Navigating the perils of artificial intelligence
a focused review on ChatGPT and responsible research and innovation

Polyportis, Athanasios; Pachos-Fokialis, N.

DOI
10.1057/s41599-023-02464-6

Publication date
2024

Document Version
Final published version

Published in
Humanities and Social Sciences Communications

Citation (APA)
Polyportis, A., & Pachos-Fokialis, N. (2024). Navigating the perils of artificial intelligence: a focused review on ChatGPT and responsible research and innovation. *Humanities and Social Sciences Communications*, *11*(1), Article 107. https://doi.org/10.1057/s41599-023-02464-6

Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-023-02464-6 OPEN

Navigating the perils of artificial intelligence: a focused review on ChatGPT and responsible research and innovation

Athanasios Polyportis [1]✉ & Nikolaos Pahos[2]

While the rise of artificial intelligence (AI) tools holds promise for delivering benefits, it is important to acknowledge the associated risks of their deployment. In this article, we conduct a focused literature review to address two central research inquiries concerning ChatGPT and similar AI tools. Firstly, we examine the potential pitfalls linked with the development and implementation of ChatGPT across the individual, organizational, and societal levels. Secondly, we explore the role of a multi-stakeholder responsible research and innovation framework in guiding chatbots' sustainable development and utilization. Drawing inspiration from responsible research and innovation and stakeholder theory principles, we underscore the necessity of comprehensive ethical guidelines to navigate the design, inception, and utilization of emerging AI innovations. The findings of the focused review shed light on the potential perils of ChatGPT implementation across various societal levels, including issues such as devaluation of relationships, unemployment, privacy concerns, bias, misinformation, and digital inequities. Furthermore, the proposed multi-stakeholder Responsible Research and Innovation framework can empower AI stakeholders to proactively anticipate and deliberate upon AI's ethical, social, and environmental implications, thus substantially contributing to the pursuit of responsible AI implementation.

[1] Faculty of Applied Sciences, Department of Biotechnology, TU Delft, Delft, The Netherlands. [2] Faculty of Technology, Policy, and Management, TU Delft, Delft, The Netherlands. ✉email: a.polyportis@tudelft.nl

Introduction

At the end of 2022, the Chat Generative Pre-Trained Transformer (ChatGPT) was unveiled, marking a significant advancement in AI. As a sophisticated chatbot, it uses deep learning to perform various language tasks with unprecedented human-like fluency. Unlike previous AI, ChatGPT's neural networks are trained on extensive data, including simulated human conversations, enabling it to offer nuanced, conceptually rich responses that closely imitate human interaction. This breakthrough is poised to revolutionize learning and information dissemination, reflecting its significant technological prowess (Dwivedi et al., 2023).

However, the question remains: Is ChatGPT an unmitigated boon, or does its deployment come with potential pitfalls? When utilized responsibly, ChatGPT might bring benefits at the individual, organization, and societal levels in fields such as customer service, education, healthcare, finance, entertainment, creative writing, digital marketing, and e-commerce (Rivas & Zhao, 2023; Stahl & Eke, 2024). Furthermore, advances in emerging AI technologies can significantly raise productivity (Yigitcanlar, 2021) and enhance administrative efficiency and public service delivery (Yigitcanlar et al., 2023) due to their economic and societal benefits (Wilson & van der Velden, 2022).

Nevertheless, with these prevalent AI advances comes the need for sustainable and responsible AI development (Dignum, 2018). The case study of ChatGPT highlights the conundrum of the AI revolution, as, apart from apparent benefits, it raises ethical concerns about the potential negative impact of advanced AI on society AI. Given the nascent stage of AI ethics, there is a tenacious need to devise frameworks for AI advancements to incorporate ethical considerations (Lo Piano, 2020). Indeed, the emergence of AI related to dangers for humanity or existential risks is viewed by the AI community as a probable scenario (Müller & Bostrom, 2016; Bostrom, 2020). This study takes as a starting point that AI innovations such as ChatGPT might challenge privacy, inclusivity, and inequality (Anshari et al., 2023; Dignum, 2018; McGee, 2023; Yuste et al., 2017), bring bias (Chen, 2023) and raise safety concerns (Kaplan & Haenlein, 2020; Rozado, 2023; Yigitcanlar et al., 2023). ChatGPT in specific, has been found to corrupt rather than improve moral judgment (Krügel, et al., 2023).

This review identifies critical research gaps regarding ChatGPT and similar AI tools. It emphasizes the urgent need to unveil potential risks at different levels and to develop an ethical framework guiding AI's lifecycle. The originality of this focused review lies in its targeted synthesis of current literature to address specific, pertinent research questions (Alderman et al., 2012; Cowan et al., 2005; Taylor et al., 2016). Specifically, we identify the perils rising at different societal levels and advocate for adopting a *Responsible Research and Innovation* approach (Owen et al., 2013; Stilgoe et al., 2013). Responsible Research and Innovation encourages a transparent and collaborative approach in which innovators and societal actors work together to ensure that innovations are sustainable, ethically acceptable, and societally desirable. Responsible Research and Innovation promotes a participative process, ensuring AI developments align with societal values and ethical standards from inception, proactively addressing potential social and ethical issues.

Our approach will, therefore, focus not on investigating how to support AI in general, but rather on suggesting a framework of ethical consideration after mapping the perils rising from irresponsible AI implementation at the individual, organizational, and societal levels. Specifically, we define two research questions:

What are the perils of irresponsible ChatGPT (and similar AI tools) development and implementation (RQ1)?

In what ways can a multi-stakeholder Responsible Research and Innovation framework guide the sustainable development and use of chatbots (RQ2)?

Research methods

This study employed a focused review of the literature (Alderman et al., 2012; Cowan et al., 2005; Taylor et al., 2016) to synthesize and analyze the concepts and theories most relevant to the development and use of advanced AI technologies within the context of Responsible Research and Innovation. A focused review is characterized by its concise examination of a particular set of literature most relevant to specific research questions, hence providing targeted insights into a specific aspect of the research area. It *does not aim for an exhaustive analysis of the entire research domain* (Alderman et al., 2012) but instead concentrates on identifying relevant articles and synthesizing key findings on specific research questions, as in the present study. Furthermore, as our research focuses on synthesizing concepts and theories in the context of advanced AI technologies, rather than conducting a quantitative analysis of citation patterns or publication trends, we did not utilize methods such as bibliometric analyses (Luo et al., 2022). Through this focused literature review, we distill critical concepts, insights, and recommendations from the selected literature, mapping the challenges of AI tools and building a targeted Responsible Research and Innovation framework for the sustainable use of emerging chatbots.

We conducted a literature search on databases Scopus and Google Scholar, using keyword combinations including: "ChatGPT AND responsible research and innovation"; "artificial intelligence AND responsible research and innovation"; "chatbot AND responsible research and innovation"; ChatGPT AND responsible development"; "ChatGPT AND perils"; "ChatGPT AND challenges"; "ChatGPT AND implications". Such keywords enable us to augment the scope of our search. This search resulted in 1892 records.

We then defined the retaining and discarding criteria to filter the searched literature, including (1) literature published in the English language; (2) literature published between 2011–2023; (3) literature with full-text articles assessed for eligibility; (4) articles published in academic journals and conference papers. Articles in the specific keyword were referred to without any deeper explanation or further development or mentioned only in the references were excluded. Furthermore, papers published in non-international journals or uploaded as preprints or without having been peer-reviewed were excluded from the subsequent analysis. Of the publications retained, only those which, in the authors' opinion, had the most relevance to the research question were considered for further review. This process resulted in 118 records.

