



## Information Systems Research

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems

Anjana Susarla, Ram Gopal, Jason Bennett Thatcher, Suprateek Sarker

To cite this article:

Anjana Susarla, Ram Gopal, Jason Bennett Thatcher, Suprateek Sarker (2023) The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems. Information Systems Research 34(2):399-408.  
<https://doi.org/10.1287/isre.2023.ed.v34.n2>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.





For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# The Janus Effect of Generative AI: Charting the Path for Responsible Conduct of Scholarly Activities in Information Systems

Anjana Susarla,<sup>a</sup> Ram Gopal,<sup>b</sup> Jason Bennett Thatcher,<sup>c</sup> Suprateek Sarker<sup>d,\*</sup>

<sup>a</sup>Department of Accounting and Information Systems, Eli Broad College of Business, Michigan State University, East Lansing, Michigan 48824; <sup>b</sup>Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom; <sup>c</sup>Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122; <sup>d</sup>McIntire School of Commerce, University of Virginia, Charlottesville, Virginia 22901

\*Corresponding author

Contact: [asusarla@msu.edu](mailto:asusarla@msu.edu),  <https://orcid.org/0000-0001-7482-1213> (AS); [ram.gopal@wbs.ac.uk](mailto:ram.gopal@wbs.ac.uk),  <https://orcid.org/0000-0003-4241-9355> (RG); [jason.thatcher@temple.edu](mailto:jason.thatcher@temple.edu),  <https://orcid.org/0000-0002-7136-8836> (JBT); [eic-isr@virginia.edu](mailto:eic-isr@virginia.edu),  <https://orcid.org/0000-0002-8079-3121> (SS)

Published Online in Articles in Advance:  
May 26, 2023

Funding: A. Susarla was funded by an R01 grant from the National Library of Medicine, through [Grant R01LM013443].

<https://doi.org/10.1287/isre.2023.ed.v34.n2>

Copyright: © 2023 INFORMS

## 1. Introduction

Generative artificial intelligence (AI) broadly refers to a class of AI models that generate seemingly new content in the form of text, images, or other media. Although generative AI techniques have been around for some time, the release of ChatGPT unleashed an explosion of conversations in the popular press, online forums, and academic circles. ChatGPT has aroused concern because it makes the broad and open set of functionalities of generative AI models accessible through simple queries to a browser-based interface. Within two months of its beta release, ChatGPT attracted more than 100 million users.

Conversations about what generative AI tools, particularly those based on Large Language Models (LLMs), mean for human learning, knowledge creation, and the future of work have consumed traditional news media as well as social media (Dwivedi et al. 2023). Discussions have ranged from expressing concern regarding the new Promethean moment that generative AI represents for humanity (Freidman 2023) to extolling the possibilities that these tools present for renewed creativity and positive social change (Larsen and Narayan 2023). As conversations about generative AI have been echoed and amplified in offline and online forums, in parallel, there has been rapid proliferation of generative AI-powered applications and corporate actions, resulting in what some are calling an AI arms race.

In our collective experience, we have never seen such an explosion of applied and academic interest in an information technology (IT) tool. We find this surge of interest exhilarating because we appear to be on the cusp of change in academic systems—disconcerting because

generative AI and its capabilities are so often mischaracterized, and highly promising because appreciating the changes that these tools might bring requires an understanding of not only the technology and the data that inform it but also of the ways in which academics will construct meaning, rules, and routines around the application of generative AI.

Academic conversations about generative AI have moved beyond bemoaning the decline of learning to thinking more critically about how scholars will and should use generative AI tools. Although they may be able to help automate aspects of writing or serve as an aid to the creative process, can they replace our accumulated experience or insights into the problems with which we grapple? Serious questions have also surfaced about the ability of humans to detect work supported by such tools. *Nature* reported on a recent study that tested humans' ability to spot artificial research paper abstracts generated by ChatGPT (Else 2023). These abstracts were evaluated by a plagiarism detector, an AI-output detector, and a group of blinded human reviewers (Gao et al. 2022). Soberingly, the blinded human reviewers were only able to correctly identify about 68% of the abstracts that were generated by ChatGPT but incorrectly identified 14% of the original (human-generated) abstracts as being generated by ChatGPT. As AI-generated content becomes indistinguishable from human-generated content, it is imperative that we establish norms, policies, and procedures regarding the use of these tools in the research process.

Recognizing that generative AI can be a double-edged sword for scholarly work, this editorial<sup>1</sup> seeks to start a

conversation about how we believe generative AI tools, including but not limited to ChatGPT, should be used by the information systems (IS) community to conduct and publish research. In doing so, we neither adopt a naively optimistic perspective—one that would normalize such technologies as nothing more than “calculators,” at least in the short run—nor do we adopt a position of deep pessimism, as espoused by notable scholars such as Noam Chomsky, who believes that ChatGPT-style systems “may have some value for something,” but “it’s not obvious what,” and characterizes ChatGPT as “basically high-tech plagiarism” that will lead to avoidance of learning (Chomsky et al. 2023). Rather, we seek to offer a balanced reflection on the changing nature of scholarly work and offer suggestions on how the IS community can cautiously embrace the changes. We acknowledge that we take some liberties in speculating about the possibilities and applications of generative AI in research and given the pace of change in the technology and tools, some of our assertions and suggested guidelines may eventually seem misplaced. However, we anticipate that by offering these preliminary thoughts based on our experience as authors, reviewers, and editors, we can contribute to a foundation for more refined applications of generative AI to support scholarship that will appear over time in journals such as *Information Systems Research (ISR)*.

