AI & SOCIETY
https://doi.org/10.1007/s00146-023-01791-1

MAIN PAPER

Friend or foe? Exploring the implications of large language models on the science system

Benedikt Fecher[1,2] · **Marcel Hebing**[1,3] · **Melissa Laufer**[1] · **Jörg Pohle**[1] · **Fabian Sofsky**[1]

Received: 3 August 2023 / Accepted: 26 September 2023

Abstract
The advent of ChatGPT by OpenAI has prompted extensive discourse on its potential implications for science and higher education. While the impact on education has been a primary focus, there is limited empirical research on the effects of large language models (LLMs) and LLM-based chatbots on science and scientific practice. To investigate this further, we conducted a Delphi study involving 72 researchers specializing in AI and digitization. The study focused on applications and limitations of LLMs, their effects on the science system, ethical and legal considerations, and the required competencies for their effective use. Our findings highlight the transformative potential of LLMs in science, particularly in administrative, creative, and analytical tasks. However, risks related to bias, misinformation, and quality assurance need to be addressed through proactive regulation and science education. This research contributes to informed discussions on the impact of generative AI in science and helps identify areas for future action.

Keywords Large language models · Science system · Delphi study · Scholarly communication

1 Introduction

The release of ChatGPT by OpenAI in November 2022 has sparked a plethora of editorials, position papers and essays, or interviews with experts, as well as some articles and preprints on the potential impacts on science and higher education. While many concerns raised relate to how ChatGPT and large language models (LLMs) will change education (e.g., Perkins 2023; Fyfe 2023), there is much less—especially empirical research—on the implications of LLMs as well as LLM-based chatbots or prompts on scholarly practices and the science system, which we understand as a collective body of all academic disciplines, including the sciences and humanities (Ribeiro et al. 2023; Chubb et al. 2022). One can, however, draw inspiration from fields that are also characterized by largely text-based or -focused, creative and knowledge work. For instance, the opinion paper by Dwivedi et al. (2023) provides a viewpoint on the potential impact of generative AI technologies such as ChatGPT in the domains of education, business, and society, based on 43 contributions by AI experts from various disciplines. However, the literature on knowledge work and the transformative effects of AI cannot account for the complexities of specific practices (Jiang et al. 2022).

In light of the limited research conducted on LLMs and their impact on the science system and scientific practice, we initiated a Delphi study involving experts who specialize in the intersection of research and AI technology. The purpose of this study was to investigate the following areas: (a) the potential applications and limitations in using LLMs, (b) the positive and negative effects of LLMs on the science system, (c) the regulatory and ethical considerations associated with the use of LLMs in science, and (d) the necessary

✉ Benedikt Fecher
fecher@hiig.de

Marcel Hebing
marcel.hebing@hiig.de

Melissa Laufer
melissa.laufer@hiig.de

Jörg Pohle
joerg.pohle@hiig.de

Fabian Sofsky
fabian.sofsky@hiig.de

[1] Alexander von Humboldt Institute for Internet and Society, Berlin, Germany

[2] Wissenschaft im Dialog, Berlin, Germany

[3] DBU Digital Business University of Applied Sciences, Berlin, Germany

Published online: 26 October 2023

competencies and capacities for effectively utilizing LLMs. Our objective in this study was to gather and structure expert opinions in an initial phase, focusing on the aforementioned categories, and subsequently evaluate and assess them in a second phase. As generative AI continues to advance, it is crucial to gather expert knowledge and informed assessments regarding its potential impact on science. This knowledge will contribute to an informed scholarly debate and help anticipate potential fields of action.

Our findings indicate that experts anticipate that the utilization of LLMs will have a transformative and largely positive impact on science and scientific practice. In LLMs, they recognize significant potential for administrative, creative, and analytical tasks. The main risks associated with LLMs pertain to issues of bias, misinformation, and overburdening of the scientific quality assurance system. Despite the perceived advantages of LLMs for science, it is imperative to acknowledge and address the associated risks. This necessitates proactive measures in regulation and science education.

2 Literature review

In the following, we provide an overview of the current state of the scholarly discourse along the aforementioned areas. While our aim was to present a comprehensive and contemporary overview of this discourse, it is, however, important to acknowledge that new and pertinent studies may have emerged by the time of the publication of this article.

2.1 Applications and limitations of LLMs in science

LLMs and LLM-based tools are expected to have a wide range of applications in scientific practice. Possible uses for researchers identified in the literature range from generating plausible research ideas (Dowling and Lucey 2023), brainstorming (Staiman 2023), transforming notes into text (Buruk 2023), creating a first draft of a paper (Dwivedi et al. 2023), assisting with grammar and language (Flanagin et al. 2023), e.g., to improve clarity (Lund et al. 2023), especially for non-native speakers (Perkins 2023), but also stylistic issues, from formatting references to complying with editing standards (Flanagin et al. 2023; Lund et al. 2023). LLM-based tools like ChatGPT may be used to generate literature reviews (Dowling and Lucey 2023), data crunching (Staiman 2023), data summaries (Lucey and Dowling 2023), even proposing new experiments (Grimaldi and Ehrler 2023). They may support the dissemination of publications and the diffusion of knowledge by helping to create better metadata, indexing, and summaries of research findings (Lund et al. 2023). They are expected to assist editors in screening submission for issues such as plagiarism or image manipulation, triaging, validating references, editing and formatting (Flanagin et al. 2023; Hosseini and Horbach 2023). Beyond scholarly writing, LLM-based tools are expected to assist with code writing, automating simple tasks, and error management (Dwivedi et al. 2023), but also in writing reports, strategy documents, emails as well as cover and rejection letters (Corless 2023). They may even be used as a replacement for human participants in psychological experiments (Dillion et al. 2023). Scientists may also use LLM-based tools for non-scholarly tasks, as a recent *Nature* poll has shown: while eighty per cent of respondents have used AI chatbots, more than half say they use them for 'creative fun' (Owens 2023).

While the fields of application appear diverse, it appears that LLMs and LLM-based tools have limitations in scholarly use. Several editorials and Op Eds have been published that point to glaring mistakes of ChatGPT, including referencing scientific studies that do not exist (Perkins 2023). The company behind ChatGPT, OpenAI, admits openly in its blog: "ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers" (OpenAI 2022). At the time of writing this article, all existing LLM-based chatbots have been trained on outdated data. As a result, they do not possess the capability to incorporate real-time data automatically, leading to a lack of updated information (Dwivedi et al. 2023). Other limitations that have been identified include flawed logical argumentation, lack of critical elaboration, and unoriginal generated content (Dwivedi et al. 2023). Errors may also occur in interpreting meaning, in particular if terms are ambiguous, have multiple meanings or consist of compound words (Lund et al. 2023). Their limitations in simulating human comprehension, reasoning, and evolving views make them unsuitable as substitutes for human participants in psychological studies (Harding et al. 2023). In addition, generated texts may lack semantic coherence and lexical diversity (Perkins 2023). Teubner et al. (2023, p. 96) state that the produced texts often "read somewhat bland, generic, and vague with a noticeable tendency to seek balance", and that a very common ChatGPT phrase is: "However, it is important to note…". Like ML-based systems in general, LLM-based chatbots are considered to lack transparency and explainability (Dwivedi et al. 2023), and reproduce or even amplify biases inherent in the information that was used to train them (Corless 2023; Hosseini et al. 2023), reproducing an "of the same old trivialities and stereotypes" (Teubner et al. 2023, p. 99). This is considered a structural issue of how these systems are trained and cannot be resolved by simply creating bigger models as size does not guarantee diversity (Bender et al. 2021). Additionally, scholars have emphasized the significance of distinguishing specific explainability requirements among users or customers (Wulff and Finnestrand 2023).

