

Opinion paper

## The ethics of ChatGPT – Exploring the ethical issues of an emerging technology

Bernd Carsten Stahl<sup>a,b,\*</sup>, Damian Eke<sup>b</sup>

<sup>a</sup> School of Computer Science, University of Nottingham, UK

<sup>b</sup> Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK



### ARTICLE INFO

#### Keywords:

ChatGPT

Ethics

Emerging technology

Generative AI systems

### ABSTRACT

This article explores ethical issues raised by generative conversational AI systems like ChatGPT. It applies established approaches for analysing ethics of emerging technologies to undertake a systematic review of possible benefits and concerns. The methodology combines ethical issues identified by Anticipatory Technology Ethics, Ethical Impact Assessment, and Ethical Issues of Emerging ICT Applications with AI-specific issues from the literature. These are applied to analyse ChatGPT's capabilities to produce humanlike text and interact seamlessly. The analysis finds ChatGPT could provide high-level societal and ethical benefits. However, it also raises significant ethical concerns across social justice, individual autonomy, cultural identity, and environmental issues. Key high-impact concerns include responsibility, inclusion, social cohesion, autonomy, safety, bias, accountability, and environmental impacts. While the current discourse focuses narrowly on specific issues such as authorship, this analysis systematically uncovers a broader, more balanced range of ethical issues worthy of attention. Findings are consistent with emerging research and industry priorities on ethics of generative AI. Implications include the need for diverse stakeholder engagement, considering benefits and risks holistically when developing applications, and multi-level policy interventions to promote positive outcomes. Overall, the analysis demonstrates that applying established ethics of technology methodologies can produce a rigorous, comprehensive foundation to guide discourse and action around impactful emerging technologies like ChatGPT. The paper advocates sustaining this broad, balanced ethics perspective as use cases unfold to realize benefits while addressing ethical downsides.

### 1. Introduction

ChatGPT, the web-based interactive system based on OpenAI's GPT large language models is expected to have transformative impact on many aspects of society. It raises huge expectations, has attracted significant investment and causes wide-spread concerns. There is much discussion in the media and a rapidly growing academic discussion on its benefits and ethical downsides. This existing discourse focuses on some high-profile topics, such as academic authorship, but currently lacks a systematic approach.

As we show in Section 2, the current discussion of ethics of ChatGPT remains one-sided, not only in its focus on specific issues but also in a lack of balance of considering ethical benefits and concerns. Rebalancing the debate is not just of academic interest but is called for in practical terms as well. Organisations making use of such technologies need reliable guidance and policymakers providing this guidance need to

develop it on the basis of a broad-based and academically reliable foundation. This article proposes such a foundation that builds on well-established research and helps overcome the ad-hoc nature of current work on ethics of ChatGPT.

The objective of this article is to provide a systematic analysis and rigorous review of the ethics of ChatGPT and, by implication, of related technologies based on large language models. This is achieved by applying established research methods that were developed to investigate ethical issues of emerging technologies. Drawing on three such methods as well as insights from the ethics of AI debate, we explore ethical benefits and concerns that likely applications of ChatGPT using a list or 93 partly overlapping possible issues identified from prior literature.

The paper is structured as follows. The next section will review the theoretical background, introducing the ethics in technology, outlining ChatGPT, its key features and their ethical implications, as well as the

\* Corresponding author at: School of Computer Science, University of Nottingham, UK.

E-mail address: [Bernd.stahl@nottingham.ac.uk](mailto:Bernd.stahl@nottingham.ac.uk) (B.C. Stahl).

<https://doi.org/10.1016/j.ijinfomgt.2023.102700>

Received 23 April 2023; Received in revised form 29 August 2023; Accepted 29 August 2023

Available online 9 September 2023

0268-4012/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ethics of emerging technologies and the ethics of AI. This background provides the basis for the methodology employed to analyse the ethics of ChatGPT. The subsequent sections then discuss the structured ethical analysis of ChatGPT, relevance of findings, and which open questions emerge from them. The article then spells out likely implications for research and practice.

## 2. ChatGPT and the ethics of emerging digital technologies

This article aims to apply ideas developed in the context of ethics-related research on emerging digital technologies to ChatGPT. It is therefore important to first understand the theoretical background of ethics that underpins ethics-related research. This section then discusses which benefits ChatGPT promises and which concerns it raises (Dwivedi, Kshetri et al., 2023) before introducing approaches to the ethics of emerging technologies.

### 2.1. Theoretical background

The current discussion of ChatGPT covers numerous technical aspects as well as current and likely applications. One crucial part of the debate relates to ethical concerns that many contributors raise about the use of ChatGPT. This article focuses on such ethical questions. Before we can explore them in more detail, it is important to provide the conceptual and theoretical basis. The concept of ethics as used in the English language has a number of related but non-identical meanings (Stahl, 2012). In general terms it refers to questions involving moral concepts such as right and wrong, or good and bad. In most familiar contexts humans intuitively know what is right and wrong, a knowledge of ethics that is acquired during human socialisation. Sometimes this intuition fails or conflicts with others' intuitions, at which point ethics appears as a set of explicit statements in the form of "you should never / always do X". Where such statements are not accepted and need justification, ethics as a set of philosophical theories is called for. Ethics can be descriptive or normative, abstract or applied.

Ethics as a philosophical discipline that is mostly concerned with examining or justifying prescriptive rules and statements has a written history of several millennia. In the European context the roots of philosophical ethics are typically traced back to Greek Antiquity. There are numerous ethical theories that form the canon of philosophical ethics. One that has its roots in Aristotle (2007) is that of virtue ethics which evaluates the ethical quality of an action or situation with reference to the character of the actor. Other high-profile ethical theories include deontology which judges the quality of an action by examining the intention of the agent undertaking it. This type of ethical reasoning is closely associated with the German philosopher Immanuel Kant (1788, 1797). For Kant an action can only count as ethically good if it is motivated by duty, which means that it must come from a purely rational insight into what should be done which Kant formulates as the so-called Categorical Imperative. A very different ethical theory is that of teleology, which uses the outcomes of an action to evaluate its ethical quality. This type of reasoning, also known as consequentialism is associated with thinkers such as Bentham (1789) or Mill (1861). They suggested that a suitable measure of ethics would be utility, which is why this type of thinking is also known as utilitarianism. One measure of utility is happiness, so that a well-known expression of this ethical position is that an action is ethically good, if it leads to the greatest happiness for the greatest number of people.

While it is important to realise that there are extensive theoretical underpinnings to ethics as a philosophical discipline, it is worth noting that there are focus areas within ethics that make use of specific aspects of this breadth of theory. Ethics of technology is typically seen as one branch of applied ethics (Nijsingh & Duwell, 2009). Applied ethics aims to understand ethical questions arising in specific fields such as medicine (Freyhofer, 2004), business (Bowie, 1999; De George, 1999), or research (OECD, 2016). It often focuses on specific sub-fields or problems and

applications. One field of applied ethics that has developed since that 1970 s is that of computer ethics (Bynum, 2001; Moor, 1985) which has strongly influenced the development of the ethics of AI (Stahl, 2021). While computer ethics is one root of the ethics of AI, there are other streams of ethical considerations that are linked to some of the features of AI. One example of this stream of work can be described using the acronym FATE which stands for fairness, accountability, transparency and ethics (Memarian & Doleck, 2023) which mirrors the subject matter of the ACM conferences series on Fairness, Accountability and Transparency (FAcCT).

For the purposes of this article that explores ethical concerns arising from ChatGPT, it is important to start with a broad understanding of ethical question that includes but goes beyond the current focus of FATE or FAcCT. ChatGPT as an example of a large language model and thus of AI may well raise questions around fairness, accountability and transparency, but its ethical concerns are likely to go beyond these established topics. The paper is therefore theoretically grounded in the broader computer ethics literature and builds more specifically on the theoretical approach to the ethics of emerging technologies as explained in detail in the methodology section below. This theoretical angle of the paper allows for broader insights than the prevailing literature on AI and ethics and therefore offers a theoretical justification for the research question. If ChatGPT turns out to be as impactful as is currently widely predicted, then we need theoretically sound and empirically rigorous and transparent research to proactively engage with its likely ethical and social consequences. The theoretical and methodological approach reported in this article are geared to meet these requirements.

In order to follow this argument, it is important to understand which aspects of ChatGPT are capable or giving rise to ethical questions, what those questions are and how they are currently addressed.

### 2.2. Literature on ChatGPT and ethics

ChatGPT is an interactive system that allows users to have conversations using natural language. According to OpenAI, the organisation that developed it, the dialogue format employed by ChatGPT allows it to answer follow-up questions, admit mistakes, challenge incorrect premises and reject inappropriate requests (OpenAI, 2022). Technically, ChatGPT is based on of OpenAI's Generative Pre-trained Transformer language (GPT) models (particularly GPT-3 and GPT-4), a large-scale neural network-based language generation model trained on a wide range of internet-based texts. It was trained using reinforcement learning from human feedback. We refer to ChatGPT as a 'technology' in the sense defined by the Encyclopaedia Britannica (2023) as "the application of scientific knowledge to the practical aims of human life." While the article focuses on ChatGPT, its underlying interest is broader than this particular example of a user-oriented technology. Competition for ChatGPT is growing rapidly, e.g. in the form of Google's Bard (Metz & Grant, 2023), Alibaba's Tongyi Qianwen (McMorrow & Liu, 2023) and many others. We use ChatGPT as the most high-profile example of an interactive large language model because of its prominence in the discourse. Our analysis is based on key features of ChatGPT and is thus applicable to competitor technologies insofar as these share the same characteristics.

Generative AI systems or Chatbots like ChatGPT are not a new phenomenon. The idea that it should be possible to interact seamlessly with a computer using natural language goes back to the beginnings of computing and AI research. The most prominent early example of such a chatbot may be ELIZA, a conversational program developed by Joseph Weizenbaum (1977) in the 1960 s. Early chatbots like ELIZA had limited functionality and adoption. With the progress of natural language processing, better hardware and connectivity, chatbots have become more prominent. Most of the large tech companies have been trying to integrate them into their service offerings. Examples include Apple's Siri, Amazon's Alexa or Microsoft Cortana.

The widespread existence, use and adoption of chatbot technologies

raises the question why ChatGPT has managed to gain the level of attention that it has been receiving since its launch in November 2022. The answer appears to be in the high quality of its outputs. ChatGPT can interact across a broad range of topics, providing answers of high quality of language and good accuracy of content. The underlying technology, the large amount of training data from the internet, and the model architecture appear to have allowed ChatGPT to pass a previously invisible threshold. Interaction with ChatGPT gives the impression that one has a conversation with an educated human who has subject expertise across a large number of subject areas.

