential of emerging AI trends in addressing issues such as poverty, health, and environmental sustainability. Furthermore, the article delves into key research questions and challenges facing the field, including the need for contextually relevant AI design, overcoming data limitations, ensuring scalable and sustainable deployments in resource-constrained environments, and establishing robust evaluation frameworks. Realizing the full potential of AI to address pressing societal needs in the coming decade and beyond will hinge on effectively navigating these challenges and fostering a deeply impact-driven approach to research and development.

Artificial intelligence for social impact (AI4SI) is a subdiscipline of AI research where measurable societal impact, particularly for vulnerable and under-resourced groups, is a primary objective, focusing on areas that have historically lacked sufficient AI research and development.[a] Unlike traditional AI research, which often prioritizes methodological advancements, AI4SI places direct social impact as a primary objective. It addresses problems that have historically lacked sufficient attention in AI research, aiming to bridge the gap between AI capabilities and real-world societal challenges in areas such as poverty, agriculture, public health, and environmental conservation. AI4SI is also tied to United Nations Sustainable Development Goals (SDGs)—the adoption of United Nations SDGs has shaped the AI for social impact agenda in the past decade.[1] The goal is to create impactful solutions tailored to real-world problems, often in resource-constrained environments.

Because of its focus on impact, AI4SI research necessitates deep engagement with domain experts and community members to identify relevant problems, design effective interventions, and rigorously evaluate them. Most of the AI4SI research is interdisciplinary, as it emphasizes the importance of understanding and addressing the specific needs of targeted communities and domains. For example, in a project that

[a]This subfield is sometimes referred to also as "AI for social good," but we will use the term AI for social impact, as it seems more popular in academic settings.

Digital Object Identifier 10.1109/MIS.2025.3565829
Date of current version 12 June 2025.

develops AI methods for a public-facing maternal health program, it is necessary that the AI team works with experts in maternal health and social work as well as stakeholders who might be directly or indirectly impacted by the AI technique.[9] These projects require a balance of technical innovation, ethical considerations, and practical feasibility.

It would be difficult to pinpoint a single founding document for AI4SI. Nonetheless, the "AI for Social Good" workshop organized by the White House Office of Science and Technology Policy in 2016 may be credited as the single event that sparked a significant interest in this topic.[2] It unified diverse efforts focused on social impact under one umbrella emerging field. This surge in interest is attributed to several key factors. First, the remarkable advancements in AI technologies, including deep learning, natural language processing, and reinforcement learning, have provided powerful tools applicable to a wide range of social issues. The availability of increased computing power and large datasets has further accelerated this progress, enabling the development of sophisticated AI models. Second, the establishment of government and industry-supported funding programs, dedicated workshops, conferences, and special tracks within major AI conferences have increased awareness and attracted researchers to AI4SI.[3,4] This increasing academic focus has led to a substantial rise in publications related to AI4SI, demonstrating the field's growing maturity.[5,10] There are also reports from international agencies, such as the United Nations, tracking the progress of AI4SI efforts.[6]

CURRENT STATE AND TRENDS

Existing efforts in AI4SI have covered a wide range of application domains,[17] such as public health,[18] agriculture,[25] environmental sustainability,[13] transportation,[26] food security,[21,22] crisis management and disaster response,[20,24] and poverty mitigation.[14,19] These efforts reflect the increasing recognition of AI's potential to address complex societal challenges, especially in low-resource settings. Notably, many AI4SI projects have moved beyond the research phase and achieved successful deployment, resulting in tangible improvements such as improved wildlife protection,[15] less food waste,[22] and better-targeted social welfare programs.[12]

As evidenced by these existing efforts, advancing interdisciplinary collaboration is becoming the norm in AI for social impact work. This necessity stems from the complex nature of societal challenges, which often require insights from diverse fields. Close partnerships with domain experts, local practitioners, and policymakers ensure that AI solutions are not only technically sound but also relevant and effective in the real world. This collaborative approach fosters a shared understanding of the problem and enables the development of solutions that are tailored to the specific needs of the communities they serve.

Emphasizing ethical AI is also critical, with fairness, transparency, and privacy as paramount considerations.[23] Major concerns, such as bias in data collection and the unintended consequences of AI deployment, must be addressed from the outset to ensure that AI systems align with societal values and avoid harm. AI4SI projects, by their very nature, work with vulnerable populations and sensitive data, making ethical considerations even more crucial. By proactively addressing potential ethical issues, researchers can build trust with communities and ensure that AI is used for good, rather than exacerbating existing issues.

Several trends are shaping the future of AI4SI in the short term. First, AI has become more broadly useful and accessible through the development of general-purpose models and the ability to serve them through scalable cloud computing. In the past, models developed for a specific task and run on a local computer predominated AI4SI projects. Because foundation models possess broad knowledge and capabilities without training,[7] they can be leveraged in AI4SI interventions to reduce the amount of task-specific data and engineering that is required. Cloud-based solutions enable the deployment of AI tools to remote or resource-constrained areas, democratizing access to AI technologies. In addition, the cloud has made maintenance and sustainability easier—the intervention can exist as a cloud service that can be more easily managed by end users. These changes have reduced the barrier to entry and increased the potential for long-term impact.

Second, the perception of AI by stakeholders has changed dramatically. AI was previously a niche topic that few, outside of the research community, possessed strong opinions about. The public increasingly believes that AI systems are capable, and it is thus easier to convince stakeholders that meaningful impact is attainable. However, many groups hold negative views about the potential consequences of the proliferation of AI systems, viewing it as a threat to jobs, a spreader of incorrect information, and unlikely to be regulated enough.[11] These views motivate work in AI4SI to go beyond effectiveness, to areas such as transparency, trustworthiness, and fairness, which are required for interventions to achieve long-term impact.

THE FUTURE OF AI FOR SOCIAL IMPACT

The long-term future of AI for social impact will be determined by the interplay between technological advancements and the human context in which technology is used. There are many possibilities for how AI and related technologies will advance over the next 15 years—especially given both rapid progress and persistent challenges faced by foundation models—and we discuss next how these uncertainties may impact the technical agenda for AI4SI. Regardless of the direction that the technology takes, developing a scientific understanding of how AI can be effectively used in a given social context will remain a key challenge for the field. Meeting this challenge will require interdisciplinary collaborations, the development of deep partnerships with communities, and investment in rigorous evaluation and empirical study, elements that we discuss in more detail next.

One key element of the AI4SI agenda will be to enable the design of AI systems that are not only technically effective but also deeply contextually relevant within social impact settings. This requires a nuanced understanding of the specific needs, cultural sensitivities, and practical constraints of the communities being served. Researchers must move beyond purely algorithmic considerations and engage in participatory design processes that prioritize the voices and experiences of end-users, ensuring that AI solutions are truly aligned with their real-world needs and challenges.