2.2 Opportunities and risks for the science system

A prevailing viewpoint in the literature anticipates positive effects of LLMs on the science system. Potential opportunities of LLMs on science include positive effects on scholarly productivity, quicker access to available scholarly resources via enhanced search engines to the automation of mundane, repetitive or tedious work such as correcting grammatical errors, allowing people to focus on creative and non-repetitive activities (Dwivedi et al. 2023; Lund et al. 2023). Foremost among these anticipated benefits is the enhancement of research productivity and the elevation of publication quality. There is an expectation that using these tools to improve their texts, researchers "can focus more on *what* to communicate to others, rather than on *how* to write it" (Pividori and Greene 2023, p. 15). Staiman (2023 n.p.), for instance, notes in a guest post for the blog of the Society for Scholarly Publishing that the writing process should be considered less an end in itself but rather "a means to an end of conveying important findings in a manner that is clear and coherent". Along these lines, Lund et al. (2023) suggest that the capability of ChatGPT and the like might lead to questioning the strong belief that 'publish or perish' is an important and valuable principle in academia and possibly change the criteria for evaluating tenure. Some scholars expect a revolution of "the whole scientific endeavor" and refer to these tools' fundamental disregard of the boundaries of scientific disciplines, which may help "bringing multidisciplinary science to new heights" (Grimaldi and Ehrler 2023, p. 879). Furthermore, these tools may also lead to the democratization of science: First, the research process might be democratized as LLM-based tools may compensate for the lack of financial resources, e.g., for "traditional (human) research assistance", as a news article puts it (Lucey and Dowling 2023 n.p.). Second, the dissemination of knowledge might be democratized as these tools can easily polish the language of a text or even translate research output to multiple languages, both of which would level the field for researchers who speak English as a foreign language (Corless 2023; Liebrenz et al. 2023), or provide it in multimodal ways, including dialogical science communication at scale (Schäfer 2023).

Among the risks for the science system identified in the literature are the adverse effects on the academic quality assurance mechanisms and, subsequently, on scientific integrity. The avalanche of AI-generated "scientific-looking papers devoid of scientific content" (Grimaldi and Ehrler 2023, p. 879) is expected by some researchers to overburden the academic review process and foster plagiarism (Dwivedi et al. 2023). Biases are expected to be reinforced and errors introduced into the scholarly debate that might be difficult to identify and correct (Lund et al. 2023), including in peer review (Chubb et al. 2022). A recent study by Liang et al. (2023) evaluating the performance of several widely used GPT detectors found that they consistently misclassify non-native English writing samples as AI-generated, whereas native writing samples are accurately identified. Several scholars expect that LLMs may lead to an increase in misinformation and disinformation and more "junk science", as an article in Wiley's Advanced Science News formulates it (Corless 2023 n.p.). Lund et al. (2023) express apprehension regarding the utilization of LLM-based tools in academia, contending that this employment not only engenders apprehensions regarding research reproducibility and transparency but also has the potential to erode trust in the scientific process (see also Van Noorden 2022). Further, in an LSE Blog post, Beer (2019) raises concerns about the diminishing prospects for scientific serendipity and unexpected discoveries.

2.3 Competencies and capacities in scientific practice

It is assumed that LLMs and LLM-based tools will mark a shift in the academic skill set. Prompt engineering, i.e., developing and producing prompts for conversational AI systems like ChatGPT, is often discussed as a new competence that is required from researchers (e.g., Teubner et al. 2023). This is believed to pose a particular challenge for individuals who already struggle with basic IT, as they will not derive much benefit from advances in AI, and this may lead to a widening productivity gap. As LLM-based tools may have better English writing skills than some people, especially non-native speakers, the focus in academic work is expected to shift from text writing to conducting research, which requires researchers to formulate interesting research questions and carry out research to find answers (Dwivedi et al. 2023). More generally, as Teubner et al. (2023, p. 98) observe, "the ability to *read* and *interpret* different text options becomes more important than the ability to *write* them." That means that researchers must be able to check the generated text for factual and citation accuracy, bias, mathematical, logical, and commonsense reasoning, relevance, and originality, as Hosseini et al. (2023) demand in an Accountability in Research editorial. That also means that researchers are expected to have the competencies to collate and combine the results that LLM-based tools generate (Floridi and Chiriatti 2020). Not surprisingly, Dowling and Lucey (2023) find that adding domain expertise greatly improves the quality of the generated results. Thus, among the key skills that researchers have to develop are critical thinking, problem solving, ethical decision-making, and creativity (Dwivedi et al. 2023).

2.4 Ethical and regulatory issues concerning ChatGPT

The existing literature frequently mentions negative implications, i.e., risks, for the science system as ethical issues, and also mixes ethical and legal aspects. Issues are raised on how to understand 'authorship' in the research context, be it as accountability, as a substantial contribution to a text, as ownership in contrast to plagiarism, and with respect to text and language improvement, as Staiman (2023) argues in a guest post for The Scholarly Kitten. Critics argue that chatbots cannot take responsibility for the content they produce and cannot be held accountable (Corless 2023; Liebrenz et al. 2023). In addition, their ability to generate quality academic research ideas "raises fundamental questions around the meaning of creativity and ownership of creative ideas" (Lucey and Dowling 2023 n.p.), which in turn sparks questions about originality, scholarly citation practices and the boundary to plagiarism (Lund et al. 2023; Tomlinson et al. 2023). It, thus, comes as no surprise that publishers like *Springer Nature* have banned ChatGPT and similar software from being given authorship on papers, a position shared by many scientists, according to Stokel-Walker (2023), and *Science* editors even have prohibited the use of any text generated by those tools. Many commentators have raised concerns about the implications of the LLMs producing inaccurate or misleading output and the potential spread of misinformation (Dwivedi et al. 2023; Liebrenz et al. 2023). Similar ethical concerns are raised regarding the potential of these tools to reproduce and amplify bias, both in the training data and the development process, and the implications of this for the integrity of science (Lund et al. 2023). There have been techno-solutionist claims that potential harms of such systems can be mitigated by watermarking their output (Kirchenbauer et al. 2023). Additional ethical considerations include the potential to replace humans in the scholarly work process (Lund et al. 2023). This includes positions that were thought to be less likely to be automated until a few years ago (Dwivedi et al. 2023). Furthermore, the commercialization of these tools would exclude scholars and institutions in low-income and middle-income countries, thus entrenching existing inequalities in knowledge dissemination and scholarly publishing (Liebrenz et al. 2023).