Subsequently, 58 records were removed from consideration in the eligibility phase because they were duplicates. A cited search was then conducted, including a snowball technique in which citations within these articles and relevant papers that cited these articles were searched and retained if they appeared relevant to the review. This process resulted in 72 articles (see Flowchart of the study selection process—Fig. 1 below).

We visited this final pool of articles to define and map the perils of irresponsible ChatGPT (and similar AI tools) development implementation at different societal levels (RQ1) and to propose a multi-stakeholder Responsible Research and Innovation framework, synthesizing meaningful implications based on this framework (RQ2).

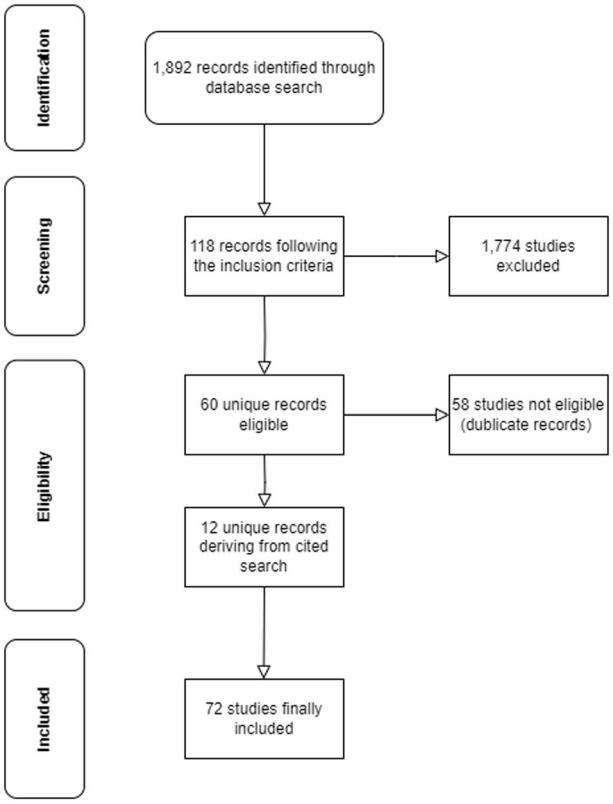

Fig. 1 Flowchart of the study selection process.

Findings

Research question 1. Since advanced AI such as ChatGPT can automate processes, deliver individualized services, and even mimic human contact, they can advance societies in anxiety-provoking and exciting ways. A few years back, Professor Stephen Hawking said that "The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not yet know which" (Cellar-Jones, 2016). Such scenarios are more realistic nowadays, given the recent AI advancements. Advanced AI tools might raise concerns of existential risk, especially as these tools close the gap towards the so-called Superintelligence level. Superintelligence refers to an intellect that surpasses human cognitive abilities in nearly all areas of interest (Bostrom, 2020). Superintelligence encompasses dominance in goal-oriented behavior, regardless of whether it is an artificial or human intelligence that possesses qualities such as self-awareness or intentionality.

Accordingly, with ChatGPT-like tools closer to the state of Superintelligence, the concept of Singularity (Vinge, 2008) is becoming possible. 'Singularity', concerning AI and human civilization, refers to a hypothetical point in the future when artificial intelligence will have surpassed human intelligence to such an extent that it can independently improve itself and evolve at an exponential rate. Singularity is based on the idea of an intelligence explosion, in which an AI system becomes capable of improving its intelligence, leading to an exponential increase in its capabilities. This concept might bring the potential of a future in which machines can address many human concerns, including poverty, illness, or environmental degradation. Still, it could also lead to loss of autonomy and economic and social upheaval.

In this context, the initial hurdle involves the collaborative and precise prediction of the impacts of advanced AI tools like ChatGPT, which are approaching the capabilities of super-intelligent systems. These tools have the potential to gradually supplant and enhance the majority of cognitive tasks that humans have historically undertaken. At the individual level, ChatGPT might raise concerns about the quality of information provided or even be used to spread misinformation or disinformation by generating text that may seem credible but is factually inaccurate (Hsu et al., (2023); Verma & Oremus, 2023). Such disinformation can be cheaper and easier to produce for even more conspiracy theorists (Hsu et al., (2023)), as chatbots have no commitment to the truth (Bell, 2023). One particularly worrying aspect is the potential for ChatGPT and similar tools to target vulnerable individuals with harmful or distressing content, subjecting individuals to emotional and psychological harm. This not only affects their mental well-being but also has broader implications for online safety and cyberbullying prevention (Mijwil & Aljanabi, 2023).

ChatGPT can also be used to gather personal information from users for identity theft or other malicious purposes, resulting in privacy and security concerns (Ahmad et al., 2023). Such concerns may lead to fear, followed by emotional uncertainty (Polyportis, 2020; Polyportis et al., 2020), hindering future experiences of human-AI interactions. Up to now, AI tools could provide a certain level of entertainment but could not understand emotions, empathize, and offer a sense of belonging that human friends can provide. Nonetheless, as AI chatbots are getting closer to Superintelligence, it is not unlikely for them to substitute human relationships, thus devaluating relationships and augmenting alienation.

At the organizational level, expanding AI tools like ChatGPT might lead to possible negative economic impacts, such as a change in the labor market structure (Rakowski et al., 2021). For example, instability in the labor market could lead to replacing full-time job positions with part-time jobs, which would cause high uncertainty to job seekers about their future career path (Polak, 2021). These changes can potentially reduce person-to-person contact in areas like customer support, personnel management, or technology roles, which could jeopardize the job security of workers in these fields. The potential displacement of human workers is driven by the efficiency and precision with which AI systems can execute tasks traditionally in the human domain. Indeed, AI technologies exemplified by ChatGPT excel at automating routine and rule-based tasks, conducting intricate calculations, and even generating code—often surpassing human speed and consistency (Bessen, 2019; Brynjolfsson & McAfee, 2014). Consider, for example, the scenario in which ChatGPT can swiftly and proficiently generate code in various programming languages. In such a context, employers may question retaining human programmers.

Furthermore, technology dependence and delegation experiences can help consumers feel empowered but can also hurt the firm (Paul et al., 2023). Indeed, recognizing AI's capability to act as a substitute for human labor can be psychologically menacing and decrease the perceived self-efficacy during human-AI interactions (Puntoni et al., 2021). Hence, replacing human-human interactions with human-AI interactions, along with ethical concerns and heightened perceptions of creepiness, the absence of personalization and empathy, could potentially result in adverse effects on the brand experience and customer loyalty (Luo et al., 2019; Rajaobelina et al., 2021; Niu & Mvondo, 2024). ChatGPT-like tools can facilitate theft of intellectual labor, copyright concerns and violations leading to unfair competition (Strowel, 2023).

Moreover, while ChatGPT offers several benefits to businesses, there are several perils related to security, loss of privacy, legal risks, and spreading false information about the firm (Dwivedi et al., 2023; Paul et al., 2023). For example, the evolution of such tools poses significant risks to cybersecurity when manipulated by

malicious actors or used for phishing attempts (Chilton, 2023). At the same time, organizations that share confidential information with ChatGPT might have severe consequences due to underscoring the importance of safeguarding sensitive data. Similarly, the evolution of AI might also threaten governments through several applications, such as using AI for surveillance purposes, which raises significant privacy concerns for individuals (Yigitcanlar et al., 2023).