A second will be to ensure the sustainability and scalability of AI deployments in resource-constrained environments. Beyond the initial prototype phase, the long-term viability of AI solutions depends on their ability to be maintained and scaled by organizations with limited resources, such as nongovernmental organizations and government agencies. This necessitates the development of sustainable software architectures, the creation of user-friendly interfaces, and the provision of adequate training and support for local stakeholders. Moreover, funding these efforts, often because of the lack of commercial viability, is in itself a major concern. Currently, many AI4SI projects struggle to translate from a prototype to a system that is regularly used and maintained. An essential part of the field's future will be to develop methods and infrastructure to support users in taking ownership of AI systems.

Third, robust evaluation frameworks are crucial for assessing the impact of AI solutions in field settings and building stakeholder trust. There is significant room for AI researchers to expand their ability to design and implement field trials and other mechanisms for program evaluation. Currently, relatively few AI4SI projects have been the subject of rigorous evaluation and there is significant opportunity for the field as a whole to learn "what works" through the development of a broader base of empirical evidence. Evidence on the limits of AI, or social settings where prediction is ineffective or inappropriate (e.g., Salganik et al.[16]) is also critical. Beyond aggregate measures of impact, these evaluation frameworks must also incorporate the perspectives of users and impacted populations. The accessibility community's mantra, "Nothing about us without us,"[8] serves as a powerful reminder of the importance of involving stakeholders in all stages of the evaluation process.

Tackling all of these challenges will require the field to address the gap between traditional AI education and the specific skills required for impactful social work. Standard AI curricula primarily focus on algorithm design and analysis, often emphasizing theoretical concepts and performance on benchmark datasets. This approach, while essential for advancing core AI methodologies, leaves students ill-equipped to address the complex, real-world problems that AI4SI tackles. Effective AI4SI research demands a broader skillset, extending beyond purely technical expertise. It requires the ability to collaborate effectively with domain experts, such as public health officials, environmental scientists, or social workers, and to engage meaningfully with community members whose lives are directly affected by the technology. Understanding the nuanced socio-economic and cultural contexts of social challenges, and translating technical advancements into practical, user-centered interventions, are crucial competencies.

In addition to education for AI researchers, another important direction for advancing AI4SI is to empower a broader set of stakeholders—beyond AI researchers—to actively participate in identifying, designing, and deploying AI solutions by themselves. This requires the development of accessible, user-friendly AI tools that can support domain experts, practitioners, and community organizations in recognizing opportunities where AI can be beneficial and in prototyping simple AI-powered solutions without deep technical expertise. Closely intertwined with this is the need for widespread AI literacy education. As AI technologies become more pervasive, equipping the general public and professionals in fields such as public health, education, agriculture, and social work with the skills to understand, evaluate, and use AI tools is essential. Building capacity across sectors not only democratizes innovation but also helps ensure that

AI solutions are grounded in real-world expertise and values. Together, accessible tooling and inclusive education form a critical foundation for scaling the impact of AI in diverse social contexts.

While these skills will remain essential regardless of how AI develops more broadly, many of the key technical questions for the future of AI4SI will depend on the manner in which general AI capabilities advance. AI research has seen a period of intense consolidation, where an increasing portion of the field has shifted its focus to foundation models, which have seen a rapid expansion in their capabilities. However, there are still key limitations in the application of foundation models in AI4SI. First, the underrepresented groups that AI4SI aims to benefit are often underrepresented in foundation model training data. Second, optimizing interventions often requires specialized mathematical tools from areas such as optimization and statistics. Foundation models currently struggle to perform complex reasoning tasks, either directly, or by leveraging external tools that perform these functions.

We see the next 15 years as playing out differently depending on the extent to which foundation models are able to overcome these challenges. In the scenario where foundation models are most effective, future work in AI4SI will diminish its emphasis on the mathematical and engineering challenges that must currently be overcome to build such systems. Instead, the social-systems challenges discussed previously, related to context-sensitive design, evaluation, and capacity-building, will only grow in importance. In the most extreme version of this world, the use of AI tools proliferates every field to such an extent that work we now consider to be AI4SI occurs without reference to the term AI at all.

In the scenario where these gaps in foundation models' capabilities persist, the design of task-specific systems (that perhaps use foundation models internally) will continue to play a central role. When designing such systems, overcoming the limitations of data will be a key technical concern. AI for social impact projects frequently grapple with scarce, low-quality, or biased data, which can significantly impact the performance and fairness of AI models. Developing robust data collection strategies, employing techniques for data augmentation and bias mitigation, and exploring alternative data sources are essential for building reliable and equitable AI systems. This also requires a careful consideration of the cultural context in which data are gathered, ensuring that AI data collection methods are adapted to local practices, rather than imposing external standards.

CONCLUSION

AI4SI is a growing field leveraging AI to address societal challenges and aid vulnerable populations. Fueled by technological advancements and interdisciplinary collaboration, AI4SI has seen successful deployments in areas like health and environment, with increasing attention on ethical considerations and new AI trends. Looking into the future, there are many challenges to overcome: designing contextually relevant and sustainable systems, improving data quality, establishing robust evaluations, and fostering necessary interdisciplinary skills. Successfully navigating these hurdles is crucial to fully harness AI's potential for positive social change in the coming years. The future trajectory of AI4SI will ultimately hinge on how technological innovations are integrated within—and shaped by—the social, cultural, and institutional contexts in which they are deployed.

REFERENCES

1. "Transforming our world: The 2030 agenda for sustainable development," United Nations, New York, NY, USA, 2015. [Online]. Available: https://sdgs.un.org/2030agenda
2. "Artificial intelligence for social good," Computing Research Association (CRA), Washington, DC, USA, 2016. [Online]. Available: https://cra.org/ccc/events/ai-social-good/
3. "Call for the special track on AI for social impact," The Association for the Advancement of Artificial Intelligence, Washington, DC, USA, Feb. 2024. [Online]. Available: https://aaai.org/aaai-24-conference/call-for-the-special-track-on-ai-for-social-impact/
4. "Call for papers and projects: Multi-year track on AI and social good (special track)," IJCAI, California, USA, 2025. [Online]. Available: https://2025.ijcai.org/call-for-papers-and-projects-multi-year-track-on-ai-and-social-good-special-track/
5. Z. R. Shi, C. Wang, and F. Fang, "AI for social impact: A survey," 2020, *arXiv:2001.01818*.
6. "Digital agriculture: A standards snapshot," International Telecommunication Union, Geneva, Switzerland, 2025. [Online]. Available: https://aiforgood.itu.int/newsroom/publications-and-reports/
7. Y. Zhao, N. Boehmer, A. Taneja, and M. Tambe, "Towards foundation-model-based multiagent system to accelerate AI for social impact," 2024, *arXiv:2412.07880*.
8. J. I. Charlton, *Nothing About Us Without Us: Disability Oppression and Empowerment*. Berkeley, CA, USA: Univ. of California Press, 1998. [Online]. Available: http://www.jstor.org/stable/10.1525/j.ctt1pnqn9

