TYPE Perspective
PUBLISHED 15 May 2024
DOI 10.3389/frsps.2024.1392128

OPEN ACCESS

EDITED BY
Michal Kosinski,
Stanford University, United States

REVIEWED BY
Jason Jeffrey Jones,
Stony Brook University, United States
Jize Zhang,
Hong Kong University of Science and Technology, Hong Kong SAR, China

*CORRESPONDENCE
David M. Markowitz
dmm@msu.edu

RECEIVED 27 February 2024
ACCEPTED 30 April 2024
PUBLISHED 15 May 2024

CITATION
Markowitz DM, Boyd RL and Blackburn K (2024) From silicon to solutions: AI's impending impact on research and discovery. *Front. Soc. Psychol.* 2:1392128.
doi: 10.3389/frsps.2024.1392128

From silicon to solutions: AI's impending impact on research and discovery

David M. Markowitz[1]*, Ryan L. Boyd[2] and Kate Blackburn[3]

[1]Department of Communication, Michigan State University, East Lansing, MI, United States, [2]Department of Computer Science, Stony Brook University, Stony Brook, NY, United States, [3]Department of Psychology, University of Texas at Austin, Austin, TX, United States

The social sciences have long relied on comparative work as the foundation upon which we understand the complexities of human behavior and society. However, as we go deeper into the era of artificial intelligence (AI), it becomes imperative to move beyond mere comparison (e.g., how AI compares to humans across a range of tasks) to establish a visionary agenda for AI as collaborative partners in the pursuit of knowledge and scientific inquiry. This paper articulates an agenda that envisions AI models as the preeminent scientific collaborators. We advocate for the profound notion that our thinking should evolve to anticipate, and include, AI models as one of the most impactful tools in the social scientist's toolbox, offering assistance and collaboration with low-level tasks (e.g., analysis and interpretation of research findings) and high-level tasks (e.g., the discovery of new academic frontiers) alike. This transformation requires us to imagine AI's possible/probable roles in the research process. We defend the inevitable benefits of AI as knowledge generators and research collaborators—agents who facilitate the scientific journey, aiming to make complex human issues more tractable and comprehensible. We foresee AI tools acting as co-researchers, contributing to research proposals and driving breakthrough discoveries. Ethical considerations are paramount, encompassing democratizing access to AI tools, fostering interdisciplinary collaborations, ensuring transparency, fairness, and privacy in AI-driven research, and addressing limitations and biases in large language models. Embracing AI as collaborative partners will revolutionize the landscape of social sciences, enabling innovative, inclusive, and ethically sound research practices.

KEYWORDS
artificial intelligence, large language models, chatbots, collaboration, scientific process

Introduction

> "*... success in creating AI, could be the biggest event in the history of our civilization. But it could also be the last, unless we learn how to avoid the risks.*"—Stephen Hawking, October 19, 2016

Artificial Intelligence (AI)—specifically, generative Large Language Models (LLMs) that power chatbots and other agents—was believed by most to be more theoretical than practical until very recent history (Huang, 2023; Prather et al., 2023). Now, as a result of sufficient intellectual and financial investment by researchers and major tech firms, the AI revolution is underway and has captured the collective consciousness. As an analog: one year after COVID-19 began, nearly 200 million people around the world received life-saving vaccines (Cohn et al., 2022) that were built upon mRNA technology developed nearly 20 years earlier (Karikó et al., 2005). This rapid transition from theoretical to practical—from academic interest to must-have technology—is an apt parallel for the expeditious rise and use of AI: the foundations of AI and LLMs are rooted in decades of computational research, yet they have only been used *en masse* for about 1 year at the time of this writing. If AI pioneers and advocates are correct, AI is poised to advance discoveries about the natural world and the human condition at a rate faster than we can presently anticipate (e.g., Shah, 2023).

As computational social scientists, and as scholars who are interested in how AI comports with our scientific traditions, we wrestle with the consequences—positive and negative, real and hypothetical—of using AI in our daily work. The current perspective piece attempts to introduce questions and formalize several key considerations that will arise when integrating emerging AI technologies into the field of computational social science and in the social sciences, more generally. Moreover, we offer a defense of prospective uses for AI across scientific disciplines, adding balance to the current AI discourse that is largely dominated by discussion of what AI *cannot do* to the exclusion of what AI already *can do* and *will be able to do* in the near future. Our primary goal is to expand the increasingly dynamic conversation currently surrounding the topic of how AI can become a collaborative partner in science. We will have succeeded with this article if readers reflect on their current practices with (or without) AI in the scientific process, engage in discussions with other scholars about how AI fits (or does not fit) within their workflow, and if researchers and practitioners consider more deeply the ethical and societal impacts of using AI in everyday inquiry.

In this article, we paint a picture of a not-too-distant future where AI has outgrown its current shortcomings, and we imagine the revolutionary potential of "emerging" AI technologies for its many impacts on how research will be conducted. Against this backdrop, we organize the current paper in the following manner: first, we begin with a basic description of AI and LLMs to orient the reader, including basic definitions, use cases, and evaluation strategies.[1] This is followed by a general review of how a majority of social scientists conduct their work, focusing on those with a quantitative orientation and who are interested in human social and psychological processes. We specifically examine how AI and LLMs might play a critical role in developing major parts of an academic paper, from pre-work (e.g., idea generation, rationale identification) to the actual mechanics of scientific discovery and evaluation (e.g., data collection, data analysis). Finally, we raise questions and offer considerations as to how AI and LLMs may facilitate positive and negative outcomes for science with a special focus on ethics, inclusivity, and trust.

1 While other papers have outlined basic definitions related to AI and LLMs (Demszky et al., 2023), we also believe it is important to briefly restate them here in the spirit of establishing common ground for uninitiated readers.

Artificial intelligence and large language models: basic conceptualizations

Artificial intelligence is the field of study interested in the development of intelligent, "thinking" machines and related technologies (McCarthy, 2007). A popular subfield of AI, which is the primary focus of the current article, is *generative AI*. Today, the domain of generative AI uses large, "foundational" models to represent and create communication messages that most often take the form of images, words, or numbers. Chatbots like Open AI's ChatGPT, Google's Gemini, and others are front-end interfaces that use back-end LLMs for task completion. For example, ChatGPT as a platform can be understood as a chat-like interface that connects with any one of several pre-trained LLMs (e.g., GPT-4) that are designed to respond in a "dialogue" format, allowing users to provide prompts to the model in a format that it can then "respond" to. Ultimately, the power and utility of these technologies rest on their underlying models, which are trained using massive amounts of data to make probabilistic estimates of the most likely (or "appropriate") next tokens (e.g., words). That is, if you ask ChatGPT (running GPT-4) to complete the rest of the following sentence—"Coffee and ..."—it provides different options based on what is most probable to follow.[2] The upshot of generative AI and LLMs is a communication output that is often indistinguishable from responses offered by humans when evaluated by other humans (Köbis and Mossink, 2021; Kreps et al., 2022; Jakesch et al., 2023)—and through a process that is able to complete tasks in a relatively efficient manner relative to humans.

To date, much of the public discourse surrounding the burgeoning generative AI industry has focused on the *quality* of AI-generated media, ranging from stern, finger-wagging admonishments about its intellectual deficits (e.g., Bogost, 2022) to tongue-in-cheek demonstrations intended to highlight its shortcomings and social pitfalls (e.g., Franken and Proft, 2023). The same "comparative" trend—establishing degrees of similarity and difference between the cognitive, social, and emotional capabilities of generative AI and humans—has rapidly permeated the social scientific literature as well, with a nontrivial amount of academic brainpower being dedicated to *comparative studies*. Such studies most often compare how generative AI performs relative to humans across a range of more scientific tasks, often taken from classical cognitive and social psychology. For example, research questions have included: to what degree does generative AI have theory of mind like humans (Kosinski, 2023)? How well can generative AI perform coding tasks like humans (Koubaa et al., 2023; Rathje et al., 2023)? To what degree can generative AI detect deception like humans (Markowitz and Hancock, 2023)? How does AI respond to personality and behavioral questionnaires as compared to humans (Mei et al., 2024)? These examples merely scratch the surface of

2 ChatGPT offered five possible answers: "Coffee and tea are popular morning beverages," "Coffee and donuts make a classic pairing," "Coffee and conversation go hand in hand," "Coffee and books are a perfect match for a relaxing afternoon," and "Coffee and cream blend together for a smooth flavor."

the rapidly expanding body of comparative studies in the social, psychological, and computational social sciences.

Evidence from such studies has, naturally, been mixed, with few large "take-home" conclusions aside from nuanced differences that often fail to meet a level of public significance for the non-academic public. Such early comparative studies have served as natural curiosities for scholars seeking to extend their scholarly disciplines to the emergence of non-humans engaging in human-like behavior, evaluating how well such non-humans could perform basic (e.g., math) or advanced operations (e.g., deception detection). These studies have fueled an expanded conversation about the ways in which AI can (or perhaps, *should*) be evaluated, extending beyond traditional "benchmarking" studies that have traditionally shown a strong bend toward an engineering perspective of performance on tasks that are both highly prescriptive and highly stilted from social scientific and humanities perspectives (e.g., Mitra et al., 2023; Sclar et al., 2023).