There is a perceived lack of regulation, or at least clear regulatory guidance for LLMs and related tools, on issues such as privacy, security, accountability, copyright violations, disinformation, misinformation, and other forms of abuses and misuses of LLMs and LLM-based tools (Dwivedi et al. 2023; Khowaja et al. 2023; Lund et al. 2023). After the Italian Data Protection Authority imposed an immediate temporary limitation on the processing of Italian users' data by OpenAI in late March 2023 to enforce demands on the protection of data subjects' rights, as outlined in their press release (GPDP 2023), other national data protection authorities in Europe have followed suit and opened proceedings against OpenAI, reports Sokolov (2023). European data protection authorities have even set up a task force to cooperate and exchange information on enforcing EU laws on OpenAI, according to a news report (Goujard 2023). At the same time, the European Parliament called for expanding the potential reach of the proposed EU AI Act by including ChatGPT-like systems to the list of high-risk categories of AI systems (Helberger and Diakopoulos 2023). Furthermore, Hacker et al. (2023) call for specific regulation of LLM-based tools, "large generative AI models", under the EU Digital Services Act and provide four concrete, workable suggestions that include transparency obligations, mandatory yet limited risk management, non-discrimination data audits, and expanded content moderation.

3 Methodology

To address our research objective, we employed the Delphi method. First developed in the 1960s, the Delphi method is a technique used to establish consensus among a group of experts on complex issues (Landeta 2006) and in some cases used to forecast future developments (Linstone and Turoff 1975). In its basic form, this method can be described as a communication process that involves engaging experts at various stages, such as through surveys and qualitative interviews. The initial stage is open and exploratory, with the information gathered analyzed and used to inform subsequent data collections. This process continues until consensus is reached among experts, for example, in defining concepts and/or trends or weighing different viewpoints. In this light, the Delphi method is a fitting technique to investigate our objective of exploring the impact of ChatGPT and LLMs on scientific practices and the science system.

In our Delphi approach, we conducted two surveys. In the first survey, we mainly used open questions along the research questions to derive a category system. This category system was the basis for the second survey, where we used the sub-categories to create closed questions for better comparison and quantitative evaluation. Our target audience were researchers working on topics that crosscut science, technology, and society, who had an interest in LLMs. Using convenience sampling, we recruited participants via our professional networks, including using institutional mailing lists and associations, e.g., the Network of Centers, an international association of internet research centers.

The first survey consisted of 12 open questions with the goal of understanding the impact of ChatGPT and LLMs on academic work, scientific practices, and the science system. In total, we collected 72 responses from researchers holding various positions and from diverse disciplinary backgrounds

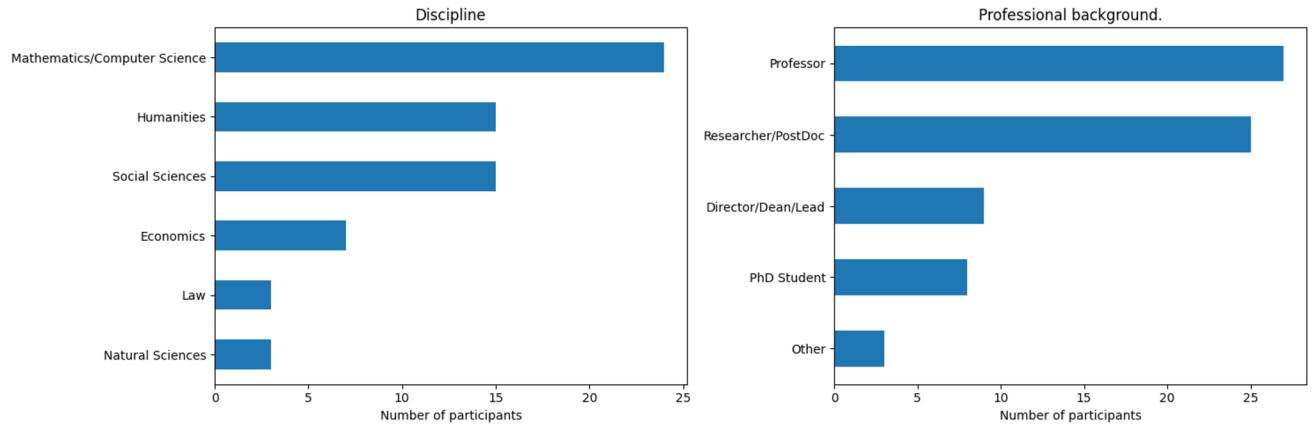

Fig. 1 Initial sample (round one) of our Delphi study. Overview of participants by discipline and professional status ($n = 72$)

(see Fig. 1). The responses of the first survey were primarily coded by two authors. In a first step, they examined 25% of the responses to generate a codebook through a combination of inductive and deductive coding (Bazeley 2009). In a second step, the codebook was then evaluated by all authors and adjustments were made when needed and the rest of the material was coded (for codebook see Online Appendix tables 7 and 8).

Based on the results and, in particular, the codebook from the first round, the second survey was created consisting of 11 questions, the majority of which were ranking questions featuring the identified codes for applications and limitations, risks and opportunities for the science system and the competencies needed for using LLMs, as well as general opinion questions on LLMs impact on science and scientific practice. Furthermore, the survey instrument contained two open questions on future scenarios.

The survey was sent to the same experts, yielding 52 responses (72% of the participants from the first round). A statistical analysis was conducted on the opinion and ranking questions. In the result tables (see Tables 2–6 in the Online appendix), we provide the individual frequencies for each item and rank, as well as two scores. The first score (sum) is a simple sum of the preceding frequencies, the rank is a weighted sum, where the first rank is weighted by factor four and the second rank by factor two. The rank questions are followed by a set of statements, which the participants could evaluate on a five-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree). We combined agree and strongly agree to sort the items and will also refer to the combination of both, when reporting it in the text. The open questions were analyzed with a combination of inductive and deductive coding, carried out jointly by the authors. Our Delphi approach allowed us to identify and refine various implications of LLMs on the science system; however, it was not without its limitations. For example, we were unable to track long-term implications as the interval between the data collections were relatively short.

We sought consent prior to each survey phase to publish the responses, aiming to enhance the transparency of our results and enable future research and educational use. The data (including the survey instruments) is published under a CC-BY-license and can be accessed via the following link.

4 Results

Below, we present the Delphi study results based on the defined aspects, i.e., applications and limitations, risks, and opportunities for the science system, competencies as well as legal and ethical implications. In each section, we begin by presenting the coded findings from phase one and use the results of the ranking and opinion questions to contextualize and weigh these results, when applicable. Figure 2 displays the results of the opinion questions, which we will refer to in the subsequent result sections. The results of the ranking questions analysis can be found in the Online appendix (Tables 2–6).