The question why ChatGPT has risen to prominence is intimately linked to the question of its social and ethical evaluation. Such an evaluation draws at least partly on real or expected practical consequences of technology use which, in turn, are driven by the technology's capability. We believe that the following characteristics and capabilities of ChatGPT may influence its use and applications and contribute to possible ethical concerns:

- Production of high-quality text in response to human input that is often difficult to identify as the output of an AI (Zhang et al., 2023);
- Ability to engage in a dialogical interaction on a very broad array of topics (Gilson et al., 2023);
- Ability to tailor its output to specific language styles (Short & Short, 2023);
- While currently text based, we assume that it can be integrated into other modalities of communication which would, for example, enable it to engage in voice communication (Ali et al., 2023);
- Ability to learn from interaction leading to allow it to further improve the content quality and acceptability (Sallam, 2023);
- Based on a large language model trained on large but limited datasets (Radford et al., 2019).

One consequence of these characteristics is that interventions from ChatGPT may be difficult to distinguish from original human interventions. Where prior instantiations of chatbots could easily be identified as such, this appears to be more difficult for ChatGPT. This ability to interact with technology without a simple way of ascertaining that it is a technology may revolutionise the way we interact with technology. We discuss the ethical issues that arise from it below, but we concede that it may lead to an acceleration of research and technical progress which can quickly give rise to new capabilities which can change the ethical evaluation again.

This brings us to the question of which specific ethical benefits or concerns ChatGPT are currently discussed. These are typically linked to possible applications. At a basic level, ChatGPT produces texts. It can thus potentially be applied across a vast range of applications that involve the production of text. An ethical analysis of any novel technology is typically well advised to look at benefits as well as concerns. At present there is relatively little discussion of ethical benefits of ChatGPT, such as the potential benefits of using ChatGPT in promoting teaching and learning Baidoo-Anu and Owusu Ansah (2023). There is, however, a wide-spread perception in the media that it is likely to be disruptive and transformational, which in terms of technology discourses typically means that it will lead to economic benefits, be these in the form of start-up companies, optimisation of existing industries or accrual of further growth in the tech sector. Such economic growth can provide improved wellbeing for many people and thus can count as a moral benefit.

Key to an ethical analysis of ChatGPT is an understanding of likely applications. One expectation is that ChatGPT will lead to fundamental changes in the internet search sector. Google is currently the entrenched market leader in this sector. ChatGPT may change the way people search, leading to more competition in this economically highly lucrative sector (Grant, 2023; Metz & Grant, 2023). Microsoft has already announced the launch of a new and improved search engine *Bing* that will run on ChatGPT and GPT-3.5 that they claim to be faster, more

accurate and more capable (Mehdi, 2023). In more general terms, the benefits of ChatGPT are sometimes described as providing additional intelligence to improve operations. One way of capturing this is the metaphor of AI interns (Nast, 2023a), suggesting that it will be easier to draw upon the intelligent but non-expert support that human interns can offer to organisations. This sounds rather incremental but given the potential ubiquity of the infinite number of AI interns, the cumulative effect may still be transformational. This may be one reason why Sam Altman, the CEO of Open AI can claim that he can envisage ChatGPT to "break capitalism" (Bove, 2023). An ethical evaluation of such a fundamental change of the socio-economic environment would be far from straightforward but it is easy to see how it would be beneficial for some.

While the moral benefits of ChatGPT thus remain somewhat fuzzy, there are more clearly defined concerns. The most prominent set of those has to do with the fact that ChatGPT texts can be difficult to distinguish from human output which can lead to problems of attribution of authorship. This is seen as a significant problem for student assessment, where such assessment is based on essays and the fear is that students can gain unfair advantages (Eke, 2023; Stokel-Walker, 2022; Weale, 2023). Much work is undertaken to explore concerns about the role of ChatGPT in research where it is seen as a threat to transparency in science (Nature editorial, 2023). Some scientists have attempted to pre-empt this issue by adding ChatGPT as an author to publications, but the practice does not seem to find acceptance in the academic community (Stokel-Walker, 2023).

In addition to these concerns about authorship and attribution, there are various other concerns that are discussed. One of these is the possible impact on employment (Frederick, 2023), in particular with regards to jobs that involve producing output that have close affinity to digital technologies, such as computer programming (Castelvecchi, 2022). As the technology becomes more advanced, there is a concern that it could have a significant impact on the jobs human programmers do (Marr, 2023a).

The fact that the origin of a text is difficult to discern may have further morally problematic consequences. One of these is that it may exacerbate the widely discussed problem of misinformation and disinformation (Hsu & Thompson, 2023). Another well-established ethical issue of AI, namely the possibility of using it for unwarranted political intervention may be exacerbated by ChatGPT as highlighted by (Sanders & Schneier, 2023) who develop a plausible scenario of the use of chatbots for political lobbying that may overwhelm existing scrutiny mechanisms.

In addition, there is a rapidly growing number of commentaries that pick up on other ethical concerns related to digital technologies and AI and that trace how these may materialise in the context of ChatGPT. Such concerns include the opacity of the underlying model (van Dis et al., 2023), environmental pollution (Nast, 2023b), the fact that chatbots still have no real understanding of the texts it produces (Hutson, 2021) but also the increasing affinity to the human mind which allows ChatGPT to pass tests designed for humans (Wilkins, 2023). There is also growing evidence that to make ChatGPT more sensitive to cultural values and sentiments, OpenAI relied on exploited and underpaid data labellers in Africa and other low-income countries (Perrigo, 2023).

This list of currently discussed issues highlights key concerns that are currently discussed. This article aims to provide a more systematic and comprehensive account of the ethics of ChatGPT. It therefore now introduces the discourse on the ethics of emerging technologies, which will provide the conceptual basis of the analysis and inform the methodology that will be described in the next section.

### 2.3. ChatGPT and the Ethics of Emerging Technologies

The previous section has shown that there is already significant research on and discussion of ethical aspects of ChatGPT. However, this

discourse is mostly ad hoc and lacks a systematic approach. Much of it is published in news media and focuses on specific high-profile topics. We therefore suggest that a more systematic and rigorous approach to the ethics of ChatGPT is called for that can be used to better understand possible issues and serve as basis for organisational and societal policy development. Such a more systematic approach should be based on peer reviewed and established knowledge that has proven to be reliable when applied to similar technologies. One body of knowledge and theory that was developed for this very purpose is represented in the literature on the ethics of emerging technologies. This article therefore draws on this literature and applies it to ChatGPT. Before we outline how this can be done in the methodology section, we first need to demonstrate that this body of knowledge can usefully be applied to ChatGPT which requires us to show that ChatGPT is an emerging technology which we do by looking at the components of the term, at ‘technology’ and when it can be considered to be ‘emerging’.

Technology has its etymological root in the Greek *tekhne*, which stands for “art, skill, craft in work; method, system, an art, a system or method of making or doing” and the ending ‘-logy’ which refers to discourse, theory or science which combines to the definition of technology as the “study of mechanical and industrial arts” ([Online Etymology Dictionary, 2022](#)). Technology is pervasive in human societies and its role in shaping humans and our society has long been discussed ([Ellul, 1973](#); [Heidegger, 1953](#); [Ihde, 1990](#); [Spengler, 1931](#)). The case of ChatGPT shows that these questions are not just idle speculation but affect the way we engage with current phenomena. One can distinguish between technology as a paradigm and a device, both of which reflect an important facet of technology ([Moor, 2008](#)). An alternative categorisation distinguishes between three levels: the top level of the technology, which can be implemented in the meso level of different artefacts, which can lead to different applications at the most basic level ([Brey, 2012](#)). In this schema ChatGPT would be best characterised as an artefact which forms part of the technology of natural language processing, which in turn is part of artificial intelligence, which is part of computing or digital technologies. ChatGPT forms part of the family of emerging digital technologies that call for ethical reflection ([Kazim & Koshiyama, 2021](#)). It can furthermore be seen as emerging, because it is still developing fast as indicated by the transition from GPT-3 to GPT-4 ([Marr, 2023b](#)). At the same time, the current iteration of ChatGPT is still very open in terms of intended and likely applications.

Dealing with ethical questions of emerging technologies is not straightforward. A key reason for this is the fundamental and unavoidable uncertainty of the future. We simply do not know with certainty what the future holds and how technology will develop ([Ellul, 1973](#)). An ethics of emerging technologies that claims full knowledge of future developments and their ethical consequences is thus impossible. However, while the future cannot be comprehensively known, it is not entirely unknowable either. Modern societies are based on strong assumptions about the future (e.g. future tax revenue or demographic development) that allow for planning and policy development and – to some degree – help pre-empt foreseeable future problems. It is this ambiguity of future knowledge that motivates calls for research and technology development to accept partial responsibility for the ethics of emerging technologies ([InterAcademy Partnership, 2016](#)). This ambiguity also shapes the claims that can be raised for research on the consequences of emerging technologies. [Collingridge \(1981\)](#) famously pointed out the trade-off that short-term technical developments that we can easily predict are difficult to change, whereas long-term developments that may be easy to steer in desired directions are difficult to predict ([Genus & Stirling, 2018](#)). While we thus cannot fully know the future, we can explore possible and likely futures with a view to understanding what they require from us today ([Cuhls, 2003](#)). Or, as [Guston \(2013\)](#) puts it, while we may not look into the future, we can look toward it.

With these caveats in mind, it is possible to say more about the ethics of emerging technologies. The spirit motivating most attempts to

undertake research on the ethics of emerging technologies is not to provide scientific certainties, but to act on the insight that researchers share some responsibility for future morally relevant technical developments and to undertake an honest attempt to proactively engage with those ([Cagnin et al., 2008](#); [Groves, 2009](#); [Swierstra et al., 2009](#); [Walsham, 2012](#)) and to provide orientation with regards to research activities ([van der Burg, 2014](#)). Such an endeavour calls for intellectual humility ([Jasanoff, 2003](#)) acknowledging its difficult and fallible nature which is the spirit that should inform the reading of the methodological approaches introduced in the next section.