Studies and extensive discourse on the capabilities and qualities of LLMs are important, to say the least, particularly in terms of understanding lingering engineering challenges, limitations to their application, and broader social impacts like the propagation of misinformation, prejudice, and inequality. Nevertheless, we view these problems as ultimately tractable, and the current state of overwhelmingly comparative discourse on the shortcomings of generative AI relative to human abilities is, ultimately, ephemeral. At some point—with enough innovation in machine learning methods, training data, fine tuning, and safeguards—AI technologies will continue, and ultimately succeed at, closing the gap between explicitly computational and human potential. The inevitable success of AI to overcome its current shortcomings will not only render the specific conclusions drawn from many of today's comparative studies moot, but brings to the foreground a long-understood shortcoming of comparative work itself: it tends to over-emphasize the meaning of differences at the expense of understanding similarities, sometimes at a very real societal cost (e.g., Hyde, 2005).

Critically, we argue that the scientific community has dedicated a disproportionately high amount of attention and energy to hand-wringing over small, ephemeral differences between the outputs of AI and humans. The value of identifying such differences has proven to be somewhat elusive. Where such differences are found, it is not clear what the *meaning* of such transient differences might be—are they the result of unseen confounds, nuances of study design, or artifacts of a particular chosen model? Such questions are, of course, not unique to studies that compare today's AI against humans. However, in the context of studying emerging AI technologies, the answers to such questions are especially uninformative given that AI models themselves are "moving targets." For example, dedicating time and resources to discovering subtle differences between human and AI output for the Microsoft *phi-1.0* model (Gunasekar et al., 2023) may be rendered moot when then *phi-2.0* model was released a few short months later.

Instead, we urge for a cultural shift toward embracing the positive and profound benefits that AI, as it becomes increasingly reliable with technological advancements, can contribute to our scientific processes. Adopting a more long-term perspective, then, we ask readers the following: what do the current and future states of AI mean for academic research on the human mind, human experiences, and the human condition writ large? As the potential for AI to revolutionize all domains of human output marches forward, what implications should we be considering *now*? If guardrails for the use of generative AI in academic research are not established, such technology may become powerful enough to subsume and automate many parts of the academic research experience. In other words, even the tasks that humans claim to be "our domain" in the research process (e.g., theory building, hypothesis generation, data interpretation) may be performed by AI technology if (or when) they become sophisticated enough. This exposes a new set of philosophical questions about what humans are uniquely positioned to contribute in academic research within the context of mainstream automation. Such open questions are ones that we are not equipped to address here. However, we raise this point to suggest that a utopian (or dystopian) view of one's research future must contend with the idea that AI will become a part of the epistemic and empirical pipelines. We suggest that this is not only plausible, but highly probable, for most researchers.

A new social science: AI as collaborative partners in the scientific process and exploring the human condition

The precise manner in which any given scientist conducts their work is a highly personalized, often idiosyncratic process. Nevertheless, despite discipline-specific traditions, paradigms, and best-practices that are often inherited through academic lineages, there are characteristics that constitute a broader "philosophy of science" that systematically transcend specific disciplinary boundaries, reflecting a consensus of how scholarship is done, particularly in the quantitative social sciences (e.g., Robinson et al., 2019). Here, we briefly reflect on several of these practices for the purpose of envisioning how AI might play an instrumental role in such work. These are not the only practices that can be impacted by generative AI in academic research, and we do not presume to be exhaustive in our commentary. However, we believe that these particular practices are those with the greatest potential for AI to make an everyday impact in research workflows. Our perspective is informed by seminal writing that articulates how the empirical research article is typically performed and communicated (e.g., Bem, 2003), yet we also draw on norms and conventions from our research labs to explain how parts of the research process might or might not be supported by AI.

Question formulation

In the early 1990s, only half a percent of the world's population used the Internet (Roser, 2024). In the span of only a few decades, we have reached a point at which now over 60% of the global population is connected online (Ritchie et al., 2023). This shift indicates that, in essence, the Internet has evolved into the largest and (to date) most comprehensive time capsule of human behavior, consisting of a dizzying quantity of current and historical information that reflects how people think, feel, and experience the world over time. This data is reflected predominantly through words, images, videos, and interactions, providing researchers with an extensive and flexible repository of human behavior for analysis (Lazer et al., 2009).

In fact, over the past two decades, digital corpus studies have gained prominence as a function of increasing data availability (Church, 2017), fueling many of the recent advances in language-based AI approaches (Dodge et al., 2021). It aligns with our natural inclinations to study artifacts, similar to how archaeologists might use traces of past human activity (e.g., buried human remains, ephemera, satellite imagery, radiocarbon dating) to identify and make meaning of past human migrations, technologies, and cultures (e.g., Sutter, 2021). The radical increase in digitization marks an evolution in both the quality and quantity of artifacts available to study for scholars of the human condition. Precisely *what* to study and *how* to study it amidst this treasure trove of data is, however, an entirely different question.

The early 2010s—what we view as the beginning ascent of the computational social sciences as we understand them today—witnessed several hallmark innovations in the aggregation of large corpora of human data. The most impactful innovations primarily came in the form of accessible *tools* that could be used by anyone with an internet connection. For one of the first times in human history, scholars and laypersons alike had immediate and reliable access to information about societal trends, the impact of social movements, and a "zoomed out" view of the human condition. Tools like the Hedonometer[3] provided up-to-date insights into the hedonic state of an entire society, derived exclusively from the analysis of language in social media data (Dodds et al., 2011). Other major innovations adopted a more interactive nature, allowing users to *engage* with data to explore their own curiosities about changes in society and culture over time. The Google Books Project, initiated in 2004, has digitized over 40 million books to date, preserving them for public consumption (Lee, 2019). Building on this project, researchers and engineers developed the Google N-Gram Viewer[4] to enable a comprehensive exploration of linguistic and cultural trends over a large breadth of documented human history. By exploring trends in the frequency with which various names, events, and places appear across the Google Books corpus, researchers have investigated diverse phenomena ranging from the nature of infamy to the aftermath of destructive and authoritarian censorship in Nazi Germany and Communist China (Michel et al., 2011; Aiden and Michel, 2013). Additionally, such resources have helped to identify over half a million English words overlooked in our dictionaries (Bohannon, 2010).

However, a crucial limitation of such tools is that they require researchers to intuit, and formulate, appropriate questions in order to use them effectively. Despite the wealth of data, scholars must possess a preconceived notion of what questions to ask, directing the analytical spotlight toward areas that they consider potentially revealing. In other words, the success of using these tools is contingent on the ability of researchers to frame relevant queries. However, inherent human limitations and biases in our own knowledge constrain our scope of discovery. Thus, while these tools democratize access to data, the interpretive lens remains in the hands of those asking the questions, underscoring the need for astute and insightful inquiry.

Looking ahead, emerging AI technologies are poised to address many of these human shortcomings by actively contributing to the formulation and exploration of research questions. As these technologies continue to evolve, they will help to not only assist with the collection of, and access to, ever-growing bodies of data that are relevant to our questions of interest, but also to help *formulate* the questions that we ask in the first place. For instance, advanced machine learning algorithms can identify patterns, correlations, and anomalies within vast datasets that may elude human researchers. AI offers the promise to enhance the efficiency and depth of investigations by automating certain aspects of the discovery process. It is often well-understood that the conclusions that we reach are only as good as the data that we gather, often expressed as the principle of "garbage in, garbage out" (Rose and Fischer, 2011). However, what is less commonly appreciated is that our conclusions are fundamentally constrained by the very questions that we decide to ask.

As scientists, we often acknowledge—typically with optimistic enthusiasm—the elusive and powerful role of serendipity in scientific breakthroughs, wherein researchers stumble into discoveries that they were not intentionally seeking (Edward Foster and Ellis, 2014). Oft-cited examples found across introductory science textbooks include the discoveries of penicillin, x-rays, and insulin (Krock, 2001). So powerful is the allure of "chance" discoveries that scholars in many fields seek to "harness" or "control" serendipity as a part of their research process, increasing the likelihood of unintended discovery (McBirnie, 2008; André et al., 2009; de Rond, 2014).

However, a complementary approach is to first identify the boundaries of knowledge that we have the first place, using a clear articulation of where innovations might offer the greatest advancements and impact. As with the tools described above, our current state of affairs requires the often murky judgment of experts to steer toward uncharted areas of research, limited by our own individual limitations in scientific knowledge. With the assistance of advanced AI, however, we may benefit from its high-powered associative understandings of not just *what* limitations characterize our scientific understanding of the world, but *which questions* are likely to yield the most meaningful results when asked *in particular ways*.

In the intricate landscape of contemporary scholarship, the metaphor of Plato's cave takes on heightened significance, especially within systems that often incentivize hyper-specialization of knowledge and exploration. Dedicated scholars, navigating complex disciplinary terrains, find their perspectives inevitably circumscribed, often embedded in intellectual niches that, while fostering depth in specific domains, simultaneously constrain their broader vision. The recognition of these self-imposed boundaries becomes an imperative facet of intellectual growth, urging us to transcend our intellectual myopia and embrace a holistic view that acknowledges the interconnectedness of diverse fields. Overcoming this limitation involves not only deepening expertise but also fostering interdisciplinary

3 hedonometer.org
4 books.google.com/ngrams

dialogues, promoting a more comprehensive understanding of complex phenomena.

