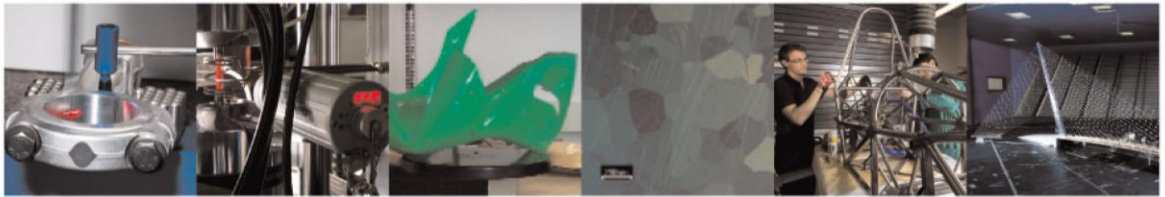




POLITECNICO
MILANO 1863

DIPARTIMENTO DI MECCANICA



How Can IJDS Authors, Reviewers, and Editors Use (and Misuse) Generative AI?

Shmueli, Galit; Maria Colosimo, Bianca; Martens, David; Padman, Rema; Saar-Tsechansky, Maytal; R. Liu Sheng, Olivia; Street, W. Nick; Tsui, Kwok-Leung

This is a post-peer-review, pre-copyedit version of an article published in INFORMS JOURNAL ON DATA SCIENCE. The final authenticated version is available online at:

<http://dx.doi.org/10.1287/ijds.2023.0007>

This content is provided under [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) license




How Can IJDS Authors, Reviewers, and Editors Use (and Misuse) Generative AI?

Galit Shmueli,^{a,*} Bianca Colosimo,^b David Martens,^c Rema Padman,^d Maytal Saar-Tsechansky,^e Olivia Sheng,^f W. Nick Street,^g Kwok Tsui^h

^aInstitute of Service Science, National Tsing-Hua University College of Technology Management, Hsinchu 30013, Taiwan; ^bMechanical Engineering, Politecnico di Milano, Milan, Italy; ^cDepartment of Engineering Management, University of Antwerp, Antwerp, Belgium; ^dInformation Systems and Public Policy, Carnegie Mellon University Heinz College, Pittsburgh, Pennsylvania; ^eIROM, University of Texas System, Austin, Texas; ^fOperations and Information Systems, University of Utah, Salt Lake City, Utah; ^gBusiness Analytics, The University of Iowa Henry B Tippie College of Business, Iowa City, Iowa; ^hIndustrial and Systems Engineering, Virginia Tech, Blacksburg, Virginia

*Corresponding author

Contact: galit.shmueli@iss.nthu.edu.tw,  <https://orcid.org/0000-0002-0820-0301> (GS); BiancaMaria.Colosimo@polimi.it (BC); david.martens@uantwerpen.be (DM); rpadman@andrew.cmu.edu (RP); Maytal.Saar-Tsechansky@mcombs.utexas.edu (MS-T); gkmcolumbia@gmail.com (OS); nick-street@uiowa.edu (WNS); klttsui@vt.edu (KT)

Received: April 21, 2023

Accepted: April 21, 2023

Published Online in *Articles in Advance*:

■■■■■ ■■■, 2023

<https://doi.org/10.1287/ijds.2023.0007>

Copyright: © 2023 INFORMS

Abstract. We have all been subjected to the latest pattern: A conversation on any topic, from data science to travel or food, ends up (with high probability) mentioning ChatGPT. OpenAI's latest free "toy" has captured the attention and imagination of so many people within and far beyond the data science community. Communities and individuals are expressing feelings of excitement, fear, optimism, dread, and curiosity regarding AI's ability to generate new forms of what we used to think of as human expression. Generative AI creates text, images, audio, and video at qualities that surpass many of our expectations. Future versions will inevitably surpass current abilities.

History: XXX served as the senior editor for this article.