In the following pages, we first provide a brief overview of the development of LLMs that are at the heart of tools such as ChatGPT,<sup>2</sup> and can be adapted to a variety of downstream tasks. We note that the overview and ideas that follow are descriptive and are not intended to advocate for the broader adoption of such models in research. Next, based on the published literature and our “analysis” of selected papers, we highlight some of the possible benefits of using these tools for developing, reviewing, and reading/understanding papers. Then, we highlight concerns about using current generative AI tools to support scholarly research activities in ways that could be dysfunctional or misleading. We end our editorial with some reflections and some thoughts for the future.

## 2. Language Models and Their Evolution: A Brief Overview

In this section, we provide a brief introduction to the language models (LLMs), the type of generative AI that is particularly relevant to scholarly work, and to our editorial. Although some readers may find this section useful, readers with advanced knowledge in this area may wish to skip this section.

The origins of language modeling can be traced back to the early days of statistical natural language processing (NLP), which emerged in the 1950s and 1960s. Early language models typically used statistical techniques to estimate the probability of a given word or string of words

occurring in a language (Russell and Norvig 2003), using methods such as *n-grams*, or sequences of *n* words. The evolution from *n-gram*-based models to *word embeddings* began in the mid-2000s with the development of the Word2Vec algorithm (Mikolov et al. 2013), which introduced a vector representation to capture the semantic meaning of words. Such vector representation of words introduced a new era in language modeling as it can be the input the recurrent neural network and later the encode-decoder architecture, which greatly transformed downstream NLP tasks such as text summarization and machine translation (Sutskever et al. 2014). A further breakthrough in language modeling came with the introduction of the *transformer* architecture (Vaswani et al. 2017). The transformer architecture paved the way for the introduction of *large language models* (LLMs) (Wei et al. 2022), which are trained on massive amounts of text data from diverse sources such as books, news articles, web pages, and social media posts. With advances in model building and model training, we have entered a new era with the emergence of what some have dubbed *foundation models*—models that are trained on massive amounts of data and tuned with billions of hyperparameters that can be adapted to a wide range of tasks, including some tasks that the model was not originally trained on (Wei et al. 2022). Generative models are now capable of creating realistic and creative content from various inputs, such as text, images, audio, and video. In the future, such models may be able to produce novel and complex content, such as interactive stories, immersive simulations, or personalized experiences. For example, a large language model for protein engineering was trained on 280 million protein sequences from more than 19,000 proteins families and used to identify protein sequences with a predictable function across large protein families (Madani et al. 2023), which researchers have likened to a language model that generates grammatical and semantically correct natural language sentences on diverse topics.

Given the increasingly general purpose nature of these models, adapting models such as ChatGPT or DALL-E to new tasks can be done with little to no coding and can be as easy as describing a task in simple language. Using a chatbot such as ChatGPT or an image generation tool such as DALL-E is as simple as entering requirements into a browser interface. Furthermore, humans can extend the functionality of these models by fine-tuning prompts that are phrased as instructions. Cognitive neuro-linguists have questioned whether the capabilities of generative AI truly demonstrate a human-like understanding of language (Mitchell and Krakauer 2023), in that they lack concepts such as connectionism (Pinker and Prince 1988) and compositionality (Marcus 1998). Experts on AI fairness have also warned that general purpose AI models trained on vast repositories of online data are little more than “stochastic parrots” (Bender et al. 2021), with large volumes of training data gleaned

from the Internet potentially encoding hegemonic and ethnocentric worldviews, and that people using these models may be misled by the efficiencies of these models to overlook the societal biases and other problematic issues arising from the training data. In this regard, it is important to heed the warnings of AI pioneers such as Norbert Wiener, an early pioneer of AI, who cautioned about the lack of AI alignment (Wiener 1960) that arises when the purpose of AI differs from the uses it is deployed for.

### 3. Generative AI as a Tool for Enabling High-Quality Scholarly Work

In the near term, generative AI does seem to offer opportunities to enhance specific areas of research, namely (i) problem formulation and research design, (ii) data collection and analysis, (iii) interpretation and theorization, and (iv) composition and writing. We will discuss the opportunities and challenges associated with generative AI in the next two sections. We briefly review each of these possibilities in this section and the next and illustrate them with examples.

#### 3.1. Generative AI Use in Problem Formulation and Research Design

Generative AI tools can be used as a sounding board for emerging ideas. When ideating on research questions or hypotheses, one can enter key questions, concepts, and arguments as prompts, which can then contribute to an iterative process of idea formulation; assessment of the quality, novelty, or state of related research; and idea refinement. For example, if one were to extend the introspective study of Tarafdar et al. (2022) of *ISR* to examine the entire discipline, one could ask ChatGPT to help by listing the best journals that represent the discipline. We asked ChatGPT: “What is the best MIS Journal?” It returned a list of 10 journals, missing two of the Association for Information Systems Senior Scholars’ elite journals and adding an additional, lesser-known journal. We tuned our query to ask: “What is the best MIS journal to start a literature review?” It returned a subset of five journals, with some explanation, opining that a journal other than *ISR* was a better starting point. When asked what each journal was good at, ChatGPT opined that

*ISR* was best for technology adoption and implementation research. Although the answers were not necessarily correct or complete, *ChatGPT can be used to support finding a starting point for research.*

ChatGPT was also able to offer suggestions for future research related to a recently published paper in *ISR* (Fugener et al. 2021), incorporating multiple perspectives that could be accommodated within the study’s setting and analysis. When prompted to redesign the study of Fugener et al. (2022) to consider algorithm aversion, a theoretical perspective that was not considered in the original paper, ChatGPT responded by proposing an experimental manipulation detailed in Table 1. The experimental manipulation suggests a set of nudges, such as (i) giving participants the option to decline assistance from the AI versus not, (ii) manipulating the degree of delegation to the AI agent, and (iii) measuring participants’ attitudes toward AI. Similar patterns were observed with different prompts asking ChatGPT to redesign the study based on alternative theoretical traditions. The responses in Table 1 illustrate some of the ways in which these tools can augment the process of literature review and research design.