4.1 Applications and limitations in use: LLMs as enhancement tools

The first phase yielded six distinct applications that can be effectively addressed by LLMs and LLM-based applications. These include (1) *text improvement*, which involves the rephrasing and optimization of textual content, (2) *text summary*, which involves the summarization of information, (3) *text analysis*, such as the use of sentiment analysis or qualitative coding, (4) *code writing*, which involves assistance in programming tasks, (5) *idea generation*, which involves generating new ideas through the combination of concepts, and (6) *text translation*, which includes the translation of a

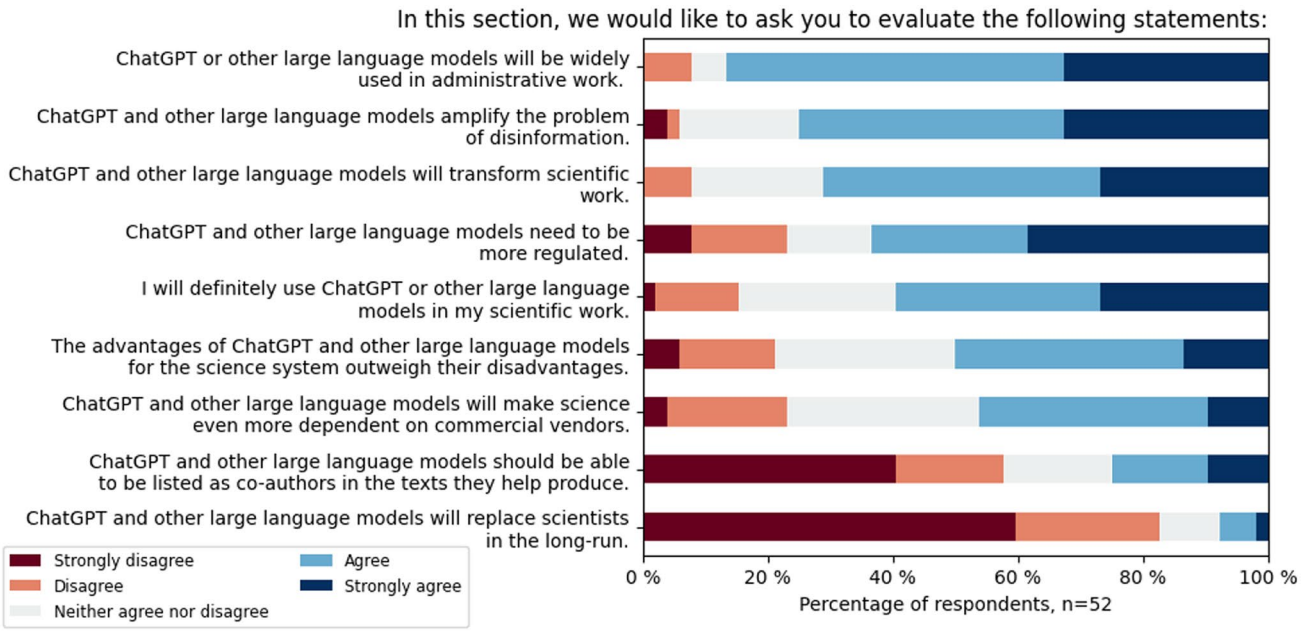

Fig. 2 Statements on LLMs, formulated based on the results from the first round of our Delphi study and quantified in the second round

text entered into the LLM in different languages. Ideally, one respondent argued, this 'time-saving' potential could be harnessed to jumpstart the writing process:

> "I think if we can figure out how to use it properly, it can be a good writing aid, for instance to get over writer's block. It can also be good at explaining things holistically, since it synthesizes a lot of material, and can thus be a tool to explore a subject."

Notably, the identified applications of LLMs extend beyond conventional text-based tasks in scientific publishing, although such tasks remain a dominant practice in the responses.

In the second round of the Delphi survey, we asked the experts to prioritize the identified applications. Our results show that *text improvement* is considered the most important application, followed by *text summary* as the second most important, and *code writing* as the third most important application. Most (59.6%) of the experts either already use or express their intention to use LLMs in their own work (Fig. 2). A significant portion (86.5%) of the experts perceive LLMs as valuable for administrative tasks, confirming the assumption that time savings are expected for researchers through LLM utilization (Fig. 2).

Asked about the limitations of LLMs in scientific work, five distinct types of limitations were mentioned. We observed (1) *lack of transparency*, as it is unclear on which data the model's outputs are based on, (2) *incorrectness*, especially regarding literature references and biographical information, which may affect the reliability of the generated text, (3) *lack of creativity*, as ChatGPT relies heavily on existing patterns and may struggle to generate entirely new content, (4) *outdatedness*, particularly as the version of ChatGPT used in this study relies on a database that only goes up to 2021, and (5) *unspecificity*, i.e., LLMs produce superficial texts that do not address topics in depth or detail. One respondent recounted their experimentation with ChatGPT which illustrates the unreliability of LLM-generated text:

> "I have played with ChatGPT a lot recently, and I have tried to ask it to perform various scientific tasks of different complexities, for example explaining in plain words a very complicate scientific topic or, conversely, explaining it in a very "sciency" way, including equations and references to corroborate explanations. If on one hand, it performed remarkably well in simplifying hard science into plain words (often naively, but that's ok), it performed really poorly when explaining a topic in detail, providing the wrong/incorrect/incomplete set of equations as well as MAKING UP references. Interestingly, ChatGPT uses names of real people well known in the subject field, it mix[es] and match[es] them, creat[ing] a fake reference title…"

In the second round of the study, participants were requested to rank the limitations. The highest-ranked limitation was *incorrectness*, followed by *non-transparency* and *unspecificity* in the responses. The incorrectness of LLMs was a dominant and recurring issue mentioned by the experts. As one expert stated, "The largest problem I

see are the factual mistakes, often given with confidence, which make it hard to trust ChatGPT and similar technology outputs without further research or prior knowledge".

The results indicate that the potential benefits of LLMs lie not only but primarily in text-based work, which is significant because scientific value creation in most disciplines is text based. There is also evidence to suggest that LLMs are relevant for ideation, conception, and programming, the latter of which is an increasingly important scientific practice. Taken together, it is not surprising that a majority of the respondents assume that ChatGPT and other LLMs will transform scientific practice, although this might—at this stage—relate primarily to the textuality of academic work. The limitations mentioned can be essentially explained by the databases that existing LLMs were trained on, and it can be assumed that many of these limitations can be addressed in newer models, as some respondents pointed out. However, the non-transparency in the training data remains problematic and was viewed by some as inconsistent with scientific principles of quality.

4.2 Risks and opportunities for the science system: advantages trump disadvantages

According to the experts, the use of LLMs provides the science system with four opportunities: (1) LLMs can promote *efficiency* by automating and supporting text work, (2) LLMs may promote *reflection* by identifying biases and new research areas, (3) LLMs may *reduce administrative workload*, (4) LLMs can promote *inclusiveness* by leveling the playing field between researchers from different backgrounds and institutions, such as those who lack resources for grant writing or those who are non-native English speakers, and (5) LLMs promote *productivity* by freeing up time for researchers to conduct more analyses or produce more scientific articles. In the second phase of the Delphi study, the experts ranked these, with the *reduction of administrative tasks* ranked first, followed by more *efficiency* and *inclusiveness* (see Table 4). These results indicate that researchers see LLMs primarily as a tool to relieve and simplify their workload. Hence, a large majority of the experts disagrees that LLMs could replace researchers (82.7%, Fig. 2).