At the societal level, education and research are among the most controversial fields regarding the implications of ChatGPT (von Garrel & Mayer, 2023), as it may exhibit critical thinking and generate highly realistic text, potentially resulting in decreased levels of students' creativity. The notion that students can efficiently complete academic tasks without producing original work contradicts essential educational values concerning student growth, development, and the practical use of knowledge. At the same time, a rapid evolution of ChatGPT-like tools might jeopardize the integrity of examinations. It came as no surprise that ChatGPT, which is still in its infancy, was recently able to receive a B-grade on the final exams of an MBA course that an Ivy League business school offered. In addition, the effects of such tools in the research field might raise concerns regarding the ethicality and integrity of publishing among the research community. ChatGPT has already been listed as a co-author on recent academic articles (e.g., ChatGPT & Zhavoronkov, 2022; O'Connor, ChatGpt (2023)), leading to reactions from a significant part of the research community (Thorp, 2023) that suggests a regulation in its use (Stokel-Walker, 2023).

Another potential societal implication associated with the accessibility of ChatGPT and similar AI tools. Although such tools are free today, "it is only a matter of time before they become paid services" (Dwivedi et al., 2023, p. 10). One can easily imagine that potential digital inequities in education might lead to tensions if learners from different socioeconomic statuses and cultural backgrounds do not have equal education opportunities.

Advanced AI tools like ChatGPT are also poised to significantly impact the healthcare sector (Ali et al., 2023; Patel et al., 2023; Sallam, 2023). Inherent risks primarily stem from system errors (Srivastava & Rossi, 2019; Dwivedi et al., 2021), while patient privacy concerns can hinder data accessibility. Additionally, the integration of AI into healthcare alters traditional decision-making practices based on epistemic probability and prudence (Parviainen, Rantala (2022)) and raises ethical, legal, and medical dilemmas when it comes to making critical decisions about human lives and medical conditions (Shaban-Nejad et al., 2021; Vilaza & McCashin, 2021).

The necessity of ethical considerations. Based on the above discussion, we identified and categorized significant risks of emerging ChatGPT-like tools at the individual, organizational, and societal levels. Such adverse consequences of advanced AI tools such as ChatGPT have triggered an ongoing debate about the need to establish a set of principles to effectively regulate and monitor AI development and use (Arrieta et al., 2020; Fjeld et al., 2020; Mikalef et al., 2022). However, prior research has yet to converge on how a multi-stakeholder Responsible Research and Innovation framework can be designed to optimally tackle the abovementioned risks and encourage the sustainable development and use of chatbots. Furthermore, previous literature has yet to analytically examine how the rapid advancement of AI and its effects on multiple industries demand a comprehensive assessment of its potential influence on aligning with ethical and sustainable goals (Guo & Polak, 2021).

In general, people are increasingly worried about the risks of AI at a societal level (Araujo et al., 2020). The concept of AI and the discussion about its ethical concerns dates back to the 1950s (Stahl et al., 2022a). However, the importance of deliberative engagement for framing responsible innovation in scientific research and technological development did not spring like Dionysus formed from Zeus' thigh. Instead, it has been recently highlighted and discussed in the business and technology ethics literature (Owen et al., 2013). The rise of voices of concern regarding ChatGPT (Tlili et al., 2023) emphasizes that developers must ensure that chatbots are built with moral considerations in mind, while users and governments should be aware of the potential risks and ethically use chatbots and AI technologies.

Nonetheless, ethical guidelines do not necessarily alter the individual (developers') decision-making (Hagendorff, 2020). For instance, in the study of McNamara et al. (2018), software engineering students and developers were presented with eleven software-related scenarios on ethical decision-making, such as responsibility to report, intellectual property, or honesty to customer to time. Some of the participants were exposed to established codes of ethics. McNamara et al. (2018) found no statistically significant difference in the participants' responses, regardless of whether they saw the code of ethics or not and their student or professional status.

At the same time, the innovators who hire the developers of these emerging AI technologies may become structures of power (Dwivedi et al., 2023) with unprecedented societal implications if a robust framework of ethics and social responsibility is lacking. Such powerful technologies are being produced by specific organizations (apart from OpenAI's ChatGPT, equally powerful systems such as Google's Bard have emerged). Naturally, one would ask if such innovators are, and should, be in the position to heavily influence who will 'win' and who will 'lose' in society due to the emerging AI technologies. Nonetheless, according to the basic principles of neoclassical economics, the AI innovator(s) may prioritize maximizing their profit and utility (e.g., Agboola, 2015) instead of aligning within an ethical framework. Hence, there is a need to address the responsible development of AI technologies by considering the AI ecosystem stakeholders.

Research question 2. The synthesis of the findings of the focused review highlights the paramount importance of an ethical framework for the development and ethical implementation of AI technologies, exemplified by ChatGPT, aligned with previous research on the ethics of artificial intelligence (e.g., Bostrom & Yudkowsky, 2018; Etzioni & Etzioni, 2017). Notably, the findings unveil the need for a multi-stakeholder approach inclusive not only of the innovator company but also of the other societal actors, including the regulators and assessors (e.g., government and national organizations) of the AI technology, as well as the direct and indirect stakeholders, such as the academia, industry, Non-Governmental Organizations (NGO's), consumers and eventually the society at large. This proposal is based on two main arguments. First, it is consistent with stakeholder theory. This well-established theoretical framework provides an understanding of the relationships between the focal organization (i.e., AI innovator) and the stakeholders they interact with (Freeman, 2010). The stakeholder theory speculates that an organization is not only accountable to its shareholders but, in essence, to all societal stakeholders that may have an interest in or are affected by the actions of the focal organization. Therefore, applying stakeholder theory to a Responsible Research and Innovation framework can assist in identifying the diverse stakeholders involved within the ecosystem and their respective interests. Second, stakeholder engagement is critical for achieving social and ethical goals in the innovation process. Importantly, multi-stakeholder collaboration can lead to more equitable outcomes,

enhance transparency and accountability, and foster trust and legitimacy (Mitchell et al., 1997; Owen et al., 2013).

Theoretical grounds of responsible research and innovation.

To craft a framework for the responsible development of AI, it is essential to thoroughly investigate the conceptual foundations of Responsible Research and Innovation, as laid out by scholars like Burget et al. (2017), Owen, Von Schomberg et al. (2021), Stahl et al. (2017), Stilgoe et al. (2013), Stilgoe & Guston (2016), Sutcliffe (2011), and Von Schomberg (2011), and extract key learnings.

Responsible Research and Innovation ensures that innovators and societal actors work together with transparency to ensure that innovation processes and their products are sustainable, ethically acceptable, and societally desirable. Indeed, various authors have brought out ethical issues and concerns (e.g., Forsberg et al., 2015; Gianni, 2016) under the lens of responsible research and innovation. Hence, a responsible research and innovation approach to the continuous development of ChatGPT-like tools and AI technologies should be adequate to encourage the involvement of all relevant stakeholders and the consideration of diverse perspectives through collective decision-making and management of such scientific and technological advancements.

Interestingly, the term responsible research and innovation was not initiated by researchers but by policymakers within the European Commission from a top-down perspective. Back in 2013, a European Commission policy document named "Options for strengthening responsible research and innovation" defined Responsible Research and Innovation as the comprehensive approach of proceeding with research and innovation in ways that enable all stakeholders involved in the processes of research and innovation, especially at an early stage (a) to gain relevant knowledge on the consequences of their outcomes and actions and on the variety of options open to them, (b) to effectively appraise both outcomes and options in terms of societal needs and moral values and (c) to use these considerations (a and b) as essential input for design and development of new research, products and services (European Commission, 2013).