### 3. Methodology

We use the term ‘methodology’ in this paper in the commonly accepted meaning of “a set of methods used in a particular area of study or activity” ([Cambridge Dictionary, 2023](#)). A methodology for undertaking research on the ethics of emerging technologies can be understood as an attempt to respond to the problem of uncertainty of the future ([Brooks et al., 2021](#)). This does not mean that they can overcome this problem, but they can provide a rigorous, transparent and reproducible way of engaging with it, which allows for highlighting uncertainties and calling them into question ([Umbrello et al., 2023](#)). The probably most comprehensive review of ethics in research and innovation ([Reijers et al., 2018](#)) has identified eight methods of what it calls “ex ante” methods, which align with the ethics of emerging technologies discussed here. A review and critique of “ethical foresight analysis” ([Floridi & Strait, 2020](#)) which is similarly aligned with this article’s research objective identified six existing methods, which significantly overlap with [Reijers et al. \(2018\)](#) findings. In the following subsection we provide an overview of relevant methodologies that can be used to study the ethics of emerging technologies before we then explain in detail our research protocol that explains how we have adapted, developed, and implemented these for the purpose of this article.

#### 3.1. Methodologies for Studying the Ethics of Emerging Technologies

Most of these methodologies aim to bridge the gap between methodologies from future and foresight studies ([Sardar, 2010](#)) and the ethics of technology ([Royackers & Poel, 2011](#)). They typically use approaches from future and foresight studies such as Delphi studies ([Adler & Ziglio, 1996](#); [Someh et al., 2019](#)) or scenario research ([Gray & Hovav, 2007](#)) and put special emphasis on questions of ethics.

This article employs a sub-set of the established methodologies of research on the ethics of emerging technologies, notably the subset that allows for the identification of ethical issues. Much effort across various methodologies is expended on identifying emerging technologies. This work is not needed in this article because it focuses on a defined technology, namely ChatGPT. Furthermore, the article’s aim is to broaden awareness of the ethics benefits and challenges raised by ChatGPT, i.e., it aims to enumerate such issues in as comprehensive a manner as possible. It therefore does not need to pay attention to what is typically the next step proposed by several of the methodologies, which is the question of how to address them.

The methodology employed in this article is therefore based on three of the established methodologies of ethics of emerging technologies that include proposals for identifying ethical issues. This is complemented by steps that undertake a similar approach in the ethics of artificial intelligence (AI). The three methodologies used here are referred to as “anticipatory technology ethics” (ATE, ([Brey, 2012](#))), the “framework for the ethical impact assessment of information technology” (EIA, ([Wright, 2011](#))) and “ethical issues of emerging ICT applications” (ETICA, ([Stahl, 2011](#); [Stahl et al., 2017](#))). These three were all developed in the early 2010 s. They are aware of each other and mutually cite one another. This can serve as an indicator that combining them, as is proposed in this article, is a legitimate activity. They also take a similar approach to the identification of issues which is the core interest in this



article.

All three agree that ethical concerns do not arise in a vacuum and that one can learn about likely future ones by looking at established ethical discussions and by asking which established issues are likely to be relevant in a novel technology or application area and how such issues can be categorised. They each produce a list of issues which shows significant overlap. They differ in how these lists are constructed. ATE's list is derived from philosophical ethics of technology and has the main categories of harms and risks, rights, (distributive) justice, and well-being and the common good. In each of these categories a number of issues are located. In some cases these issues are then subdivided further. For example, under rights there is the issue of freedom, which is broken down in freedom of movement, speech and assembly. EIA starts with its main categories by using the principles of biomedical ethics (Beauchamp & Childress, 2009): autonomy, nonmaleficence, beneficence, and justice. These are complemented by the category of privacy and data protection. Again, each of these have a number of issues associated with each of the categories. For each of the issues (or in some case values), EIA offers a number of questions that support the reflection of the issues in a specific context. The ETICA method, final, offers a similar list of issues. However, the individual issues were arrived at differently. They were identified on the basis of the analyses of 10 different emerging technologies and the categories were constructed in a bottom-up manner. The categories listed in (Stahl et al., 2017) are: conceptual issues and ethical theories, impact on individuals, consequences for society, uncertainty of outcomes, perceptions of technologies and role of humans. In this methodology each of the categories or issues are also linked to guiding questions that are meant to help researchers understand and identify the nature of the potential ethical concern.

The aspect of these three approaches to the ethics of emerging technologies that is of most interest to this article is the list of likely issues they produced. For the purposes of this article the idea was to generate the most comprehensive overview of possible issues to then interrogate this list with a view to identifying issues related to ChatGPT that are most in need of discussion. The lists provided by the three methodologies were therefore merged into one list. For this purpose, the lists were reduced to a one-dimensional list, i.e., the categories and, in the case of ATE, the sub-issues, were either included, where they offered additional insights, or excluded, if they were covered by the issues. Additional information, notably the guiding questions from EIA and ETICA were retained as comments linked to the issues to guide the more detailed subsequent analysis steps. The table in appendix A contains the list of all the issues as identified across the three methodologies.

This list of issues in appendix A provides the basis of the analysis of the ethical issues of ChatGPT undertaken for this article. They represent the ethical issues identified by the three methodologies and need to be read in conjunction with the supporting documentation, notably the definitions of the issues and the guiding questions accompanying them. We believe that they constitute a good starting point for an analysis of the ethics of emerging technologies. However, we realise that this list of issues dates back to the early 2010 s and was not specifically compiled with a view to AI. In light of the rapid development of AI and the discussion of the ethics of AI in the intervening years, we decided to calibrate this list by adding a set of issues that were identified specifically with AI in mind. As ChatGPT is an example of an AI application, we aimed to ensure that established ethical concerns arising from the ethics of AI discourse were properly included.

The inclusion of the ethics of AI raised further methodological questions. While there is a relatively settled landscape of research on the ethics of emerging technologies as outlined above, the same cannot be said about the ethics of AI (Ashok et al., 2022; Dwivedi et al., 2019). This discourse continues to mushroom. It has a strong emphasis on ethical guidelines and principles (Fjeld et al., 2020; Jobin et al., 2019) which includes a number of high profile interventions (AI HLEG, 2019) but it also legislation such as the EU's AI Act (European Commission, 2021), various approaches to standardisation (IEEE, 2017; NIST, 2022),

certification and many others. There are several attempts to identify and categorise AI-specific ethical issues, e.g. by Müller (2020) who distinguishes between ethical issues of AI systems as objects which includes privacy, manipulation, opacity, bias, human-robot interaction, employment, and the effects of autonomy and ethical issues of AI systems as subjects which include machine ethics, artificial moral agency, and finally the issues arising from an AI superintelligence. Vesnic-Alujevic et al. (2020) distinguish between individual issues which includes autonomy, dignity, privacy and data protection, and societal issues which includes fairness and equity; the good life and diversity; responsibility and accountability; transparency; surveillance and datafication; governance of the AI.

These examples demonstrate that there is significant overlap between the ethics of emerging technologies and the ethics of AI which is not surprising, as AI can still be considered an emerging technology. It raises the question, however, which list to include in our analysis to avoid possible blind spots. We decided to include a list published by the European Parliament (2020) which was compiled by a group of academics with a view to providing broad coverage of issues of interest to the European Parliament and hence, by implication, to European citizens. This list is categorised in terms of AI impact on different aspects of society where individual issues are then listed as shown in Appendix B below.

### 3.2. Research protocol

In our research we based the analysis of ChatGPT on the list of issues that are displayed in Appendix A and B. These formed the theoretically derived starting point for us to explore possible ethical issues. We used these lists of ethical issues as shown in appendices A and B to find out what we can learn about the ethical benefits and challenges raised by ChatGPT by applying existing approaches to the ethics of emerging digital technologies. In order to achieve this, we combined the lists of issues into one spreadsheet which also included the guiding questions from EIA and ETICA to help us understand the issues in more detail. We then wanted to explore whether ChatGPT is likely to have an impact on the various issues. For each issue we therefore explored how ChatGPT might impact it, both in positive and negative terms. This was done by providing a short narrative outlining our reasoning and then allocating a likelihood of this impact arising (1 =low, 2 =medium, 3 =high) and a measure of severity (1 =low impact, 2 =medium impact, 3 =high impact). This allowed us to calculate an expected impact measure, separated between positive and negative impacts, which we arrived at by multiplying likelihood and severity, giving us two scores between 1 and 9 (one for benefits, one for damages) for each of the issues. To allow for independent scrutiny of our approach, we make the full spreadsheet including the narrative justifications and the scores for each issue we make the complete spreadsheet available here.<sup>1</sup>

A key challenge when using this approach was to identify likely areas and mechanisms of impact. Our aim was to offer an analysis of ethical issues of ChatGPT that cuts across application areas. However, the number of possible applications is close to infinite, and we may not be aware of some that are already being worked on, even less of ones that are currently unexplored. We therefore focused on the key ethically relevant features of ChatGPT, namely its abilities to produce text that is difficult to recognise as machine-generated, its ability to learn and interact seamlessly with humans. In order to provide plausibility of our evaluations we used the narrative part of the analysis to indicate which application examples we had in mind before scoring them. Furthermore, in order to provide some rigour to the analysis the two authors co-reviewed each other's analysis to ensure consistency and plausibility.

As all future and foresight research methods, this methodological approach does not promise scientific exactness. Its purpose is to surface

<sup>1</sup> <https://tinyurl.com/5fawk8yc>

possible issues, in particular those that have not yet been looked at in much detail and to do so in a systematic and transparent way. We understand that our analysis is somewhat idiosyncratic and driven by our prior knowledge in particular our understanding of ethics of technology and that researchers with different backgrounds might highlight other issues and use different justifications for this. From our perspective this is not majorly problematic, and we would indeed welcome alternative voices that come to different conclusions from ours, as it would strengthen the overall aim of the exercise which is to support a public discourse on the ethical issues of ChatGPT.

As a final remark on the article's methodology, we would like to highlight that we interacted with ChatGPT during the writing of this article and incorporated some of the insights into the narrative. As the use of ChatGPT in academia is currently highly contested and at the centre of public attention, we chose to collect all of our relevant interactions with the system and make them accessible to scrutiny on this document.<sup>2</sup>

#### 4. Findings and discussion

By following the methodology described above, we aimed to arrive at a comprehensive understanding of the ethics of ChatGPT that is transparent and rigorous. However, when implementing the method, we had to make assumptions and choices that are relevant to the analysis and need to be highlighted. When assessing the likelihood and severity of both positive and negative impacts, we had to make use of our understanding and possible scenarios of use. We only considered immediate consequences based on the core characteristics and capability of ChatGPT, i.e., its abilities to produce text that is difficult to recognise as machine-generated, its ability to learn and interact seamlessly with humans. For example, we assumed that it can be used for telephone conversations where the respondent is not aware of speaking to a machine. This, in turn can be used for data collection, which may have for legitimate and illegitimate purposes, such as online or telephone marketing but also social engineering or political persuasion. Our key challenge was to use a consistent boundary of future exploration which is anchored in the actual technical capability of ChatGPT and avoid general speculation. Where possible and relevant we therefore used the textual description of benefits and risks to explain how we arrived at our evaluation of the various issues.