Once a manuscript has been written, authors can use ChatGPT to evaluate key elements of the paper’s quality and completeness and to identify additional studies that may be relevant. To illustrate this point, we describe our experience when we entered the text of Susarla et al. (2020) and asked ChatGPT to summarize the key points. We reviewed the summary provided by ChatGPT to see if the key points extracted by ChatGPT were consistent with our understanding of the paper’s thrust. Had the paper still been in development, we could have used the summary to review why key points were missed and others (over)emphasized.

Furthermore, supervised prompt fine-tuning could be used to elicit insights from ChatGPT on how IS researchers with different theoretical and paradigmatic perspectives might respond, thereby providing opportunities to address possible misunderstandings or discrepancies of the work, as perceived from different perspectives, prior to entering the peer review process. For completeness, one could use prompt fine-tuning to seek possible explanations or missing papers relevant to the manuscript being prepared. To identify possible missing explanations, we

**Table 1.** Example Prompt and Response for a Published *ISR* Paper (Fugener et al. 2021)

Prompt: “how would you redesign the study to consider algorithm aversion”

Edited and condensed response.

To redesign the study to consider algorithm aversion, the researchers could introduce a manipulation where participants are informed about the potential benefits of using the AI algorithm in the collaborative task, but they are also given the option to decline the AI’s assistance and complete the task on their own. This would allow the researchers to measure the extent to which participants are reluctant to use the AI algorithm and whether this reluctance affects the collaborative performance between humans and AI. ...

To test whether algorithm aversion affects delegation decisions and performance, the researchers could manipulate the delegation role of the AI and participants in the task. For example, in one condition, the AI could delegate tasks to the participants, and in another condition, the participants could delegate tasks to the AI. The researchers could then measure the delegation performance and efficiency for each condition and compare them to the control condition where there is no delegation ...



entered the abstract of Wong et al. (2021) and asked ChatGPT for perspectives that might help explain the findings. The tool responded with suggestions that were expected, such as social learning theory and the technology acceptance model, but also added social identity theory and diffusion of responsibility. When we compared this to the research model, we found elements of each theory in the whistleblowing framework that appeared in the paper. This is quite interesting, given that the paper of Wong et al. was not part of the ChatGPT training data.

3.2. Generative AI Use in Data Collection and Analysis

Generative AI can provide valuable support in assembling data sets or in identifying deep patterns in the data. Generative AI is well suited to help researchers match data across archival sources. Manual, time-consuming tasks such as matching names, companies, or other variables can be somewhat automated—somewhat, in the sense that once the records are matched, the researcher should verify the quality of the results. For example, when compiling a unique data set for their work on corporate social performance and data breaches, D’Arcy et al. (2020) drew information from several sources to match firm data breaches, corporate social performance ratings, and related factors. In doing so, they had to match records that used sometimes inconsistent firm names, which is a detailed and time-consuming process. Generative AI tools such as ChatGPT are well suited for such pattern-matching exercises. They can be used to identify possible names to search across heterogeneous archival sources, assess the likelihood of a match between records, or assemble a data set that takes each permutation of each company’s name into account. Again, human supervision is required to ensure the search and matching exercise is properly executed because the generative AI may lack the context or information in the training data set to correctly match all the records. Researchers (Mollick 2023) have reported examples where they input data and a data dictionary into ChatGPT, along with a code interpreter, and asked ChatGPT to analyze the results and provide some preliminary conclusions. With a code interpreter plugin, ChatGPT can offer extensive visualization functionalities that may prove useful to researchers investigating a variety of topics.

Beyond simple pattern matching, generative AI can be applied to explore text data. Researchers can explore

how different assumptions and definitions affect the type, content, and quality of insights derived from patterns manifested in the corpus text. In doing so, they can identify ways to complement analysis. For example, Zheng et al. (2022) applied generative AI to surface patterns and critically evaluated those patterns, creating opportunities to confirm or disconfirm the findings suggested by the data. When queried about a potential study using word embedding-based methods to study an IS-specific phenomenon (Table 2), ChatGPT suggested comparing across alternative methods such as latent semantic analysis (LSA) or, better still, a hybrid approach that combines different techniques like LSA, topic modeling, sentiment analysis, and network analysis to get a more comprehensive understanding of the nuances of the problem domain. Although this is no substitute for human reasoning and cognition, generative AI can clearly serve as an aid for researchers seeking to identify and evaluate alternative approaches to solving problems.

With primary data collected directly through methods such as experiments or observational data curated through scraping websites or calls to an Application Programming Interface, generative AI can be used to suggest approaches to analyzing and interpreting the data. Although generative AI is not a substitute for having a deep understanding of statistical or visualization tools, it can speed up the process of creating the syntax, estimating the results, and evaluating the quality of the results. Because one has far more control over the data manipulation and analysis with text-based commands than with a point-and-click interface, augmenting human ability to craft and evaluate such code for statistical analysis can provide researchers with opportunities to conduct more granular analysis. For example, a recently published paper in *ISR* (Oh et al. 2022) uses clickstream and social media data to examine the complementary relationship between social media and news sites. Although the analysis of Oh et al. (2022) appropriately focused on text and numbers relevant to their research question, using generative AI could have afforded opportunities to integrate additional data found in social media posts such as image content, geospatial data, and more.