Similarly, Von Schomberg (2011) defined Responsible Research and Innovation as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products" (p. 9). Von Schomberg hence refers to interaction between the stakeholders ("societal actors and innovators") beyond, for instance, the company that develops and commercializes the AI technology, with transparency being the principal value. Von Schomberg views this procedure as a safeguard that the innovation process and its products satisfy specific essential criteria, namely *ethical acceptability, sustainability*, and *societal desirability*. These criteria, subsequently coined as "normative anchor points" (Boenink & Kudina, 2020) can be operationalized in alignment with the European Union declaration on human rights and include the *public values of safety, privacy, sustainability, quality of life*, and *gender equality* (Von Schomberg, 2011, p. 9–10).

Stilgoe et al. (2013) proposed an alternative approach to value identification and established four dimensions of Responsible Research and Innovation: anticipation, inclusiveness, reflexivity, and responsiveness. Concerning OpenAI, the innovator that commercialized ChatGPT, and similar companies that develop and commercialize similar AI tools within their respective multi-stakeholder ecosystems, these dimensions can be described as follows:

- *Anticipation* involves proactively assessing the potential effects of the relevant innovation as integrated into the research and innovation process (Fraaije & Flipse, 2020; Owen et al., 2013; Stahl et al., 2017), preferably at the early stage of development. Anticipation is thus related to a priori recognizing a chatbot's social, ethical, and environmental implications. This dimension can be applied to the broader society, including direct and indirect stakeholders, by promoting the proactive identification and assessment of the AI technology impact. For example, foreseeing the influence of chatbots on employment conditions, privacy, or social interaction can be beneficial to mitigate any negative effects.
- *Inclusiveness* pertains to engagement with relevant stakeholders at the early stage of innovation processes, such as the involvement of public values (Bozeman et al., 2015), with the byproducts of such engagement being integrated into the research and innovation process. Public values influence the outcomes important to society, including users, regulators, and civil society organizations. Also, an innovator may be encouraged to collaborate with experts (Baba & Walsh, 2010). For instance, engaging specialists in human-AI interaction can augment the probability that AI systems are usable and accessible to diverse users.
- *Reflexivity* is relevant to reflecting on its impact on society, along with its purposes, motivations, and values (Burget et al., 2017; Stilgoe et al., 2020). Reflecting on the values and assumptions that underlie AI technology development can be applied not only to the innovator(s) but also to the direct and indirect stakeholders and, in essence, to society. Reflexivity can promote transparency and accountability in chatbot development and deployment, urging researchers and developers to be open about the limitations, biases, and uncertainties associated with chatbots and fostering trust among users and stakeholders.
- *Responsiveness*: Is the Responsible Research and Innovation process responsive to social needs? Moreover, is this process organized to respond to new challenges, insights, and emerging contingencies? Overall, it is paramount for companies to be responsive to societal needs and concerns to ensure that chatbots are designed and deployed in responsible and ethical ways to mitigate risks and seize opportunities (Wiarda et al., 2022). In addition, the responsiveness dimension entails considering stakeholders' feedback and concerns (Nielsen, 2016; Schuijff & Dijkstra, 2020) during the development and deployment of chatbots. This, for instance, may require engaging in dialog with stakeholders, including users, customers, and the public.

Both Von Schomberg (2011) and Stilgoe et al. (2013) converge that Responsible Research and Innovation should serve the process and the products of innovation and that a conceptualization of values as ready-made entities, being available for deliberation between stakeholders, is needed. In the context of AI and ChatGPT-like tools, the Responsible Research and Innovation approach enables the consideration of the ethical implications of chatbot technology. This approach ensures that the development and use of chatbots align with the multi-stakeholder ecosystem values (i.e., safety, privacy, sustainability, quality of life and gender equality; Owen & Pansera et al., 2021; Von Schomberg, 2011) and needs of the multi-stakeholder system and addresses potential ethical issues based on the four dimensions recognized above (anticipation, inclusivity, reflexivity, responsiveness; Burget et al., 2017; Fraaije & Flipse, 2020; Stilgoe et al., 2013; Stilgoe et al., 2020). Furthermore, extending Von Schomberg (2011), we propose that the values that guide innovation are considered from the beginning of the innovation process within the multi-stakeholder ecosystem.

Guidelines and proposed framework. Based on the above Responsible Research and Innovation grounds and synthesizing on broader literature on artificial intelligence, ethics, and policy, we argue on specific guidelines that stakeholders, such as the AI innovators, but also the regulators, assessors, direct and indirect stakeholders, and the society at large, should follow to ensure the sustainability, ethical acceptability, and social desirability of ChatGPT and similar prevalent AI tools. Given that applying AI guidelines can be a complex task (Atkins et al., 2021), we delve into these difficulties with precision, drawing from our review findings to enhance the practical enactment of these guidelines and practices.

First, developers must emphasize building and/or advancing robust AI tools by proactively tackling safety and security issues. Promoting transparency and accountability can lead to more equitable and responsible development of this technology and also help build trust and confidence in using such tools. In addition, ChatGPT-like tools could be developed with the primary aim of improving living conditions and shared prosperity, while global governance structures should be facilitators in that process. Specifically, in AI governance, a multi-layered model is advocated, involving governments, civil society, the private sector, and academia to collectively discuss and implement governance mechanisms (Gasser & Almeida, 2017). This collaborative effort is vital to minimize risks associated with AI while maximizing its potential benefits. Responsible AI also directly contributes to achieving several of the UN's Sustainable Development Goals, such as gender equality, decent work, economic growth, industry innovation, infrastructure, and the reduction of societal inequalities. Ensuring that AI is fair and equitable across different demographics is a crucial concern within this framework. Moreover, the first-ever global agreement on the ethics of AI, adopted by 193 countries, aims to utilize AI as a force for good, emphasizing the promotion of human rights and contributing to the Sustainable Development Goals. This includes addressing transparency, accountability, and privacy (United Nations, 2021). Finally, the ethical transformation of AI in the public sector requires an open dialog among developers, decision-makers, deployers, end-users, and the public. This dialog is essential for developing persuasive government strategies that guide the responsible development and deployment of AI technologies (Leikas et al., 2022).

Second, proper incentivization is needed so that the AI innovator is aligned with the illustrated (ework and adopts responsible practices and values in product development and commercialization. To this end, companies should be incentivized to invest in responsible AI as a priority, understanding that responsible AI not only mitigates risks but also propels innovation and competitive advantage, fostering a more robust and trustworthy AI ecosystem. This aligns with Responsible Research and Innovation practices, which are gaining traction across various research disciplines (Chen et al., 2022). In general, incentives are defined as a motivating force to incite action. An established categorization of incentives is grounded on the monetary aspect. Based on the Bartik's (1991) perspective on policy, direct incentives for the AI innovator can be either financial incentives (e.g., tax relief, industrial revenue bonds, and loans) or non-financial incentives (e.g., regulatory relief, training, prestige, or praise). Such incentives mostly pertain to the power that the regulator or direct stakeholders (e.g., creditors and suppliers) may exert on the AI innovator. Nonetheless, consumers can also pressure the innovators to conduct research and innovation in a responsible way as part of their Corporate Social Responsibility (Van de Poel et al., 2017); hence they are a crucial reference group for companies to better align their products and services to the market expectations (Gurzawska et al., 2017).