In order to allow for a general overview and comparability of benefits and risks we used a simple method of quantifying these, providing measures of likelihood and impact. We use these measures for the discussion below, as they allowed us to highlight what we perceive to be key topics worthy of attention. We believe that this is a reasonable approach that mirrors widely used ways of dealing with risks and benefits of emerging technology through risk management or impact assessments processes (Stahl, Antoniou et al., 2023). We fully realise, however, that the numerical scores we allocated are expressions of our interpretations rather than objective truths. Before we return to the reflection on and evaluation of our approach, let us take a look at what the analysis unearthed.

##### 4.1. Benefits and concerns of ChatGPT

One intention behind this article was to arrive at a balanced view of ChatGPT that takes into account both benefits and concerns linked to the technology. We therefore start the discussion of our substantive findings by looking at the concepts that promised benefits that scored highest. We evaluated the following concepts as having a high likelihood of being realised and a high level of social or ethical benefit: collective human identity and the good life; perceptions of technology, the role of humans, beneficence, sustainability, health and bodily harm, the ability to think

one's own thoughts and form one's own opinions, animal rights and animal welfare, support of vital social institutions and structures, the labour market, and impact on the financial system. This very eclectic group of benefits is defined by its broad scope and reach. These are not specific issues or topics but mostly comprised of higher-level aggregate concepts. Human identity, the role of humans, beneficence, the labour market, the financial system or health are all high-level topics that depend on the collaboration of many individuals across large parts of society. This implies that they rely heavily on successful communication between different types of stakeholders. Such communication and the translation of meaning across different vocabularies of different stakeholder groups is something that is core to the capabilities of ChatGPT. Being high-level, these concepts also are open to interpretation and offer scope for the analyst to fill them with life.

We identified 11 concepts that scored the maximum in terms of benefits. The above benefits are highlighted in literature on how ChatGPT and other generative AI systems can be leveraged in hospitality and tourism industry (Dwivedi, Pandey et al., 2023), in healthcare (Javaid et al., 2023), in transportation (Du et al., 2023), and agriculture (Ray, 2023). In contrast with the benefits, we identified 30 concepts that scored the maximum (i.e. were deemed to have a high likelihood of materialising and a high impact) in terms of negative impact. These can be grouped as issues related to social justice and rights, individual needs, culture and identity as well as environmental impacts as shown in Fig. 1.

This figure suggests that the key issues that have the potential to raise the most significant concerns can be divided into four main categories. These are social justice and rights where ChatGPT is seen as having a potentially detrimental effect on the moral underpinnings of society, such as a shared view of justice and fair distribution as well as specific social concerns such as digital divides or social exclusion. The second group pertains to individual needs, such as safety and autonomy which are also reflected in informed consent and the avoidance of harm. We included environmental harms as one group of issues and added topics of culture and identity as a fourth group. This categorisation is one of many possible ones. The dividing lines are not as clear-cut as the figure may suggest. Some of the issues could fit into more than one of the boxes. Its main purpose here is to show the breadth of issues and concerns which cover pretty much any aspect of life.

A further perspective on our analysis that may be helpful with regards to identifying key ethical concerns worthy of further study and intervention is provided by the difference between benefits and downsides. We calculated the total impact by subtracting the total negative impact (probability times damage) from the positive impact (probability times benefit). The total impact score is positive if the positive impact outweighs the negative impact. The higher the score, the more the benefits outweigh risks and the reverse is true for negative impacts.

A look at the net positive concepts shows that these are mostly linked to the external world, such as with regards to the rights of animals (which could be promoted) or the promotion of physical health (which could benefit from telemedicine, clinical decision support, disease surveillance etc.) In all of these cases ChatGPT has the potential to improve processes. As a language-based system, ChatGPT can have negative impacts in these areas as well, but these would not be immediate, as they might be in the case of robotic systems, and we felt that on the balance of probability, the positives would dominate the negatives in these cases.

On the other end of the scale, we have those concepts that promise a higher negative than positive impact. Many of these are related to the changes in social relationships that ChatGPT may engender. This includes questions of responsibility and accountability where there is no doubt a possibility that a better way of understanding complex situations may help and the ability to generate texts may strengthen responsibility and accountability regimes. However, we felt that the much more likely consequences of ChatGPT use will be that strong agents can use the technology to obfuscate and confuse public discourse with a view to evading responsibility and being held accountable.

Further broad issues having to do with social relationships follow a

<sup>2</sup> <https://tinyurl.com/yuk4w2yk>

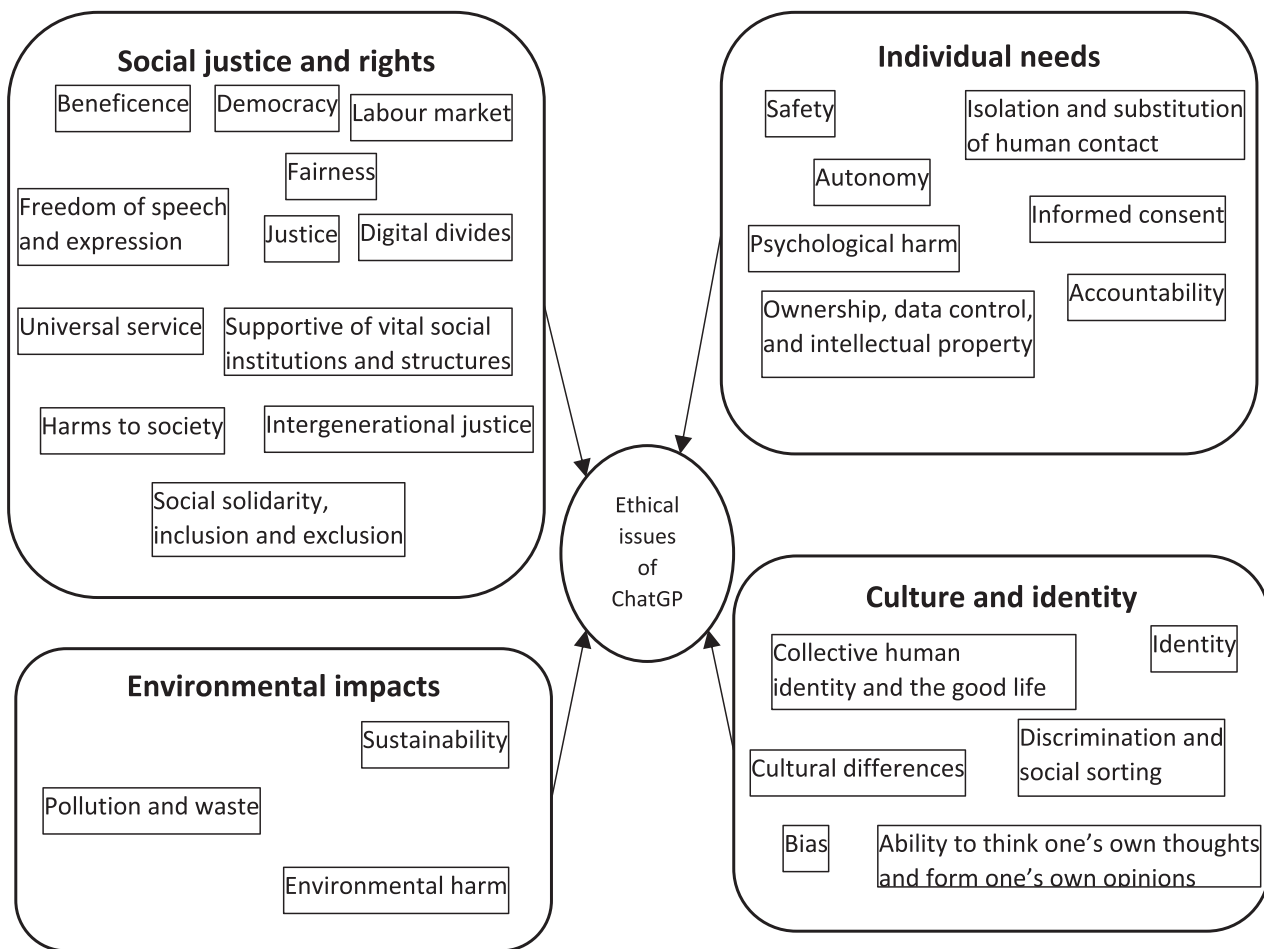


Fig. 1. Ethical issues with highest negative impacts of ChatGPT.

similar logic. We included social solidarity, inclusion and exclusion, intergenerational justice, biases and digital divides in the same category. In all of these cases it is perfectly plausible that ChatGPT will do good and help those who are disenfranchised to gain a voice. But in practice and based on past experience in the way emerging digital unfold in western society, we assume that the negative implications will vastly overshadow the positive ones. The divides between those who have access to ChatGPT and are able to leverage it and those who do not will likely follow similar lines as existing digital and social divides and the ones who will be able to draw most benefit will continue to be the ones who are advantaged already, i.e. the large companies who have the resources to make use of new technologies and the individuals who have the education, financial and intellectual independence to understand the opportunities and realise them.

#### 4.2. Relevance of the findings

The objective of this article is to explore ethical benefits and challenges raised by ChatGPT by applying existing approaches to the ethics of emerging digital technologies. A key question in assessing the value and contribution of the article is the degree to which it corresponds or contradicts existing scholarship. It is therefore worth comparing our findings with the current discourse on the ethics of ChatGPT.

Due to the novelty of the technology, there is relatively little academic and peer-reviewed work on the topic available, even though the number of articles is growing rapidly. At the time of writing a search of the Scopus research databased using the search string "ethics AND ChatGPT" returns 25 results. At present the academic literature focuses on the set of topic that are prominent in the public domain and media

that was discussed earlier. A key topic is that of authorship and the influence of ChatGPT on research and education (Lee, 2023; Salvagno et al., 2023), a topic that our analysis is sensitive to. There are contributions that aim to capture a broader set of issues such as Dwivedi, Kshetri et al. (2023) including ethical concerns but they remain relatively superficial.