Within this dynamic intellectual ecosystem, the role of AI will emerge as a transformative force that will not only widen, but deepen, the scope of exploration and inquiry. AI will function as a metaphorical seer and knower, possessing a cognitive capacity that extends beyond the confines of individual and collective human limitations. It acts as a beacon, illuminating uncharted territories of knowledge and guiding researchers through the labyrinth of information. The integration of AI into scientific inquiry becomes more than a technological augmentation; it becomes a collaborative journey with intelligent systems that broadens the scope of exploration and inquiry. By leveraging the multifaceted capabilities of AI, scholars will be able to entertain a symbiotic relationship, navigating the shadows cast by current understanding. AI's ability to discern patterns, uncover subtle correlations, and identify anomalies will empower researchers to unearth new dimensions of knowledge. This harmonious interplay between human and artificial intelligence holds the promise of unlocking profound insights into the intricacies of the human condition, transcending the limitations inherent in traditional methodologies and propelling us toward a deeper and more nuanced comprehension of our complex existence.

Consider the initial phase of a project when one is digging into existing research. Instead of spending hours manually searching databases like Google Scholar, one can efficiently request AI, such as ChatGPT, to provide a literature review. Although not infallible, as it might miss certain research articles or require additional refinement for topic-specific content, AI significantly speeds up the process. It serves as a valuable starting point, streamlining the review of prior research. An empirical journal article reviews the body of literature that underlies the authors' interests. This comes in the form of reviewing theory, primary studies that relate to the theoretical ideas in question, and meta-analyses if they exist. Determining precisely what literature to review, and how much literature to present to an intended audience, is a tricky process. Disciplinary norms and journal expectations often matter here. Take, for example, the idea of understanding folk theories of social media. Folk theories are "non-authoritative conceptions of the world that develop among non-professionals and circulate informally" as they relate to social networking sites (Eslami et al., 2016). A communication scholar may rely on this conceptualization because it is rooted in traditions of human-computer interaction and media effects (DeVito et al., 2017, 2018; French and Hancock, 2017). Scholars in other disciplines might call folk theories by another name, however. For example, a related idea in psychology (i.e., "mental models") describes a similar phenomenon as "selected concepts, and relationships between them, and uses those to represent the real system" (Forrester, 1971; Johnson-Laird, 2010). Schemas are also a related concept (Piaget, 1952). Altogether, this example marks the complexities of how authors decide to theoretically and conceptually ground their work. It is an imperative and challenge of the social scientist to describe such conceptual boundaries and to articulate what they are studying (and, by association, what they are not studying). AI is poised to assist with drawing such conceptual boundaries between theories to properly motivate research and to identify their inherent boundaries.

As we progress beyond the initial stages of the scientific process, we can leverage AI as a collaborative tool, akin to working with our best and brightest peers. AI can become a scientific companion that not only aids in resolving literature disputes and hypothesis formation, along with potentially suggesting novel ideas that researchers might not have considered, but also assisting in the refinement of experimental parameters. This capability is remarkable, as AI can simulate the presence of multiple scholars in the room (e.g., offering perspectives on how to manipulate a particular construct, offering alternate versions of experimental stimuli), collectively contributing to a project through their diverse resources.

Data curation

Once a research question has been formulated, the work of collecting adequate and appropriate data to test our hypotheses can begin. We view the integration of AI technologies into the data curation process to be the next major horizon in navigating the intricacies of contemporary data landscapes specific to our respective research questions. The sheer expanse of available data, particularly within the expansive realms of the internet, necessitates advanced tools for efficient curation and analysis. In this dynamic landscape, the role of AI's data processing capabilities becomes indispensable, providing computational social scientists the means to navigate and extract meaningful insights from the vast digital expanse. The infusion of AI into data curation represents a paradigm shift, expediting what would otherwise be a time-consuming and intricate process of data collection. Machine learning algorithms, powered by the prowess of AI, will emerge as indispensable allies in executing data collection efficiently and effectively.

Moreover, and perhaps more importantly, the role of AI transcends assistance with data gathering and sifting; it will actively contribute to the refinement of research methodologies within the computational social sciences. We detailed above the synergistic relationship that will emerge between humans and AI insofar as we decide *what* questions to ask. In this phase of research, AI will likely help us determine *where* and *how* to look for answers. Today, we often have limited perspectives on the optimal *sources* of information that can best address our research questions. Currently, several computational disciplines are coming to grips with the consequences of dataset monocultures, wherein "gold standard" datasets are effectively a shared component numerous programs of research, leading to something of a collapse in generalizability and outcome differentiation (Bommasani et al., 2022). We believe that the infusion of not just a greater *quantity* of data, but data with a greater diversity of *qualities*, broadly defined, will serve to ameliorate many such issues. By harnessing its computational prowess, AI can suggest innovative approaches to data gathering, introducing a level of creativity and efficiency that can serve to enhance the overall research design that maximizes equality and equity among stakeholders ranging from other researchers to the general public.

Consider a research question in the field of online political discourse, formulated to explore the dynamics of political sentiment during election campaigns across various social media platforms. While the question is clear, the challenge lies in identifying the most suitable and representative data sources to address it. Social media platforms like Twitter, Facebook, and Reddit are often initial considerations for collecting data on political discourse due to their vast user bases and real-time nature. However, these platforms inherently harbor biases and lack certain demographic representations that might not be immediately evident to researchers. Twitter and Facebook, for instance, tend to host users whose demographics, ideologies, and political behaviors are not representative of the diverse political opinions of the entire population (Mellon and Prosser, 2017). Similarly, issues like political polarization can happen in real-time through even small changes made on platforms, and "filter bubbles" may be introduced on such platforms in ways that provide critical, but unseen, contours to the data that we might collect from such platforms (Chitra and Musco, 2020). Identifying these implicit biases and gaps in representation becomes crucial for ensuring the research findings are not skewed or limited by the selection of data sources.

This is where AI can play a transformative role in guiding the data collection strategy. Advanced machine learning algorithms can analyze existing datasets, demographic information, and engagement patterns to identify potential biases in different social media platforms. Moreover, AI can suggest alternative sources of data that may offer a more comprehensive and representative view of online political discourse. For instance, the AI system might recommend integrating data from specialized political forums, community blogs, and news commentary sections to capture a broader spectrum of political opinions and diverse demographic perspectives. By doing so, the research not only diversifies its data sources but also addresses implicit biases present in more mainstream social media platforms.

Furthermore, AI can guide the actual data collection strategy by recommending specific keywords, hashtags, or topics to track. It can dynamically adjust the data collection approach based on real-time insights, ensuring that the dataset remains representative and unbiased throughout the research process. The system may even propose incorporating sentiment analysis to discern nuances in political expressions, contributing to a more nuanced understanding of online political discourse. In essence, AI acts as a strategic guide, not only pointing researchers toward the most suitable and representative data sources but actively shaping the data collection strategy to address biases and ensure a more comprehensive exploration of the formulated research question in the realm of computational social science.

Data analysis

The advent of AI companions in the analysis landscape is reshaping the contours of research practices, as evidenced by a global survey of over 1,600 researchers (Van Noorden and Perkel, 2023). The survey underscores the pivotal role that AI is expected to play in producing information across various modalities found in their training data, encompassing text, images, and code. In several domains across the natural sciences, advances in AI have already shown great promise in creating new ways to model and operationalize phenomena that were unimagined just a few years ago (Chen et al., 2022). In the social sciences, such potential could be applied to help code human behavior in a number of ways that are difficult to imagine. In the most immediate future, it is likely that AI will be used to at least partially replace subjective human coders for the purpose of inferring psychologically and socially meaningful phenomena from data. Today, research demonstrates the degree to which AI and humans judge content similarly (Gilardi et al., 2023; Kocoń et al., 2023; Zambrano et al., 2023), strongly implying the most immediate areas of interest and excitement for how AI can be used to automate highly common, but tedious, tasks that have already long been a part of the "human only" territory in the social science workflow.

This transformative influence extends beyond quantitative analysis of datasets for the purpose of quantifying data features (e.g., sentiment in facial expression and language, object and scene detection, body posture and movement) to the realm of statistical methodologies that are already integral to the training of graduate students in the quantitative social sciences. As AI grows increasingly capable of replicating and enhancing statistical methods, there is a potential to reduce oversights and errors in the analytical process, offering systematic evaluations of data and ensuring adherence to prescribed checklists. Such "compound AI" is already emerging (Zaharia et al., 2024), wherein our understanding and application of AI is not limited to specific models that can perform only highly specific or highly general tasks but, rather, is composed of multiple models interacting to accomplish more complex and abstract behaviors with already-impressive results.

Today, a nontrivial amount of graduate training is dedicated to learning basic (e.g., ordinary least-squares regression) and more complex statistical methods (e.g., cross-lagged panel models, structural equation modeling, etc.). However, once the rationale and application of such methods are learned, they often are relegated to rote tasks that are more reflective of one's technical ability than their intellectual prowess. Statistical analyses are often prescriptive—and highly so—constituting a cascading set of decisions that often have well-entrenched resolutions. Data are inspected for their various distributional qualities (e.g., center, variance, skew, kurtosis) and, depending on the data's properties and intended statistical method, various preprocessing and "cleaning" procedures may be adopted according to relatively dogmatic standards established by experts (e.g., "roping in" outliers, skew adjustment). Given the (often) deterministic manner in which statistical analyses unfold, particularly in a world of pre-registrations and analytic transparency, the quantitative social scientist is therefore typically trained in a range of statistical methods that can also be reproduced by emergent AI technologies according to best practices and contemporary knowledge of statistical theory—and, in some cases, perhaps creating new innovations in computing and statistical practice (e.g., Hutson, 2023; Mankowitz et al., 2023). AI will soon be able to reduce oversights and errors in the analytic process—in its simplest form, this may come in the form of a simple checklist to be followed while doing routine, systematic evaluations of the data to ensure that the analytic strategy is most appropriate given the latest insights into modeling strategies and limitations.