9. A. Mate et al., "Field study in deploying restless multi-armed bandits: Assisting non-profits in improving maternal and child health," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 11, pp. 12,017–12,025, Jun. 2022, doi: 10.1609/aaai.v36i11.21460.
10. L. Floridi, J. Cowls, T. C. King, and M. Taddeo, "How to design AI for social good: Seven essential factors," in *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Cham, Switzerland: Springer-Verlag, 2021, pp. 125–151.
11. C. McClain, B. Kennedy, J. Gottfried, M. Anderson, and G. Pasquini, "How the U.S. public and AI experts view artificial intelligence," Pew Research Center, Washington, DC, USA, Apr. 2025. [Online]. Available: https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/
12. M. Tambe and E. Rice, Eds. *Artificial Intelligence and Social Work*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
13. C. Gomes et al., "Computational sustainability: Computing for a better world and a sustainable future," *Commun. ACM*, vol. 62, no. 9, pp. 56–65, 2019, doi: 10.1145/3339399.
14. N. Jean, M. Burke, M. Xie, W. M. Alampay Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016, doi: 10.1126/science.aaf7894.
15. F. Fang, M. Tambe, B. Dilkina, and A. J. Plumptre, Eds. *Artificial Intelligence and Conservation*. Cambridge, U.K.: Cambridge Univ. Press, 2019.
16. M. J. Salganik et al., "Measuring the predictability of life outcomes with a scientific mass collaboration," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 117, no. 15, pp. 8398–8403, 2020, doi: 10.1073/pnas.1915006117.
17. M. Tambe, F. Fang, and B. Wilder, Eds. *AI for Social Impact*, 2022. [Online]. Available: https://ai4sibook.org/
18. A. Yadav, H. Chan, A. Jiang, H. Xu, E. Rice, and M. Tambe, "Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, May 2016, vol. 16, pp. 740–748.
19. E. Aiken, S. Bellue, D. Karlan, and J. E. Blumenstock, "Machine learning and phone data can improve targeting of humanitarian aid," *Nature*, vol. 603, no. 7903, pp. 864–870, 2022, doi: 10.1038/s41586-022-04484-9.
20. S. Nevo et al., "Flood forecasting with machine learning models in an operational framework," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 15, pp. 4013–4032, 2022, doi: 10.5194/hess-26-4013-2022.
21. K. Peters et al., "UN world food programme: Toward zero hunger with analytics," *Informs J. Appl. Analytics*, vol. 52, no. 1, pp. 8–26, 2022, doi: 10.1287/inte.2021.1097.
22. Z. R. Shi, L. Lizarondo, and F. Fang, "A recommender system for crowdsourcing food rescue platforms," in *Proc. Web Conf.*, Apr. 2021, pp. 857–865.
23. S. Vollmer et al., "Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness," *BMJ*, vol. 368, Mar. 2020, Art. no. l6927, doi: 10.1136/bmj.l6927.
24. M. Madaio et al. "Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 185–194.
25. J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian process for crop yield prediction based on remote sensing data," in *Proc. 31st AAAI Conf. Artif. Intell. (AAAI-17)*, 2017, vol. 31, pp. 4559–4565, doi: 10.1609/aaai.v31i1.11172.
26. Y. Li, Y. Zheng, and Q. Yang, "Dynamic bike reposition: A spatio-temporal reinforcement learning approach," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 1724–1733.

MILIND TAMBE is the Gordon McKay Professor of Computer Science and Director of Center for Research on Computation and Society at Harvard University, Cambridge, MA, 02134, USA. Contact him at milind_tambe@harvard.edu.

FEI FANG is an associate professor at the Software and Societal Systems Department in the School of Computer Science at Carnegie Mellon University, Pittsburgh, PA, 15213, USA. Contact her at feifang@cmu.edu.

ANDREW PERRAULT is an assistant professor of computer science and engineering at The Ohio State University, Columbus, OH, 43210, USA. Contact him at perrault.17@osu.edu.

BRYAN WILDER is an assistant professor at Carnegie Mellon University, Pittsburgh, PA, 15213, USA. Contact him at bwilder@andrew.cmu.edu.

More info

UTM UNIVERSITI TEKNOLOGI MALAYSIA | UTM Library

SDi@JPUTM

SELECTIVE DISSEMINATION OF INFORMATION

library.utm.my

ARTICLES FOR UTM SENATE MEMBERS

"AI for the Greater Good: Harnessing Intelligence for a Better Society"

TITLE	SOURCE
10. AI: what do we fear? What do we hope for? Perception of the societal impact of AI in a European transnational cross-sectional study (2024)	Proceedings of the 30th ICE IEEE/ITMC Conference on Engineering, Technology, and Innovation (Article from : Institute of Electrical and Electronics Engineers Inc.

24th SEPTEMBER 2025
SOURCE: UTM LIBRARY

AI: What Do We Fear? What Do We Hope For? Perception of the Societal Impact of AI in a European Transnational Cross-Sectional Study

Andrés del Álamo Cienfuegos
Fundación Cibervoluntarios
Madrid, Spain
0000-0003-4438-5312

Piotr Lis
School of Economics, Finance and Accounting
Coventry University
Coventry, UK
0000-0001-6060-2423

Olha Zadorozhna
Kozminski University
Warsaw, Poland
0000-0002-8560-6428

Óscar Espiritusanto
Fundación Cibervoluntarios
Universidad Carlos III of Madrid
Madrid, Spain
0000-0003-2285-9265

Bogna Gawronska-Nowak
Institute of Urban and Regional Development
Krakow University of Economics
Krakow, Poland
0000-0003-3651-7637

Anita Zarzycka
Institute of Urban
and Regional Development
Warsaw, Poland
0000-0002-8560-6428

2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC) | 979-8-3503-6243-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICE/ITMC61926.2024.10794321

***Abstract*—This paper presents the results of a qualitative transnational and cross-sectional study of the attitudes, fears, hopes and uses of AI in the framework of the KT4D HORIZON 2024 project [1]. It contributes to the literature exploring social factors behind the discourse and practice surrounding AI. Focus group meetings with laypersons in Spain and Poland explored the impact of AI on healthcare, entertainment, education, employment and democracy. The expected changes in healthcare were perceived as most positive, whereas the impact on civil rights and democracy brought up the most fears regarding privacy, misinformation, manipulation and invasive surveillance. We conclude that digital literacy levels and trust influence heavily in the attitudes towards AI. We also suggest that media narratives should not be underestimated as a determining factor in building attitudes towards technology.**

***Index Terms*—AI, digital literacy, trust, TAM, generational gap, user acceptance**

I. Introduction

In the two year period between 2022 and 2024, the world has experienced a tremendous increase in the use of AI by the general public [1], [2]. Yet, the adoption of this technology is still uneven and heavily reliant on social factors. The discourse around AI often involves a display of strong opinions about the perceived threats and opportunities around it, whereby highly optimistic revolutionary paradigms coexist with dystopic imagined futures of surveillance and dependency. This paper presents a qualitative study of this discourse as perceived by laypersons: their expectations, hopes and fears associated with AI. The empirical work was conducted in the form of focus group discussions in Spain and Poland in the autumn of 2023, and involved participants from diverse backgrounds, age groups and with the inclusion of representatives of vulnerable groups. As AI continues to permeate various aspects of society, it is crucial for policymakers and researchers to comprehend the nuanced perspectives, hopes, and apprehensions of the general public towards this transformative technology.