One way of assessing whether our approach meets the objective of providing a rigorous and comprehensive account of ethical aspects of ChatGPT is successful is to explore whether it reflects and is consistent with recent scholarship. There is currently much work that aims to provide guidance on the use of ChatGPT in the academic publication system. A good example is provided by Shmueli et al. (2023) who explore how authors, reviewers and editors in data science can responsibly leverage and address challenges of using generative AI tools like ChatGPT to improve scholarship. The editorial covers ethical issues around authorship, accountability, methodological rigor, bias, fairness, accuracy, gaming the system, privacy, data exposure, interpretability, intellectual property, availability, review reliability, and responsible use as ethical issues arising from the application of generative AI like ChatGPT in data science research and publishing. All of these concerns can be located in our ethical analysis as well. Another recent example of engagement with ethics of ChatGPT is given by Susarla et al. (2023) who discuss guidelines and challenges for the responsible use of generative AI like ChatGPT in IS research, including concerns around bias, accountability, intellectual property, and depth versus potential benefits for problem formulation, data analysis, and writing. Again, these concerns are well covered by our analysis. As a final indication whether our analysis is relevant to the current discussion of ChatGPT, it is worth looking at a publication that is not primarily interested in ethical

questions. Dwivedi et al. (2023) explore practices, challenges, and a research agenda for implementing generative AI tools like ChatGPT in the hospitality and tourism industry. While not focused on ethics, the paper acknowledges ChatGPT raises important ethical issues around privacy, bias, transparency, governance, labour impacts etc. that require thoughtful policies and practices by users. By referring to these three publications we can confirm that our analysis engages with the ethical issues that are currently discussed and that it has a broader scope than the current discourse.

In addition to the current general discussion, it is worth looking more specifically at the way in which the community of scholars working on large language models are perceiving the issues arising from the class of technologies that includes ChatGPT. The most prominent contribution to this discourse comes from the Google-owned AI company DeepMind (Weidinger et al., 2021). In their report they report a total of 21 risks of harm that can be expected from language models, categorised in the following six risk areas: I. Discrimination, Exclusion and Toxicity, II. Information Hazards, III., Misinformation Harms, IV. Malicious Uses, V. Human-Computer Interaction Harms, VI. Automation, Access, and Environmental Harms. The report does not provide a detailed methodology of how the risks were identified which makes a detailed comparison with the ethical issues derived from the emerging technology literature difficult. However, there is significant similarity and overlap between the 21 risks identified by Weidinger et al. (2021) and the issues that we list in the appendix.

Maybe even more interesting than the general issues related to language models is the observation of how OpenAI is currently engaging with the ethical issues of ChatGPT. While we were writing this article, OpenAI launched GPT-4, the next version of the language model that ChatGPT is based on. As part of this launch the company published a technical report (OpenAI, 2023b) and a “system card” document (OpenAI, 2023a). These documents provide some technical background to GPT-4 that allows an insight into the ethical issues that OpenAI has recognised and is attempting to address in the development of GPT-4. Both documents contain an identical figure (OpenAI, 2023a, p. 8) showing example prompts (i.e. input from external users) and the response of an earlier version and how GPT-4 responds when it was launched. The prompts covered in the table refer to ethically problematic information that ChatGPT could conceivably provide, including advice on immoral activities such as murder, production of weapons, money laundering, self-harm, threats, racism and purchase of weapons. In an earlier version GPT-4 provided responses to these prompts whereas the launched version refuses to provide input. When tested in April 2023, ChatGPT gave similar answers to the launch version of GPT-4, suggesting that these safeguards had already been implemented. All of these issues that OpenAI is addressing in GPT-4 are part of or implied in the list of issues we identified in our methodology, thus confirming that our ethical analysis is valid. We thus think there is evidence that our research objective was met and that the approach we offer in this article is rigorous and comprehensive. Despite this success in moving towards our research objective, open questions remain.

#### 4.3. Open questions

While we believe that our analysis is valid, robust and transparent, we do not claim that it is exclusive. By this we mean that others who would follow the same methodology could plausibly come to different insights. The value of doing the analysis is not so much in the substantive items and concerns that we have looked at but in the deeper understanding of the overall landscape of ChatGPT use that it provides. By undertaking the analysis, we have come to a better understanding of how ethical issues of ChatGPT are conceptualised and framed in public discourse which raises questions that need to be discussed for reasonable and realistic responses to emerge.

One key question that pervades many of the issues and is highly relevant to the evaluation of ethical issues is that of ownership and

control of use. In our analysis we have assumed that ChatGPT will be freely available to all possible users. This is the model that OpenAI has chosen to initially bring to potential users, and it is consistent with access models of other types of general digital technologies, such as popular search engines. However, it is far from obvious whether this will remain the case. If benefits of ChatGPT can be monetised, then this is likely to happen. A paid-for version called ChatGPT plus already exists which requires a monthly payment of US\$20 and has a more advanced architecture, is based on more training data and likely better performance (Salunke, 2023). Open AI which started out as non-profit organisation created a capped-profit subsidiary, ostensibly with a view of allowing employee share options. The non-profit OpenAI Inc is the controlling shareholder of the for-profit OpenAI LP which may provide an avenue for continuation of the organisation’s aim of democratising QI (“OpenAI, 2023c). However, there have been some high-profile links to corporate players, notably a partnership with Microsoft which is using OpenAI technology for its chat business which have been critically evaluated by some observers (Xiang, 2023). It is not the purpose of this article to discuss corporate structures of OpenAI, but it seems likely that these corporate structures will have an influence on the ethical issues related to ChatGPT. As we have seen, many of the key ethical concerns involve broader question of social cohesion, inclusion and exclusion. Profit-oriented ownership of technology has been described as a key concern of other technologies (Zuboff, 2019) and ChatGPT is likely to be covered by these concerns.

Where profit is the driver of the organisations that develop and make available ChatGPT and related technologies, it is easy to foresee that profitable applications will be privileged over non-profitable ones. This means that many of the benefits that we identified, those that will allow underserved communities to strengthen their position and make their voices heard, are less likely to be supported than those that promise a return on investment.

A further open question that has some link to ownership but also to the broader social context of ChatGPT is that of governance and oversight, in particular with regards to misuse of the technology. This raises difficult questions of what would count as use or misuse, e.g. in fields like the persuasion of voters to vote for a particular party or of customers to buy products. These may well be deemed legitimate uses by some but misuse by others. However, there are more clear-cut cases of misuse, notably the use of ChatGPT for illegal purposes such as scamming, fraud etc. The capabilities of ChatGPT clearly render it a useful tool for such illegal uses, a point that is already well recognised (Europol, 2023; Sweeney, 2023). The current technical and organisational model where ChatGPT is hosted centrally and allows for interactions via the web may offer mechanisms for tracing illegal uses and countering them. If and when ChatGPT or its successors or competitors become available as stand-alone systems, this may open additional avenues for illegal use.

The ownership model is also likely to have implications for the broader understanding of the technology and thus of its robustness and reliability. In our analysis of the different ethical issues, we assumed that ChatGPT can provide human-like input into conversations with a high level of reliability. However, there is the well-discussed issue that machine learning models can learn biases from its training data and replicate these in interactions with users. This has been a key short-coming of previous models such as Microsoft’s chatbot Tay (Wolf et al., 2017). OpenAI seems to have spent considerable effort in avoiding ChatGPT from generating racist or similarly offensive output, but it is open to which degree this will remain successful. A related challenge in terms of reliability and robustness of the technology is the transparency of its sources and outputs comes from the technology’s propensity to “hallucinate”. This term refers to mistakes that ChatGPT makes when generating text that is semantically correct but factually incorrect or even nonsensical (Smith, 2023). Such hallucinations are a general problem for ChatGPT but depending on the context of their occurrence could raise significant ethical problems, for example in healthcare settings.



In addition to these questions that are focused on the technology itself and the way people can interact with it, there are further open questions about how a long-term use of this type of technology changes our collective view of what it means to be human and to interact with technology. We posit in this article that one of the crucial features of ChatGPT is its ability to produce human-like texts which implies that it may be difficult to assess whether a contribution to a conversation comes from a human or a machine. This is traditionally seen as problematic, as we assume that interactions between humans and machines and among humans differ. The European General Data Protection Regulation (GDPR, 2016) stipulates in Article 22 that any data subject “shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.” This could be read as implying that many interactions with chatbots will have to be flagged as such. What is unclear, however, is whether and how long this position will be upheld. If ChatGPT paves the way for widespread use of chatbots for all sorts of purposes, then it may well be that our moral preferences change and the distinction between interacting with a human and a machine are seen as decreasingly crucial. The opposite development is also conceivable, where more emphasis will be put on the identification of artificial agents in conversations.

In the longer term these developments can also touch on the broader ethical questions of the impact of artificial general intelligence, a highly contested term (Mitchell, 2019) but one that OpenAI openly promotes. This points to larger philosophical question around the nature of reality and how AI can change our view of it, which are beyond the scope of this article that focuses on immediate capabilities and impacts of ChatGPT but which may well turn out to be the most significant ethical implications of ChatGPT.

At a philosophical level, there is also the open question of whether ChatGPT’s ability to produce humanlike responses affirms Jacques Derrida’s critique of logocentrism of giving privileges to speech over writing. Derrida’s position was that meaning in natural language is determined by the ‘play’ of differences between words and not by ideas and intentions of a speaker (Sallis, 1987). There is no reason therefore to consider that speech comes before writing. For some, it may be taken that ChatGPT ‘writes without speech’. If this is true, it confirms Derrida’s position that there is a disconnect between the living voice of the speaker and written words. However, does ChatGPT write without speech? Or is it trained on living voices of many speakers and subsequently predicts patterns of words? Does ChatGPT ‘write’ without speech or ‘predicts’ from living voices? This is an open question that requires more thought and research.

## 5. Implications

The analysis of the ethical issues of ChatGPT has implications for research as well as practice that are spelled out below.

### 5.1. Implications for research

The article has demonstrated that a rigorous and comprehensive approach to the ethics of ChatGPT is possible and that existing methodologies can be used to gain an overview of the ethical issues. At the same time, the previous section on open questions has given an indication of avenues for further research that can be pursued to further strengthen the approach and go beyond it.

An initial limitation of the present research is that it was undertaken by the authors of the paper which allowed for a controlled analysis of ethical issues but is limited by the worldviews and experiences of the researchers. Undertaking a similar study using the ethics of emerging technologies but broadening expected benefits and concerns by including more stakeholders will very likely bring to light further insights that would likely change the evaluation of benefits and risks.