The analysis of the first phase of the Delphi study reveals the existence of seven distinct risks associated with the use of LLMs in scientific work. These risks include (1) reinforce *bias* / *dominant voices*, because statistical systems favor mainstream opinions, (2) *overburden academic quality assurance mechanisms* with semi-automated papers, (3) reinforce *inequalities* between researchers who have access to LLMs and those who do not, (4) increase *dependence* on commercial providers, (5) encourage academic *misconduct*, either intentional or unintentional by researchers, (6) lead to a *decrease in originality* due to the generic nature of LLM-generated text, and (7) the possibility to an increase in *disinformation*, which could potentially challenge scientific truths in the public domain. One respondent summarized these key problems for scientific practices in the following quote:

> "ChatGPT and possibly other large language models may make production of plausibly looking, but false content easy and low-cost. This presents significant risks and could lead to overload of the peer-review system. It could also homogenise scientific outcomes, reducing breakthrough innovations."

In the second phase, the experts ranked these, indicating that *bias* is seen as the biggest threat, followed by *disinformation* and *overburdening academic quality assurance mechanisms* (see Online appendix Table 5). These risks are significant as they touch on fundamental pillars of scientific ethics and good practice, such as scientific freedom regarding the dependence on commercial publishers, scientific quality assurance concerning the handling of highly generic publications, as well as the public legitimation of science, which could be put into question by plausible and seemingly scientific nonsense produced by LLMs—large majority of the experts (75.0%) regard LLMs as a catalyst for disinformation (Fig. 2). Notwithstanding the gravity of the aforementioned risks, the majority of experts perceive the benefits of LLMs to outweigh the drawbacks (Fig. 2), which explains why most of them already use or intend to use LLMs in their work. This, however, can also be attributed to the sampling strategy employed in this study, possibly involving technology-proficient experts. This result is noteworthy nonetheless and supports the hypothesis that generative AI will change scientific work in the long run.

4.3 Competencies in usage: scientists need to learn to (re)think

In the inquiry regarding the competencies required for researchers to utilize ChatGPT and other LLMs, the respondents pointed out four distinct competencies, namely (1) *technical know-how* to comprehend the inner workings of LLMs, (2) the *ability to contextualize results* utilizing the outcomes generated by LLMs in practical scenarios, (3) a *reflective mindset* to consider the feedback effects on scientific practice, and (4) *ethical understanding* to responsibly employ LLMs. In the second phase, they ranked a *reflective mindset* first, followed by the *ability to contextualize results* and *ethical understanding*. One respondent succinctly put it scientists need "common sense", understanding that:

> "This is a tool that paraphrases its original knowledge and has (as of yet) no evaluation of the quality of its own answer. Which is not an real issue when you use

it for valentine cards (well, actually...) , but is when you take that output as factual knowledge (publication, protocols, ...) The more the result is meant to be formal factual knowledge, the more it should be considered suspect and reviewed by humans."

The results indicate that the experts anticipate feedback effects on science, while also suggesting that the responsible application of knowledge will become even more paramount in the future.

It can be argued that reflexivity highlights the ethical implications of AI on scientific practices and ways to proactively address them, while contextuality focuses on the practical use of AI-supported findings and strategies for maximizing their utility. Our findings suggest that generative AI should be incorporated in scientific training and science education, specifically in relation to scientific ethics and effective communication of AI-driven results in their appropriate context.

4.4 Ethical and legal implications: clear need for regulation

The answers in the first phase allow to discern five ethical implications, namely (1) the *need for accountability in relation to the outcomes produced by LLMs*, (2) the question of *originality with regards to human creativity* (e.g., concerns of plagiarism arise), (3) the *sustainability issue* regarding the environmental effects of LLMs, (4) the *potential exclusion of researchers* who lack access to LLMs, raising concerns about universalism, and (5) the *issue of autonomy*, in which researchers may become overly dependent on (commercial) AI tools. The comments show clearly that the majority deem ChatGPT unfit for authorship due to its inability to assume responsibility for the results.

The experts perceive legal implications regarding (1) *copyright*, due to the unclear infringement of intellectual property by LLMs, (2) *data protection*, due to the ambiguity of the data used and how OpenAI utilizes input data, and (3) *liability*, due to the uncertainty of the extent to which LLMs can be held responsible for criminal errors. A large majority of the experts (63.5%) believe that LLMs should be subject to stronger regulations (Fig. 2).[1] However, all these issues, one respondent pointed out, need to start with a discussion on how we view AI tools:

"Do we consider the models as a human-like something or as a tool? This affects all ethical and legal aspects."

[1] We did not ask the participants to rank the ethical and legal implications in the second round of the Delphi survey.

The initial round of the Delphi survey revealed that the ethical implications discussed frequently underscore the significance of the human element in scientific endeavors. This includes the responsibility and accountability of individuals for their contributions, the value of creativity and generating novel ideas, ensuring equitable access to science and the scientific community, and addressing the potential risk of dependency on LLM-based tools that may hinder individual skills and capabilities in scientific work. The amount of energy that is necessary both for training models and running inference and the CO_2 footprint are mentioned as primary examples for the ecological sustainability issues ChatGPT and the like present. Taking into account that LLMs are trained on works produced by others and produce (or co-produce) works, both of which almost certainly fall under copyright law, it is not surprising that a large majority of the experts identify issues with copyright law as a pressing legal implication. The lack of transparency regarding the personal data on which the LLMs were trained, but also the further possible uses of personal data generated by the use of the tools, certainly explains why many respondents identify privacy and data protection law issues as considerable legal challenges. Whereas accountability is identified by many experts as a key ethical challenge, this does not carry over to the legal principle of liability that builds on it, which is mentioned by relatively few respondents.

4.5 Transformative and deformative scenario

In the first phase of the Delphi, we consulted with experts to ascertain the potential impact of LLMs on scientific practice within the next 5–10 years. In the subsequent phase, we investigated the potential influence of generative AI on the relationship between science and society. Based on the answers to these questions, our study reveals two possible scenarios, namely (1) a utopian transformative scenario and a (2) dystopian deformative scenario. It is noteworthy that the negative scenario is almost a negation of the positive scenario and vice versa. However, overall, there are significantly more indications (in terms of the number of codes) for a positive scenario, which was also confirmed by the opinion battery in Phase 2.

In the utopian scenario, integrating generative AI into scientific practices offers transformative potential, overcoming path dependencies in scientific practice and accelerating scientific and societal progress. Our analysis identifies three key aspects of its impact on science: (1) streamlining repetitive tasks, (2) promoting inclusivity, and (3) facilitating interdisciplinary research. The experts propose that generative AI could automate administrative and generic tasks, freeing up time for critical reflection, analysis and innovation. In this light, one respondent explained:

> "Generative AI will have a tremendous impact on the way that research is done. Mostly with regard to improving productivity and efficiency. More repetitive and monotonous tasks will be outsourced to the AI models, while humans can focus more strongly on contextualising, research design, and creative thinking."