The primary objective of this study is to identify perceived threats and opportunities for democracy and civil rights that stem from the advent of AI. To that purpose, the research team ran focus group meetings with a selected sample of participants, in which their perceptions were explored. The discourse centred, among others, on the relationship between the rapid development of technology and civic engagement, and the role of values and norms in this complex interaction.

The meetings were organised in a two-step approach. In the first step, a pilot study meeting was held in Cracow, Poland, which enabled us to test the meeting assumptions and verify practical solutions. In the second step, two concurrent use case meetings were held in Warsaw and Madrid, which form the basis of the analysis in this paper.

All focus group participants, irrespective of their age or other characteristics, indicated that they had encountered AI in their personal lives. They tended to have a positive attitude towards AI, and saw it as a tool that could benefit societies in multiple ways. Thus, there seems to be a good amount of good will towards AI in society. They saw potential efficiency and effectiveness gains in healthcare, which in their opinion unequivocally would benefit from AI. They also indicated the benefits of greater access, choice, content creation and

[1]The Knowledge Technologies for Democracy (KT4D) has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302. It is a project that aims to foster civic participation in democracy by capitalising on the benefits of developments in AI & Big Data Technologies by developing and validating tools, guidelines and a Digital Democracy Lab Demonstrators platform as well as conducting multidisciplinary research on the use and perception of this cutting-edge technologies. More information can be found here: https://kt4democracy.eu/

customisation in the fields of entertainment and education. Although they were aware that there would be losers in the job market, they did not appear overly worried about that as they expected that the creation of new jobs and safety improvements will lead to cumulative net gains from AI. In the sphere of civil rights and democracy, although noticing potential gains, participants had the most fears regarding privacy, misinformation, manipulation and invasive surveillance. However, they agreed that those risks could be reduced with appropriate regulation and education that raises individual digital literacy.

The final aim of this research is also to contribute to the young but thriving body of literature about the attitude towards AI, where relatively small-sized studies in different countries are being carried out to conform a constellation of knowledge that serve us to not simply complete the puzzle about the nature of the attitudes towards AI but also to understand the reasons and cultural/demographic factors behind such attitudes. The originality of this work lies in its exploration of the perceived threats and opportunities for democracy and civil rights arising from the proliferation of AI. Unlike previous studies that often highlight either optimistic or dystopian narratives, this research presents a nuanced understanding of the discourse surrounding AI, shedding light on the coexistence of optimistic and pessimistic viewpoints of lay citizens in different cultural contexts. Policy-makers can benefit from the findings of this research by gaining a deeper understanding of public sentiments towards AI, thereby informing the development of regulations and policies that address concerns such as privacy, misinformation, and surveillance while harnessing the potential benefits of AI for society. Additionally, researchers can build upon this study to further explore the cultural and demographic factors influencing attitudes towards AI, contributing to a more comprehensive understanding of this evolving field. Overall, this paper not only adds to the growing body of literature on AI attitudes but also offers practical implications for policy-making and future research endeavors.

II. Relation to the existing work

A. Attitudes towards AI, a constelation of small-scale studies

Along with the wider adoption of AI in the last year, a new and nourished body of literature about the attitudes towards AI has been growing over the last few years. [3] , [4] or [5] are examples of it. Usually these studies review medium-scale papers localised in a geographical area that explore the attitudes towards AI.

When it comes to comparing small-scale studies centred on the acceptance of AI, [6] conducts a systematic literature review with more than 60 small studies about the acceptance or rejection of AI. Several factors are shown as determinants across studies for the acceptance of AI. Across most of these studies, systematically analysed trust in AI and the complexity of use are among the common key hurdles towards AI adoption. Similarly, these studies find common ground in framing the adoption and likelihood towards acceptance to the overall digital skills, conceived as a global form of familiarity with the software and hardware that make possible the interaction of any user with the AI. These studies usually structure the attitudes towards artificial intelligence in regard to five axles (healthcare, education, recreational, labour and politics/civil society framed in the General Attitudes towards Artificial Intelligence Scale (GAAIS) [7] [8] which is an analytical framework aimed at unifying qualitative and quantitative studies about attitudes towards AI. The aim of this paper is to contribute from a transnational European comparative and inclusive perspective to the constellation of literature in this field.

Along these reports, we can see common ground: they often show results of how negative and positive opinions and attitudes are heavily fragmented depending on the thematic and age group asked. In particular, these studies focus on (1) discourse and (2) use of AI technologies which in conjunction form the attitude towards AI [7], [8] . They also show that the attitudes towards AI are polarised and polarising, and point at several causes, including the overall digital literacy, knowledge about the functioning of AI, and most importantly demographics [4], with the older age groups having commonly a far less favourable attitude towards AI than younger users. This fact is particularly worth paying attention to when observing the aforementioned narratives towards AI absorbed by such generations. Other studies focus on particular fields, notably healthcare where acceptance is mostly linked to trust and knowledge of the functioning of AI [9], [10]. When coming to these specific professional fields, the need for a deep understanding of the functioning and purpose of AI is vital at the time of establishing the necessary trust for adoption, but in the end it is trust that matters the most. Age and digital literacy stay relevant as indicators for AI adoption in these fields but with less relevance than when research is about the general population.

B. Attitudes towards AI in context

A sensible interpretation of the results of these studies should assess the fact that unlike other ground-breaking technologies (e.g., the Internet), AI does not come to us uncharted, at least from the attitudinal point of view. We must not underestimate fiction, and more specifically 20th century dystopian narratives as the initial ground to build the attitudes towards AI in particular [11]. However, the ways in which science fiction has categorised AI do not normally convey the actual functionality of the modern-day AI, usually exceeding the actual capabilities of the technology [12]. Rather than AI being autonomous, everyone engaged in an AI undertaking is part of the AI system, which includes, in addition to researchers, “those who set up the institutional arrangements in which AI systems operate, and those who fill roles in those arrangements by monitoring, maintaining, and intervening in AI systems” [13]. Ignoring all of these actors leads to a ”sociotechnical blindness” that allows for the belief “that AI systems got to be the way they are without human intervention [. . .] which facilitates futuristic thinking that is misleading” (ibid.: 587). Hence, sociotechnical blindness obscures the fact

that AI systems follow human interests and are embedded in social power structures set up by humans. These perceptions may have an important impact in trust towards AI, particularly among the elderly. Trust has been classically identified as one of the most important indicators of technology acceptance in general [14].