Another avenue of further research would be to undertake a direct

comparison with between the abstract analysis of ethical aspects offered here and the specific work that is already undertaken by the owners and developers of ChatGPT and other large language models. The previous section suggests that these questions are taken seriously by developers but that work on ethical issues is focused on immediate harm and its avoidance, thus excluding many of the broader issues such as fairness, inclusion, access, labour markets etc. Furthermore, research appears to focus heavily on ethical concerns and tends not to be explicit about the difficult process of comparing and balancing benefits and concerns.

Future studies will furthermore be called for to better understand detailed empirical impacts of ChatGPT. In this paper we focused on the generic characteristics of the technology and inferred possible uses and consequences from those characteristics. A more detailed understanding would be gained by a differentiated analysis looking at specific application areas or industries. Such detailed analysis would also help research keep abreast of the rapidly developing technical landscape and the capability of large language models. This question of technical development is also likely to have implications for the more abstract questions around the potential for the development of AGI or other long-term philosophical questions. It is likely that such questions will not be confined to ChatGPT or large language models but need to include other developments of AI and its enabling technologies. Elsewhere we have suggested that a suitable way of framing such research is to explore the ethics of the socio-technical innovation ecosystem in which AI is realised (Stahl, 2022, 2023). This perspective is important when determining the implications of ChatGPT because it shifts the focus from individuals (e.g. developers or users) or organisations (notably OpenAI) to the broader ecosystem (Chae, 2019; Senyo et al., 2019). The question of who is responsible for specific ethical consequences then transforms into the question of how should the ecosystem be structured to promote beneficial and prevent detrimental effects of the technology.

The key implication for research is thus that further research is called for. While this is a typical and often self-serving position taken in many research publications, its relevance and legitimacy in the case of ChatGPT is hard to deny. The potential impact of the technology combined with its rapidly evolving nature combine into a strong argument for ongoing research-based reflection of its ethical implications. Such research is in the public interest but also in the interest of the AI industry and forms part of the implications for practice.

### 5.2. Implications for practice

The insights developed in this paper and the availability of a more comprehensive overview of ethical concerns related to ChatGPT are relevant to the practice of several stakeholder groups including policy-makers, research funders and companies developing or using large language models.

One can observe a rapidly growing recognition among policymakers that the nature of recent AI developments is such that a passive approach to legislation and regulation is difficult to justify. While some regulatory developments such as the EU AI Act precede the arrival of ChatGPT, one can argue that ChatGPT has contributed to the push for the need of policy intervention, notably in the USA. It is now broadly recognised that policy initiatives, potentially including legislation, regulation or the creation of new regulators (Graham & Warren, 2023) is required to ensure responsible development and use of generative AI. Such initiatives may be called for to mitigate impacts, for example where harm arises due to job losses, inequality, social biases and mitigate identified harms. These are not fundamentally new insights (Executive Office of the President, 2016), but the arrival of ChatGPT has highlighted the need for such interventions.

The insights arising from this paper show that at present there is a lack of balance between the attention paid to ethical benefits and downsides of ChatGPT. Relevant policy should pay attention to both of these aspects as well as to the question who receives the benefits and who is exposed to the downsides. This implies that policies are called for

that promote diversity, equity, and inclusion in access to generative AI to prevent exacerbating divides. This may include subsidies or public access programmes. Policy can be used to create incentives through government procurement practices, grants etc. to steer development towards social good applications. ChatGPT and related technologies are not locally bound with risks and benefits potentially having global reach. This means that any relevant policy intervention needs to move beyond the local and national level and establish reliable policy frameworks on the international level (UNESCO, 2022; Wallach & Marchant, 2018). Some of the ethical issues point to the fundamental structures of our societies. ChatGPT arises from the current context of what Zuboff (2019) calls surveillance capitalism. A long-lasting response to the ethics of ChatGPT may well call for fundamental reconsiderations of the structure of current socio-economic systems.

Some of the more immediate implications of our research arise for research funders. The objective of our research was to come to a more comprehensive understanding of the ethical issues of ChatGPT. The follow-on research activities outlined in the implications for research will need to be funded. This not only includes the broadening of the research by involving stakeholder perspectives but also research on possible and likely mitigation measures. These can range from the use of open, transparent and unbiased foundations for generative models that can overcome some of the limitations of proprietary systems. Key to the success of engaging with ethics of ChatGPT will be to bring in diverse perspectives which may call for funding to support diverse participation in generative AI development and preventing homogeneous thinking. This calls for interdisciplinary research that can inform educational initiatives to develop expert and public understanding of responsible development and use. The broadening of the discourse furthermore calls for stronger interdisciplinary collaboration.

Policy and funding support will only have a chance of promoting ethically beneficial outcomes of ChatGPT use, if the organisations driving the development of large language models actively engage with ethical questions. Implications for industry thus include a need to prioritise ethical considerations early, e.g. through ethical review boards and risk assessment processes. Responsible design principles that are being developed across the AI landscape like transparency, accountability and bias testing should become standard in training generative AI models. Companies should consider the use of impact assessment methods (Stahl et al., 2023) prior to deploying their technologies. It is likely that a risk-based approach will be applied to AI (NIST, 2022) and companies need to consider how high risk applications of large language models can be governed. But even for lower risk applications, the balance of benefits and concerns should be considered. This is likely to call for companies to develop access and pricing models that prevent the exclusion of typically underserved communities, to allow them to benefit from large language models.

In developing appropriate responses, companies need to work with governments, regulators and policymakers. This certainly appears to be happening at present where OpenAI and other large tech companies are working with the US legislature and executive to address concerns around AI. It is important to keep in mind, however, that the oligopolistic structure of the tech industry is part of the underlying problem and the leading companies in this space are unlikely to wish to change the structure that they benefit from. Furthermore, companies will focus on

those markets that are most profitable for them, thus leading to limits to international collaboration and exacerbating the risk that the downsides of new technologies will be pushed into areas where they do not have their economic focus.

It would thus be naïve to expect industry to solve these issues, but at the same time, they will not be solved without the contribution of industry. A public commitment of the tech giants stating that they have ethical duties arising from the products they distribute cannot replace stronger regulatory interventions, but it can serve as a useful reminder in controversial debates that at least some companies accept that they have ethical commitments beyond the maximisation of profits and that they recognise their social responsibilities.

## 6. Conclusion

This article explores the ethical issues raised by ChatGPT beyond current debates on authorship, demonstrating the value of applying existing approaches to the ethics of emerging technologies. Our analysis highlights a broad range of relevant ethical concerns going significantly beyond the current discourse. This is useful for AI ethics scholars, developers, and anyone interested in suitable ethics methodologies. The analysis also provides insights into implications and interventions, emphasizing the need to consider the whole socio-technical AI ecosystem, not just specific issues. Fully addressing the challenges will likely require multi-level societal engagement, from individual software engineers to international policies, to ensure the benefits of AI, in particular large language models such as those underpinning ChatGPT, are realized while harms are avoided. Overall, the article advocates applying a holistic ethics perspective to guide ongoing discourse and action around such impactful emerging technologies.

### Responsible innovation ecosystems

Ethical implications of the application of the ecosystem concept to artificial intelligence.

### Funding source

This research was funded by the European Union's Horizon 2020 Research and Innovation Programme Under Grant Agreement No. 945539 (Human Brain Project SGA3). This work was furthermore supported by the Engineering and Physical Sciences Research Council [Horizon Digital Economy Research 'Trusted Data Driven Products: EP/T022493/1].

### CRedit authorship contribution statement

**Bernd Carsten STAHL:** Conceptualization, Data collection and analysis, Methodology design, Writing, Funding acquisition. **Damian EKE:** Data collection and analysis, writing.

### Declaration of Competing Interest

None.

## Appendix A. Cumulative list of ethical issues from ATE, EIA and ETICA

---

ETICA	Conceptual issues and ethical theories
	Privacy
	Autonomy
	Treatment of humans
	Identity
	Security

(continued on next page)

(continued)

---

	Digital divides
	Collective human identity and the good life
	Ownership, data control, and intellectual property
	Responsibility.
	Surveillance
	Cultural differences
	Uncertainty of outcomes
	Perceptions of technology
	Role of humans
EIA	autonomy
	dignity
	informed consent
	safety
	Social solidarity, inclusion and exclusion
	Isolation and substitution of human contact
	Discrimination and social sorting
	Beneficence
	Universal service
	Accessibility
	Value sensitive design
	Sustainability
	Justice
	Equality and fairness (social justice)
	Collection limitation (data minimisation) and retention
	Data quality
	Purpose specification
	Use limitation
	Confidentiality, security and protection of data
	Transparency (openness)
	Individual participation and access to data
	Anonymity
	Privacy of personal communications: monitoring and location tracking
	Privacy of the person
	Privacy of personal behaviour
ATE	Health and bodily harm
	Pain and suffering
	Psychological harm
	Harm to human capabilities
	Environmental harm
	Harms to society
	Freedom
	Freedom of movement
	Freedom of speech and expression
	Freedom of assembly
	Autonomy
	Ability to think one's own thoughts and form one's own opinions
	Ability to make one's own choices
	Responsibility and accountability
	Informed consent
	Human dignity
	Privacy
	Information privacy
	Bodily privacy
	Relational privacy
	Property
	Right to property
	Intellectual property rights
	Other basic human rights as specified in human rights declarations (e.g., to life, to have a fair trial, to vote, to receive an education, to pursue happiness, to seek asylum, to engage in peaceful protest, to practice one's religion, to work for anyone, to have a family etc.)
	Animal rights and animal welfare
	Just distribution of primary goods, capabilities, risks and hazards
	Nondiscrimination and equal treatment relative to age, gender, sexual orientation, social class, race, ethnicity, religion, disability, etc.
	North-south justice
	Intergenerational justice
	Social inclusion
	Supportive of happiness, health, knowledge, wisdom, virtue, friendship, trust, achievement, desire-fulfillment, and transcendent meaning
	Supportive of vital social institutions and structures
	Supportive of democracy and democratic institutions
	Supportive of culture and cultural diversity

---

## Appendix B. Categories and ethical issues of AI according to (European Parliament, 2020)

Impact on society	The labour market Inequality Privacy, human rights and dignity Bias Democracy Relationships Personhood
Impact on human psychology	
Impact on the financial system	
Impact on the legal system	Criminal law Tort law
Impact on the environment and the planet	Use of natural resources Pollution and waste Energy concerns Ways AI could help the planet
Impact on trust	Why trust is important Fairness Transparency Accountability Control