Keywords: *IJDS* • editorial • editors • computers/computer science • artificial intelligence

Introduction

As a data science researcher with a background in statistics, my notion of "generating data" or "synthetic data" was monopolized by the idea of creating data for the purpose of evaluating or improving the properties of a method. One example is *simulation*, where we generate data from an underlying distribution and evaluate the performance of a method on these data. We can then modify the simulation parameters to assess sensitivity to different aspects of the data. Simulation is extremely common in *IJDS* papers for evaluating the performance of a proposed methodology and its sensitivity to different manipulable factors. A second example is *data imputation*, where missing data are "filled in" using different approaches. A third example is *resampling methods*, where multiple samples are randomly drawn from a data set. Popular resampling methods include the bootstrap and cross-validation techniques. Bootstrap is useful for estimating the distribution of an estimator without making parametric assumptions. Cross-validation is useful for evaluating out-of-sample performance. What is common to all these "data generation" methods is the end goal: We want our data science method and its implications to generalize: from a sample to a population, to new observations, or to other data contexts. A second commonality is the type of data

generated: quantitative (mostly numerical) data. Third, we generate the data by writing computer code. In short, we create data that we know how to analyze, by using coding skills, so that we can expand our understanding of a model/method.

Yet, generative artificial intelligence (AI) seems different: First, it creates "unstructured data" such as text, images, audio, and video. The input can be human language (e.g., text prompts). Second, current applications such as OpenAI's ChatGPT (based on a large language model (LLM)) and DALL-E (for creating images) are already used by the public for purposes other than methodological or applied research. Third, rather than generating a sample of observations, the tools are often used to generate a single "data point" (a response to a question, a single "conversation," or a single image). (Are you wondering why Monte Carlo simulation software is not so hot among the general public?)

Although we expect a burst of new data science research on generative AI, as well as on the impact of generative AI on society, business, and industry, in this editorial, we decided to explore how *IJDS* authors, reviewers, editorial members, and management might use (or abuse) these new applications. For this purpose, the entire group of *IJDS* senior editors put their heads together and proposed several uses, challenges, and

general recommendations. Although some of these uses are possible with current tools, some are speculative, awaiting the improved capabilities of future LLMs. We came up with the following.

Opportunities for Using Generative AI in *IJDS* Submissions

Authors

Authors generate, consume, and interact with text and image data of different forms. They write their manuscript using English language and LaTeX, create and test computer code, read literature and other resources in written form, generate images for their manuscript, and perhaps even analyze text, images, or other nonnumerical information as their data.

Authors can use text-based tools as a *writing assistant* while writing their manuscript similar to a professional human editor. Such uses can include language editing, translation, customizing the readability level to the *IJDS* audience, suggesting paper structure, and helping create a first rough draft. The summarization capabilities of tools such as ChatGPT can be used for shortening the manuscript to meet length requirements (Dwivedi et al. 2023, p. 32). They can also generate ideas for a suitable title and keywords the authors can consider. Summarization can also be useful for helping craft a succinct, but sufficiently informative abstract within the journal's word limit.

As a *creative assistant*, image generation tools can be used for creating visuals such as illustrative diagrams, replacing current uses of paid design services.

As most *IJDS* papers include code that implements the proposed methodology, generative AI tools can support authors as *coding assistants*, helping during code writing and for validating human-written code. Code assistance can be done via prompts to an application such as ChatGPT, code editors, or integrated development environments. For example, ChatGPT can be used for debugging and writing LaTeX code: a sometimes daunting and frustrating task. Another example is Github Copilot by Microsoft (<https://github.com/features/copilot>), which can “suggest code and entire functions in real-time, right from your editor” and can use natural language prompts.

Authors can use LLM-based tools as a *prereviewing assistant* for reviewing their manuscript before submission. Such an assistant might help detect unclear arguments, suggest missing references, and who knows what else!

Authors proposing new methodologies suitable for nonnumerical data (e.g., text or image) could use generative AI tools for generating data, which can then be used, similar to numerical simulated data, for evaluating the methodology under different assumptions and conditions and

for providing reproducible illustrations (especially when the real data cannot be disclosed or shared).

Reviewers

Reviewers provide invaluable feedback on submitted manuscripts. Reviewers interact and generate written text by reading the manuscript, related documents, and communications from the editors, writing a referee report, and sometimes examining submitted code.

Reviewing consumes significant time and energy, and therefore we have tried to make the reviewing process more efficient by creating a structured reviewer form with five questions (<https://pubsonline.informs.org/page/ijds/reviewer-guidelines>). Generative AI tools hold the promise to further improve the efficiency, rigor, and effectiveness for creating referee reports and communicating the feedback to authors. Similar to authors, reviewers can use the language and summarization capabilities of LLM-based tools for crafting their referee report: editing their answers for the five questions, clarifying and prioritizing their further comments and suggestions, and improving the report's communication style so that authors perceive it as professional and constructive.