3.3. Generative AI Use as an Aid to Composition and Writing

Generative AI can also be used to support the writing process. Researchers can enter a draft text and ask it to

Table 2. Example Prompt and Response for a Recently Accepted Paper in *ISR* (Zheng et al. 2022)

Edited and condensed responses for a prompt on how to use machine learning to study the same phenomenon	
1. Classification:	A supervised learning algorithm like logistic regression or decision trees could be used to classify sellers into high-quality and low-quality categories based on their refund policies and platform insurance. This classification model could then be used to predict the quality of new sellers based on their policies.
2. Clustering:	An unsupervised learning algorithm like k-means clustering or hierarchical clustering could be used to group sellers based on their refund policies, reputation, popularity, or other characteristics. This could help identify patterns in seller behavior and reveal potential correlations between refund policies and seller quality ....

**Table 3.** Example of Using ChatGPT in Developing a Manuscript

Writing tasks	ChatGPT outputs
Overall writing style	Reviewing for inconsistent word usage
Proofreading for consistency and clarity	Reviewing for clarity and organization
References	Formatting in accordance with approved styles

assess the structure of arguments or the quality of writing. We asked ChatGPT to assess the quality of our first paragraph in this section. It returned four substantive suggestions for improving the clarity, organization, evidence, and acknowledgment of limitations. Generative AI tools can also provide holistic reviews of a paper’s grammar, punctuation, complexity, and voice, and can even assist with reference formatting. Unlike widely used line-by-line tools such as Grammarly, generative AI can provide global support for the technical aspects of papers, making it more accessible to a broader audience. For example, as we were polishing this paper, we used ChatGPT to suggest consistent word choices among the four authors, ensuring a consistent voice in describing the opportunities and challenges of ChatGPT for IS research. Improving these fundamental elements of a paper reduces the demands placed on peer reviewers and can facilitate teamwork by allowing authors to improve the quality of the text before it is shared with collaborators. The examples listed in Table 3 below, though not exhaustive, provide pointers on using ChatGPT to improve manuscript development.

We would like to emphasize that our support for using generative AI to assist with writing should not be interpreted as support for having generative AI conduct the research or write the paper in its entirety. Some suggest using tools like ChatGPT to generate draft sections, especially for more structured sections like methods and results. We believe that ceding control of the writing and knowledge-creation process to ChatGPT is a mistake, at least at this point in a sociotechnical discipline such as IS; we elaborate more on this in the next section.

#### 4. Concerns About Generative AI Tools Detracting from High-Quality Scholarly Work

In this section, we discuss challenges posed by generative AI in supporting scholarly activities, including but not limited to (i) problem formulation, research design, and methodological critique, and (ii) manuscript summaries and literature reviews. We critically examine undesirable outcomes that may result from applying generative AI tools. Concerns that emerge include *limitations and biases in the training data*, *hallucinations that lead to veracity and*

*reliability concerns*, *violation of intellectual property rights*, and *the lack of appropriate depth and coherence in the outputs*.

##### 4.1. Generative AI Use in Problem Formulation, Research Design, and Methodological Critique

Birhane et al. (2022) reviewed a large corpus of published work in machine learning research and found that values such as performance, generalization, quantitative evidence, and efficiency dominate as justifications for research. However, these publications rarely mention the negative potential of these technologies or question whether there is a societal need for them. Indeed, there is growing concern that the large-scale models underlying generative AI may prioritize predictive accuracy and performance over transparency, equity, and open access. Furthermore, training on digital traces and Internet sources may replicate existing societal biases without adequate safeguards or protections. Word embedding methods, for example, tend to encode significant gender biases (Yang and Feng 2020). These biases can propagate to downstream tasks that rely on biased word vectors, such as analyzing user-generated content. More broadly, biases in machine learning models can arise from biases in training data sets and users’ perceptions of model outputs.

Beyond these well-known concerns about machine learning, generative AI tools may have additional undesirable implications for scholars involved in developing, reviewing, and summarizing research papers. Because generative AI is trained using published papers, when asked to suggest research questions, these tools may suggest known questions found in the established corpus of a research community’s work rather than forward-looking state-of-the-art questions at the leading edges of a topic. Similarly, when asked to design a research study, ChatGPT may overlook relevant, emerging approaches. Reviews conducted with these tools may selectively prioritize variables, methods, and findings from dominant subdisciplines and traditions in the training data. For example, reviews provided by ChatGPT may apply standards associated with a dominant method, such as econometrics, to a paper that comes from a different research tradition, such as interpretivism. There may also be significant factual errors in the output of these tools, which we address in our discussion on “hallucinations.”

Generative AI models also present interpretability concerns due to the type of architecture and the amount of data used to train them. Unlike predictive models commonly associated with AI, generative models learn a probability distribution over an extremely high-dimensional space, making it difficult to interpret their outputs. This lack of interpretability can make it difficult to understand research design suggestions and critiques, as there is virtually no explanation provided for why a particular option is considered suitable, whereas others are not.

As another example, ChatGPT, in response to the prompt, “How do we use mixed methods research design to study value destruction in information technology ecosystems?,” proposed the sequential use of quantitative and qualitative data. Although the proposed approach is valid, it is generic and lacks innovation. Mixed methods are often portrayed as using qualitative and quantitative data (Venkatesh et al. 2013), but they can involve multiple *approaches* using similar or different types of data. Indeed, a recent study by Sahaym et al. (2023) used the hermeneutic circle to integrate analytical modeling and interpretive case study (involving qualitative data) to shed light on the issue of value destruction—there was no quantitative data involved.

With respect to refereeing conducted with such tools, such reviews are likely to selectively prioritize variables, methods, and findings from certain subdisciplines and traditions that are dominant in the training data. These reviews are thus likely to “parrot” criticisms that are no longer considered valid or legitimate. For example, we asked ChatGPT: “Can you provide a critique of the methods used in the study ‘Toward a Theory of Information Systems Security Behaviors of Organizational Employees: A Dialectical Process Perspective’ published in *ISR* by Karjalainen et al. (2019)?” Excerpts from the response received are presented in Table 4 (problematic issues are in bold; hallucinations, which we discuss later, are underlined).