## References

- Adler, M., & Ziglio, E. (1996). *Gazing into the Oracle: The Delphi Method and its Application to Social Policy and Public Health*. Jessica Kingsley.
- AI HLEG. (2019). Ethics Guidelines for Trustworthy AI. European Commission - Directorate-General for Communication. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- Ali, S. R., Dobbs, T. D., Hutchings, H. A., & Whitaker, I. S. (2023). Using ChatGPT to write patient clinic letters. *The Lancet Digital Health*, 5(4), e179–e181. [https://doi.org/10.1016/S2589-7500\(23\)00048-1](https://doi.org/10.1016/S2589-7500(23)00048-1)
- Aristotle. (2007). *The Nicomachean Ethics*. Filiquarian Publishing, LLC.
- Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, 62, Article 102433. <https://doi.org/10.1016/j.ijinfomgt.2021.102433>
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning (SSRN Scholarly Paper 4337484). <https://doi.org/10.2139/ssrn.4337484>.
- Beauchamp, T.L., & Childress, J.F. (2009). *Principles of Biomedical Ethics* (6th ed.). OUP USA.
- Bentham, J. (1789). *An introduction to the principles of morals and legislation*. Dover Publications Inc.
- Bove, T. (2023, February 3). OpenAI founder Sam Altman says he can imagine ways that ChatGPT “breaks capitalism.” *Fortune*. <https://fortune.com/2023/02/03/openai-sam-altman-chatgpt-break-capitalism/>.
- Bowie, N. E. (1999). *Business ethics: A kantian perspective*. Blackwell Publishers.
- Brey, P. A. E. (2012). Anticipating ethical issues in emerging IT. *Ethics and Information Technology*, 14(4), 305–317. <https://doi.org/10.1007/s10676-012-9293-y>
- Brooks, L., Cannizzaro, S., Umbrello, S., Bernstein, M. J., & Richardson, K. (2021). Ethics of climate engineering: Don't forget technology has an ethical aspect too. *International Journal of Information Management*, Article 102449. <https://doi.org/10.1016/j.ijinfomgt.2021.102449>
- Bynum, T. W. (2001). Computer ethics: Its birth and its future. *Ethics and Information Technology*, 3(2), 109–112. <https://doi.org/10.1023/A:1011893925319>
- Cagnin, C., Keenan, M., Johnston, R., Scapolo, F., & Barré, R. (2008). *Future-oriented technology analysis*. Springer. <https://doi.org/10.1007/978-3-540-68811-2>
- Cambridge Dictionary. (2023). Methodology. Cambridge Dictionary. <https://dictionary.cambridge.org/dictionary/english/methodology>.
- Castelvecchi, D. (2022). Are ChatGPT and AlphaCode going to replace programmers? *Nature*. <https://doi.org/10.1038/d41586-022-04383-z>
- Chae, B. (Kevin) (2019). A General framework for studying the evolution of the digital innovation ecosystem: The case of big data. *International Journal of Information Management*, 45, 83–94. <https://doi.org/10.1016/j.ijinfomgt.2018.10.023>
- Collingridge, D. (1981). *The social control of technology*. Palgrave Macmillan.
- Cuhls, K. (2003). From forecasting to foresight processes—New participative foresight activities in Germany. *Journal of Forecasting*, 22(2–3), 93–111. <https://doi.org/10.1002/for.848>
- De George, R.T. (1999). *Business Ethics* (5th edition). Prentice Hall College Div.
- Du, H., Teng, S., Chen, H., Ma, J., Wang, X., Gou, C., Li, B., Ma, S., Miao, Q., Na, X., Ye, P., Zhang, H., Luo, G., & Wang, F.-Y. (2023). Chat with ChatGPT on intelligent vehicles: An IEEE TIV perspective. *IEEE Transactions on Intelligent Vehicles*, 8(3), 2020–2026. <https://doi.org/10.1109/TIV.2023.3253281>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koochang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & 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. *International Journal of Information Management*, 71, Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Dwivedi, Y. K., Pandey, N., Currie, W., & Micu, A. (2023). Leveraging ChatGPT and other generative artificial intelligence (AI)-based applications in the hospitality and tourism industry: Practices, challenges and research agenda. *International Journal of Contemporary Hospitality Management*. <https://doi.org/10.1108/IJCHM-05-2023-0686>
- Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity. *Journal of Responsible Technology*, 13, Article 100060. <https://doi.org/10.1016/j.jrt.2023.100060>
- Ellul, J. (1973). *The technological society*. Inc. USA: Random House.
- European Commission. (2021). Proposal for a Regulation on a European approach for Artificial Intelligence (COM(2021) 206 final). European Commission. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence>.
- European Parliament. (2020). The ethics of artificial intelligence: Issues and initiatives (PE 634.452). EPRS | European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS\\_STU\(2020\)634452\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf).
- Europol. (2023). ChatGPT - the impact of Large Models on Law Enforcement [Tech Watch Flash]. Europol. <https://www.europol.europa.eu/publications-events/publications/chatgpt-impact-of-large-language-models-law-enforcement>.
- Executive Office of the President. (2016). Preparing for the Future of Artificial Intelligence. Executive Office of the President National Science and Technology Council Committee on Technology. [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf).
- 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. <https://dash.harvard.edu/handle/1/42160420>.
- L. Floridi A. Strait Ethical foresight analysis: What it is and 2020.
- Frederick, B. (2023, January 16). Will ChatGPT Take Your Job? *Search Engine Journal*. <https://www.searchenginejournal.com/will-chatgpt-take-your-job/476189/>.
- Freyhofer, H.H. (2004). *The Nuremberg Medical Trial: The Holocaust and the Origin of the Nuremberg Medical Code* (2nd Revised edition edition). Peter Lang Publishing Inc.
- GDPR. (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119/1.
- Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, 47(1), 61–69. <https://doi.org/10.1016/j.respol.2017.09.012>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), Article e45312. <https://doi.org/10.2196/45312>
- Graham, L., & Warren, E. (2023, July 27). Opinion | Lindsey Graham and Elizabeth Warren: When It Comes to Big Tech, Enough Is Enough. *The New York Times*. <https://www.nytimes.com/2023/07/27/opinion/lindsey-graham-elizabeth-warren-big-tech-regulation.html>.