The data analysis was made through a discourse analysis approach [16] selecting representative and/or unique segments or components of language use (e.g., several lines of a focus group transcript) and then analysing them in detail to examine how versions of elements such as the society, community, institutions, experiences, and events emerge in discourse [17]. More specifically, [18] conceptualised that discourse analysis operates on three fundamental assumptions: antirealism (i.e., people's descriptions cannot be deemed true or false portrayals of reality), constructionism (i.e., how people's constructions are formed and undermined), and reflexivity. Discourse analysis depends on the researcher's sensitivity to language use, from which an analytic tool kit is developed that includes facets such as rhetorical organisation, variability, accountability, positioning, and discourses paying special attention top the "way versions and descriptions are assembled to perform actions" [19]. Along with that and to coordinate the different backgrounds of the transnational research team we followed the "adversarial collaborations" approach, a methodological procedure in which disagreeing scholars work together to resolve their empirical disputes, are the next necessary science reform for addressing lingering weaknesses in social scientific norms. Adversarial collaborations will further minimise false positives, expedite scientific corrections, stimulate progress for stalemated scientific debates, and ultimately improve the quality of social scientific outputs [20] .

III. Research approach

The empirical method of this study was focus group discussions and deliberations. Aided by the existing literature, we selected our participant sample in a way that ensure representation by (1) gender (2) age, (3) native/migrants and (4) participants with disabilities. with a total of 69 participants across the study being a 16% of them people from vulnerable collectives (See Table I)[2].

Since the sample chosen is not based on statistical probability but on reflecting human sense and expression, it did not require an extensive number of participants [21] (101-112). As established in [22], it is estimated that with an approximate sample of 7 participants, conclusive results can be obtained if saturation point of responses is reached and the questions are focused on a specific theme. However in this particular case this sample of 7 participants (representative of the different target groups) was replicated 10 times to get more consistent results within the qualitative sample. These 10 groups of 7 participants were the conformants of the different focus groups. Focus groups have been chosen as a research approach since they are powerful tool towards understanding the cohesiveness in attitude of a group towards a certain topic in a social setting [16] understanding not only opinions and ways in which people understand certain items, but also the way in which this knowledge is exchanged and how the cohesiveness or disjoint of the group interact with opinion [23] in a safe, isolated environment [24] Regarding the specific composition and sample selection of focus groups we follow Madriz and Merton's approach [25], [26] to identify "salient dimensions of complex social stimuli" approach. For that reason the focus groups themselves are organised by mixing participants of different age groups and backgrounds regarding vulnerability to bolster and enrich discussion through difference rather than likelihood.

These choices depart from two main reasons. One is the gaps of knowledge about AI between the diversity of profiles among participants, something that can work in detriment of other methodological approaches such as the interview or the survey where some participants may lack knowledge to answer certain questions. Focus groups give a space of knowledge sharing among participants thus constituting an adequate space for the expression of opinion gained from that knowledge acquisition. Another reason is the polarising attitudes commonly found toward cutting-edge technologies, particularly from a generational point of view [27], [28]. The mixed composition of these focus groups was designed to let those polarisations erupt for its analysis.

A. Focus group approach

The team ran focus group sessions in which the perceptions of citizens were explored. The meetings were organised in a two-step approach. In the first step, a pilot study meeting was held in Cracow, Poland, on 6th October 2024, which enabled us to test the meeting assumptions and verify practical solutions. In the second step, two concurrent use case meetings were held in Warsaw (9th November 2024) and Madrid (31st October 2024), which form the basis of the analysis in this paper. In both countries, the meetings were conducted in local languages, i.e., Polish and Spanish, respectively. As mentioned previously, the participants were selected based on several socio-demographic criteria that were set to ensure group diversity. These included gender, age, belonging to a vulnerable group, politically active/inactive, or having diverse digital skills. In total, 69 people took part in the sessions: 8 in the pilot use case in Cracow, and then 20 in Warsaw and 41 in Madrid. Although we achieved the goal of having representatives of different social groups (see Table I), considering the age dimension, the younger demographic was over represented. On one hand, this might have biased the deliberations towards the perspectives, experiences and expectations of younger people. But on the flip side, it gave us an opportunity for a deeper exploration of the attitudes and opinions on AI of that social group which will see their lives

[2]We can define a vulnerable group as an umbrella term that groups together, different intersectionalities, due to the interrelation of various categories [29], such as sex, gender, age [30] ethnicity, disability, human or social capital, socioeconomic level, access to available resources, etc [31]. This reality must also be perceived as such at the same time by the members of the same society [32]. In this sense, it is the intersectionality of these elements that shapes vulnerable individuals or groups.

TABLE I
PARTICIPANTS CHARACTERISTICS

Characteristic	Warsaw	Madrid
Number of participants	20	41
Female	9	14
Male	11	27
Student	12	27
Employed	6	6
Retired	2	4
Immigrant	2	0
Disabled	1	4
Minimum age	18	18
Maximum age	78	72

affected by the advent of AI and related technologies the most. However, it should be emphasised that we also had participants from the middle-aged and elderly groups, and their input also turned out to be significant, frequently driving the key points of the conversations. That inter-generational contrast was the primary factor fuelling the previously mentioned adversarial collaboration.

IV. FINDINGS

The goal of the focus group meetings was to explore participants' perceptions of AI, their perceived fears and opportunities, as well as explore solutions within the realms of policy and education that would alleviate some of the risks and boost opportunities that come with AI. Particular focus, due to the Knowledge Technology for Democracy (KT4D) Project requirements, was placed on the impact of AI on widely understood democracy and civil society.

Each meeting started with introductions of the research team and participants, and an overview of the KT4D project. As a warm-up exercise, before starting the discussion, the researchers asked participants to write down on a piece of paper whether in their opinion AI was a positive, negative or neutral development for humans. The exercise was repeated at the very end of the session to see whether the deliberations might have changed individuals' opinions. Participants in Warsaw saw AI as something positive already at the onset of the meeting, with only one person seeing it as a negative development. The opinions moved even more in favour of AI when the exercise was repeated at the end of the session. The number of individuals who thought of it as a good thing has increased even further. However, some individuals noticed the more nuanced effect of AI as they said that it would have both negative and positive impacts. In Madrid, in an open introductory discussion, the majority of participants also agreed that AI would have a major positive impact on society. In both countries the shift towards the positive perception was more acute among the older participants, particularly those who had lesser knowledge about the topic and acquired more knowledge during the session. This is a strong indication of how acquiring down-to-reality knowledge about this technology can be a powerful tool to break preconceived fears.

A. What is AI and where do we already encounter it in our lives?

Subsequently, there was a moderated discussion about the areas in which people interact with AI in daily life and their understanding of the term "AI". Participants indicated that they already encountered AI in chatbots, text editing tools, home appliances, search engines, image and sound editing tools, and virtual assistants such as Siri. In both countries, the use of AI applications was determined by two factors in particular. The first one was age, whereby younger individuals were more prone to integrate AI into their daily tasks. The second one was the employment status, whereby those in employment indicated increasing use of AI in their job activities. There was a huge gap in not only the use of AI but also the perception and self-consciousness of the usage of it among the disabled individuals.

