Generative AI in Scientific Research: Opportunities, Challenges, and Ethical Implications

António Luís Lopes
antonio.luis@iscte-iul.pt
Iscte – Instituto Universitário de Lisboa

Resumo:

A Inteligência Artificial Generativa (IAG) emergiu rapidamente como uma ferramenta transformadora na investigação científica. Os grandes modelos de linguagem (LLMs) podem produzir texto e outros conteúdos semelhantes aos dos humanos, oferecendo novas formas de gerar ideias, rever a literatura, formular hipóteses, conceber estudos, analisar dados e ajudar na redação científica. Estas capacidades apresentam oportunidades significativas para aumentar a criatividade e a eficiência em todo o fluxo de trabalho de investigação. Ao mesmo tempo, a utilização da IAG acarreta riscos e desafios assinaláveis. Muitas vezes, os modelos alucinam com informações incorretas, refletem enviesamentos nos dados de treino e suscitam preocupações sobre a integridade académica, a privacidade e a potencial desqualificação ou deslocação dos investigadores. Este documento apresenta uma panorâmica abrangente, mas concisa, da evolução e do funcionamento interno da IAG e examina a sua aplicação em fases-chave do processo científico. Cada caso de utilização é discutido de forma crítica, destacando tanto os benefícios como as limitações, tal como referido na literatura recente. Uma discussão unificada das implicações éticas e sociais sublinha a importância da utilização responsável, da supervisão e da política de integração da IAG no meio académico. O documento conclui com uma reflexão sobre as perspetivas da IAG na investigação e a necessidade de diretrizes que garantam que estas ferramentas poderosas são utilizadas para aumentar, e não para prejudicar, a integridade e o progresso científicos.

Palavras-chave: Responsabilidade académica; IA generativa; grandes modelos de linguagem; fluxo de trabalho de investigação.

Abstract:

Generative Artificial Intelligence (GenAI) has rapidly emerged as a transformative tool in scientific research. Large Language Models (LLMs) can produce human-like text and other content, offering new ways to generate ideas, review literature, formulate hypotheses, design studies, analyse data, and assist in scientific writing. These capabilities present significant opportunities to enhance creativity and efficiency across the research workflow. At the same time, the use of GenAI brings notable risks and challenges. Models often hallucinate incorrect information, reflect

biases in training data, and raise concerns about academic integrity, privacy, and the potential deskilling or displacement of researchers. This paper provides a comprehensive yet concise overview of GenAI's evolution and inner workings and examines its application in key stages of the scientific process. Each use-case is critically discussed, highlighting both the benefits and the pitfalls as reported in recent literature. A unified discussion of ethical and societal implications underscores the importance of responsible use, oversight, and policy in integrating GenAI into academia. The paper concludes by reflecting on prospects for GenAI in research and the need for guidelines that ensure these powerful tools are used to augment, not undermine, scientific integrity and progress.

Keywords: Academic Responsibility; Generative AI; Large Language Models; Research Workflow.

1. Introduction

Generative Artificial Intelligence (GenAI) refers to models capable of producing novel content (such as text, images, or audio) that mimics human-generated data. The advent of the Transformer architecture (Vaswani et al., 2017) enabled models to capture long-range dependencies in text more effectively, paving the way for today's large language models. OpenAI's GPT series exemplifies this progress. GPT-3, with 175 billion parameters, demonstrated unprecedented prowess in generating coherent text and performing NLP tasks with minimal examples (Brown et al., 2020). Such models are often called foundation models or Large Language Models (LLMs), which are trained on massive corpora and adaptable to many tasks.

They operate by predicting the most probable next token in a sequence, an approach that can generate impressively fluent and contextually relevant outputs. However, this probabilistic nature also means LLMs lack a true understanding of facts, and they can produce plausible-sounding, but incorrect or unfounded, information (Bender et al., 2021). For example, an LLM might generate a scientific-sounding statement or a citation that has no basis in the training data or real world. As Lee et al. (2025) note, tools like ChatGPT can produce "*false or misleading information, known as 'hallucination'*," even while being helpful in other respects. Hallucinations arise because the model's goal is to produce *likely* text, not *true* text. It has no built-in mechanism to verify facts (Ji et al., 2023). The risk is especially acute in scientific applications, where accuracy is paramount. Mitigating hallucinations is an active area of research; strategies include retrieval-augmented generation (RAG), where the model consults an external knowledge base for facts, and fine-tuning models on domain-specific data to

improve factuality. Nonetheless, experts acknowledge that some level of hallucination may be an inherent limitation of current LLMs (Ji et al., 2023).