Perhaps more exciting is the prospect of AI to fully transform our understanding and appreciation of data analysis itself by facilitating the discovery of new ways to collect, quantify, and model important and causal relationships in data, both existing and yet to be created. This transformative influence of AI extends not only to the realm of quantitative analysis but also to the very fabric of data analysis itself, promising to redefine how we collect, quantify, and model important and causal relationships within data. As AI capabilities evolve, we envision a future where these systems not only replicate established statistical methods but actively contribute to the development of innovative approaches in computing and statistical practice. One can imagine AI-driven algorithms continuously adapting and refining analytical strategies based on the latest insights into modeling strategies and limitations. This dynamic interplay between AI technologies and human researchers not only ensures adherence to best practices but may also lead to the emergence of novel methodologies, pushing the boundaries of what we thought possible in data analysis.

Consider the scenario where AI, operating in conjunction with human researchers, is tasked with exploring complex social phenomena. In the realm of sentiment analysis, for instance, AI could go beyond the conventional identification of emotions in textual data and evaluate nuanced aspects of facial expressions and body language. By integrating advanced image recognition and natural language processing, AI systems could decipher intricate patterns of human behavior captured concurrently in images and videos. This not only enriches the depth of qualitative insights but also introduces a multidimensional understanding of how sentiments manifest across various modalities. The potential applications span from gauging public reactions to social events in real-time to providing nuanced insights into the emotional dynamics of interpersonal interactions. In essence, AI becomes a catalyst for the evolution of data analysis, propelling us toward a future where the synergy between human intellect and artificial intelligence results in a more nuanced, comprehensive, and innovative approach to understanding complex phenomena within our datasets.

Moreover, AI's transformative influence is not confined solely to the quantitative realm but also extends seamlessly into qualitative analysis, marking a paradigm shift in how researchers glean insights from textual sources. Today's natural language processing algorithms already serve as formidable tools, enabling the exploration of nuanced and intricate qualitative aspects embedded within vast datasets. These algorithms excel in deciphering the intricacies of language, discerning patterns, attentional processes, and contextual nuances that might elude traditional human analysis. In the context of social sciences, the integration of AI-driven NLP into qualitative analysis signifies a revolution in understanding human behavior, opinions, and societal trends. Imagine a research project evaluating public perceptions of climate change, where AI not only sifts through an extensive corpus of text from social media, news articles, and academic publications but also identifies underlying themes, sentiments, and the evolution of discourse over time. Through a complex and compound analysis, the system may discern shifts in public attitudes, identifying emerging concerns, controversies, or consensus within the discourse. Simultaneously, advanced AI may contribute to a more holistic, qualitative understanding of social and psychological phenomena by categorizing and extracting latent themes that align with our existing understanding of the discourse landscape, allowing researchers to uncover not only what is being discussed but also the intricate interplay of various scientific ideas, notions, and perspectives.

This multidimensional analytical approach, fusing quantitative and qualitative insights, holds the potential to elevate the depth and richness of research findings. As AI seamlessly navigates through the vast sea of textual data, it not only aids in summarizing and categorizing information but also discerns subtle nuances, cultural variations, and evolving linguistic expressions. This holistic understanding of textual sources equips researchers with a more comprehensive grasp of the phenomena under investigation, transcending traditional boundaries and paving the way for a more nuanced and inclusive interpretation of complex societal dynamics. The symbiotic relationship between human researchers and AI technologies, bridging the gap between quantitative and qualitative analysis, thus propelling the social sciences into an era of unprecedented depth and sophistication in research methodologies.

Beyond aiding in research design, data collection, and analyses, AI is well-positioned to streamline the preparatory process required to conduct research within institutions, including the production and proofing of research materials, Institutional Review Board (IRB) packets, survey questions, and grant writing. This is perhaps a great relief to many researchers who are already overburdened with bureaucracy and pre-work processes. These are typically time-consuming tasks that follow established patterns, and AI has the potential to simplify and expedite these processes. As a part of this process, AI can be used to anticipate ethical needs, quandaries, and risks, providing insights and solutions to challenges and obstacles that are likely to appear throughout the execution of research.[5]

New empirical frontiers

It is clear that there are numerous empirical tasks that currently consume human resources and appear to be ready-made for emerging AI technologies to take on, either directly or in collaboration with human researchers. Such tasks ranging from entry-level data analysis like running correlations to high-level conceptual disentanglement like resolving literature disputes are within the range of AI in the near- and slightly-less-near terms. We

5 Of course, we are not suggesting that AI be given *carte blanche* across all aspects of research design, analysis, and ethical considerations. Rather, we suggest that seeking multiple, well-informed perspectives for the identification of opportunities, challenges, and ethical concerns is nearly always a necessary pursuit for researchers within and outside of the social sciences. In particular, there are several ongoing efforts to make AI more inclusive of diverse cultures, perspectives, and human needs (e.g., Porayska-Pomsta and Rajendran, 2019; Kinnula et al., 2021; Park, 2022), and the inclusion of such broader considerations that humans typically lack—in spite of their best efforts and intentions—will likely help to reduce unintended harm and negative consequences that might result from the conduct of our research.

also envision a future where AI and LLMs aid in the unanticipated discovery of new phenomena of which we currently are completely unaware. The rest of this section opines the idea that AI can help facilitate new knowledge.

Much of social science is organized around a hypothetico-deductive model wherein scholars make hypotheses about how variables will be related, and evaluate if such hypotheses are supported. This top-down approach to social and psychological inquiry is supported by confirmatory analyses, and represents a relatively straightforward path to understand if variables are related or if an intervention worked. Bottom-up or exploratory analyses are also common and just as useful for pushing our basic understanding of social scientific theory forward (e.g., Tukey, 1962). One difficulty associated with exploratory research, however, is the potential for HARKing (Kerr, 1998), the possible lack of care around open science research practices (Nilsen et al., 2020), and the lack of a "backstop" to understand when exploration starts and ends (e.g., "How much exploration is too much exploration?"). Our perspective is that while such concerns related to exploration are valid and noteworthy, the potential of acquiring new knowledge about the human condition—as a result of such exploration—cannot be ignored, either. Humans have a finite ability to forecast all possible relationships between variables, and we do our best to examine such relationships. AI offers an opportunity for researchers to be hyper-exploratory, and reveal unexpected and valid relationships between variables that can lead to incredible discoveries, prompting more in-depth, cautious, and meticulous follow-up studies.

This form of hyper-exploration with AI is common in other fields like medicine (Bhinder et al., 2021). For example, AI and machine learning models have been used in the classification of cancers that have an unknown origin (Moon et al., 2023) and screening for diabetes (Guan et al., 2023). In anecdotal cases, providing LLMs with a list of symptoms has also led to proper diagnoses compared to seeing human doctors who could not isolate a cause of a medical condition (Holohan, 2023). We use these examples to signal that science, in general, is open to and is expanded by exploration, embracing it when the costs or unknowns associated with purely hypothetico-deductive or confirmatory research are high. Scholars in the social sciences may benefit from using AI as a tool for both confirmation and exploration, to the degree that they abide by best practices for scientific reporting and conduct (c.f., Acion et al., 2023).

A future with AI as a collaborative partner in exploratory research may look like giving LLMs or other tools a set of parameters or variables, and requesting those that best predict an outcome of interest. This is a relatively simple use case. However, imagine integrating several research processes we have discussed and having an AI perform them, simultaneously, to create a social and psychological *understanding* of individuals based on such data. In this envisioned future where AI becomes an integral collaborator in exploratory research, imagine a scenario where researchers provide a complex dataset to a language model alongside the nuanced *meanings* and *interpretations* of each variable. This marks a departure from conventional input, as the AI system not only processes the numerical or categorical values but also grasps the semantic intricacies associated with each variable.

For instance, in a study focused on mental health, variables may include traditional quantitative measures like survey scores alongside more qualitative elements such as personal narratives, cultural contexts, and linguistic nuances extracted through natural language processing. The researchers, in collaboration with the AI system, set out to uncover hidden patterns and relationships within this multifaceted dataset. Unlike traditional approaches that hinge on predefined hypotheses, the AI is tasked with exploring the intricate interplay between variables, both known and novel, guided by the nuanced meanings provided. For instance, the AI may discern subtle linguistic shifts in personal narratives that correlate with specific mental health outcomes, revealing the emotional nuances embedded within language that were previously overlooked. As the AI system goes deeper into the data, it not only identifies statistically significant associations but also generates hypotheses grounded in the semantic understanding of variables. For instance, it might propose that the confluence of certain linguistic expressions, survey responses, and cultural factors could be indicative of a previously unrecognized mental health resilience trait. This proposition arises not only from statistical patterns but from a semantic synthesis of the meanings attributed to each variable.