In discussing how participants understood AI, they typically described it as a tool that performs tasks automatically ("*does tasks for you*"), uses large data sets and algorithms to order and exploit information, and simulates human abilities. Participants associated the following words with AI: speed, innovation, replacement of human work and thinking. They were aware that AI is not affected by emotions and feelings as humans are, and that the effectiveness and reliability of AI depend on how it is trained. They said that the more and more reliable data it is fed, the better results it produces. Moreover, participants thought that AI was still not good at distinguishing between false and true information. Notably, there was consensus regarding this description among different groups of participants with one exception: those who had technical background - they tended to modulate strong opinions and definitions of AI. For example, a participant in Spain with such a more technical profile defined AI as a "tool that let you code with common language" which consequently served as a convincing enough narrative to shift the tone of the group discussions for the rest of the session. This reveals how providing a small amount of knowledge of the functioning of AI can influence commonly attributed narratives about AI.

In the next step, the participants selected and discussed AI technologies they thought were most impactful in everyday life. They listed such tools as facial recognition, chatbots (ChatGPT), the Internet, and *deepfakes*. They debated pros, such as logging in easily, versus cons, such as invasive surveillance with facial recognition. Overall, the participants seemed acutely aware of the risks linked to the facial recognition technology. Some saw chatbots as useful learning tools for education. However, others worried that AI made cheating at school too easy, although it was hinted that this sense of threat was accompanied by a higher conformity with the traditional education system. Most participants across countries recognized the increased ease of accessing relevant information online thanks to AI, but at the same time, they were aware of potential issues and risks of bias, misinformation and manipulation. Deepfakes were indicated as a potentially destabilising means for democracy whereby rogue or hostile

agents could use the technology to influence political debate in the country, discredit public figures and deepen mistrust between various parts of the society. Table II summarises the AI-powered technologies and uses that are expected to have the greatest impact on society indicated according to the participants.

B. The impact of AI on the world in 10 years

The above warm-up activities and discussions, when participants' minds were already focused on AI and related issues, in the most crucial part of the meeting we asked participants to imagine the world in 10 years' time. Their task was to foresee AI's effects on five domains of life: health, workplace, education, entertainment, democracy and civil society. The discussion was moderated using the common framework described in the literature review section. The participants were issued with flipchart sheets to write down the expected effects. Initially, we did not ask them for any value judgments. However, when the expected effects were listed on the paper, we asked them to consider whether those were likely to be overall positive, negative, neutral or ambiguous. Table III and Table IV summarise respectively the positive and negative effects as envisaged by the participants.

In the first domain, healthcare, the outlook presented by our participants is decisively positive across all backgrounds and age groups. Imagining the future, participants envisaged that AI could aid healthcare via more accurate, speedier and cheaper diagnosis. They saw AI as a tool that would give more power to patients. This could be through access to self-diagnosis and self-treatment tools, a better understanding of medical information, and greater independence for disabled people, in particular those with mobility issues. They were also hopeful that AI would help humanity by inventing new medicines and even designing artificial vital organs for transplants. Overall, they thought that we would see great leaps in efficiency and effectiveness within the healthcare sector. However, they were worried that healthcare staff could lose some of their core knowledge as their reliance on AI would increase and expressed concerns about the security of sensitive data and nefarious purposes it could be used for.

The second considered domain was the workplace. Although the participants saw the risk and negative consequences of job displacement as AI replaces humans in some sectors, and those of older age might particularly struggle to adjust, they thought that AI would enable new efficiencies and have a net positive impact on both employment and work-life balance. The latter would be helped particularly by shorter working hours or work week. Here they saw an important role for governments to be agile in updating labour laws in a way that would protect workers and minimise the negative consequences of the AI revolution. Again, AI was viewed as a technology that could improve the access of disabled persons to employment by overcoming existing barriers. They also thought that the demand for and value of jobs requiring direct human contact would increase as people would feel a greater need to interact with other humans in the world where they are increasingly surrounded and influenced by omnipresent technology. Notably, those participants who already participated in the job market were less concerned about AI replacement than those who were not. The biggest concerns arose in the participants who were retired.

Education was the third debated domain. Unsurprisingly, students immediately pointed out that AI could do homework as well as enable cheating. However, when we delved deeper into the topic, the participants foresaw benefits that could be attributed to enhanced access to knowledge and fine-tuned personalisation in education. For example, teaching content creation could be automated, and it could be delivered in various formats, versions, and communication channels to meet the specific needs of each student. That would also improve the educational attainment of neurodiverse individuals or those with disabilities. Each student could have a personal AI advisor that would provide them with frequent feedback, advice and motivation. However, two substantial risks were flagged. First, AI-developed teaching materials and tools could be biassed or misused as the individuals or organisations behind them could seek to influence or manipulate the recipients. This is particularly worrying as AI is perceived by some as a "black box" and the users of content tend to be unable to verify its origin or processing. The second worry was that the AI revolution in education could increase inequalities and diverge in access to quality education as individuals from poorer backgrounds or less developing countries would struggle to obtain access to the latest tools. The latter worry was emphasised by the elderly participants.

Entertainment was another area where participants were unanimous about the net positive effects of AI. As in the case of education, the key change in the participants' opinion would be greater and more fine-tuned customization of content. For example, digital books and films could be immediately available in multiple languages as AI would take care of the translation, and possibly voice-overs in films. The advent of virtual reality would make remote and multi language visits to museums and other points of interest cheap and easily accessible to everyone. This would also have an inclusive effect on people with disabilities as physical barriers they often face would become irrelevant. Finally, AI was seen, mostly by younger profiles and those more digitally skilled as an upcoming art creator and enhancer of special effects, scenarios and scripts in gaming and other visual arts. Here, some participants wondered about intellectual property, the value of AI-created artworks, and whether they could be as valuable as those delivered by humans. The conclusion was that AI would create new and exciting avenues within the entertainment industry, but some rules would be required not to price out human creators. For instance, the participants emphasised that they would always want to see a full disclosure of whether an artwork is a result of AI or human work.