The critique offered by ChatGPT of this qualitative, interpretive paper has characteristics similar to those expected of a novice, unqualified reviewer who may not grasp the nuances of the literature in the discipline. The critique misses the idea of “theoretical sampling,” frames the study as one involving “measurement” and “multidimensional scale,” and repeats concerns of “bias” and lack of “generalizability” that are not directly applicable to interpretive work. A reviewer untrained in the interpretive tradition and instead relying on ChatGPT to assist in the review is likely to submit a report that is inconsistent with the standards associated with the research tradition to which the paper belongs. In addition, there are factual errors (underlined text), which we delve into next.

4.2. Generative AI Use in Summarizing Manuscripts or Conducting Literature Reviews

The most pressing challenges arise from hallucinations. In *natural language processing* (NLP), a hallucination is often defined as “generated content that is nonsensical or unfaithful to the provided source content”; it is a confident response by an artificial intelligence system that does not seem to be justified by its training data (Ziwei et al. 2022, p. 3). Hallucinations can occur because the AI has been trained on vast amounts of data containing a wide range of information, including both factual and fictional content. However, the phenomenon by which generative AI produces hallucinations is still not well understood. Since these models have been developed by private organizations using proprietary datasets, we do not know the provenance of the data used to train these models, and cannot adequately generalize answers to questions for which data is not available. Worryingly, generative AI models, even those that are designed to be less prone to hallucinations, have been shown to amplify hallucinations (Dziri et al. 2022).

The problem of hallucination was evident in responses to many of our prompts. For instance, in the critique of “Toward a Theory of Information Systems Security Behaviors of Organizational Employees...” (Table 4), the underlined text represents hallucinations. Only one (not two) organization was studied, this was not a mixed methods study involving a survey, and SEM was not applied to survey data! As another example, consistent with the findings of past studies (Eysenbach 2023), ChatGPT provided several fabricated references along with legitimate references in response to the prompt: “Can you list the most important papers looking at the dark side of information technology use on employee well-being and productivity?” We were more troubled by the fact that when asked to summarize a recently published paper in *ISR* (Wang et al. 2021), ChatGPT not only made up an entirely fictitious paper title and set of authors (Silver et al. 2020) but also an entire set of fabricated claims about how the paper provided an overview of reinforcement learning (RL) as applied to robotics and discussed the challenges of applying RL to robotics, such as the need for high-dimensional state and action spaces,

Table 4. Critique by ChatGPT of a Qualitative Study

1. ....
2. Sampling: The study uses a <b>convenience sample</b> of employees from <u>two organizations</u> , which may <b>limit the generalizability</b> of the findings to other organizations and contexts.
3. Data Collection: The study uses a <u>mixed-methods approach</u> with a <b>survey and interviews</b> to collect data. While this approach provides rich data and allows for triangulation of findings, the data may be subject to <b>social desirability bias</b> and the interviews may not fully capture the nuances of security behaviors.
4. Operationalization of Variables: The study operationalizes security behaviors <b>using a multidimensional scale that includes cognitive, affective, and behavioral aspects of security</b> . While this scale provides a comprehensive view of security behaviors, <b>the measurement of security behaviors may still be subject to biases</b> and limitations.
5. Data Analysis: <u>The study uses structural equation modeling (SEM) to analyze the survey data</u> and thematic analysis to analyze the interview data. While these are appropriate methods for the research questions, the findings may be <u>subject to the assumptions of SEM</u> and the limitations of thematic analysis.



the long-horizon nature of many robotic tasks, and the need for robust and reliable algorithms.

Apart from the effects of hallucinations, literature reviews conducted using generative AI can be problematic in that the output may lack scholarly depth consistent with the expectations of a given research community. For example, when prompted to provide a literature review on the information systems security behaviors of organizational employees, the resulting output from ChatGPT provided brief summaries of key themes and findings but lacked the theoretical framing and critical analysis expected in an academic literature review. We were curious to know how two leading IS security researchers (both serve on the board of *ISR*) would react to the review. One of them assessed the review as follows:

*... It is throwing out concepts or factors studied, without telling [us] what they mean, or what is the state of the art. The taxonomy or categories are overlapping, so the categorization does not make sense. Moreover, [the] IS audience often discusses things based on theories. This is missing.... Just listing concepts randomly without their theory, does not tell what these concepts or factors really mean. Citations are missing.*

The other colleague was even more severe:

*I don't see this as a review at all... I see no depth for review purposes. In fact, in looking at the key constructs in this area, it only covers a few of them and glosses over the more interesting ones. No terms are defined, no citations are provided, no meaningful statistics are given, and I have no sense of who is writing this. So I would personally find this useless from a review perspective.... A key problem here is that when we review the literature, we're not just trying to get the high-level obvious stuff, we're trying to find the interesting tensions, metaphors, paradoxes, controversies, unresolved problems, and so forth.*

Although the literature review by ChatGPT did provide a high-level summary of the topic, the summary lacked depth, citations, and theoretical grounding. Citations are crucial for readers and authors to judge the credibility and trustworthiness of the ideas presented. The lack of references and explanation of why certain themes were included or excluded made the review less credible and potentially violated intellectual property rights.

Until we develop a reliable base of knowledge on a topic that is fine-tuned to a given specific research community, it is crucial for scholars to prioritize their understanding of the relevant body of work over the suggestions provided by generative AI tools. We think this is an important point to keep in mind because the inherent biases of AI models such as ChatGPT often reflect the corpus of documents on which they are trained. These biases can manifest as misclassification of ideas or hallucinations when processing technical or literature-specific language. For example, in the context of the literature on fake news,

specific terminology is used to describe false information shared on the Internet (Moravec et al. 2022). If the generative AI algorithm omits or replaces this language, it could lead a reader to misunderstand the paper's connection to broader research on fake news and online misinformation. Furthermore, in cases where the generative AI algorithm fails to find relevant content, it tends to generate incorrect citations or link the author to inappropriate sources. In both situations, the author must carefully review the results of all GPT instantiations to identify and correct any problematic results.