In interacting with GenAI, users have discovered that *prompt engineering* can significantly influence the quality and reliability of outputs. *Prompt engineering* is the craft of designing input queries or instructions that guide the model to produce the desired response. Early LLM usage often involved naive prompts (most likely a consequence of being used to performing keyword-based search engine queries) that led to generic or unpredictable answers, but researchers soon found that formulating more precise, contextual, or role-playing prompts yields better results. For instance, explicitly asking the model to behave as a critical reviewer or to think step-by-step can induce more rigorous responses. Such techniques build on findings that certain models can follow complex instructions and even perform reasoning when prompted appropriately (Wei et al., 2022). One notable approach is *chain-of-thought prompting*, where the prompt encourages the model to generate intermediate reasoning steps before giving a final answer (Wei et al., 2022). This method has been shown to improve performance on multi-step problems by making the model's latent reasoning process explicit. Effective prompts often include role directives (e.g., "*You are a knowledgeable domain expert…*"), specific tasks, and constraints or format requirements. As GenAI becomes more integrated into workflows, *prompt engineering* is evolving from an ad-hoc art into a documented practice, with researchers sharing prompt templates and best practices to reduce irrelevant output and minimize misunderstandings between human and machine.

The remainder of this paper examines how GenAI can be leveraged in the research workflow, from the spark of a new idea to the writing of a polished manuscript. Each subsection highlights practical applications reported in the literature, along with an appraisal of opportunities and risks. This is followed by a discussion of ethical and societal considerations, recognizing that technology adoption in academia must be paired with responsibility and foresight. Through this analysis, the aim is to provide a balanced perspective: GenAI in research holds immense promise, but realizing its benefits safely will require both technical improvements and thoughtful policy.

2. Applications of Generative AI in Research

In this section, we traverse the entire process on which the scientific method is based. Each subsection addresses a step in the scientific method, highlighting how GenAI can facilitate

and contribute to increase the productivity of researchers, but also taking into account its limitations and challenges.

2.1. Idea Generation

In the early stages of research, defining a compelling question or novel idea is often a creative challenge. GenAI has shown potential as a brainstorming partner for researchers seeking to develop or refine research ideas. Conversational LLMs like ChatGPT can produce a stream of possibilities around a given topic, helping users explore angles they might not have immediately considered. For example, a researcher can prompt an AI with a broad theme and ask for possible research questions or emerging trends in that area, to which the AI might respond with a list of pertinent questions, each with a brief rationale. This kind of ideation support can be especially useful for junior researchers or when venturing into an unfamiliar field, as the AI can quickly summarise common ideas from the literature and suggest where gaps might exist. Consider the following prompt, which follows the best practices for prompt engineering, as an example for such a task:

"You are a senior scholar in [specific field/subject] and I would like you to help me generate research ideas that explore [specific theme/aspect]. Consider potential gaps in existing literature, emerging trends, or areas with practical implications. Provide a brief overview of each idea, highlighting its significance and potential contributions to the field."

Beyond listing ideas, generative models can also help develop and critique them. One strategy highlighted in practical settings is to use an AI like ChatGPT in iterative roles: first as an idea generator (as shown above), then as a sceptic. Instructing the AI to act as a "*devil's advocate*" or critical reviewer, helps the researcher scrutinize each idea's weaknesses or implicit assumptions. For instance, the following prompt can help surface pitfalls that strengthen the researcher's understanding or reasoning:

"Operate as a fact-based sceptic with a focus on technical accuracy and logical coherence. Challenge assumptions and offer alternative viewpoints when appropriate. Prioritize quantifiable data and empirical evidence. Maintain an organized structure in your responses."