Third, it is important that regulatory bodies, policymakers, and democratic frameworks collaboratively strive to harmonize the functions of ChatGPT with the core tenets of human autonomy, agency, and privacy. This endeavor necessitates the establishment of universally recognized principles of trustworthy AI (Floridi, 2021; Shneiderman, 2020; Smuha, 2019; Wing, 2021). It involves not only crafting and enforcing more rigorous regulations for ethical AI systems but also establishing continuous oversight mechanisms. Monitoring and assessing the impact of AI instruments akin to ChatGPT is critical. This can be achieved through systematic policy impact evaluations (Stahl et al., 2023) and feedback from a wide array of stakeholders across various sectors such as education, healthcare, high technology, finance, and non-governmental organizations. Such comprehensive oversight could benefit from applying tools like UNESCO's Ethical Impact Assessment, which evaluates AI algorithms for alignment with human rights and transparency principles (UNESCO, 2023). A robust approach to AI governance should encompass the development of ex-ante and ex-post regulatory frameworks (Malgieri & Pasquale, 2022), ensuring that AI systems are fair and effective from inception to deployment. Moreover, the AI Act proposed by the European Union mandates impact assessments for high-risk AI systems, emphasizing the need for a preemptive and corrective oversight mechanism. The involvement of diverse stakeholders, including civil society and subject matter experts, in the policymaking process can enhance the democratic legitimacy of AI governance (Roberts et al., 2023). AI developers and procurers should also employ the Readiness Assessment Methodology (RAM) put forth by UNESCO, which assists governments in evaluating the readiness of their legal, policy, and institutional frameworks to address the risks associated with AI. This diagnostic tool can guide targeted capacity building, thereby stimulating the institutional and human capacities required to navigate the complexities of AI effectively.

Fourth, investing in research on the societal, economic, and ethical impact of ChatGPT, concerning transparency, fairness, diversity, inclusivity, and prevention of harm, is pivotal for industry and academic partners. It is also recommended that chatbots be co-created with the feedback of consumers or end-users in a participatory design process (e.g., Danieli et al., 2021) to ensure that the chatbot is secure, inclusive, and fair. Taking healthcare as an example, careful consideration should be given to potential biases that could arise in treatment recommendations or diagnostics from ChatGPT, which would significantly affect patient outcomes. For example, collecting feedback from healthcare professionals and patients can contribute to designing healthcare chatbots that align with their needs and adhere to privacy regulations (Hasal et al., 2021). Likewise, in academia, a critical assessment of the use of ChatGPT is crucial to uphold the integrity of academic standards (Perkins, 2023). For example, involving students and educators in the design of chatbots can help customize the learning material towards the promotion of equal access to educational resources.

Below, we illustrate the proposed multi-stakeholder Responsible Research and Innovation framework, which can be applied in the case of the ChatGPT chatbot and emerging AI tools. Inspired by Van de Poel (2020), this framework aims to ensure that the innovation process and its products satisfy the criteria of ethical acceptability, sustainability, and societal desirability, named Responsible Research and Innovation performance. The framework is inclusive of (a) the AI innovator, who is responsible for the development and commercialization of the AI tool and primarily for the Responsible Research and Innovation performance, (b) the regulator(s) of AI, and (c) the direct and indirect stakeholders. There is a direct influence exerted by the regulator on the innovator through policy and incentivization.

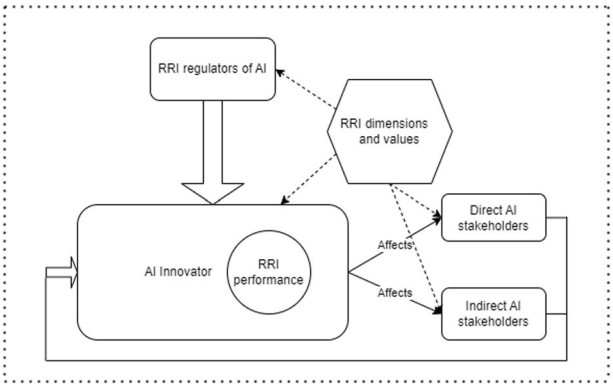

Fig. 2 Illustration of proposed multi-stakeholder responsible research and innovation framework to safeguard the ethical development and use of artificial intelligence.

Furthermore, through the Responsible Research and Innovation performance outputs, the innovator affects all direct and indirect stakeholders. These stakeholders, in turn, provide pressure (feedback, demand of AI technology, etc.) as input to the AI innovator, thus closing the loop. Notably, a prevalent set of Responsible Research and Innovation dimensions (anticipation, inclusiveness, reflexivity, responsiveness) and values (safety, privacy, sustainability, quality of life, gender equality, with transparency being the overarching value; Von Schomberg, 2011) are proposed to affect each one of the societal actors, and eventually the performance of the entire multi-stakeholder Responsible Research and Innovation ecosystem (see Fig. 2).

The proposed multi-stakeholder responsible research and innovation framework is persuasive since its theoretical grounds derive from established theories, specifically the stakeholder theory (Freeman, 2010) and the Responsible Research and Innovation principles, while tailored to the unique context of AI development. The framework is also aligned with recent literature on Innovation ecosystems (e.g., Stahl, 2022), focusing on the actors that constitute such ecosystems (Adner, 2017). Its utility is evident through its systematic incorporation of diverse societal dimensions, active feedback loops between stakeholders, and alignment with recognized Responsible Research and Innovation values, ensuring its holistic effectiveness in guiding ethical and sustainable AI innovation.

Discussion and implications

This focused review examines the AI revolution's dilemmas, addressing two key research questions. The first question investigates ChatGPT's potential hazards, detailing the adverse outcomes for individuals, businesses, and society, such as diminished human connections, job loss, privacy invasion, algorithmic bias, misinformation, and digital divide. These perils build upon existing research into AI's negative impact on business and the labor market (Chen, 2023; Rakowski et al., 2021). Our analysis concurs that AI's swift progress and widespread industrial impact necessitate an exhaustive evaluation to ensure alignment with ethical and sustainable objectives (Guo & Polak, 2021).

The second research question delves into the appropriateness of the Responsible Research and Innovation framework for developing and deploying chatbots like ChatGPT. Given the nascent stage of AI ethics, there is a need to devise frameworks for AI advancements to incorporate ethical considerations (Lo Piano, 2020). Our findings endorse a Responsible Research and Innovation framework involving the direct and indirect stakeholders of the AI innovation ecosystem to assess and ethically integrate ChatGPT's potential. This strategy highlights the need to anticipate potential ethical dilemmas and proactively address them by engaging with relevant parties early in the development cycle. By embracing Responsible Research and Innovation, the creation and application of chatbots may align with an ecosystem of stakeholders' values, covering safety, privacy, sustainability, life quality, and gender parity (Von Schomberg, 2011), ensuring ethical concerns are recognized and acted upon through anticipation, inclusivity, reflexivity, and responsiveness. Such alignment guarantees that the AI innovation ecosystems adhere to ethical standards, uphold human dignity and promote the common good, reducing risks and maximizing societal and individual benefits.

This focused review contributes significantly to the realms of AI, ethics, and Responsible Research and Innovation, being one of the first to apply such an approach to the usage and development of ChatGPT. It opens up avenues for future research, inviting scholars to empirically explore the ethical challenges of ChatGPT.

Our discussion proposes implications based on this framework and future pathways for the evolution of AI and Responsible Research and Innovation in academia, industry, and society. Practically, our insights offer guidelines for responsibly crafting AI solutions, urging developers to ensure transparency, reduce biases in AI, like ChatGPT, and conduct periodic evaluations (Arrieta et al., 2020). We advocate for active stakeholder engagement and partnerships for a more inclusive AI progression.

The proposed framework can serve as a guide for the responsible usage of AI in specific fields, such as education, which is poised to undergo significant changes due to ChatGPT (Dwivedi et al., 2023). Addressing ethical concerns in educational AI technologies is imperative (Ahmad et al., 2023), suggesting the integration of regulatory templates and institutional policies to overcome the limitations of ChatGPT. A collaborative effort among AI creators, regulators, educators, and students will ensure an educational experience that keeps pace with technological advances.