## 5. A Path Forward

The rise of generative AI, including LLMs, has significantly lowered the cost of knowledge acquisition and, as many believe, democratized knowledge. With generative AI, it is now possible to discover new insights and learn about complex topics with ease. It is also possible to develop texts that span genres from memos to poetry and to disseminate knowledge on a wide range of topics with ease. The optimistic vision holds that the use of generative AI tools can lead to more efficient knowledge creation, assessment, and consumption and eventually to an informed and empowered society. Yet as we have seen in our earlier section, the role of generative AI in the creation, assessment, and summarizing of knowledge is not unproblematic—it requires a baseline of knowledge to appropriately assess and fact-check outputs.

The *raison d'être* of top journals such as *ISR* is to publish original and credible research that contributes to the advancement of knowledge. Such knowledge creation involves a rigorous process that may include theorizing, hypothesis testing, data collection, analysis, interpretation, and thoughtful evaluation, which cannot yet be replicated by generative AI alone. As a result, the human endeavor associated with papers published by journals such as *ISR* remains critical. Although generative AI tools do offer noteworthy benefits with respect to supporting scholarly activities, such as those outlined in the third section of this editorial, they cannot at this time replace the role of top journals and human engagement in the process of knowledge creation, assessment, and dissemination, which are deeply embedded in the social context and thus involve many tacit rules and conventions. We also sense the tendency to offload human cognition and intelligence to generative AI which can have potentially dysfunctional consequences that are, at this time, largely unknown. In fact, we believe that normalizing the tools such as ChatGPT by thinking of them as being similar to calculators may be misplaced, given the issues related to bias, hallucinations/fabrications, lack of interpretability and consistency, lack of appropriate depth expected in a community, and potential violations of intellectual property, which we discuss in the fourth section of this editorial. These are not issues posed by the



calculator, which is a nonstochastic technology used to support or automate certain straightforward computing tasks. Our view is that the technology and the social process of scholarly work will need to evolve significantly before using generative AI becomes as routine and unproblematic as using a calculator.

We also contend, at this time, that human supervision and oversight may be necessary for almost all stages of research because generative AI lacks the intuition, perspective, and rich understanding of the sociotechnical context related to scholarly work in our discipline. Furthermore, given that generative AI models are rapidly changing, we believe that scholars who use such tools to support their research must, at a minimum, declare the tool(s) used, the version, and how it was used to support their work. This is required by ISR currently. For certain research activities, it would be advisable to provide more detail than a simple acknowledgment of the tool used, such as choices of the different prompts made to elicit suggestions.

In addition, any use of these models should be done with a demonstrated awareness of possible biases, with debiasing methods carefully adopted and explained. There are additional issues as illustrated in our section on concerns regarding generative AI tools. The use of these tools could discourage deeper scholarly engagement because of the lack of theoretical depth in the outputs and encourage scholars to seek breadth rather than depth of understanding. Finally, in light of the problems of bias, fabrication, and poor contextualization that can affect

scholarly activities and consequently the quality of papers published in journals, we echo recent comments (Van Dis et al. 2023) urging authors to take responsibility for fact-checking their text, results, data, code, and references and expressing the need for discipline-specific norms and investment in truly open models. We provide a summary of some of the takeaways in the form of challenges, suggestions, and principles (Table 5).

Although we are hopeful that this editorial will encourage the adaptation of our research practices in the era of generative AI and lead to a refinement of norms, there are clearly many fundamental questions that our community needs to think about, including the following. *What will scholarship look like in the age of generative AI? Will and can it remain the same? What tasks might be automated, augmented, or untouched by generative AI?* Table 6, adapted from Nature (VanDis et al. 2023), provides a useful set of questions that might be abstracted to the future of work in general to generate valuable research studies.

Looking to the future use of generative AI models in journals in IS, we note that, although these models are pretrained on a diverse range of Internet text sources, they offer limited knowledge specific to *ISR* and other journals in the IS discipline. Although some papers in the field may have been included in the training data through open access repositories, preprint servers, or personal websites, the extent to which generative AI models incorporate research articles, particularly those not publicly available and easily accessible, remains