Such an exercise imitates a brainstorming session where one participant proposes and another critiques, helping to refine research questions before any empirical work begins. It is important to note, however, that the diversity and quality of AI-suggested ideas depend heavily on the prompt and the model's knowledge base. If the question is very novel or specialized beyond the AI's training, the suggestions may be generic or off-target. Moreover, an AI can only remix existing knowledge, it does not truly invent radically new concepts that aren't in its data. Thus, human creativity and intuition remain central, with GenAI serving as a catalyst or sparring partner to stimulate the ideation process rather than replace it.

2.2. Literature Review

Conducting a thorough literature review is a foundational research activity that GenAI can augment in several ways. LLMs are adept at summarizing and synthesizing large bodies of text, which aligns well with the needs of scholars scanning numerous papers. One immediate application is using AI to summarise academic articles. Given the full text of an article, an LLM can produce a concise summary highlighting the study's objectives, methods, and key findings. This capability can save researchers time in identifying which papers are most relevant to their work. Some specialized tools have emerged to facilitate this process: applications like ChatPDF and SciSummary allow users to upload PDFs of research papers and interact with them through a chat, asking questions like "*What problem does this paper address?*" or "*How did the authors conduct the experiment?*" The AI's ability to understand context and retrieve specific answers from the text can streamline the literature review process, especially in fields where one must quickly get up to speed on a vast existing literature.

GenAI can also assist in the discovery of relevant literature. Traditional keyword searches in databases have limitations, often returning too many results or missing papers due to vocabulary mismatches. In contrast, an LLM-based system can accept a natural language query or even a paragraph describing a research idea and then suggest pertinent references or findings. For instance, a researcher could ask, "*Have there been studies on the impact of social media use on adolescent cognitive development?*" and an AI might respond with a summary of known studies in that area, possibly citing seminal works or recent reviews (if it has access to that information). LLMs combined with search engines (like in the case of tools such as Perplexity.ai or Elicit) can serve as a kind of AI research assistant, returning not just a list of relevant works but synthesised insights across sources.

However, a critical caveat emerges when it comes to accuracy and factuality. LLMs, when they lack direct access to a database of verified publications, may hallucinate references, producing titles of papers or author names that sound plausible but are not real. Haman and Školník (2024) found that when asked to provide references, ChatGPT often generated legitimate-sounding citations that on closer inspection turned out to be fabricated. Relying on AI alone for literature searches can thus be perilous; any references or quotes provided by a model must be cross-checked against authoritative databases.

Despite these concerns, when used carefully, GenAI can accelerate literature reviews by serving as a starting point to gather themes and pointers, which the researcher can then verify and build upon. Indeed, a recent survey reported that more than half of polled researchers believe LLMs will have a major positive impact on tasks such as literature review (Mishra et al., 2024). The consensus is that AI can handle the heavy lifting of summarization and initial information gathering, while human experts must ensure completeness, accuracy, and critical appraisal of the literature.

2.3. Hypothesis Generation

Formulating a clear, testable hypothesis (which involves articulating a statement or prediction that can be empirically examined) is a creative and logical step that GenAI is beginning to assist with. AI's role here can be thought of as a bridge between raw idea brainstorming and experimental design.

The work by Banker et al. (2023) provides an illustrative case of AI-assisted hypothesis formulation, in which the authors used GPT-3 to generate hypotheses in the field of social psychology by fine-tuning it with a body of literature and tasking it with proposing new hypotheses. These AI-generated hypotheses – along with human-generated hypotheses – were then (blindly) evaluated by human experts. The findings were encouraging: the AI-generated hypotheses were deemed novel, scientifically interesting, and indistinguishable from the human-generated hypotheses, demonstrating that AI can combine existing knowledge in ways that spark new lines of inquiry.

However, caution is warranted in this application. A hypothesis generated by AI is only as good as the information it has seen and the logical paths it can follow. AI may also generate hypotheses that replicate biases or blind spots in the existing literature (e.g., overlooking certain populations or explanations that have historically been understudied). Moreover, an AI lacks the intuition about plausibility that an experienced researcher has (it might propose a

relationship that is practically nonsensical or unethical to test). Therefore, human vetting is crucial to assess the validity of each suggestion coming from an AI.