On a societal scale, our study unveiled the pivotal role of technology and the necessity for AI to progress in a favorable direction. Public policymakers need to craft regulations and strategies that ensure the ethical and advantageous deployment of AI (Stahl et al., 2022b). We call for a collective endeavor to leverage the benefits of AI while safeguarding against its risks, encouraging innovators to raise awareness among policymakers and the public about the technology's functioning and its potential adverse effects (Hedlund & Persson, 2022).

Moreover, our analysis suggests the necessity for robust AI governance to maintain the responsible efficacy of AI innovation ecosystems amidst the growing complexity of new AI technologies. There is a call for a unified understanding of AI governance, which varies across academic and policy contexts (Ulnicane et al., 2021, p. 78; Zuiderwijk et al., 2021). Gahnberg (2021) offers a technologically centered definition of AI governance that could simplify the emerging complexities of AI by focusing on general elements such as performance, environment, effects, and sensors. This method could provide unique governance structures for varying AI technological aspects (Sigfrids et al., 2022). Furthermore, as there is a pressing need for policies that foster diversity, equity, and inclusion in generative AI to avoid societal gaps, AI governance measures could include financial support or initiatives for public access. Additionally, policy mechanisms such as government procurement standards and funding can motivate development in directions that serve the public interest. The influence of ChatGPT and similar technologies transcend local boundaries, with the potential for risks and benefits to be felt worldwide. As a result, policy measures must extend beyond local and national confines to form effective and reliable frameworks at the international level (Wallach & Marchant, 2018).

Future research should use the 'responsible research and innovation keys' to examine the geographical and disciplinary interpretations of Responsible Research and Innovation, such as the differences between the EU's and the UK's AREA approaches (Owen et al., 2021). Also, we emphasize the need for empirical studies in the greater area of Responsible Research and Innovation ecosystems to unveil further the underlying mechanisms (e.g., co-evolution and mutual learning of actors; Stahl et al., 2022a; Weiss & Spiel, 2022) of the responsible development of emerging AI technologies.

Conclusion

The swift progression of AI has been at the center of debates concerning its impacts, advantages, risks, and societal consequences (Diaz-Rodriguez et al., 2023). It was inevitable that the release of ChatGPT at the end of 2022 would fuel this ongoing debate and elicit diverse perspectives.

Is, eventually, ChatGPT a good bot or a bad bot? This question is a rhetorical one. It is crucial to remember that ChatGPT, being a human creation, has the potential to be employed for both positive and negative purposes, thus we need to ensure that it is utilized responsibly. As Hauer (2022, p. 7) mentions, "we need to believe this and not lose optimism".

This study attempts to convey this optimism by proposing a multi-stakeholder and responsible approach to using and developing ChatGPT and AI tools in general under the lens of stakeholder theory and Responsible Research and Innovation. The results of this focused literature review provide a comprehensive framework and highlight the need for a collaborative approach between innovators, direct and indirect stakeholders, and other societal actors. Specifically, AI innovators, regulators, academics, practitioners, and the wider public should engage in continuous dialogs to ensure that ChatGPT aligns with societal values and ethical standards. Such an approach would ensure that ChatGPT is developed with a broader understanding of its ethical implications and societal impact based on anticipation, inclusivity, reflexivity, and responsiveness, aligning with the values of the multi-stakeholder ecosystem. It is essential to understand that a responsible continuation of such AI tools does not only rest with the innovators but also with all those who benefit from their use.

The integration of values and ethics within AI systems such as ChatGPT is becoming increasingly critical given our growing dependence on such technologies in societal and economic contexts (Kaplan & Haenlein, 2020). This article is a reference for practitioners and policymakers, mapping the pitfalls of irresponsible AI development and delineating actionable guidelines for ethical AI creation. It also carves out vital paths for further research in Responsible Research and Innovation, a research area that is ripe for empirical inquiry. Future studies are encouraged to shed light on the efficacy of diverse stakeholder engagement and to assess the role of varied actors—industry specialists, end-users, and policymakers—in shaping ChatGPT. Moreover, evaluating the impact of regulatory measures on promoting responsible innovation could provide insights into how such frameworks ensure equity and clarity in the evolution of nascent AI systems.

Data availability

This is a review paper wherein the authors have analyzed various articles by different authors who have been cited at required places. The cited information belongs to the authors mentioned in the reference section. The generated data is included in the paper.

Received: 23 May 2023; Accepted: 27 November 2023;
Published online: 15 January 2024

References

Adner R (2017) Ecosystem as structure: an actionable construct for strategy. J Manag 43(1):39–58
Agboola AO (2015) Neoclassical economics and new institutional economics: An assessment of their methodological implication for property market analysis. Prop Manag 33(5):412–429
Ahmad SF, Han H, Alam MM, Rehmat M, Irshad M, Arraño-Muñoz M, Ariza-Montes A (2023) Impact of artificial intelligence on human loss in decision making, laziness and safety in education. Humanit Soc Sci Commun 10(1):1–14
Atkins S, Badrie I, van Otterloo S (2021) Applying ethical AI frameworks in practice: Evaluating conversational AI chatbot solutions. Comput Soc Res J 1:1–6
Alderman L, Towers S, Bannah S (2012) Student feedback systems in higher education: A focused literature review and environmental scan. Qual High Educ 18(3):261–280
Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MAA, Dwivedi YK (2023) A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. J Innov Knowl 8(1):100333
Anshari M, Hamdan M, Ahmad N, Ali E, Haidi H (2023) COVID-19, artificial intelligence, ethical challenges and policy implications. AI Soc 38(2):707–720
Araujo T, Helberger N, Kruikemeier S, De Vreese CH (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 35(3):611–623
Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115
Baba Y, Walsh JP (2010) Embeddedness, social epistemology and breakthrough innovation: the case of the development of statins. Res Policy 39(4):511–522
Bartik TJ (1991) Who benefits from state and local economic development policies? W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan
Bell E (2023) "A fake news frenzy: why ChatGPT could be disastrous for truth in journalism." Guardian. Retrieved on 1/10/2023 from: https://www.theguardian.com/commentisfree/2023/mar/03/fake-news-chatgpt-truth-journalism-disinformation
Bessen JE (2019) AI and jobs: the role of demand. NBER Working Paper No. 24235
Boenink M, Kudina O (2020) Values in responsible research and innovation: from entities to practices. J Responsible Innov 7(3):450–470
Bostrom N, Yudkowsky E (2018) The ethics of artificial intelligence. In: Artificial intelligence safety and security. Chapman and Hall/CRC. pp. 57–69
Bostrom N (2020) Ethical issues in advanced artificial intelligence. In: Machine Ethics and Robot Ethics. Routledge. pp. 69–75
Bozeman B, Rimes H, Youtie J (2015) The evolving state-of-the-art in technology transfer research: Revisiting the contingent effectiveness model. Res Policy 44(1):34–49
Brynjolfsson, E, & McAfee, A (2014) *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company
Burget M, Bardone E, Pedaste M (2017) Definitions and conceptual dimensions of responsible research and innovation: a literature review. Sci Eng Ethics 23:1–19
Cellar-Jones R (2016) Stephen Hawking—will AI kill or save humankind? Retrieved on 15/4/2023 from http://www.bbc.com/news/technology-37713629
ChatGPT, Zhavoronkov A (2022) Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. Oncoscience 9:82–84
Chen J, Nichele E, Ellerby Z, Wagner C (2022) Responsible research and innovation in practice: driving both the 'How' and the 'What' to research. J Responsible Technol 11:100042
Chen Z (2023) Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humani Soc Sci Commun 10(1):1–12
Chilton J (2023) The new risks ChatGPT poses to cybersecurity. Harvard Bus Rev. Retrieved on 10/5/2023 from: https://hbr.org/2023/04/the-new-risks-chatgpt-poses-to-cybersecurity
Cowan DT, Norman I, Coopamah VP (2005) Competence in nursing practice: a controversial concept–a focused review of literature. Nurse Educ today 25(5):355–362
Danieli M, Ciulli T, Mousavi SM, Riccardi G (2021) A conversational artificial intelligence agent for a mental health care app: evaluation study of its participatory design. JMIR Formative Res 5(12):e30053
Diaz-Rodriguez N, Del Ser J, Coeckelbergh M, de Prado ML, Herrera-Viedma E, Herrera F (2023) Connecting the dots in trustworthy Artificial Intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation. Inf Fusion 99:101896