The researchers, armed with these AI-generated hypotheses, embark on a new phase of exploration, conducting targeted follow-up studies to validate and refine the emergent findings. The collaboration between human intuition, domain expertise, and the AI's capacity for hyper-exploration results in the illumination of a novel aspect of mental health that was obscured by traditional research paradigms. This organic integration of data organization, analysis, and exploration, guided by the semantic nuances of variables, showcases the potential for AI not just as a tool for confirmation but as a catalyst for the discovery of previously unknown phenomena, fundamentally altering the landscape of social and psychological inquiry.[6]

Considerations for an AI-infused scientific future

So far, we have articulated the ways that AI can become a part of the research process for social scientists. This relatively positive and forward-thinking overview suggests there are several opportunities for academics to leverage LLMs and become more efficient, creative, and flexible with their work. There are underlying, structural issues that appear downstream in the research process that academics need to wrestle with and resolve, however. We articulate these issues in the current section, encouraging readers to internalize such ideas and have conversations with their colleagues to work through such evolving ideas.

6 How, precisely, we deal with the ethics of such potentially transformative power is far beyond the scope of the current paper. As others have recognized and elaborated upon for centuries, vigilance for new technology's potential to be used for noble and horrific purposes alike is paramount to the ongoing, and rapidly-evolving, discussion about the development and deployment of AI technologies.

Productivity

The first issue relates to research productivity and what *counts* as a contribution to one's home discipline or field. Some departments operate on a publication points system, where a certain number of publications award the scholar credit toward their tenure and promotion dossier. Papers published in especially prestigious journals may be awarded more points than less prestigious journals. Using AI as part of one's research process will likely make certain tasks more efficient, which may facilitate greater research output and possibly more published papers (e.g., increasing the denominator of possible papers to be published in a given year). That is, what once took a scholar 10 min to write analysis code may now take them 30 s. By being 20 times more efficient, does this now suggest researchers should be 20 times more productive with their time and produce work at a substantially higher rate (while, of course, maintaining quality and rigor)?

The current example may seem hyperbolic but, in truth, it is not unrealistic. Outside of academia, the less-than-utopian impact that automation has played on work life has been discussed extensively. Rather than creating utopian work conditions that lead to increased leisure and personal time, more efficient production has driven several wholly unanticipated workforce outcomes, including complete reorganization of several sectors, with fiscal gains/benefits being reaped not by the employees who produce higher quality and quantities of output but, rather, by the institutions that house them (Novak, 2014; Hayes, 2017). The potential dilemma of increased research efficiency through AI integration leads us to a broader question: how should academia adapt to such changes in research productivity? While the prospect of producing more work in less time is enticing, the existing framework for evaluating scholarly contributions may not be equipped to handle this accelerated pace.

Promotion and tenure committees will be faced with comparing scholars who have used AI for their work against those who have not. Whether a scholar is required to disclose the use of AI as part of their research process is another issue to contend with altogether. Is this an equivalent comparison in productivity and scholarly impact (see: Biagioli, 2018)? Traditional metrics, like the number of publications, are deeply ingrained in academic evaluation systems. The current emphasis on prestigious journals and the quantity of published papers reflects a longstanding approach to assessing academic success. Moreover, the potential for AI-driven efficiency raises concerns about the quality and depth of research. Does a higher volume of publications *necessarily* translate to an equivalent increase in the generation of profound insights or groundbreaking knowledge? The risk of prioritizing quantity over substance looms large in this scenario. We do not have a clear way out of this conundrum nor an answer to this question, but departments and academic institutions need to be forthcoming with their AI policies for how academics do their work, and how their work will count in the face of emerging AI tools for academic research.

Beyond the individual researcher's perspective, there is the broader societal impact to consider. If academia shifts toward a more output-centric model, driven by AI-driven efficiency gains, how does this influence the dissemination of knowledge and the overall progress of research fields? Could it inadvertently add fuel to the fire of an increasing culture of academic "fast food," where quick and numerous publications—often benefitting from arguably superficial "style over substance" findings—take precedence over the meticulous and thoughtful cultivation of ideas (Woolston, 2014; Berezow, 2019; Mapes, 2020)? Such questions, when zoomed out to their logical conclusions, converge on greater questions of responsibility:

> Who, ultimately, is to be held accountable for the production and dissemination of research, findings, and publications that lack intellectual merit or scientific rigor when AI is involved in the ecosystem?

As we navigate the intersection of AI and academia, it becomes imperative to address these fundamental questions. Balancing the advantages of efficiency with the need for substantive contributions will likely require a reevaluation of the traditional metrics and an open dialogue among academics, institutions, and the wider academic community. Only through thoughtful consideration and collaborative efforts can we ensure that the integration of AI enhances, rather than diminishes, the core values of academic research.

Resource allocation, equity, and equality of access

A second issue relates to resource allocation in academia. Offer letters for faculty positions often include help from research assistants or graduate educators as part of one's start-up package. Will universities become less incentivized to provide human assistance to faculty if AI can demonstrate its research utility? This question ignores the essential and necessary training that both faculty and students receive from being a part of an advisor-advisee relationship, but departments that need to decide where to dedicate resources and how to allocate them might turn to AI as a cheaper option for faculty support. We are not necessarily advocating for this approach, as we believe that knowledge transfer from teacher to student and student to teacher would suffer immensely if AI became the default and sole research assistant for one's research program. However, we are not far from a reality where AI subscriptions may become part of one's start-up package, and human support and training therefore become deprioritized. The degree to which this impacts graduate training, faculty mentoring abilities, and research quality is an open question that humans must resolve.

Beyond resource allocation, a broader societal concern arises concerning the potential exacerbation of existing inequalities. If AI-driven solutions become a predominant force in academia, there is a risk that institutions with more substantial financial resources may disproportionately benefit, widening the gap between well-funded and less affluent academic entities. Affordability and accessibility to advanced AI tools can become a determining factor in the academic success of individuals and institutions. Smaller universities or those with limited financial means might find themselves at a disadvantage, struggling to keep pace with their more affluent counterparts. This not only impacts the quality of education for currently-underserved populations and research output for scholars without ample funding, but also perpetuates a system where disparities in resources lead to disparities in academic success which, in turn, fuels existing inequality in communities that are already marginalized in the educational landscape.

Furthermore, as AI becomes deeply ingrained in the academic world, questions of data ownership, privacy, and ethical considerations come to the forefront. The potential reliance on AI platforms could result in the concentration of data in the hands of a few major entities, raising concerns about the democratic dissemination of knowledge and the equitable representation of diverse perspectives. To navigate these challenges, a proactive and inclusive approach is essential. Academic institutions must prioritize policies that promote equal access to AI tools, ensuring that financial constraints do not impede the educational and research opportunities available to students and faculty. Collaborative efforts within the academic community and beyond are crucial to developing ethical guidelines, data governance frameworks, and ensuring that AI is harnessed responsibly.

Research evaluation

A third issue concerns peer review and AI's ability to serve as a reviewer for social scientific papers. Academics reading this work will likely have a list of experiences where they believe referees simply missed, misinterpreted, or ignored key elements of a paper that ultimately led to a negative editorial decision. It is reasonable that this happens. Reviewers are often unpaid for their time, limited in how available they are for peer review, and may not be content experts for the topics they are asked to review. Editors also face an uphill battle trying to find suitable reviewers for papers. Could (and should) AI serve as a reviewer on academic papers? As some top journals openly wrestle with policies on AI as coauthors at the foreground of a submission (Berdejo-Espinola and Amano, 2023; Thorp, 2023; Thorp and Vinson, 2023), it is important to consider that AI may play a critical role throughout the lifespan of a submission and during the review process as well. To the degree that AI can offer an objective and informed review, it is possible that AI may be able to resolve reviewer disputes on paper evaluations. For example, if Reviewer 1 suggests a paper should receive a revision, while Reviewer 2 suggests a paper should be rejected, the editor might solicit an AI review. In theory, a well-trained AI should have domain-level knowledge about a paper and assess its fit based on other submissions that have been accepted to a particular journal. The possible benefits of this approach are speed and presumed objectivity. The possible downsides of this approach are review validity and reliability (which, to be clear, are concerns with human reviewers as well). How much the human editor should weigh an AI evaluation is an open question, but an AI review may provide another piece of evidence to facilitate a timely and accurate decision on a paper.

Furthermore, AI is touted to have nearly all of human knowledge in its training data, suggesting it might be suited to assess certain parts of a manuscript. For example, if authors claim novelty in their work (e.g., "We are the first to examine..."), this is likely verifiable with AI. If authors state the assumptions of a theory incorrectly, this is likely verifiable with AI as well. We imagine, perhaps in a utopian manner, a world where AI can serve as an objective third reviewer alongside two humans. AI, if fine-tuned to the aims, scope, interests, and conventions of a journal and discipline, might assist editors in making an informed decision about a manuscript. This raises a host of other issues, including the degree to which authors should be notified that AI were involved in the review process of their article. However, consistent with our stated objectives of the current piece, it is important to raise nontrivial issues regarding academic workflows and processes where AI may become central. AI reviewing academic papers needs to be taken seriously just as much as AI writing academic papers.