2.4. Study Design

Once a hypothesis is defined, the next major step is to design a study to test it. GenAI can assist here as well, by choosing an appropriate methodology, defining variables, determining sample size, planning data collection procedures, and anticipating potential pitfalls. Because models like GPT-4 have been trained on vast scientific corpora, they often can enumerate common designs (e.g., randomized controlled trial, longitudinal survey, simulation study) and discuss their pros and cons. In this sense, GenAI can function like a statistical advisor or an experienced colleague who points out standard practices and considerations in study planning.

Another emerging application is using AI for simulations and synthetic data generation to aid study design. In a study by Jansen et al. (2025), the authors introduced an AI-assisted platform that generates analysis code for biomedical data. While their focus was data analysis, the underlying idea that AI can produce and execute code based on a researcher's objectives extends to simulation and design tasks. The AI effectively translates a high-level experimental plan into concrete code, which can be used to probe the design's properties (e.g., expected variance, potential biases).

However, AI suggestions are based on general knowledge and may not account for specific practical constraints of a researcher's situation (such as limited resources, unique population access, or ethical considerations that are highly context-specific). The model might confidently suggest using a certain instrument or recruitment strategy without realizing it's not feasible in that context. Thus, while AI can provide a broad template of ideas, the fine-tuning of a study design requires human expertise that can balance scientific rigor with real-world feasibility. GenAI holds promise for making the study design phase more efficient and informed by accumulated knowledge, but it should be used as an aid and sounding board rather than an oracle. The researcher remains responsible for critically evaluating AI's suggestions and tailoring them to the specific scientific question at hand.

2.5. Data Analysis

The analysis of data, whether quantitative or qualitative, is a phase where precision and correctness are essential. GenAI (with tools such as Julius.AI) is increasingly playing a role here, particularly through its ability to generate and explain code for data analysis (in languages

like Python or R) as well as to interpret statistical results in plain language. Researchers can describe the data and the analysis they want to perform, and an advanced LLM-based assistant can then output a code snippet (and run it) to perform exactly that task. This has tremendous potential to help researchers who are not coding experts or to save time for seasoned analysts by automating boilerplate scripting tasks.

Jansen et al. (2025) demonstrated this concept by developing an AI-powered system named *mergen* that integrates an LLM to produce and execute R code for genomics data analysis. Their results showed that the LLM could successfully generate working code for many standard bioinformatics tasks and even handle some troubleshooting by refining code in response to errors. This kind of tool could democratize data analysis, enabling researchers to conduct complex analyses by describing their goals in natural language and letting the AI handle the programming intricacies.

Another domain of data analysis where GenAI shows promise is in handling unstructured data or performing textual analysis for qualitative research. For example, LLMs can be used to analyse interview transcripts, identifying common themes or sentiments. A researcher can prompt the model with an excerpt of qualitative data and ask for a summary of key points or the emotional tone. While this doesn't replace the nuanced interpretive work of a human qualitative researcher, it can serve as a preliminary step to organize data or generate codes/labels for content analysis. Mishra et al. (2024), in their global survey of researchers, suggest that researchers foresee LLMs helping in diverse analysis-related tasks (such as data cleaning, coding, and result summarization) are expected to be positively impacted by AI tools.

GenAI can accelerate and support data analysis by handling technical tasks and offering interpretations, but it works best as a collaborator under human supervision. When the human analyst and AI each contribute their strengths – domain understanding and oversight from the human, and speed and breadth of knowledge from the AI – data analysis can become more efficient without compromising accuracy.

2.6. Scientific Writing

Perhaps the most visible and debated use of GenAI in research is in the realm of scientific writing. LLMs have proven to be surprisingly proficient at generating formal text, which generates both opportunities and concerns for scientific writing. On the positive side, AI writing assistants can help authors produce and polish text more efficiently. Researchers who are not writing in their first language might use tools like ChatGPT to rephrase sentences in more fluent

academic English, improving readability while preserving the original meaning. LLMs can also help structure an article by suggesting logical flows of sections or even drafting an outline given a summary of the study. For example, an author could input a list of findings and have the AI draft a possible conclusion section that contextualizes those findings in the literature. This can serve as a starting draft, which the human author can then refine. The survey by Mishra et al. (2024) found that among researchers aware of LLMs, a large proportion had used them for writing assistance, especially for improving grammar and formatting. This indicates a clear demand for AI in easing the mechanical aspects of writing.