Civil society, politics, and democracy was the final domain of deliberation. The participants expected that AI technology and improved search engines would enable them to better fact-check information. It would also allow users to dive deeper

TABLE II
AI TECHNOLOGIES THAT WILL HAVE THE GREATEST IMPACT ON DEMOCRACY AND CIVIL RIGHTS

Tool / use	Positive impacts	Negative impacts (risks)
AI chatbots (e.g., ChatGPT)	Augmenting learning (education),	Cheating – doing work for students, AI cannot be impartial, lack of critical thinking and inability to write autonomously. Long term takeover by biased AI in decision-making
Deepfakes	Uses in the entertainment industry	Bias, misinformation, manipulation, social media and traditional media mistrust
Facial recognition	Easily logging into devices and services Ease of accessing, filtering and ordering relevant information	Intrusive surveillance, access from corporations and governments to data related to the body
Search engines	Recommendations based on your interests in any topic Entertainment and and informative value	Exacerbating information bubbles, undermining critical thinking
Content generation in social media	Advancing research and efficient decision-making	Privacy concerns if data falls into the wrong hands (governments and large corporations)

TABLE III
A SUMMARY OF THE PERCEIVED POSTITIVE EFFECTS OF AI ON LIFE IN 10 YEARS' TIME

Healthcare	Workplace	Education	Entertainment	Civil society, politics, and democracy
Improved and speedier diagnostics	Improved efficiency of some jobs	Access to greater amounts of knowledge	Customization of content	Enhanced tools for verifying information / fact checking
Accurate self-diagnosis of minor ailments	New jobs related to communicating with AI	Education will become more specialised and personalised	e-books and films available instantly in multiple languages	Online voting – more efficient and involving democracy, more frequent elections/referendums
Freeing up capacity to deal with more serious cases	Faster and better data processing and analysis	More accessible to disabled and neurodiverse individuals	Remote and multi-language realistic museum visits	Increased access to information
Development of new medicines and treatments	Shorter working week and remote work will improve work-life balance	Increased automation in content creation	More animated art in streaming services	Increased ease of expressing one's views and building connections with like-minded individuals
Remote surgeries will become more advanced and accessible	Greater importance of jobs requiring direct human contact	AI will do homework for students	Expansion of social media	Improved lawmaking and court rulings
Letting disabled be more independent	Improved safety in dangerous professions	Greater opportunities for self-learning	AI as an art creator	Better tools to detect corruption
Accessible and understandable information/instructions for patients	Overcoming barriers in access to job markets for disabled		Film: improved special effects, scripts, production assistance	Crime reduction through detection of criminal plots and prevention

into topics of their interests and foster connections with like-minded individuals on social networks. The effects of AI on the latter were a key angle in the discussion. The participants pointed out that AI would promote further developments in social media, which in turn would increase the ease of expressing individual views and lead to greater pluralism. However, extensive customisation of online content could also come with a high risk of reducing plurality by limiting people's exposure to new or opposing views. In addition, the use of AI would allow greater anonymity and microtargeting online, exacerbating information bubbles and thus could lead to a greater manipulation of individuals, groups or even whole societies.

A key theme was making democracy more responsive to people's expectations by advancing online voting which could mean more frequent elections and referendums. However, as

TABLE IV
A SUMMARY OF THE PERCEIVED NEGATIVE EFFECTS OF AI ON LIFE IN 10 YEARS' TIME

Healthcare	Workplace	Education	Entertainment	Civil society, politics, and democracy
Data security: loss or misuse of data	Job displacement	Discouraging learning, enabling cheating	Replacing traditional entertainment businesses such as nightclubs, no need for DJs	Enabling the state to closely monitor citizens – decrease in individual freedom; risks of abuse of power
Healthcare staff heavily reliant on knowledge from AI instead of their own	Disappearance of certain jobs	Increasing inequalities and divergence in access to education	Replacing referees in sport	Corporations eliciting private and sensitive information and using it for profit
Healthcare will be more effective but there will be more restrictions	Threat to methodical work	AI-developed teaching tools may be biased or misused	Reducing need/work for waiters	Risk of vote rigging or hacking if AI oversees elections
				Increased anonymity in media enabling manipulation and misinformation
				Deepfakes: impersonating politicians and opinion-makers
				Extensive customisation of online content will reduce plurality and people's exposure to new or opposing views
				Facilitating widespread misinformation and microtargeting
				Increase in phishing

the process would be automated, the risk of vote rigging or hacking by rogue agents or enemy states would be much greater if AI oversees elections.

Participants also worried that the AI technology would enable governments and private corporations to extract private and sensitive information about individuals, although they could not picture exactly how. This was a way more present worry in older generations and could track down to the mentioned fictional depiction of AI in media, in which these scenarios are common. These worries were exacerbated by the perceived opacity of AI technology and the inability to understand its functioning by laypersons. The misuse of AI could lead to intrusive monitoring of citizens for political gains by governments, or profits by corporations. As in other cases, participants spotted some opportunities also here. They thought that AI could help reduce crime and corruption. In both cases, it could improve prevention by spotting early signs or risk flags, or unusual patterns linked to certain activities or individuals. Overall, the participants worried that in the area of civil rights, the risks of AI may outweigh its benefits and indicated that more (intergovernmental) oversight would be necessary in this area. Interestingly, participants in Poland feared more the abuse of power by the government rather than corporations.

C. Educating for the world with AI

The participants were keen to point out ethical concerns related to AI. They saw a need to regulate the technology, impose legal or self-policing limits, and come up with a code of conduct for tech companies developing and using AI. A key conclusion of the use case meetings was that for societies to benefit from AI and minimise the risks of negative effects, more education of citizens is required. This should give them the knowledge and tools that would enable individuals to navigate the world with AI in it.

Therefore, the final part of each session explored features of good educational materials about AI, such as their format, diversity, credible sources of such materials, and accessibility. The participants urged educators to recognise that no one format will work for all recipients, and a multimodal approach is needed to reach diverse groups across generations and layers of society. They favoured diversified formats, interactivity, consideration of accessibility measures, and involvement of credible experts and thought leaders to communicate the knowledge. They saw room for involving show business celebrities as this could help match messengers to audiences, but they cautioned against relying only on celebrities in any case. Some suggested AI could even be involved in teaching people about AI.

The issue of trust in educational materials was strongly linked to the sources of their funding, which according to

the participants should be clearly disclosed and transparent so that users can assess potential bias. Here, some saw a role for governments that should promote digital skills and competencies and lend credibility to AI educational campaigns. This view, however, brought up a heated debate as other participants raised concerns that some citizens may distrust government-sponsored messages. Some consensus was reached that EU-level financing could be better as it should seem more neutral and trustworthy to promote common understanding across member states and citizens. Although it was acknowledged that those critical of the EU may perceive the content as biassed. When considering whether educational materials could be sponsored by corporations such as tech firms, participants were unanimously sceptical and agreed that this could be viewed as promoting a particular agenda instead of impartial education. In the end, a weak consensus emerged that a mix of public and private funding from diverse sources could help signal neutrality and balance concerns of bias. But this would have to be with a strong caveat of full transparency of who the sponsors were, and potential conflicts of interest should be disclosed. The majority of participants also thought that whenever possible, ways to verify or fact-check information should be provided as this would give people confidence and ability to assess content credibility themselves.