Lastly, several dystopian possibilities arise from the questions around not just "if" AI is included in the scientific process, but additionally the questions of "who" and "how" when it comes to AI research applications. Currently, the majority of high-powered AI models are trained and deployed by a small number of massive, for-profit global corporations like Google, Microsoft, and Meta. In the near future, might a model for which we do not have absolute transparency into the training and guard railing process be positioned in such a way as to secretly encode corporate interests in the responses that they give? Might an AI model whose development was funded, in part, by an oil company more harshly review a paper demonstrating the public health risks of fossil fuel? Might an AI model lead researchers down a path of study that finds only positive aspects of living under authoritarian rule as a function of who trained it? History shows us, consistently and repeatedly, that those with power possess boundless ingenuity when it comes to promoting their own self-interests at the expense of public wellbeing, dignity, and human life itself (see, e.g., Brownell and Warner, 2009; Gates et al., 2015; Kovarik, 2021). To some degree, the democratization of access to AI may serve to eschew such conflicts of interest; we believe that equal access to emerging AI technologies is necessary, but not sufficient in and of itself, for preventing such issues from becoming embedded in AI's application to the next generation of research questions. We urge for ongoing, and increased, vigilance and discourse on such matters.

Presently, there exists a consensus within academic communities regarding the imperative nature of upholding complete transparency in the use of AI. There is an increasing demand for researchers to elucidate the incorporation of AI within academic manuscripts and in the evaluation of research proposals. Esteemed publications such as *Science* have initiated the acceptance of manuscripts integrating AI, underscoring its potential to augment the accessibility of scholarly content to a broader audience. We advocate for meticulous disclosure of AI use in academic discourse during this nascent phase. Nonetheless, we anticipate a future wherein the integration of AI in such endeavors will become as commonplace as employing spell-check functionalities, thereby obviating the necessity for disclosure.

This prompts a broader reflection: as AI grows increasingly integrated into our scientific methodologies, novel standards are bound to emerge. It is imperative not only to divulge the use of AI in disseminating research through manuscripts and other documentation but also to disclose additional AI attributes, such as source code, training parameters, training interventions (e.g., "guard-railing" procedures) and training data. Analogous to the principles of open science as well as broader interoperability concerns in the world of free and open-source software, which advocate for transparent dissemination of all research components, encompassing data origins and software utilities, we posit that these standards should encompass AI tools as well. Similar to how open science proponents advocate for streamlined research endeavors by mitigating duplication of resources among researchers, embracing such practices with AI can likewise foster collaboration and efficacy in scientific undertakings. However, it is noteworthy that resistance may arise not only from certain segments of the academic community but, in particular, from external stakeholders: namely, giants of the tech industry, who stand to forfeit a sizeable profit resulting from a "walled garden" model of access to AI (see, e.g., Doctorow, 2023).

Public engagement and policy

Finally, we also believe that AI provides an opportunity for better and more approachable public understanding of science. A host of evidence from communication research and psychology suggests people perceive the writers of complex texts to be less intelligent, more difficult to understand, and less warm and moral than the writers of simple texts (Oppenheimer, 2006, 2008; Markowitz et al., 2021). Given how complex most science papers are written for the average person, it is therefore in scientists' best interest, and perhaps a scholarly imperative, to communicate research in simple terms. This process can be assisted or facilitated by AI. For example, LLMs are often effective information summarizers because they can condense massive amounts of text to the essential pieces of information. We took the abstract of this paper and prompted ChatGPT (model GPT-4) with the following text: "Provide a summary of this paragraph for a lay audience in three bullet points that each have <20 words." It replied with: *(1) Social sciences, traditionally based on comparisons, are now integrating AI as key collaborators in research. (2) AI is envisioned as a versatile tool for both high-level discoveries and routine research tasks in social sciences. (3) Ethical AI use is emphasized for transparent, fair, and inclusive advancements in social science research.* These responses, we believe, are quite accurate and fair even without any fine-tuning. Therefore, AI and LLMs offer an immense opportunity for journals to become more inclusive and approachable for the masses.

Outlets like the *Proceedings of the National Academy of Sciences* and *Psychological Science* require significance statements that are intended to engage the public in the science using simple language. This worthwhile requirement can now become commonplace for journals, at scale, with little cost to the authors or publishers. Imagine that for each online publication, there is dedicated space for an AI-generated public summary of the scientific paper that summarizes its main findings—and perhaps more importantly—its implications for lay audiences. We believe that this openness and transparency might increase public goodwill and trust in science (e.g., Song et al., 2022), which are unfortunately decreasing among many Americans (Kennedy and Tyson, 2023). Journals may be incentivized to experimentally test if such summaries engage readers within and outside of academia, and if they help to increase credibility of the science that is published, ideally with a principled approach to avoid becoming a "race to the bottom" for maximizing page clicks and other metrics that have become key drivers of the attention economy (Huberman, 2013).

In the realm of public policy, the intersection of AI and scientific research will likely find much of its impact when traced to more "upstream" sources of public policy and funding decisions. As AI becomes increasingly integrated into various aspects of the research process, the question arises: How will granting agencies leverage AI to effectively evaluate the greatest needs and potential benefits of research proposals? What type of philosophical frameworks should guide such decisions (e.g., a utilitarian view that emphasizes the greatest gains for the public in spite of the costs to a few vs. a Kantian view of what acts themselves are right vs. wrong?; for two separate thought-provoking reads, see, e.g., Sison and Redín, 2023; Volkman and Gabriels, 2023). Granting agencies play a pivotal role in shaping the research landscape by allocating funds to projects with the potential for significant contributions to knowledge and societal advancement. However, we note the lack of objectivity in the broad array of definitions that may be offered for "significant contributions" to each of these domains. On the one hand, AI could enhance the objectivity and efficiency of various aspects of the grant review process, providing a more comprehensive and thorough analysis of strengths and weaknesses for any given proposal. On the other hand, estimations and recommendations by AI itself may suffer from the typical survivorship biases found in the funding landscape, prioritizing work that conforms to "traditional" success at the negligence of potentially groundbreaking projects that break the established mold. In the case of an especially advanced AI that might be able to offer forecasts of success in various domains, resource limitations will still require humans to ultimately make decisions about which domains are *most* pressing and which problems require the most urgent attention or pose the most existential threats to humanity (e.g., coronal mass ejection mitigation, global climate change, nuclear holocaust, world hunger, etc.), and what the philosophical implications of engineering human behavior and society in such a fashion as to tackle such concerns would entail (Skinner, 1971).

Conclusion

The technological innovations associated with AI and LLMs have been remarkable, pushing research forward in various fields of computation, automation, language studies, and beyond. These innovations have encouraged scholars to wrestle with the idea that AI may also find its way into the academic research process, and to consider what this means for our everyday workflows and understanding of scientific inquiry. In the present piece, we offered several points for consideration about how AI and LLMs can be used in everyday academic work. We couched our discussion in the admission that many of our ideas may sound hyperbolic or unrealistic, but perhaps only in the short term. Having an AI serve as a capable and reliable research assistant, for example, will take significant training and fine-tuning, but it is unclear how much will actually be needed to reach competency (especially, relative to human trainees).

Ultimately, we believe it is up to academic communities, and leadership within such communities, to articulate norms and ground rules for the use of AI in academic work. A *laissez faire* or "anything goes" approach may appear haphazard or unthoughtful to academics who crave direction and structure for how new tools should be used. In this scenario, an AI-infused scientific community would be like "the wild west," where scholars cannot understand human-level contributions to knowledge creation and it is unclear how AI tools may be used in the first place. We are cautiously against this approach, and encourage a more thoughtful, intentional, and problem/task-specific use of AI in academic research. At the very least, the search for a healthy and productive balance or equilibrium between multiple approaches and paradigms will be essential to the successful adoption of emerging AI technology. Authors are often looking for direction and best practices with AI (Jobin et al., 2019), and we encourage academic leadership to take this issue seriously. Even should guardrails be provided, there will be scholars who naturally bypass such rules. This reality is unavoidable. Textual and visual detection algorithms may therefore need to be implemented in a manner that cannot be eluded (e.g., plagiarism detection, watermarking technology, etc.; see, e.g., Saberi et al., 2024).

Further, like the open science movement has successfully labeled papers as adhering to particular research practices, journals may consider labeling papers as "AI-Contributed." It is unclear how people may perceive such papers—on the one hand, scholars may believe that the authors were "cheating" by having AI do some of the work; on the other hand, scholars may believe that the authors were creative and smart for using tools to their advantage. We encourage all authors—users and non-users of AI, alike—to identify how norms around the use and acceptance of AI develop, and may change over time.

In the ever-evolving landscape of academic research intertwined with AI, we find ourselves at a pivotal juncture where the promise of innovation and the need for ethical considerations converge. As we envision a future where AI seamlessly integrates into scholarly workflows, one must grapple with the profound transformations it may usher in. The idea of AI serving as a capable research assistant, while currently laden with challenges, beckons us to question the very nature of human contributions to knowledge creation. What might seem hyperbolic today may become commonplace tomorrow, as technology advances and AI undergoes continuous refinement.

Amidst the many uncertainties, the responsibility lies with academic communities—in collaboration with the general public—to chart a course that balances the allure of AI-driven efficiency with the enduring values of intellectual rigor. A thoughtful embrace of AI, anchored in clear guidelines and proactive measures, can illuminate a path toward a harmonious coexistence. Yet, this journey demands not just adherence to rules but a dynamic adaptation to the evolving interplay between human intellect and artificial intelligence. It prompts us to envision a future where AI augments the scholarly pursuit, not overshadowing it, and where each stride forward is guided by an unwavering commitment to excellence.