However, using AI to generate substantive content for a scientific paper crosses into ethically fraught territory. There have already been instances of authors using AI to write parts of manuscripts and not disclosing it, leading to detectable anomalies in the text (Conroy, 2023). In one high-profile case, a published paper was found to contain tell-tale phrases (like "Regenerate response") indicative of ChatGPT's involvement, which had slipped past peer review (Conroy, 2023). This has raised alarm in the academic community about transparency and responsibility and also the question of authorship. Major publishers and ethics bodies have been quick to assert that an AI cannot be listed as an author on a scientific paper, since it does not meet the criteria of an author, like having no ability to take responsibility for the work or rights to the content (Goto & Katanoda, 2023). Many journals (e.g., Science and Nature journals) now explicitly require that any use of AI in writing or analysis must be disclosed in the manuscript, and some have banned AI-generated text that cannot be attributed to a source.

From an academic integrity standpoint, indiscriminate use of AI in writing blurs the line between original writing and mechanized plagiarism. If a researcher were to let an AI write large portions of a paper, even based on their ideas, is that still the researcher's authentic voice and intellectual contribution? Opinions vary, but a common stance is that AI can be a tool for editing and refining, but the core content should come from the humans conducting the research.

AI can also be used to check the consistency of a paper, for example, asking it to read the manuscript and highlight any inconsistencies or unclear passages. This can be a helpful final review step, catching issues that the authors (as domain experts deeply involved in the work) might overlook due to familiarity. Prompts such as the one depicted in section 0 (playing "*devil's advocate*") can help the AI act as a reviewer of the manuscript.

3. Ethical and Societal Considerations

The integration of GenAI into scientific research raises several ethical and societal issues that must be carefully navigated. AI models learn from vast datasets that often reflect historical and cultural biases and, as a result, GenAI can inadvertently perpetuate stereotypes or unfair biases in its outputs (Weidinger et al., 2021). For example, an LLM might systematically produce examples or scenarios that assume Western contexts or male default subjects, if those were prevalent in its training data. In research usage, this could translate to biased hypothesis generation or literature summaries that overlook contributions from underrepresented groups or regions. Such biases could reinforce existing inequalities in academia, where access to AI tools and large datasets might already be skewed toward well-resourced institutions. Moreover, not all researchers have equal access to cutting-edge GenAI (some models are behind subscription services or require significant computational resources). This inequity of access could widen the gap between wealthy research institutions and those with fewer resources, as the former can more readily leverage AI to accelerate their work.

Ensuring more democratic access to AI tools and emphasizing diversity in the development of these models (including training data from a wide range of languages and contexts) are important steps to mitigate these disparities. Some efforts, such as open-source multilingual models, aim to address this by providing more inclusive AI systems. Nonetheless, vigilance is needed: researchers should be aware of the potential for bias in AI outputs and actively check and correct for it.

Privacy is another critical ethical issue when using GenAI in research. Many research projects involve sensitive data such as medical records, personal interviews, or proprietary experimental data. Using AI tools often means inputting text or data into a system that may be cloud-based or whose inner workings are not fully transparent. Some AI companies also state in their terms and services that any data shared with their products can be used to train future models. This is a privacy issue on its own (when these external companies have access to this sensitive data), but there have also been demonstrations showing that LLMs can, in some cases, regurgitate portions of their training data verbatim (Nasr et al., 2023), which is alarming if that training data contained private information.

Therefore, academic institutions should issue guidelines advising caution on using GenAI with any information that is classified, confidential, or identifiable unless the tool is locally hosted and secure. The academic community must also guard against AI being used to fabricate scientific outputs that pollute the literature, a phenomenon that could erode trust in research if

not kept in check. Tools for verification and fact-checking will become increasingly important and AI itself can be harnessed to detect AI-generated content or anomalies, but these are not foolproof solutions and carry their own ethical ambiguity.