It may democratize access to scientific resources, foster diversity of voices and collaboration, and aid in discovering connections across different schools of thought. The integration of generative AI tools aligns research with societal challenges, driving technological development and supporting evidence-based decision-making. Effective science communication and education are enabled through AI-driven tools. This collaborative approach propels scientific advancements toward innovative solutions.

In a dystopian scenario, the anticipated positive impacts of generative AI are largely negated, as our analysis reveals three crucial aspects: (1) a decline in research quality due to plausible yet flawed results, compromising reliability and validity; (2) a loss of research diversity through amplifying mainstream voices, resulting in missed opportunities for novel perspectives; and (3) a decrease in scientific integrity, as the ease of producing AI-generated content raises risks of reinforcing predatory publishing practices and disseminating false information, leading to confusion and distrust.

> "…. there's a greater deluge of predatory publishing practices being driven by LLM, that is that trust in science will decrease. People will then have to become sufficiently familiar with the scientific system to be able to judge the scientific merit of a paper, which can be a real challenge. Surely some scientific journalists will be able to help filter out (some) problematic papers from reaching newspapers and from reaching the broader public. But with material being increasingly openly available, and directly engaged with by society, if there are more such problematic papers, this makes it increasingly difficult for the lay public to assess the paper's merits…"

The perpetuation of plausible nonsense could further have negative consequences for society when policy decisions or public opinions rely on unreliable information.

Additionally, dependence on commercial providers for generative AI tools raises concerns among experts about the lack of independence and control over scientific research, potentially leading to conflicts of interest and biases in research results.

5 Discussion

The aim of this study was to investigate the impact of ChatGPT and other LLMs on the science system and scientific practices by examining their potential applications, limitations, effects, ethical and legal considerations and the necessary competencies needed by users. To date, scholars have primarily focused on the implications of LLMs on education (e.g., Perkins 2023; Fyfe 2023) with limited attention being paid to their impact on science and scientific practices (for exceptions, see Chubb et al. 2022; Ribeiro et al. 2023). The overnight popularity ChatGPT experienced since its debut in November 2022 stressed even more the necessity to evaluate the implications of LLMs for science and scientific practice. To examine these implications, we employed a two-stage Delphi method, which included inviting experts, researchers working in the fields of science, technology and society to participate in two surveys as means to identify and refine the impact of LLMs on the science system and scientific practices.

At the time of the second round of our Delphi method, less than half a year had passed since the first preview of ChatGPT. Accordingly, it is difficult to make concrete predictions about the potential capabilities of future versions of LLMs like ChatGPT. Nevertheless, our study presents a consistent picture from experts which can further our understanding of future expectations of LLMs. We were also able to identify patterns emerging regarding potential opportunities and risks. It is important to note the majority of the experts saw no danger that LLMs will replace researchers in the foreseeable future. They share this expectation with researchers in the field of AI and work, who in their optimistic scenarios expect a shift in tasks, the creation of new tasks and an emergence of new work profiles rather than the replacement of workers (Deranty and Corbin 2022).

Overall, the experts in our study were optimistic and in agreement that the advantages of this technology outweigh their disadvantages. This optimism was paired with concerns, which allow us to paint a nuanced picture of the potential positive and negative implications of LLMs. In general, ChatGPT and other LLMs were collectively understood as potential 'time-savers' to be used to improve and streamline the writing process, especially academic writing. For example, *text improvement* as in the rephrasing and optimization of textual content was considered the most important application. This outcome resonates with the scholarly discourse which highlights how generative AI can be used to enhance texts, such as with brainstorming (Staiman 2023), crafting literature reviews (Lucey and Dowling 2023), and improving text clarity (Lund et al. 2023). At the same time, experts in our study were aware

of the limitations of LLMs and cited similar apprehensions to those raised in the literature (Dwivedi et al. 2023; OpenAI 2022; Perkins 2023). The experts highlighted key shortcomings such as AI- produced texts may have incorrect information, their origin and referencing is non-transparent and that they lack specificity, shortcomings which are at odds with the principles of good scientific practice. It is not surprising that our study reinforced text-based applications and limitations for LLMs identified in the scholarly discourse, as text production is a key scientific practice. However, this focus may likely shift in the future as more usages of LLMs are explored.

In addition, our study indicates that LLMs have the potential to reshape the science system. The experts anticipate that they will lead to more efficient workflows, with the reduction of administrative tasks being ranked the highest anticipated change. This forecast supports claims made by other scholars, who argue that LLMs will help automate mundane tasks and free up space for creative thinking (Lund et al. 2023; Dwivedi et al. 2023). Other changes LLMs bring to the science system are, however, more complex. For example, our findings point to a double-edged sword embedded within the LLM constellation: this technology could serve to both promote inclusion and reinforce biases. On the one hand, LLMs can level the playing field for non-English speakers as they can provide editorial support, but on the other hand, they can also increase inequalities by drawing on mainstream opinions and widening the gap between those who have access to these technologies and those who do not. This multifaceted concern was also echoed by other scholars (Corless 2023; Liebrenz et al. 2023).

The most pressing fear we identified is that LLMs perpetuate disinformation and will overburden quality assurance mechanisms in academia. In other words, "scientific truths" could be in greater competition with plausible nonsense than they already are. Similar thoughts are discussed by other scholars (Grimaldi and Ehrler 2023; Lund et al. 2023), with these changes being described in revolutionary terms in which LLMs are positioned as the great 'game-changers' of academia. In contrast, experts in our study were more cautious with such claims seeing these changes as more incremental and pragmatic.

Moreover, our study provided insights into the competencies researchers need to be able to utilize LLMs. In line with scholars such as Teubner et al. (2023), experts in our study voiced concerns that ChatGPT and other LLMs have the potential to widen the digital divide between researchers who possess technical know-how and researchers who do not. Furthermore, the experts pointed out that the researcher's role in the writing process will shift from being the originator of ideas and texts to being required to contextualize and reflect on AI-generated results. This change will entail a new way of thinking about key scientific practices and the role the individual academic plays in them. Our experts also expressed the importance of researchers having an ethical understanding, e.g., using AI in a responsible manner. A point that was only marginally addressed in the literature (Dwivedi et al. 2023). Underlying these findings is the understanding that it is up to the individual academic to ensure that they have the skills and knowledge needed to navigate these technological changes. Such a stance can contribute to furthering digital divides due to preexisting uneven digital literacy between academics, institutions, and higher education systems.