In the pursuit of this delicate equilibrium, the role of academic leadership becomes paramount. Their task is not merely to set guardrails but to cultivate an environment that fosters innovation, integrity, adaptability, and equity. Even as we acknowledge that some may test the boundaries, the proactive implementation of foolproof algorithms becomes a safeguard against unintended pitfalls. The integration of AI into the academic tapestry is not a sprint but a deliberate marathon, demanding continuous vigilance, discourse, and a commitment to steering this transformative force toward a future where the synergy between human intellect and artificial intelligence propels us to unprecedented heights of knowledge and discovery.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DM: Conceptualization, Writing – original draft, Writing – review & editing, Funding acquisition. RB: Conceptualization, Writing – original draft, Writing – review & editing. KB: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JJ declared a shared affiliation with the author RB to the handling editor at the time of review.

Publisher's note

References

Acion, L., Rajngewerc, M., Randall, G., and Etcheverry, L. (2023). Generative AI poses ethical challenges for open science. *Nat. Hum. Behav.* 7, 1800–1801. doi: 10.1038/s41562-023-01740-4

Aiden, E., and Michel, J.-B. (2013). *Uncharted: Big data as a Lens on Human Culture*. New York: Riverhead Books.

André, P., Schraefel, M. C., Teevan, J., and Dumais, S. T. (2009). "Discovery is never by chance: Designing for (un)serendipity," in *Proceedings of the Seventh ACM Conference on Creativity and cognitIon* C&c'09 (New York, NY, USA: Association for Computing Machinery), 305–314. doi: 10.1145/1640233.1640279

Bem, D. (2003). "Writing the empirical journal article," in *The Compleat Academic: A Practical Guide for the Beginning Social Scientist, 2nd Edition,* eds. J. M. Darley and M. P. Zanna (Washington, DC: American Psychological Association).

Berdejo-Espinola, V., and Amano, T. (2023). AI tools can improve equity in science. *Science* 379, 991–991. doi: 10.1126/science.adg9714

Berezow, A. (2019). "Blame academia for junk science and media hype?," in *American Council on Science and Health*. Available online at: https://www.acsh.org/news/2019/02/19/blame-academia-junk-science-and-media-hype-13826 (accessed February 27, 2024).

Bhinder, B., Gilvary, C., Madhukar, N. S., and Elemento, O. (2021). Artificial intelligence in cancer research and precision medicine. *Cancer Discov.* 11, 900–915. doi: 10.1158/2159-8290.CD-21-0090

Biagioli, M. (2018). Quality to impact, text to metadata: publication and evaluation in the age of metrics. *KNOW J. Form. Knowl.* 2, 249–275. doi: 10.1086/699152

Bogost, I. (2022). "ChatGPT is dumber than you think," in *The Atlantic*. Available online at: https://www.theatlantic.com/technology/archive/2022/12/chatgpt-openai-artificial-intelligence-writing-ethics/672386/ (accessed February 20, 2024).

Bohannon, J. (2010). Google opens books to new cultural studies. *Science* 330, 1600–1600. doi: 10.1126/science.330.6011.1600

Bommasani, R., Creel, K. A., Kumar, A., Jurafsky, D., and Liang, P. S. (2022). Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Adv. Neural Inf. Process. Syst.* 35, 3663–3678.

Brownell, K. D., and Warner, K. E. (2009). The perils of ignoring history: Big Tobacco played dirty and millions died. How similar is Big Food? *Milbank Q.* 87, 259–294. doi: 10.1111/j.1468-0009.2009.00555.x

Chen, B., Huang, K., Raghupathi, S., Chandratreya, I., Du, Q., and Lipson, H. (2022). Automated discovery of fundamental variables hidden in experimental data. *Nat. Comput. Sci.* 2, 433–442. doi: 10.1038/s43588-022-00281-6

Chitra, U., and Musco, C. (2020). "Analyzing the impact of filter bubbles on social network polarization," in *Proceedings of the 13th International Conference on Web Search and Data Mining WSDM'20.* (New York, NY, USA: Association for Computing Machinery), 115–123.

Church, K. W. (2017). "Corpus methods in a digitized world," in *Computational and Corpus-Based Phraseology Lecture Notes in Computer Science*, ed. R. Mitkov (Cham: Springer International Publishing), 3–15.

Cohn, A. C., Mahon, B. E., and Walensky, R. P. (2022). One year of COVID-19 vaccines: a shot of hope, a dose of reality. *JAMA* 327, 119–120. doi: 10.1001/jama.2021.23962

de Rond, M. (2014). The structure of serendipity. *Cult. Organ.* 20, 342–358. doi: 10.1080/14759551.2014.967451

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., et al. (2023). Using large language models in psychology. *Nat. Rev. Psychol.* 2, 688–701. doi: 10.1038/s44159-023-00241-5

DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., and Liu, S. (2018). "How people form folk theories of social media feeds and what it means for how we study self-presentation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems CHI'18.* (New York, NY, USA: Association for Computing Machinery), 1–12.

DeVito, M. A., Gergle, D., and Birnholtz, J. (2017). ""Algorithms ruin everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems CHI'17.* (New York, NY, USA: Association for Computing Machinery), 3163–3174.

Doctorow, C. (2023). *The Internet Con: How to Seize the Means of Computation*. London, New York: Verso.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. *PLOS ONE* 6:e26752. doi: 10.1371/journal.pone.0026752

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., et al. (2021). "Documenting large webtext corpora: a case study on the Colossal Clean Crawled Corpus," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, eds. M.-F. Moens, X. Huang, L. Specia, and S. W. Yih (Online and Punta Cana, Dominican Republic: Association for Computational Linguistics), 1286–1305. 8

Edward Foster, A., and Ellis, D. (2014). Serendipity and its study. *J. Doc.* 70, 1015–1038. doi: 10.1108/JD-03-2014-0053

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., et al. (2016). "First I "like" it, then I hide it: folk theories of social feeds," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose California USA: ACM), 2371–2382.

Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technol. Forecast. Soc. Change* 3, 1–22. doi: 10.1016/S0040-1625(71)80001-X

Franken, A., and Proft, P. (2023). "I'm a hack, by ChatGPT," in *New Yorker*. Available online at: https://www.newyorker.com/magazine/2023/08/07/im-a-hack-by-chatgpt (accessed February 20, 2024).

French, M., and Hancock, J. (2017). *What's the Folk Theory? Reasoning About Cyber-Social Systems.*

Gates, G., Ewing, J., Russell, K., and Watkins, D. (2015). "How Volkswagen's 'defeat devices' worked," in *N. Y. Times*. Available online at: https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html (accessed February 27, 2024).

Gilardi, F., Alizadeh, M., and Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proc. Natl. Acad. Sci.* 120:e2305016120. doi: 10.1073/pnas.2305016120

Guan, Z., Li, H., Liu, R., Cai, C., Liu, Y., Li, J., et al. (2023). Artificial intelligence in diabetes management: advancements, opportunities, and challenges. *Cell Rep. Med.* 4:101213. doi: 10.1016/j.xcrm.2023.101213

Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C. C. T., Del Giorno, A., Gopi, S., et al. (2023). *Textbooks Are All You Need.*

Hayes, B. (2017). "Automation on the job," in *American Science.* Available at: https://www.americanscientist.org/article/automation-on-the-job (accessed February 27, 2024).

Holohan, M. (2023). "A boy saw 17 doctors over 3 years for chronic pain. ChatGPT found the diagnosis," in *TODAY.com*. Available online at: https://www.today.com/health/mom-chatgpt-diagnosis-pain-rcna101843 (accessed February 21, 2024).

Huang, H. (2023). "The generative AI revolution has begun: How did we get here?," in *Ars Tech.* Available online at: https://arstechnica.com/gadgets/2023/01/the-generative-ai-revolution-has-begun-how-did-we-get-here/ (accessed February 27, 2024).

Huberman, B. A. (2013). Social computing and the attention economy. *J. Stat. Phys.* 151, 329–339. doi: 10.1007/s10955-012-0596-5

Hutson, M. (2023). DeepMind AI creates algorithms that sort data faster than those built by people. *Nature* 618, 443–444. doi: 10.1038/d41586-023-01883-4

Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi: 10.1037/0003-066X.60.6.581

Jakesch, M., Hancock, J. T., and Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proc. Natl. Acad. Sci.* 120:e2208839120. doi: 10.1073/pnas.2208839120

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* 1, 389–399. doi: 10.1038/s42256-019-0088-2

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18243–18250. doi: 10.1073/pnas.1012933107

Karikó, K., Buckstein, M., Ni, H., and Weissman, D. (2005). Suppression of RNA recognition by toll-like receptors: the Impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* 23, 165–175. doi: 10.1016/j.immuni.2005.06.008

Kennedy, B., and Tyson, A. (2023). "Americans' trust in scientists, positive views of science continue to decline," in *Pew Research Center Sci. Soc.* Available online at: https://www.pewresearch.org/science/2023/11/14/americans-trust-in-scientists-positive-views-of-science-continue-to-decline/ (accessed February 25 2024).

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personal. Soc. Psychol. Rev.* 2, 196–217. doi: 10.1207/s15327957pspr0203_4

Kinnula, M., Iivari, N., Sharma, S., Eden, G., Turunen, M., Achuthan, K., et al. (2021). "Researchers' toolbox for the future: understanding and designing accessible and inclusive artificial intelligence (AIAI)," in *Proceedings of the 24th International Academic Mindtrek Conference Academic Mindtrek'21* (New York, NY, USA: Association for Computing Machinery), 1–4.