Another aspect that warrants discussion on the use of AI in research is the question of academic responsibility and integrity, namely the proper attribution of AI's role in the research workflow or failing to disclose AI assistance in writing or analysis. Researchers have a responsibility to ensure that all co-authors (human) are aware of and consent to AI usage in producing results or text. If an AI helped analyse data, this should be documented in the methods and if an AI wrote a draft paragraph, some journals require a notation (often in the acknowledgments or a methods/limitations section).

Nevertheless, instances of undisclosed and flagrant use of AI are showing up in scientific journals and present a worrying picture for the future of research. An example is the retracted paper (Zhang et al, 2024) that features this sentence as the first in the introduction: "*Certainly, here is a possible introduction for your topic*". The problematic part is that this misplaced sentence (which is commonly known because it is used by LLMs to begin its helpful response to a prompt) was not detected by the authors, the editors nor the reviewers. It was only when the paper was published that readers detected the weird sentence (and other elements in the paper) that indicated that the paper was heavily produced using AI. Such behaviour is unequivocally unethical and undermines the integrity of the scientific record. The community response has been to develop better checks (like journals performing more rigorous statistical or plagiarism reviews, including looking for tell-tale signs of AI text). Encouragingly, many in academia are calling for a culture of transparency where using AI is not stigmatized as long as it is disclosed and done responsibly. For example, AI can be a legitimate tool for proofreading, as long as the authors ensure the content remains their own and factual.

Another broad societal concern that arises is the impact that the use of GenAI in research and academia may have on the workforce. As AI systems become capable of performing tasks that were once the domain of research assistants, junior analysts, or technical writers, there is apprehension about job displacement. Will labs hire fewer data analysts if one AI system can do the work of many? Will students rely on AI to do coursework, impacting learning outcomes? History with other technological advances (like industrialization and the internet) suggests that while some roles change or even disappear, new roles often emerge. In the context of research, mundane and repetitive tasks (like data cleaning, basic coding) might be largely automated, potentially freeing up researchers to focus on creative, conceptual, and interpretative work – the aspects of science that truly require human insight. This could lead to a positive outcome in

which researchers spend more time thinking about research questions and theory, and less time on the mundane tasks that steal productive time. However, the transition period could reveal difficult to early-career researchers that might need to adapt to the new reality. For instance, statistical education might shift from learning how to program analyses from scratch to learning how to effectively guide and validate an AI that does the programming.

A concerning scenario is if over-reliance on AI leads to skill atrophy (or "de-skilling"), considering that future researchers may not develop a deep understanding of methods if they always lean on AI for support. This underscores a need for updated curricula and mentoring that integrate AI literacy: knowing when and how to (responsibly) use AI, and when not to. Moreover, as AI takes over some tasks, individuals in those roles will need support to up-skill or re-skill for new kinds of work (perhaps focusing on AI oversight, data ethics, critical thinking, or interdisciplinary communication, which will be increasingly valuable).

The stakes are high: if managed poorly, AI could exacerbate problems of trust and inequality in science; if managed well, it could contribute to a more efficient, inclusive, and innovative research ecosystem. The onus is on the scientific community to adapt its codes of conduct and oversight accordingly, ensuring that human values and judgment remain at the core of research, even as we welcome these powerful new AI collaborators.

4. Conclusions

GenAI has swiftly moved from the periphery to the mainstream of scientific research, offering tools that can inspire new ideas, accelerate tedious tasks, and even draft scholarly texts. This paper has examined how large language models can be woven into each stage of the research process, from the first glimpses of an idea, through literature exploration and hypothesis framing, to the nuts-and-bolts of experimentation, analysis, and ultimately communicating findings. The opportunities presented by these developments are exciting: researchers can brainstorm with a tireless and always-available assistant, distil large corpus of literature in minutes, generate code at the snap of a finger, and refine their writing with instant feedback. Early evidence suggests that these tools, when used wisely, have the potential to make research more efficient and perhaps even more creative, as they expose scientists to connections and information they might not have encountered otherwise.