- Grant, N. (2023, January 20). Google Calls In Help From Larry Page and Sergey Brin for A.I. Fight. *The New York Times*. <https://www.nytimes.com/2023/01/20/technology/google-chatgpt-artificial-intelligence.html>.
- Gray, P., & Hovav, A. Z. (2007). The is organization of the future: Four scenarios for 2020. *Information Systems Management*, 24(2), 113–120.
- Groves, C. (2009). Future ethics: Risk, care and non-reciprocal responsibility. *Journal of Global Ethics*, 5(1), 17–31.
- Guston, D. (2013). “Daddy, Can I Have a Puddle Gator?”: Creativity, Anticipation and Responsible Innovation. In R. Owen, M. Heintz, & J. Bessant (Eds.), *Responsible Innovation* (pp. 109–118). Wiley.
- Heidegger, M. (1953). Die Frage nach der Technik. <http://content.wuala.com/contents/nappan/Documents/Cyberespace/Heidegger,%20Martin%20-%20Die%20Frage%20nach%20der%20Technik.pdf>.
- Hsu, T., & Thompson, S.A. (2023, February 8). Disinformation Researchers Raise Alarms About A.I. Chatbots. *The New York Times*. <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html>.
- Hutson, M. (2021). Robo-writers: The rise and risks of language-generating AI. *Nature*, 591(7848), 22–25. <https://doi.org/10.1038/d41586-021-00530-0>
- IEEE. (2017). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. [https://standards.ieee.org/develop/indconn/ec/autonomous\\_systems.html](https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html).
- Ihde, D. (1990). *Technology and the lifeworld: From garden to earth*. Indiana University Press.
- InterAcademy Partnership. (2016). *Doing global science: A guide to responsible conduct in the global research enterprise*. Princeton University Press.
- Jasanoff, S. (2003). Technologies of humility: Citizen participation in governing science. *Minerva*, 41(3), 223–244. <https://doi.org/10.1023/A:1025557512320>
- Javaid, M., Haleem, A., & Singh, R. P. (2023). ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(1), Article 100105. <https://doi.org/10.1016/j.tbench.2023.100105>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kant, I. (1788). *Kritik der praktischen Vernunft*. Reclam, Ditzingen.
- Kant, I. (1797). *Grundlegung zur Metaphysik der Sitten*. Reclam, Ditzingen.
- Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 2(9). <https://doi.org/10.1016/j.patter.2021.100314>
- Lee, J. Y. (2023). Can an artificial intelligence chatbot be the author of a scholarly article? (Scopus) *Science Editing*, 10(1), 7–12. <https://doi.org/10.6087/kcse.292>.
- Marr, B. (2023a, January 23). How ChatGPT And Natural Language Technology Might Affect Your Job If You Are A Computer Programmer. *Forbes*. <https://www.forbes.com/sites/bernardmarr/2023/01/23/how-chatgpt-and-natural-language-technology-might-affect-your-job-if-you-are-a-computer-programmer/>.
- Marr, B. (2023b, February 24). GPT-4 Is Coming – What We Know So Far. *Forbes*. <https://www.forbes.com/sites/bernardmarr/2023/02/24/gpt-4-is-coming-what-we-know-so-far/>.
- McMorrow, R., & Liu, N. (2023, April 11). China slaps security reviews on AI products as Alibaba unveils ChatGPT challenger. *Financial Times*. <https://www.ft.com/content/755cc5dd-e6ce-4139-9110-0877f2b90072>.
- Mehdi, Y. (2023, February 7). Reinventing search with a new AI-powered Microsoft Bing and Edge, your copilot for the web. *The Official Microsoft Blog*. <https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/>.
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5, Article 100152. <https://doi.org/10.1016/j.caeai.2023.100152>
- Metz, C., & Grant, N. (2023, February 6). Racing to Catch Up With ChatGPT, Google Plans Release of Its Own Chatbot. *The New York Times*. <https://www.nytimes.com/2023/02/06/technology/google-bard-ai-chatbot.html>.
- Mill, J.S. (1861). *Utilitarianism* (2nd Revised edition). Hackett Publishing Co, Inc.
- Mitchell, M. (2019). *Artificial Intelligence: A guide for thinking humans (Illustrated edition)*. Farrar, Straus and Giroux.
- Moor, J. H. (1985). What is computer ethics. *Metaphilosophy*, 16(4), 266–275.
- Moor, J. H. (2008). Why we need better ethics for emerging technologies. In J. V. D. Hoven, & J. Weckert (Eds.), *Information technology and moral philosophy* (pp. 26–39). Cambridge University Press.
- Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Fall 2020)*. *Metaphysics Research Lab*. Stanford University. (<https://plato.stanford.edu/archives/fall2020/entries/ethics-ai/>).
- Nast, C. (2023a, January 10). Infinite AI Interns for Everybody. *Wired UK*. <https://www.wired.co.uk/article/ai-labor-interns>.
- Nast, C. (2023b, February 10). The Generative AI Race Has a Dirty Secret. *Wired UK*. <https://www.wired.co.uk/article/the-generative-ai-search-race-has-a-dirty-secret>.
- Nature editorial. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use, 612–612 *Nature*, 613(7945). <https://doi.org/10.1038/d41586-023-00191-1>.
- Nijssingh, N., & Duwell, M. (2009). Interdisciplinary, applied ethics and social science. In P. Sollie, & M. Düwell (Eds.), *Evaluating new technologies: Methodological problems for the ethical assessment of technology developments* (pp. 79–92). Springer.
- NIST. (2022). AI Risk Management Framework: Second Draft. <https://www.nist.gov/document/ai-risk-management-framework-2nd-draft>.
- OECD. (2016). *Research Ethics and New Forms of Data for Social and Economic Research [OECD Science, Technology and Industry Policy Papers]*. Organisation for Economic Co-operation and Development. <http://www.oecd-ilibrary.org/content/workingpaper/5jln7vnp32-en>.
- Online Etymology Dictionary. (2022, September 10). *Technology* | Etymology, origin and meaning of technology by etymonline. <https://www.etymonline.com/word/technology>.
- OpenAI. (2023a). GPT-4 System Card. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>.
- OpenAI. (2023b). GPT-4 Technical Report (arXiv:2303.08774). arXiv. <http://arxiv.org/abs/2303.08774>.
- OpenAI. (2023c). In Wikipedia. <https://en.wikipedia.org/w/index.php?title=OpenAI&oldid=1149269339>.
- OpenAI. (2022, November 30). ChatGPT: Optimizing Language Models for Dialogue. <https://openai.com/blog/chatgpt/>.
- Perrigo, B. (2023, January 18). Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer. *Time*. <https://time.com/6247678/openai-chatgpt-kenya-workers/>.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9.
- Ray, P. P. (2023). AI-assisted sustainable farming: Harnessing the power of ChatGPT in modern agricultural sciences and technology. *ACS Agricultural Science & Technology*, 3(6), 460–462. <https://doi.org/10.1021/acscagstech.3c00145>
- Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O’Sullivan, D., & Gordijn, B. (2018). Methods for practising ethics in research and innovation: A literature review, critical analysis and recommendations. *Science and Engineering Ethics*, 24(5), 1437–1481. <https://doi.org/10.1007/s11948-017-9961-8>
- Royakkers, L., & Poel, I. van de (2011). *Ethics, technology and engineering: An introduction*. Wiley-Blackwell.
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, 6. <https://doi.org/10.3390/healthcare11060887>
- Sallis, J. (1987). *Deconstruction and philosophy: The texts of Jacques Derrida*. University of Chicago Press.
- Salunke, M. (2023, March 15). ChatGPT vs ChatGPT Plus: A Comparison. *Medium*. <https://medium.com/@ind/chatgpt-vs-chatgpt-plus-a-comparison-e8b233165def>.
- Salvagno, M., Taccone, F. S., Gerli, A. G., & ChatGPT. (2023). Can artificial intelligence help for scientific writing? (Scopus) *Critical Care*, 27(1). <https://doi.org/10.1186/s13054-023-04380-2>.
- Sanders, N.E., & Schneider, B. (2023, January 15). Opinion | How ChatGPT Hijacks Democracy. *The New York Times*. <https://www.nytimes.com/2023/01/15/opinion/ai-chatgpt-lobbying-democracy.html>.
- Sardar, Z. (2010). The namesake: Futures; futures studies; futurology; futuristic; foresight—What’s in a name? *Futures*, 42(3), 177–184. <https://doi.org/10.1016/j.futures.2009.11.001>
- Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International Journal of Information Management*, 47, 52–64. <https://doi.org/10.1016/j.ijinfomgt.2019.01.002>
- Shmueli, G., Maria Colosimo, B., Martens, D., Padman, R., Saar-Tsechansky, M., R. Liu Sheng, O., Street, W. N., & Tsui, K.-L. (2023). How can LJDs authors, reviewers, and editors use (and Misuse) generative AI? *INFORMS Journal on Data Science*, 2(1), 1–9. <https://doi.org/10.1287/ijds.2023.0007>
- Short, C. E., & Short, J. C. (2023). The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. *Journal of Business Venturing Insights*, 19, Article e00388. <https://doi.org/10.1016/j.jbvi.2023.e00388>
- Smith, C.S. (2023, March 13). Hallucinations Could Blunt ChatGPT’s Success. *IEEE Spectrum*. <https://spectrum.ieee.org/ai-hallucination>.
- Someh, I., Davern, M., Breidbach, C., & Shanks, G. (2019). Ethical issues in big data analytics: A stakeholder perspective. *Communications of the Association for Information Systems*, 44(1), 718–747. <https://doi.org/10.17705/1CAIS.04434>
- Spengler, O. (1931). *Der Mensch und die Technik* (2007 reprint). Voltmedia, Paderborn.
- Stahl, B. C. (2011). IT for a better future: How to integrate ethics, politics and innovation. *Journal of Information, Communication and Ethics in Society*, 9(3), 140–156. <https://doi.org/10.1108/14779961111167630>
- Stahl, B. C. (2012). Morality, ethics, and reflection: A categorization of normative IS research. *Journal of the Association for Information Systems*, 13(8), 636–656. <https://doi.org/10.17705/1jais.00304>
- Stahl, B. C. (2021). From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics*, 2. <https://doi.org/10.1007/s43681-021-00080-1>
- Stahl, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, 62, Article 102441. <https://doi.org/10.1016/j.ijinfomgt.2021.102441>
- Stahl, B. C. (2023). Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. Article 1 *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-34622-w>.
- Stahl, B. C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warszo, Z., & Wright, D. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-023-10420-8>
- Stahl, B. C., Flick, C., & Timmermans, J. (2017). Ethics of Emerging Information and Communication Technologies—On the implementation of RRI. *Science and Public Policy*, 44(3), 369–381. <https://doi.org/10.1093/scipol/scw069>
- Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays—Should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>
- Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613(7945), 620–621. <https://doi.org/10.1038/d41586-023-00107-z>

- Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research*, 34(2), 399–408. <https://doi.org/10.1287/isre.2023.ed.v34.n2>
- Sweeney, M. (2023, March 8). Darktrace warns of rise in AI-enhanced scams since ChatGPT release. The Guardian. <https://www.theguardian.com/technology/2023/mar/08/darktrace-warns-of-rise-in-ai-enhanced-scams-since-chatgpt-release>.
- Swierstra, T., Demerding, D., & Boenink, M. (2009). Exploring Techno-moral change: The case of the obesity pill. In P. Sollie, & M. Düwell (Eds.), *Evaluating new technologies: Methodological problems for the ethical assessment of technology developments* (pp. 119–138). Springer.
- Umbrello, S., Bernstein, M. J., Vermaas, P. E., Resseguier, A., Gonzalez, G., Porcari, A., Grinbaum, A., & Adomaitis, L. (2023). From speculation to reality: Enhancing anticipatory ethics for emerging technologies (ATE) in practice. *Technology in Society*, 74, Article 102325. <https://doi.org/10.1016/j.techsoc.2023.102325>
- UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- van der Burg, S. (2014). On the hermeneutic need for future anticipation. *Journal of Responsible Innovation*, 1(1), 99–102. <https://doi.org/10.1080/23299460.2014.882556>
- van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bocking, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, 614(7947), 224–226. <https://doi.org/10.1038/d41586-023-00288-7>
- Vesnic-Alujevic, L., Nascimento, S., & Pólvera, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, Article 101961. <https://doi.org/10.1016/j.telpol.2020.101961>
- Wallach, W., & Marchant, G.E. (2018). An agile ethical/legal model for the international and national governance of AI and robotics. In W. Wallach (Ed.), *Control and Responsible Innovation in the Development of AI and Robot* (pp. 45–59). The Hastings Center. <https://www.thehastingscenter.org/wp-content/uploads/Control-and-Responsible-Innovation-FINAL-REPORT.pdf>.
- Walsham, G. (2012). Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *Journal of Information Technology*, 27(2), 87–93.
- Weale, S. (2023, January 13). Lecturers urged to review assessments in UK amid concerns over new AI tool. The Guardian. <https://www.theguardian.com/technology/2023/jan/13/end-of-the-essay-uk-lecturers-assessments-chatgpt-concerns-ai>.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., & Gabriel, I. (2021). *Ethical and social risks of harm from Language Models* (arXiv:2112.04359). arXiv. <https://doi.org/10.48550/arXiv.2112.04359>
- Weizenbaum, J. (1977). *Computer power and human reason: From judgement to calculation (New edition)*. W.H.Freeman & Co Ltd.
- Wilkins, A. (2023, February 15). ChatGPT AI passes test designed to show theory of mind in children. New Scientist. <https://www.newscientist.com/article/2359418-chatgpt-ai-passes-test-designed-to-show-theory-of-mind-in-children/>.
- Wolf, M. J., Miller, K., & Grodzinsky, F. S. (2017). Why we should have seen that coming: Comments on Microsoft’s tay “experiment,” and wider implications. *ACM SIGCAS Computers and Society*, 47(3), 54–64. <https://doi.org/10.1145/3144592.3144598>
- Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology*, 13(3), 199–226. <https://doi.org/10.1007/s10676-010-9242-6>
- Xiang, C. (2023, February 28). OpenAI Is Now Everything It Promised Not to Be: Corporate, Closed-Source, and For-Profit. Vice. <https://www.vice.com/en/article/5d3naz/openai-is-now-everything-it-promised-not-to-be-corporate-closed-source-and-for-profit>.
- Zhang, C., Zhang, C., Li, C., Qiao, Y., Zheng, S., Dam, S. K., Zhang, M., Kim, J. U., Kim, S. T., Choi, J., Park, G.-M., Bae, S.-H., Lee, L.-H., Hui, P., Kweon, I. S., & Hong, C. S. (2023). *One small step for generative AI, one giant leap for AGI: A complete survey on ChatGPT in AIGC Era* (arXiv:2304.06488). arXiv. <https://doi.org/10.48550/arXiv.2304.06488>
- Zuboff, P.S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st edition). Profile Books.