Dignum V (2018) Ethics in artificial intelligence: introduction to the special issue. Ethics Inf Technol 20(1):1–3
Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Williams MD (2021) Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manag 57:101994
Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Wright R (2023) "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag 71:102642
UNESCO (2023) Ethical Impact Assessment: a tool of the recommendation on the ethics of artificial intelligence. https://www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence
Etzioni A, Etzioni O (2017) Incorporating ethics into artificial intelligence. J Ethics 21:403–418
European Commission (2013) Directorate-General for Research and Innovation. Options for strengthening responsible research and innovation: report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation. Publications Office. Retrieved on 4/4/2023 from https://data.europa.eu/doi/10.2777/46253
Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication
Forsberg EM, Quaglio G, O'Kane H, Karapiperis T, Van Woensel L, Arnaldi S (2015) Assessment of science and technologies: advising for and with responsibility. Technol Soc 42:21–27
Floridi L (2021) Establishing the rules for building trustworthy AI. In: Ethics, Governance, and Policies in Artificial Intelligence. Philosophical Studies Series, vol 144. Springer, Cham. pp. 41–45
Fraaije A, Flipse SM (2020) Synthesizing an implementation framework for responsible research and innovation. J Responsib Innov 7(1):113–137
Freeman RE (2010) Strategic management: a stakeholder approach. Cambridge University Press
Gahnberg C (2021) What rules? Framing the governance of artificial agency. Policy Soc 40(2):194–210
Gasser U, Almeida VA (2017) A layered model for AI governance. IEEE Intern Comput 21(6):58–62
Gianni R (2016) Responsibility and freedom: the ethical realm of RRI. John Wiley & Sons
Guo H, Polak P (2021) Artificial intelligence and financial technology FinTech: how AI is being used under the pandemic in 2020. In: The fourth industrial revolution: implementation of artificial intelligence for growing business success. Studies in Computational Intelligence. Springer, Cham. pp. 169–186
Gurzawska A, Mäkinen M, Brey P (2017) Implementation of responsible research and innovation (RRI) practices in industry: providing the right incentives. Sustainability 9(10):1759
Hagendorff T (2020) The ethics of AI ethics: an evaluation of guidelines. Minds Mach 30(1):99–120
Hasal M, Nowaková J, Ahmed Saghair K, Abdulla H, Snášel V, Ogiela L (2021) Chatbots: Security, privacy, data protection, and social aspects. Concurr Comput Pract Exp 33(19):e6426
Hauer T (2022) Importance and limitations of AI ethics in contemporary society. Humanit Soc Sci Commun 9(1):1–8
Hedlund M, Persson E (2022) Expert responsibility in AI development. AI Soc 1–12. https://link.springer.com/article/10.1007/s00146-022-01498-9
Hsu, T, & Thompson, SA (2023) Disinformation Researchers Raise Alarms About A.I. Chatbots. *The New York Times*. Retrieved on 4/4/2023 from. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html
Kaplan A, Haenlein M (2020) Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Bus Horiz 63(1):37–50
Krügel S, Ostermaier A, Uhl M (2023) ChatGPT's inconsistent moral advice influences users' judgment. Sci Rep. 13(1):4569
Leikas J, Johri A, Latvanen M, Wessberg N, Hahto A (2022) Governing ethical AI transformation: a case study of AuroraAI. Front Artif Intell 5:13
Lo Piano S (2020) Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. Humanit Soc Sci Commun 7(1):1–7
Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: machines vs. humans: the impact of artificial intelligence chatbot disclosure on customer purchases. Mark Sci 38(6):937–947
Luo F, Li RYM, Crabbe MJC, Pu R (2022) Economic development and construction safety research: a bibliometrics approach. Saf Sci 145:105519
Malgieri G, Pasquale FA (2022) From transparency to justification: toward ex ante accountability for AI. Brooklyn Law School, Legal Studies Paper, (712). Available at SSRN: https://ssrn.com/abstract=4099657
McGee RW (2023) Is chat GPT biased against conservatives? an empirical study. SSRN Electron
McNamara A, Smith J, Murphy-Hill E (2018) Does ACM's code of ethics change ethical decision making in software development?" In: Leavens GT, Garcia A, Păsăreanu CS (eds.) Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering—ESEC/FSE 2018. ACM Press, New York
Mijwil M, Aljanabi M (2023) Towards artificial intelligence-based cybersecurity: the practices and ChatGPT generated ways to combat cybercrime. Iraqi J Comput Sci Math 4(1):65–70
Mikalef P, Conboy K, Lundström JE, Popovič A (2022) Thinking responsibly about responsible AI and 'the dark side' of AI. Eur J Inf Syst 31(3):257–268
Mitchell RK, Agle BR, Wood DJ (1997) Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Acad Manag Rev 22(4):853–886
Müller VC, Bostrom N (2016) Future progress in artificial intelligence: a survey of expert opinion. In: Fundamental Issues of Artificial Intelligence. Synthese Library, vol 376. Springer, Cham. pp. 555–572
Nielsen MV (2016) The concept of responsiveness in the governance of research and innovation. Sci Public Policy 43(6):831–839
Niu B, Mvondo GFN (2024) I Am ChatGPT, the ultimate AI Chatbot! Investigating the determinants of users' loyalty and ethical usage concerns of ChatGPT. J Retail Consum Serv 76:103562
O'Connor S, ChatGpt (2023) Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? Nurse Educ Pract 66:103537
Owen R, Stilgoe J, Macnaghten P, Gorman M, Fisher E, Guston D (2013) A framework for responsible innovation. In: Owen R, Bessant J, Heintz M (eds.). Responsible innovation: Managing the responsible emergence of science and innovation in society. Wiley, Sussex, pp. 27–50
Owen R, Pansera M, Macnaghten P, Randles S (2021) Organisational institutionalisation of responsible innovation. Res Policy 50(1):104132
Owen R, Von Schomberg R, Macnaghten P (2021) An unfinished journey? Reflections on a decade of responsible research and innovation. J Responsible Innov 8(2):217–233
Parviainen J, Rantala J (2022) Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 25(1):61–71
Patel SB, Lam K, Liebrenz M (2023) ChatGPT: friend or foe. Lancet Digit Health 5:e102
Paul J, Ueno A, Dennis C (2023) ChatGPT and consumers: benefits, pitfalls and future research agenda. Int J Consum Stud 47(4):1213–1225
Perkins M (2023) Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. J Univ Teach Learn Pract 20(2):07
Polak P (2021) Welcome to the digital era—the impact of AI on business and society. Society 58(3):177–178
Polyportis A (2020) Incidental emotions and hedonic forecasting: the role of the certainty-uncertainty appraisal dimension. PhD thesis. Athens University of Economics and Business, Greece. Retrieved on 22/4/2023 from https://www.didaktorika.gr/eadd/handle/10442/47930
Polyportis A, Kokkinaki F, Horváth C, Christopoulos G (2020) Incidental emotions and hedonic forecasting: the role of (un) certainty. Front Psychol 11:536376
Puntoni S, Reczek RW, Giesler M, Botti S (2021) Consumers and artificial intelligence: an experiential perspective. J Mark 85(1):131–151
Rajaobelina L, Prom Tep S, Arcand M, Ricard L (2021) Creepiness: its antecedents and impact on loyalty when interacting with a chatbot. Psychol Mark 38(12):2339–2356
Rakowski R, Polak P, Kowalikova P (2021) Ethical aspects of the impact of AI: the status of humans in the era of artificial intelligence. Society 58(3):196–203
Rivas P, Zhao L (2023) Marketing with ChatGPT: navigating the ethical terrain of GPT-based Chatbot technology. AI 4(2):375–384
Roberts H, Hine E, Taddeo M, Floridi L (2023) Global AI governance: barriers and pathways forward. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4588040
Rozado D (2023) The political biases of ChatGPT. Soc Sci 12(3):148
Sallam M (2023) ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare 11(6):887
Schuijff M, Dijkstra AM (2020) Practices of responsible research and innovation: a review. Sci Eng Ethics 26(2):533–574
Shaban-Nejad A, Michalowski M, Brownstein JS, Buckeridge DL (2021) Guest editorial explainable AI: towards fairness, accountability, transparency and trust in healthcare. IEEE J Biomed Health Inform 25(7):2374–2375
Shneiderman B (2020) Human-centered artificial intelligence: reliable, safe & trustworthy. Int J Hum–Comput Interact 36(6):495–504
Sigfrids A, Nieminen M, Leikas J, Pikkuaho P (2022) How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI. Front Hum Dyn 4:858108
Smuha NA (2019) The EU approach to ethics guidelines for trustworthy artificial intelligence. Comput Law Rev Int 20(4):97–106
Stahl BC, Obach M, Yaghmaei E, Ikonen V, Chatfield K, Brem A (2017) The responsible research and innovation (RRI) maturity model: linking theory and practice. Sustainability 9(6):1036