Köbis, N., and Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Comput. Hum. Behav.* 114:106553. doi: 10.1016/j.chb.2020.106553

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., et al. (2023). ChatGPT: Jack of all trades, master of none. *Inf. Fusion* 99:101861. doi: 10.1016/j.inffus.2023.101861

Kosinski, M. (2023). *Theory of Mind May Have Spontaneously Emerged in Large Language Models*.

Koubaa, A., Qureshi, B., Ammar, A., Khan, Z., Boulila, W., and Ghouti, L. (2023). Humans are still better than ChatGPT: case of the IEEEXtreme competition. *Heliyon* 9:e21624. doi: 10.20944/preprints202305.0780.v1

Kovarik, B. (2021). "A century of tragedy: how the car and gas industry knew about the health risks of leaded fuel but sold it for 100 years anyway," in *The Conversation*. Available online at: http://theconversation.com/a-century-of-tragedy-how-the-car-and-gas-industry-knew-about-the-health-risks-of-leaded-fuel-but-sold-it-for-100-years-anyway-173395 (accessed February 27, 2024).

Kreps, S., McCain, R. M., and Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *J. Exp. Polit. Sci.* 9, 104–117. doi: 10.1017/XPS.2020.37

Krock, L. (2001). "Accidental discoveries," in *PBS*. Available online at: https://www.pbs.org/wgbh/nova/article/accidental-discoveries/ (accessed February 27, 2024).

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., et al. (2009). Computational social science. *Science* 323, 721–723. doi: 10.1126/science.1167742

Lee, H. (2019). "15 years of Google Books," in *Google*. Available online at: https://blog.google/products/search/15-years-google-books/ (accessed February 27, 2024).

Mankowitz, D. J., Michi, A., Zhernov, A., Gelmi, M., Selvi, M., Paduraru, C., et al. (2023). Faster sorting algorithms discovered using deep reinforcement learning. *Nature* 618, 257–263. doi: 10.1038/s41586-023-06004-9

Mapes, D. (2020). "Spinning science: overhyped headlines, snarled statistics lead readers astray," in *Fred Hutch Cancer Cent.* Available online at: https://www.fredhutch.org/en/news/center-news/2020/02/spinning-science-overhyped-headlines-snarled-statistics-lead-readers-astray.html (accessed February 27, 2024).

Markowitz, D. M., and Hancock, J. T. (2023). Generative AI are more truth-biased than humans: a replication and extension of core truth-default theory principles. *J. Lang. Soc. Psychol.* 43, 261–267. doi: 10.31234/osf.io/hm54g

Markowitz, D. M., Kouchaki, M., Hancock, J. T., and Gino, F. (2021). The deception spiral: Corporate obfuscation leads to perceptions of immorality and cheating behavior. *J. Lang. Soc. Psychol.* 40, 277–296. doi: 10.1177/0261927X20949594

McBirnie, A. (2008). Seeking serendipity: the paradox of control. *Aslib Proc.* 60, 600–618. doi: 10.1108/00012530810924294

McCarthy, J. (2007). *What is Artificial Intelligence?*

Mei, Q., Xie, Y., Yuan, W., and Jackson, M. O. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. *Proc. Natl. Acad. Sci.* 121:e2313925121. doi: 10.1073/pnas.2313925121

Mellon, J., and Prosser, C. (2017). Twitter and Facebook are not representative of the general population: political attitudes and demographics of British social media users. *Res. Polit.* 4:2053168017720008. doi: 10.1177/2053168017720008

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science* 331, 176–182. doi: 10.1126/science.1199644

Mitra, A., Del Corro, L., Mahajan, S., Codas, A., Simoes, C., Agarwal, S., et al. (2023). *Orca 2: Teaching Small Language Models How to Reason*.

Moon, I., LoPiccolo, J., Baca, S. C., Sholl, L. M., Kehl, K. L., Hassett, M. J., et al. (2023). Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nat. Med.* 29, 2057–2067. doi: 10.1038/s41591-023-02482-6

Nilsen, E. B., Bowler, D. E., and Linnell, J. D. C. (2020). Exploratory and confirmatory research in the open science era. *J. Appl. Ecol.* 57, 842–847. doi: 10.1111/1365-2664.13571

Novak, M. (2014). "The late great American promise of less work," in *Gizmodo*. Available online at: https://gizmodo.com/the-late-great-american-promise-of-less-work-1561753129 (accessed February 27, 2024).

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: problems with using long words needlessly. *Appl. Cogn. Psychol.* 20, 139–156. doi: 10.1002/acp.1178

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends Cogn. Sci.* 12, 237–241. doi: 10.1016/j.tics.2008.02.014

Park, T. M. (2022). "Making AI inclusive: 4 guiding principles for ethical engagement," in *Partnership on AI*. Available online at: https://partnershiponai.org/paper/making-ai-inclusive-4-guiding-principles-for-ethical-engagement/ (accessed February 27, 2024).

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York, NY, US: W W Norton & Co.

Porayska-Pomsta, K., and Rajendran, G. (2019). "Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion," in *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices Perspectives on Rethinking and Reforming Education*, eds. J. Knox, Y. Wang, and M. Gallagher (Singapore: Springer), 39–59.

Prather, J., Denny, P., Leinonen, J., Becker, B. A., Albluwi, I., Craig, M., et al. (2023). "The robots are here: Navigating the generative AI revolution in computing education," in i*Proceedings of the 2023 Working Group Reports on Innovation and Technology in Computer Science Education ITiCSE-WGR'23.* (New York, NY, USA: Association for Computing Machinery), 1108–159.

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C., and Bavel, J. J. V. (2023). *GPT is an Effective Tool for Multilingual Psychological Text Analysis*.

Ritchie, H., Mathieu, E., Roser, M., and Ortiz-Ospina, E. (2023). "Internet," in *Our World Data.* Available online at: https://ourworldindata.org/internet (accessed February 21, 2024).

Robinson, B., Gonnerman, C., and O'Rourke, M. (2019). Experimental philosophy of science and philosophical differences across the sciences. *Philos. Sci.* 86, 551–576. doi: 10.1086/703553

Rose, L. T., and Fischer, K. W. (2011). Garbage in, garbage out: having useful data is everything. *Meas. Interdiscip. Res. Perspect.* 9, 222–226. doi: 10.1080/15366367.2011.632338

Roser, M. (2024). "The Internet's history has just begun," *in Our World Data*. Available online at: https://ourworldindata.org/internet-history-just-begun (accessed February 21, 2024).

Saberi, M., Sadasivan, V. S., Rezaei, K., Kumar, A., Chegini, A., Wang, W., et al. (2024). *Robustness of AI-Image Detectors: Fundamental Limits and Practical Attacks*.

Sclar, M., Kumar, S., West, P., Suhr, A., Choi, Y., and Tsvetkov, Y. (2023). "Minding language models' (lack of) theory of mind: A plug-and-play multi-character belief tracker," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, eds. A. Rogers, J. Boyd-Graber, and N. Okazaki (Toronto, Canada: Association for Computational Linguistics), 13960–13980.

Shah, S. (2023). "Sam Altman on OpenAI and Artificial General Intelligence," in *TIME*. Available online at: https://time.com/6344160/a-year-in-time-ceo-interview-sam-altman/ (accessed February 6, 2024).

Sison, A. J. G., and Redín, D. M. (2023). A neo-Aristotelian perspective on the need for artificial moral agents (AMAs). *AI Soc.* 38, 47–65. doi: 10.1007/s00146-021-01283-0

Skinner, B. F. (1971). *Beyond Freedom and Dignity*. New York, NY, US: Knopf/Random House.

Song, H., Markowitz, D. M., and Taylor, S. H. (2022). Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics. *J. Commun.* 72, 497–510. doi: 10.1093/joc/jqac017

Sutter, R. C. (2021). The pre-Columbian peopling and population dispersals of South America. *J. Archaeol. Res.* 29, 93–151. doi: 10.1007/s10814-020-09146-w

Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science* 379, 313–313. doi: 10.1126/science.adg7879

Thorp, H. H., and Vinson, V. (2023). Editor's note. *Science* 379, 991–991. doi: 10.1126/science.adh3689

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Stat.* 33, 1–67. doi: 10.1214/aoms/1177704711

Van Noorden, R., and Perkel, J. M. (2023). AI and science: what 1,600 researchers think. *Nature* 621, 672–675. doi: 10.1038/d41586-023-02980-0

Volkman, R., and Gabriels, K. (2023). AI moral enhancement: upgrading the socio-technical system of moral engagement. *Sci. Eng. Ethics* 29, 11. doi: 10.1007/s11948-023-00428-2

Woolston, C. (2014). Study points to press releases as sources of hype. *Nature* 516, 291–291. doi: 10.1038/nature.2014.16551

Zaharia, M., Khattab, O., Chen, L., Davis, J. Q., Miller, H., Potts, C., et al. (2024). "The shift from models to compound AI systems," in *Berkeley Artificial Intelligence Research Lab*. Available online at: https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/ (accessed February 27, 2024).

Zambrano, A. F., Liu, X., Barany, A., Baker, R. S., Kim, J., and Nasiar, N. (2023). "From nCoder to ChatGPT: From automated coding to refining human coding," in *Advances in Quantitative Ethnography Communications in Computer and Information Science,* eds. G. Arastoopour Irgens and S. Knight (Cham: Springer Nature Switzerland), 470–485.