This study aided in disentangling the ethical and legal implications of ChatGPT and other LLMs. The findings provide additional clarity on the matter of authorship concerning AI utilization, a subject also explored by other scholars (Lucey and Dowling 2023; Tomlinson et al. 2023). Even if our perspectives on authorship and language are not to be understood as outdated, as some scholars argue (e.g., Coeckelbergh and Gunkel 2023), they should at least be reconsidered in light of the potential impact of ChatGPT and similar LLM-based tools on our conventional understandings (Gellers 2023). The majority of the experts deem that LLMas cannot claim authorship due to its inability to assume responsibility for its actions. In this light, the experts centered on distilling the role humans play in being accountable for their usage of LLMs, taking into consideration issues such as plagiarism, copyright and data protection. Thus, they underlined that human responsibility in AI usage is both a legal and ethical challenge, a sentiment that echoes the arguments of critics who postulate that chatbots cannot take responsibility for their actions (Corless 2023; Liebrenz et al. 2023). In addition, the issue of access was highlighted as an ethical dilemma, that is, not all researchers will have equal access to such technologies, potentially furthering inequalities. Furthermore, the CO_2 emissions generated by these use of AI technologies poses environmental risks (Hao 2019). The complexities of these ethical and legal implications show the need to take diverse issues into account when it comes to regulating the usage of LLMs in academia.

Lastly, our study presents potential future pathways for AI and its impact on the science system and society in the form of future scenarios constructed from our data. In the positive transformative scenario, the integration of LLMs in scientific practice holds great potential for improving scientific productivity, efficiency, education, communication, creativity, and discovery. In other words, LLMs can automate repetitive tasks, allowing researchers to allocate more time and resources to analytical and innovative work. It is the prevailing perception of the experts in our study that suggests that this scenario is more likely to occur. However, it is crucial to acknowledge the potential negative deformative scenario. Experts raised concerns about the impact of generative AI on scientific quality, integrity, and the scientific

ecosystem. Issues such as decreased scientific rigor, reproducibility, and a potential homogenization of science were highlighted. In addition, the reliance on generative AI models without proper validation may lead to a decrease in critical thinking and creativity.

We can strive to ensure the positive scenario by addressing the concerns highlighted in our study. In this regard, striking a balance between embracing the benefits of LLMs and upholding scientific principles is crucial. Accordingly, we should remember our scientific tools, the good practices of scientific work, and create appropriate frameworks and conditions that enable us to make use of the diverse opportunities these technologies might have to offer. At the same time, we must withstand any attempt to compromise the quality standards that we as a science community have established and which distinguishes the scientific discourse.

In conclusion, while the transformative scenario holds great promise for the positive impact of LLMs on the science system and society, it is imperative to proactively address the potential risks and challenges to ensure that the integration of generative AI in science is guided by ethical considerations, scientific integrity, and a commitment to societal benefit.

6 Conclusion

Our study highlights the great potential generative AI has for transforming the science system. This adds a new and fresh direction to the LLM discussion, which has focused heavily on the impact of this technology for individual academics and learners (Fyfe 2023; Perkins 2023). Generative AI, exemplified by LLMs, presents a transformative prospect for the social organization of science. LLMs' text generation capabilities have implications for scholarly communication and knowledge dissemination, potentially redefining conventional norms such as the academic reputation or quality control system. For instance, regarding the reputation system, citation-based metrics, pivotal for scholarly progress, could conceivably diminish in value if scientific articles could be predominantly generated through automation. Potential avenues might encompass a heightened emphasis on micro-publications or alternative measures of success (e.g., peer recognition, grant acquisitions, performative accomplishments such as presentations, etc.). In the context of an increasingly automated content generation landscape, the dynamics of scholarly evaluation warrant exploration to ensure the relevance and robustness of metrics within evolving scholarly communication paradigms. Regarding quality control, it appears plausible that the quality assurance system might experience increased strain due to a potential rise in LLM-assisted articles, or conversely, that the implementation of generative AI could be employed for quality assurance purposes. Furthermore, these tools offer the prospect of ameliorating administrative burdens on researchers, thus reorienting focus toward analytical facets of scientific endeavors. However, this theoretical promise is counterbalanced by the instantaneous generation of potentially misleading content, raising concerns about scientific integrity and the nature of evidence. Consequently, the integration of generative AI necessitates nuanced evaluation and recalibration of policies to ensure judicious incorporation, thereby molding the future contours of scientific inquiry. The broader societal ramifications of integrating LLMs demand that the scientific community meet its responsibilities to society, engage in open and public discussions on the ethical considerations related to these technologies, and identify suitable proactive regulation approaches.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00146-023-01791-1.

Acknowledgements We express our gratitude to the participating experts, the majority of whom have given their consent to be named in the second phase of the Delphi. These experts are Alexander Terenin, Alison Kennedy, Anaëlle Gonzalez, André Vellino, Andrea Klein, Benjamin Tan, Brigitte Mathiak, Christian Gagné, Christian Vater, Daniel Guagnin, Debora Weber-Wulff, Ekaterina Hertog, Eva Seidlmayer, Evgeny Bobrov, Fabro Steibel, Fiona Kinniburgh, Florian Hoffmann, Georg von Richthofen, Graham Taylor, Hadi Asghari, Hendrik Send, Ingrid Richardson, Johannes Breuer, Katharina Mosene, Klaus Gasteier, Marina Gavrilova, Mark Spektor, Martin Schmidt, Maximlian Heimstädt, Mike Thelwall, Naireet Gosh, Natalie Sontopski, Philipp Mehl, Richard Boire, Robert Lepenies, Ronny Röwert, Sebastian Moraga Scheuermann, Thorsten Thiel, Tony Ross-Hellauer, Vince I. Madai, Vincent Traag, Wojciech Hardy, Zining Zhu. We also extend our thanks to the 9 experts who preferred to remain anonymous. We furthermore would like to extend our appreciation to our colleagues from the Global Network of Internet and Society Research Centers (https://networkofcenters.net/) for their invaluable assistance in the recruitment of experts and their insightful contributions in shaping initial ideas.

Funding This study was funded and primarily conducted by the Alexander von Humboldt Institute for Internet and Society.

Data availability The datasets generated during and/or analyzed during the current study are available on Zenodo repository, https://doi.org/10.5281/zenodo.8009429

Declarations

Conflict of interest The authors and affiliated institutions declare no conflicting interests.