**Table 5.** Some Guidance for the Use of Generative AI Tools (for Now)

<i>Challenges Posed by Generative AI for Scholarly Work</i>	
1. <b>Institutionalizing bias:</b>	Generative AI relies on training data and encoded rules that reflect human biases, resulting in further reification of social problems in broader society or flawed conclusions in the literature.
2. <b>Hallucinations:</b>	Generative AI fabricates false or misleading responses to prompts from scholars.
3. <b>Interpretability:</b>	Generative AI offers suggestions with scant explanation for why an output was created, creating issues related to credible inference and interpretation.
4. <b>Inappropriate Training Data:</b>	Generative AI offers inappropriate guidance because it has not been trained on data information relevant to a research topic or method.
5. <b>Misapplication:</b>	Generative AI is used to complete peer reviews, lead literature searches, or write key elements of papers without appropriate supervision.
6. <b>Unreasonable expectations:</b>	Scholars may expect access to generative AI to result in unreasonably sophisticated understanding of literature or expert applications of state-of-the-art research methods
<i>Suggestions for Mitigating Problems Posed by Generative AI</i>	
1. <b>Independent Knowledge:</b>	When studying a topic or using a method, a scholar must develop a baseline of knowledge and experience necessary to prompt as well as assess the outputs of generative AI.
2. <b>Critical Thinking:</b>	When reviewing the output of generative AI, scholars should consider the following: (a) the face validity: are the outputs consistent with the scholar's understanding of the literature? (b) falsifiability, is it possible—through logic or evidence—to critically examine and subject the outputs provided by generative AI to invalidation? This helps ensure that humans guide the application of AI.
3. <b>Awareness of Implications:</b>	As scholars consider the output of generative AI, they need to actively question the biases and their implications for people and organizations. Scholars should demonstrate in their work an awareness of multiple perspectives and interpretations so that readers understand the potential biases from these tools have been taken into consideration.
4. <b>Understanding Provenance:</b>	When using generative AI, scholars must understand the data that was used to train the tool.
5. <b>Ethical Conduct:</b>	Scholars should use generative AI as per current professional norms. Writing and reviewing papers are critical activities that the researchers need to be undertake cautiously and responsibly when using generative AI.
<i>Guiding Principles for Scholars Seeking to Apply Generative AI in Scholarly Work</i>	
1. <b>Human Primacy:</b>	Scholars retain decision rights over key elements of research.
2. <b>Responsible Reporting:</b>	Scholars faithfully report (1) which generative AI was used, (2) what data were used to train the AI (if known), and (3) how the tool was used to support scholarly inquiry.

**Table 6.** Questions for Further Exploration (Partially Adapted from Van Dis et al. 2023)

<ul style="list-style-type: none"><li>• What are the different ways in which generative AI tools can augment or assist human scholarly activities?</li><li>• What characteristics of tasks in the research process lend themselves to effective automation and augmentation with generative AI?</li><li>• How should we formulate and adopt policies to ensure research integrity?</li><li>• How can we build discipline-specific and independent open-source LLMs and other models to ensure that diverse research paradigms are well represented?</li><li>• What quality standards should be expected of AI-augmented research and reviewing (for example, transparency, accuracy, debiasing, and source crediting), and which stakeholders are responsible for enforcing the standards? What technological innovations are needed for monitoring and enforcing standards?</li><li>• How can researchers ensure that AI-augmented tools promote equity in research, and avoid risks of widening inequities in authoring, reviewing, and editing?</li><li>• What legal implications does the practice of AI-assisted research have regarding intellectual property?</li></ul>
--

uncertain. Consequently, their ability to provide accurate and up-to-date information (and guidance) on specific topics in information systems and support research activities is significantly restricted.

To effectively use generative AI, the IS community may need to develop discipline-specific applications of such tools. ArXiv, which is an open-access repository for academic papers in various disciplines such as physics, mathematics, economics, and computer science, has recently developed a bespoke language model, termed ArxivGPT, which “summarizes arXiv papers and provides key insights” (<https://twitter.com/ArXivGPT>). *ISR* has a unique opportunity to leverage its wealth of resources to create a valuable knowledge base. A phased approach could possibly begin with collecting abstracts and fine-tuning, offering broad-stroke summaries and articulations of contributions. A more ambitious endeavor would involve building a knowledge base derived from published papers and drafts with the reviews in the manuscripts’ journey in the review process. We believe this logic could be extended to capture the corpus of IS papers found in IS-focused journals and conferences to offer a broader view of the IS discipline. Although challenging, this effort could provide immense support for early-career researchers, authors, reviewers, and editors and lead to a far more efficient manuscript development and editorial process.

In closing, the potential for generative AI in revolutionizing IS research lies in our ability to harness its strengths, address its limitations, and forge a path forward that illuminates the profound impact of our work. The future beckons and it is our responsibility to seize it.

**Acknowledgments**

The authors acknowledge the constructive input and assistance of Monica Birth, Sutirtha Chatterjee, Dirk Hovorka, Brent Kitchens, Kai Larsen, Xiao Liu, Paul Lowry, Abhay Mishra, and Mikko Siponen.

**Endnotes**

<sup>1</sup> Although this piece constitutes an editorial, it diverges from the customary structure, format, and content typically found in *ISR*. The staggering pace of recent AI advancements has already begun to instigate seismic shifts in research conduct and communication. It is evident that all journals, including *ISR*, will experience the reverberations of this paradigm shift. As a community, it is incumbent upon us to unite

and formulate guidance for the future, ensuring that *ISR* remains the preeminent platform for publishing and promoting innovative, transformative research within information systems. The impetus for composing this editorial in its current format and structure is the pressing need to furnish IS scholars, particularly junior researchers, with direction regarding the appropriate employment of generative AI in their *ISR*-targeted research. In doing so, we endeavor to outline potential trajectories for the journal to continue fulfilling its core mission. We encourage readers to perceive this piece not as the definitive word on the subject, or as advocacy for the use of generative AI in research, but as the genesis of conversations that will shape our collective future.

<sup>2</sup> We do not sharply distinguish between LLMs and generative AI models because both offer the promise of broad societal deployment. For instance, it is not inconceivable that researchers might use tools such as DALL-E or Midjourney to study image analytics or use video generation tools that can augment existing data sets and might even serve as an alternative to current models of synthetic control. It is also possible to use generative AI tools to assist in coding tasks or in data collection. The same set of principles that we articulate henceforth in the context of LLMs will apply to such modes of generative AI as well.

**References**

Bender EM, Gebru T, McMillan-Major A, Mitchell M (2021) On the dangers of stochastic parrots. *Proc. ACM Conf. on Fairness, Accountability, and Transparency* (ACM, New York), 610–623.

Birhane A, Kalluri P, Card D, Agnew W, Dotan R, Bao M (2022) The values encoded in machine learning research. 2022 *ACM Conf. Fairness, Accountability, Transparency (FAccT '22)* (ACM, New York). <https://doi.org/10.1145/3531146.3533083>.