V. Conclusions

Popular media discussions about AI often showcase strong opinions on both potential threats and opportunities, ranging from optimistic visions of revolution to dystopian fears of surveillance and dependence. The methodology employed, including a two-step approach with a pilot study followed by concurrent use case meetings, demonstrates a rigorous and systematic approach to data collection and analysis. By examining laypersons' perspectives on the relationship between technology development, civic engagement, and societal values, this study offers valuable contribution into the understanding of complex interplay between AI and democratic principles.

At the onset of the meetings, participants expressed a positive, although not necessarily critical, view of AI. The deliberations during the sessions, which explored the impact of AI on life spheres such as healthcare, entertainment, education, employment and democracy, fostered a critical thinking approach as the participants identified both positive and negative impacts AI could have on individuals and society. The expected changes in healthcare were perceived as most positive. On the other end of the spectrum was the impact on civil rights and democracy, where although noticing potential gains, participants had the most fears regarding privacy, misinformation, manipulation and invasive surveillance. Although the overall perceptions of AI remained overwhelmingly positive at the end of the sessions, the participants presented a more nuanced and critical assessment of the effects of the technology.

We conclude that digital literacy levels are significant drivers of attitudes towards AI. Thus, from the societal perspective, efforts should be made to develop and implement solutions through policy and education to address people's concerns and promote understanding of and trust towards AI. Our participants unanimously agreed that more regulation and education are needed to make them feel safe, minimise the risks and maximise the benefits of AI. This should happen through credible sources of information with transparent funding to avoid bias, misconceptions and distrust.

Acknowledgment

Special thanks for the research and recruitment of participant's team in both in IRMIR and Cibervoluntarios who has worked in the gathering and processing of data for this focus groups. Special thanks in particular to Lucía García Larrauri who was essential in the execution of the research, particularly in the execution of the focus group methodology and data gathering.

References

[1] Lee, H. (2023). The rise of ChatGPT: Exploring its potential in medical education. Anatomical sciences education.

[2] McGrath, Q. (2024). Responding to the Sharp Rise in AI in the 2023 SIM IT Trends Survey. MIS Quarterly Executive, 23(1), 8.

[3] Brauner, P., Hick, A., Philipsen, R., & Ziefle, M. (2023). What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. Frontiers in Computer Science, 5, 1113903.

[4] Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O., Demir Kaya, M. (2024). The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. International Journal of Human–Computer Interaction, 40(2), 497-514.

[5] NASK (2019). Sztuczna Inteligencja w społeczeństwie i gospodarce. Analiza wyników ogólnopolskiego badania opinii polskich internautów. NASK – National Research Institute, Warsaw, Poland. https://www.nask.pl/pl/raporty/raporty/2594,Sztuczna-inteligencja-w-oczach-Polakow-raport-z-badan-spolecznych.html

[6] Kelly, S., Kaye, S. A., Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. Telematics and Informatics, 77, 101925.

[7] Schepman, A., Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. Computers in Human Behavior Reports, 1, 100014.

[8] Schepman, A., Rodway, P. (2022). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. International Journal of Human–Computer Interaction, 1–18.

[9] Hamedani, Z., Moradi, M., Kalroozi, F., Manafi Anari, A., Jalalifar, E., Ansari, A., ... Karim, B. (2023). Evaluation of acceptance, attitude, and knowledge towards artificial intelligence and its application from the point of view of physicians and nurses: A provincial survey study in Iran: A cross-sectional descriptive-analytical study. Health Science Reports, 6(9), e1543.

[10] Lambert, S. I., Madi, M., Sopka, S., Lenes, A., Stange, H., Buszello, C. P., Stephan, A. (2023). An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digital Medicine, 6(1), 111.

[11] Hermann, I. (2023). Artificial intelligence in fiction: between narratives and metaphors. AI society, 38(1), 319-329.

[12] Hermann, I. (2020). Künstliche intelligenz in der science-fiction: zwischen magie und technik. FIfF-Kommun, 4(2020), 12-17.

[13] Johnson, D. G., Verdicchio, M. (2017). Reframing AI discourse. Minds and Machines, 27, 575-590.

[14] Silva, P. (2015). Davis' technology acceptance model (TAM)(1989). *Information seeking behavior and technology adoption: Theories and trends*, 205-219.

[15] Silva, P. (2015). Davis' technology acceptance model (TAM)(1989). Information seeking behavior and technology adoption: Theories and trends, 205-219.

[16] Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., Zoran, A. G. (2009). A qualitative framework for collecting and analyzing data in focus group research. International journal of qualitative methods, 8(3), 1-21.
[17] Phillips, L. J., Jorgensen, M. W. (2002): Discourse analysis as theory and method. Thousand Oaks, CA: Sage.38(5), 869-875.
[18] Cowan, S., McLeod, J. (2004). Research methods: Discourse analysis. Counselling Psychotherapy Research, 4, 102.
[19] Potter, J. (2004). Discourse analysis. *Handbook of data analysis*, 607-624.
[20] Clark, C. J., & Tetlock, P. E. (2023). Adversarial collaboration: The next science reform. In *Ideological and Political Bias in Psychology: Nature, Scope, and Solutions* (pp. 905-927). Cham: Springer International Publishing.
[21] Ladner, S. (2016). *Practical ethnography: A guide to doing ethnography in the private sector*. Routledge.
[22] Handwerker, W. P., & Wozniak, D. F. (1997). Sampling strategies for the collection of cultural data: an extension of Boas's answer to Galton's problem. *Current Anthropology*, *38*(5), 869-875.
[23] Peters, D. A. (1993). Improving quality requires consumer input: Using focus groups. Journal of Nursing Care Quality, 7, 34–41.
[24] Vaughn, S., Schumm, J. S., & Sinagub, J. (1996). Focus group interviews in education and psychology. Thousand Oaks, Sage.
[25] Madriz, E. (2000). Focus groups in feminist research. In N. K. Denzin & Y. S. Lincoln (Eds.), Handbook of qualitative research (2nd ed., pp. 835–850).
[26] Merton, R. (1987). The focused group interview and focus groups: Continuities and discontinuities. Public Opinion Quarterly, 51, 550–566.
[27] Elias, S. M., Smith, W. L., & Barney, C. E. (2012). Age as a moderator of attitude towards technology in the workplace: Work motivation and overall job satisfaction. *Behaviour & Information Technology*, *31*(5), 453-467.
[28] Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *53*(5), P329-P340.
[29] Ruof, M. C. (2004). Vulnerability, Vulnerable Populations, and Policy. Kennedy Institute of Ethics Journal, 14(4), 411–425. doi:10.1353/ken.2004.0044
[30] Bozzaro C., Boldt J., & Schweda M. Are older people a vulnerable group? Philosophical and bioethical perspectives on ageing and vulnerability. Bioethics. 2018; 32: 233–239.
[31] Medeiros, P. . (2019). A Guide for Graduate Students: Barriers to conducting qualitative research among vulnerable groups. Contingent Horizons: The York University Student Journal of Anthropology, 3(1), 37–52.
[32] European Institute for Gender Equality (EIGE) Marginalized groups. (n.d.).