References

Bazeley P (2009) Analysing qualitative data: more than 'identifying themes.' Malays J Qual Res 2(2):6–22

Beer D (2019) Should we use AI to make us quicker and more efficient researchers?. Impact of Social Sciences—LSE Blog. https://blogs.lse.ac.uk/impactofsocialsciences/2019/10/30/should-we-use-ai-to-make-us-quicker-and-more-efficient-researchers/

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big?. In: Proceedings of the 2021 ACM Conference on fairness, accountability, and transparency, pp 610–623. https://doi.org/10.1145/3442188.3445922

Buruk O "Oz" (2023) Academic writing with GPT-3.5: reflections on practices, efficacy and transparency. arXiv http://arxiv.org/abs/2304.11079

Chubb J, Cowling P, Reed D (2022) Speeding up to keep up: exploring the use of AI in the research process. AI & Soc 37(4):1439–1457. https://doi.org/10.1007/s00146-021-01259-0

Coeckelbergh M, Gunkel DJ (2023) ChatGPT: deconstructing the debate and moving it forward. AI & Soc. https://doi.org/10.1007/s00146-023-01710-4

Corless V (2023) ChatGPT is making waves in the scientific literature. Advanced Science News. https://www.advancedsciencenews.com/where-and-how-should-chatgpt-be-used-in-the-scientific-literature/

Deranty JP, Corbin T (2022) Artificial intelligence and work: a critical review of recent research from the social sciences. AI & Soc. https://doi.org/10.1007/s00146-022-01496-x

Dillion D, Tandon N, Gu Y, Gray K (2023) Can AI language models replace human participants? Trends Cognit Sci 27(7):597–600. https://doi.org/10.1016/j.tics.2023.04.008

Dowling M, Lucey B (2023) ChatGPT for (Finance) research: the Bananarama Conjecture. Financ Res Lett 53:103662. https://doi.org/10.1016/j.frl.2023.103662

Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, Baabdullah AM, Koohang A, Raghavan V, Ahuja M, Albanna H, Albashrawi MA, Al-Busaidi AS, 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. Int J Inform Manag 71:102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL (2023) Nonhuman "Authors" and implications for the integrity of scientific publication and medical knowledge. JAMA 329(8):637. https://doi.org/10.1001/jama.2023.1344

Floridi L, Chiriatti M (2020) GPT-3: its nature, scope, limits, and consequences. Mind Mach 30(4):681–694. https://doi.org/10.1007/s11023-020-09548-1

Fyfe P (2023) How to cheat on your final paper: Assigning AI for student writing. AI & Soc 38(4):1395–1405. https://doi.org/10.1007/s00146-022-01397-z

Gellers JC (2023) AI ethics discourse: a call to embrace complexity, interdisciplinarity, and epistemic humility. AI & Soc. https://doi.org/10.1007/s00146-023-01708-y

Goujard C (2023) European data regulators set up ChatGPT task force. Politico. https://www.politico.eu/article/european-data-regulators-set-up-chatgpt-taskforce/

GPDP (2023) Intelligenza artificiale: Il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori. https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870847

Grimaldi G, Ehrler B (2023) AI et al.: machines are about to change scientific publishing forever. ACS Energy Lett 8(1):878–880. https://doi.org/10.1021/acsenergylett.2c02828

Hacker P, Engel A, Mauer M (2023) Regulating ChatGPT and other Large Generative AI Models. arXiv https://arxiv.org/abs/2302.02337

Hao K (2019) Training a single AI model can emit as much carbon as five cars in their lifetimes. MIT Technology Review. https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Harding J, D'Alessandro W, Laskowski NG, Long R (2023) AI language models cannot replace human research participants. AI & Soc. https://doi.org/10.1007/s00146-023-01725-x

Helberger N, Diakopoulos N (2023) ChatGPT and the AI act. Internet Policy Rev. https://doi.org/10.14763/2023.1.1682

Hosseini M, Horbach SPJM (2023) Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other Large Language Models in scholarly peer review. Res Integr Peer Rev 8:4. https://doi.org/10.1186/s41073-023-00133-5

Hosseini M, Rasmussen LM, Resnik DB (2023) Using AI to write scholarly publications. Acc Res. https://doi.org/10.1080/08989621.2023.2168535

Jiang M, Breidbach C, Karanasios S (2022) How does artificial intelligence transform knowledge work? PACIS 2022 Proceedings, 312. https://aisel.aisnet.org/pacis2022/312

Khowaja SA, Khuwaja P, Dev K (2023) ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) evaluation: a review. arXiv http://arxiv.org/abs/2305.03123

Kirchenbauer J, Geiping J, Wen Y, Katz J, Miers I, Goldstein T (2023).A watermark for large language models. arXiv https://arxiv.org/abs/2301.10226

Landeta J (2006) Current validity of the Delphi method in social sciences. Technol Forecast Soc Chang 73(5):467–482. https://doi.org/10.1016/j.techfore.2005.09.002

Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J (2023) GPT detectors are biased against non-native English writers arXiv 2304.02819. http://arxiv.org/abs/2304.02819

Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A (2023) Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. Lancet Digit Health 5(3):e105–e106. https://doi.org/10.1016/S2589-7500(23)00019-5

Linstone HA, Turoff M (1975) The Delphi method. Addison-Wesley, Reading

Lucey B, Dowling M (2023) ChatGPT: our study shows AI can produce academic papers good enough for journals—just as some ban it. The Conversation. https://theconversation.com/chatgpt-our-study-shows-ai-can-produce-academic-papers-good-enough-for-journals-just-as-some-ban-it-197762

Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z (2023) ChatGPT and a new academic reality: artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. J Assoc Inform Sci Technol. https://doi.org/10.1002/asi.24750

OpenAI (2022) Introducing ChatGPT. https://openai.com/blog/chatgpt

Owens B (2023) How Nature readers are using ChatGPT. Nature 615(7950):20–20. https://doi.org/10.1038/d41586-023-00500-8

Perkins M (2023) Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. J Univ Teach Learn Pract. https://doi.org/10.53761/1.20.02.07

Pividori M, Greene CS (2023) A publishing infrastructure for AI-assisted academic authoring [Preprint]. Sci Commun Edu. https://doi.org/10.1101/2023.01.21.525030

Ribeiro B, Meckin R, Balmer A, Shapira P (2023) The digitalisation paradox of everyday scientific labour: how mundane knowledge work is amplified and diversified in the biosciences. Res Policy 52(1):104607. https://doi.org/10.1016/j.respol.2022.104607

Schäfer MS (2023) The Notorious GPT: science communication in the age of artificial intelligence. J Sci Commun 22(02):Y02. https://doi.org/10.22323/2.22020402

Sokolov DA (2023) ChatGPT: Deutschlands Datenschützer eröffnen Verfahren gegen OpenAI . Heise Online. https://heise.de/-8974708

Staiman A (2023) Guest Post—academic publishers are missing the point on ChatGPT. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2023/03/31/guest-post-academic-publishers-are-missing-the-point-on-chatgpt/

Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. Nature 613:620–621. https://doi.org/10.1038/d41586-023-00107-z

Teubner T, Flath CM, Weinhardt C, Van Der Aalst W, Hinz O (2023) Welcome to the era of ChatGPT et al.: the prospects of large language models. Bus Inform Syst Eng 65(2):95–101. https://doi.org/10.1007/s12599-023-00795-x

Tomlinson B, Torrance AW, & Black RW (2023) ChatGPT and Works Scholarly: Best Practices and Legal Pitfalls in Writing with AI. arXiv http://arxiv.org/abs/2305.03722

Van Noorden R (2022) How language-generation AIs could transform science. Nature 605(7908):21–21. https://doi.org/10.1038/d41586-022-01191-3

Wulff K, Finnestrand H (2023) Creating meaningful work in the age of AI: explainable AI, explainability, and why it matters to organizational designers. AI & Soc. https://doi.org/10.1007/s00146-023-01633-0