Chomsky N, Roberts I, Watumull J (2023) Noam Chomsky: The false promise of ChatGPT. *New York Times*, <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>.

D’Arcy J, Adjerid I, Angst CM, Glavas A (2020) Too good to be true: Firm social performance and the risk of data breach. *Inform. Systems Res.* 31(4):1200–1223.

Dwivedi YK, Kshetri N, Hughes L, Louise Slade E, Jeyaraj A, Kumar Kar A, Baabdullah AM (2023) “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Internat. J. Inform. Management* 71:1–63.

Dziri N, Milton S, Yu M, Osmar Z, Reddy S (2022) On the origin of hallucinations in conversational models: Is it the datasets or the models? Carpuat M, de Marneffe M-C, Vladimir Meza Ruiz I, eds. *Proc. Conf. of the North Amer. Chapter of the Assoc. for Comput. Linguistics: Human Language Technologies* (Association for Computational Linguistics, Seattle), 5271–5285.

Else H (2023) Abstracts written by ChatGPT fool scientists. *Nature* 613:423.

Eysenbach G (2023) The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Medical Ed.* 9:e46885.

- Freidman T (2023) Our new Promethean moment. *New York Times*, <https://www.nytimes.com/2023/03/21/opinion/artificial-intelligence-chatgpt.html>.
- Fugener A, Grahl J, Gupta A, Ketter W (2021) Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Inform. Systems Res.* 33(2): 678–696.
- Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, Pearson AT (2022) Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. bioRxiv Preprint, submitted December 27, <https://www.biorxiv.org/content/10.1101/2022.12.23.521610v1>.
- Karjalainen M, Sarker S, Siponen M (2019) Toward a theory of information systems security behaviors of organizational employees: A dialectical process perspective. *Inform. Systems Res.* 30(2):687–704.
- Larsen B, Narayan J (2023) Generative AI: A game-changer that society and industry need to be ready for. *World Economic Forum* (January 9), <https://www.weforum.org/agenda/2023/01/davos23-generative-ai-a-game-changer-industries-and-society-code-developers/>.
- Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, Olmos JL Jr (2023) Large language models generate functional protein sequences across diverse families. *Nature Biotech.* <https://doi.org/10.1038/s41587-022-01618-2>.
- Marcus GF (1998) Rethinking eliminative connectionism. *Cognitive Psych.* 37(3):243–282.
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. Preprint, submitted September 7, <https://arxiv.org/abs/1301.3781>.
- Mitchell M, Krakauer DC (2023) The debate over understanding in AI's large language models. *Proc. National Acad. Sci. USA* 120(13): e2215907120.
- Mollick E (2023) @emollick. Retrieved from Twitter, <https://twitter.com/emollick/status/1653945049275670528>.
- Moravec PL, Kim A, Dennis AR, Minas RK (2022) Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media. *Inform. Systems Res.* 33(3):887–907.
- Oh H, Goh K-Y, Phan TQ (2022) Are you what you tweet? The impact of sentiment on digital news consumption and social media sharing. *Inform. Systems Res.* 34(1):111–136.
- Pinker S, Prince A (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28(1–2):73–193.
- Russell SJ, Norvig P (2003) *Artificial Intelligence: A Modern Approach*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ).
- Sahaym A, Vithayathil J, Sarker S, Sarker S, Bjørn-Andersen N (2023) Value destruction in information technology ecosystems: A mixed-method investigation with interpretive case study and analytical modeling. *Inform. Systems Res.* 34(2):508–531.
- Susarla A, Holzhaecker M, Krishnan R (2020) Calculative Trust and Interfirm Contracts. *Manage. Sci.* 66(11):5465–5484.
- Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. *Adv. Neural Inform. Processing Systems* 27:3104–3112.
- Tarafdar M, Shan G, Thatcher JB, Gupta A (2022) Intellectual diversity in IS research: Discipline-based conceptualization and an illustration from *Information Systems Research*. *Inform. Systems Res.* 33(4):1490–1510.
- Van Dis E, Bolen J, van Rooij R, Duidema W, Bockting CL (2023) ChatGPT: Five priorities for research. *Nature* 619:224–226.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, et al. (2017) Attention is all you need. Preprint, submitted December 6, <https://arxiv.org/abs/1706.03762>.
- Venkatesh V, Brown S, Bala H (2013) Bridging the qualitative–quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quart.* 37(1):21–54.
- Wang Y-Y, Guo C, Susarla A, Sambamurthy V (2021) Online to offline: The impact of social media on offline sales in the automobile industry. *Inform. Systems Res.* 32(2):582–604.
- Wei J, Yi T, Bommasani R, Raffel R, Zoph B, Borgeaud S, Yogatama D, et al. (2022) Emergent abilities of large language models. *Trans. Machine Learn. Res.* <https://openreview.net/forum?id=yzkSU5zdWd>.
- Wiener N (1960) Some moral and technical consequences of automation: As machines learn they may develop unforeseen strategies at rates that baffle their programmers. *Science* 131(3410): 1355–1358.
- Wong RYM, Cheung CM, Xiao B, Thatcher JB (2021) Standing up or standing by: Understanding bystanders' proactive reporting responses to social media harassment. *Inform. Systems Res.* 32(2): 561–581.
- Yang Z, Feng J (2020) A causal inference method for reducing gender bias in word embedding relations. *Proc. Conf. AAAI Artificial Intelligence* 34(05):9434–9441.
- Zheng J, Wang Y, Tang Y (2022) Platform refund insurance or being cast out: Quantifying the signaling effect of refund options in the online service marketplace. *Inform. Systems Res.*, ePub ahead of print August 31, <https://doi.org/10.1287/isre.2022.1162>.
- Ziwei J, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, et al (2022) Survey of hallucination in natural language generation. *ACM Comput. Surveys* 55(12):1–38.