Stahl BC (2022) Responsible innovation ecosystems: ethical implications of the application of the ecosystem concept to artificial intelligence. Int J Inf Manag 62:102441
Stahl BC, Antoniou J, Ryan M, Macnish K, Jiya T (2022a) Organisational responses to the ethical issues of artificial intelligence. AI Soc 37(1):23–37
Stahl BC, Rodrigues R, Santiago N, Macnish K (2022b) A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. Comput Law Security Rev 45:105661
Stahl BC, Antoniou J, Bhalla N, Brooks L, Jansen P, Lindqvist B, … Wright D (2023) A systematic review of artificial intelligence impact assessments. Artif Intell Rev 56:12799–12831
Stahl BC, Eke D (2024) The ethics of ChatGPT–exploring the ethical issues of an emerging technology. Int J Inf Manag 74:102700
Stilgoe J, Owen R, Macnaghten P (2013) Developing a framework for responsible innovation. Res Policy 42(9):1568–1580
Stilgoe J, Guston D (2017) Responsible research and innovation. In Felt U, Fouché R, Miller CA, Smith-Doerr L (eds) The handbook of science and technology studies. The MIT Press. pp. 853–880
Stilgoe J, Owen R, Macnaghten P (2020). Developing a framework for responsible innovation. In: The ethics of nanotechnology, geoengineering, and clean energy. Routledge. pp. 347–359
Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. Nature 613(7945):620–621
Strowel A (2023) ChatGPT and generative AI tools: theft of intellectual labor? IIC Int Rev Intellect Prop Compet Law 54(4):491–494
Srivastava B, Rossi F (2019) Rating AI systems for bias to promote trustable applications. IBM J Res Dev 63(4/5):1–9
Sutcliffe H (2011) A report on responsible research and innovation. MATTER and the European Commission
Taylor L, Watkins SL, Marshall H, Dascombe BJ, Foster J (2016) The impact of different environmental conditions on cognitive function: a focused review. Front Physiol 6:372
Thorp HH (2023) ChatGPT is fun, but not an author. Science 379(6630):313–313
Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, Agyemang B (2023) What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learn Environ 10(1):15
Ulnicane I, Eke DO, Knight W, Ogoh G, Stahl BC (2021) Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. Interdiscip Sci Rev 46(1-2):71–93
United Nations (2021) 193 countries adopt first-ever global agreement on the ethics of artificial intelligence. UN News. https://news.un.org/en/story/2021/11/1106612#:~:text=AI%20as%20a%20positive%20contribution,and%20privacy%2C%20with%20action
Van de Poel I (2020) RRI measurement and assessment: Some pitfalls and a proposed way forward. In: Assessment of responsible innovation. Routledge. pp. 339–360
Van de Poel I, Asveld L, Flipse S, Klaassen P, Scholten V, Yaghmaei E (2017) Company strategies for responsible research and innovation (RRI): a conceptual model. Sustainability 9(11):2045
Verma P, Oremus W (2023) "ChatGPT invented a sexual harassment scandal and named a real law prof as the accused." Washington Post. Retrieved on 30/9/2023 from: https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/
Vilaza GN, McCashin D (2021) Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. Front Digit Health 3:689736
Vinge V (2008) Signs of the singularity. IEEE Spectr 45(6):76–82
von Garrel J, Mayer J (2023) Artificial intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. Humanit Soc Sci Commun 10:799
Von Schomberg R (2011) Towards responsible research and innovation in the information and communication technologies and security technologies fields. A report from the European Commission Services. Publication Office of the European Union, Luxembourg
Wallach W, Marchant GE (2018) An agile ethical/legal model for the international and national governance of AI and robotics. Association for the Advancement of Artificial Intelligence
Weiss A, Spiel K (2022) Robots beyond science fiction: mutual learning in human–robot interaction on the way to participatory approaches. AI Soc 37(2):501–515
Wiarda M, van de Kaa G, Doorn N, Yaghmaei E (2022) Responsible innovation and de jure standardisation: an in-depth exploration of moral motives, barriers, and facilitators. Sci Eng Ethics 28(6):65
Wilson C, Van Der Velden M (2022) Sustainable AI: an integrated model to guide public sector decision-making. Technol Soc 68:101926
Wing JM (2021) Trustworthy AI. Commun ACM 64(10):64–71
Yigitcanlar T (2021) Greening the artificial intelligence for a sustainable planet: an editorial commentary. Sustainability 13(24):13508
Yigitcanlar T, Li RYM, Beeramoole PB, Paz A (2023) Artificial intelligence in local government services: Public perceptions from Australia and Hong Kong. Gov Inf Q 40(3):101833
Yuste R, Goering S, Arcas BAY, Bi G, Carmena JM, Carter A, Wolpaw J (2017) Four ethical priorities for neurotechnologies and AI. Nature 551(7679):159–163
Zuiderwijk A, Chen YC, Salem F (2021) Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda. Gov Inf Q 38(3):101577

Acknowledgements

This research received no external funding.

Author contributions

AP: conceptualization, investigation, methodology, visualization, project administration, writing—original draft, writing—review and editing. NP: investigation, methodology, resources, writing—original draft, writing—review and editing.

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Athanasios Polyportis.