However, this exploration also underscores a fundamental reality: GenAI's very strengths (fluency, speed, breadth of knowledge) come with mirror-image weaknesses (propensity to fabricate data, lack of true understanding, and reflection of biased training data). Thus,

alongside the success stories are cautionary tales, in which an AI can suggest a brilliant hypothesis one moment and a nonsensical one the next. It can summarise a paper accurately in one instance and hallucinate a completely fake citation in another. The integration of GenAI into research workflows therefore demands a heightened sense of scrutiny and responsibility from the scientific community. In practical terms, this means maintaining rigorous human oversight over any AI-assisted output. Researchers must act as editors, fact-checkers, and ethical guardians for their AI collaborators: verifying every reference an AI provides, cross-checking every analysis it performs, and ensuring that no sensitive information is mishandled. It also means developing new norms and guidelines so that the use of AI in research is transparent and does not undermine the trustworthiness of scientific publications.

It is expected that GenAI will become even more capable by continuing to improve in accuracy and in their ability to justify their answers or cite sources. Such advances could further reduce the friction in knowledge work, allowing scientists to focus on high-level ideas and interpretation. But they will also pose new questions: for instance, if an AI becomes significantly better at drafting certain sections of a paper than a human, how do we train students in that skill, or do we at all? The forward-looking reflection here is that the role of a researcher is likely to evolve. Mastery in the future may be measured not just by one's ability to carry out tasks personally, but by one's ability to effectively leverage AI and integrate its contributions with human insight. In that sense, AI literacy will become a crucial component of scientific training. Despite the uncertainties, one principle rises above others: the core values of scientific research (curiosity, rigor, integrity, and the quest for truth) must guide the adoption of any new tool, AI included. By harnessing its strengths and remaining vigilant about its weaknesses, the scientific community can ensure that this instrument amplifies human potential and propels research forward responsibly, ushering in an era of augmented discovery.

Acknowledgments

In line with the guidelines exposed in this work and in the spirit of transparency and integrity standards, the author declares that this manuscript was written with the assistance of an AI product, namely, OpenAI's ChatGPT. No part of the manuscript was fully generated by AI. The author declares

that all AI-generated or AI-assisted material was thoroughly validated, reviewed, edited, and approved by the author, who takes full responsibility for the final text.

References

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). 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).

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.

Banker, S., Chatterjee, P., Mishra, H., & Mishra, A. (2024). Machine-assisted social psychology hypothesis generation. *American Psychologist*, *79*(6), 789.

Conroy, G. (2023, September 8). Scientific sleuths spot dishonest ChatGPT use in papers. *Nature News*. https://doi.org/10.1038/d41586-023-02477-w

Goto, A., & Katanoda, K. (2023). Should we acknowledge ChatGPT as an author?. *Journal of epidemiology*, *33*(7), 333-334.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys*, *55*(12), 1-38.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., ... & Lee, K. (2023). Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*.

Haman, M., & Školník, M. (2024). Using ChatGPT to conduct a literature review. *Accountability in research*, *31*(8), 1244-1246.

Mishra, T., Sutanto, E., Rossanti, R., Pant, N., Ashraf, A., Raut, A., ... & Zeeshan, B. (2024). Use of large language models as artificial intelligence tools in academic research and publishing among global clinical researchers. *Scientific Reports*, *14*(1), 31672.

Zhang, M., Wu, L., Yang, T., Zhu, B., & Liu, Y. (2024). RETRACTED: The three-dimensional porous mesh structure of Cu-based metal-organic-framework-Aramid cellulose separator enhances the electrochemical performance of lithium metal anode batteries.

Jansen, J. A., Manukyan, A., Al Khoury, N., & Akalin, A. (2025). Leveraging large language models for data analysis automation. *PloS one*, *20*(2), e0317084.

Lee, Y., Oh, J. H., Lee, D., Kang, M., & Lee, S. (2025). *Prompt engineering in ChatGPT for literature review: practical guide exemplified with studies on white phosphors*. Scientific Reports, 15(1), 15310.

Author Profile

António Luís Lopes has received a PhD on Artificial Intelligence from the Department of Information Science and Technology of Iscte – Instituto Universitário de Lisboa – Portugal. He is currently the Director of the Information Systems Development Service at Iscte and an Artificial Intelligence researcher at Instituto de Telecomunicações – IUL. His research interests are on the application of Artificial Intelligence in other areas of research.