Using AI tools as a *writing assistant* can help reviewers better mask their identity (*IJDS* uses a single-blind system where reviewer and associate editors (AEs) are anonymous to authors).

Reviewers can use LLMs as a *reading assistant* by summarizing the main points of different parts of the paper. Many reviewers first skim a paper before reading it in depth, to understand its structure and overall approach. Section-by-section summaries by an LLM can help identify the main points during such a first browse.

Reviewers often point authors to key references that are missing, or to other resources that can strengthen the paper's contribution (e.g., a data set the authors should apply their method to). To do this, the reviewer must identify the exact reference or the resource's location. Although tools such as Google Scholar are powerful for locating references to share with authors, LLMs can further help the reviewer identify additional relevant references/resources that they might have forgotten, or even suggest ones the reviewer was not aware of (in which case the reviewer should carefully check the proposed references/resources before recommending them to the authors!).

For submissions where authors use the output from LLMs as part of their data, analysis or arguments, reviewers can use LLMs as a *verification assistant* to validate these outputs. To allow this kind of “reproducibility check,” authors must provide the exact system specification and the prompt used for generating the output.

When reviewing submitted code, reviewers can use LLMs as a *coding assistant* for understanding the purpose of some part of the code (if not clearly written or

documented), and even for modifying the code to evaluate some aspect.

Finally, at the *IJDS*, authors and reviewers are asked to consider implications of the proposed methodology. Given that data science ethics is a relatively new and fast-changing area, LLMs can help reviewers identify potential ethical issues that were not considered by the authors (prompt: “what are possible ethical issues with a method that proposes...”). Reviewers using LLMs for such uses should disclose to the authors their use of an LLM and the exact system specification and prompt used.

Editors

The *IJDS* editorial board includes the editor-in-chief (EIC), senior editors (SEs), AEs, managing editor, and the reproducibility editor and team. They all encounter and generate text of some form as part of the review process and other journal-related tasks.

AEs at the *IJDS* have the critical task of guiding authors through the expert review team’s feedback. The AE reads the manuscript, and the referee reports (including the report from the reproducibility team) and creates an AE report that summarizes the key points, guiding the authors on the way forward (or explaining a rejection recommendation). The AE report is critical for authors to understand, so they can follow the guidance toward successful publication (or if rejected, so they can understand the reasons for rejection and how they might move forward with another outlet). The AE can therefore benefit from the language and summarization capabilities of LLM tools for integrating the key points from the referee reports into the AE report (an *integrating assistant*). However, such a summary is by no means an appropriate AE report! (This is a critical point we share with *IJDS* AEs during onboarding and training.) The AE must then use human expert judgment to carefully guide the authors by prioritizing which recommendations the authors should follow, and in what form. Hence, the AE should carefully scrutinize the “integration assistance” by the LLM, and substantially inject their expertise and active guidance to arrive at a useful AE report.

Like reviewers, AEs can use LLMs for improving their report in terms of language, readability, and professional communication style. Such editing would also better mask AE identity.

AEs perform another critical role: selecting and inviting reviewers for a manuscript. Identifying and recruiting a sufficient number of suitable reviewers is a difficult task. Although AEs have a good network of experts, and the journal’s database includes potential reviewers that can be searched by expertise keywords, LLMs can further help AEs expand their network and identify new potential reviewers based on adequate text prompts. This expansion is likely much

broader than the limits of our current community and can also help further diversify our reviewers by specifically directing the search toward certain geographies, gender, or other groupings that are outside the AE’s familiarity.

SEs at the *IJDS* guide the reviewing process and make decisions about manuscripts. They communicate with the EIC and AEs and, importantly, convey decisions, explanations, and guidance to authors. The SEs are not anonymous. SEs can also use LLMs as *writing assistants* for writing and personalizing their decision letters. Using appropriate prompts, they can make sure their report conveys sufficient information to the authors about the strengths and weaknesses of the manuscript, and about the SE’s expectations of a revision. Because SEs also read the submissions, they can use LLMs as *reading assistants*, helping to more efficiently identify the contribution for the purpose of making decisions and selecting suitable AEs. Like AEs, the SE can benefit from the summarization capabilities of LLMs for summarizing the review team’s reports.

The EIC encounters and generates text through a variety of roles: written communications with authors and editorial board members regarding submissions and with external entities for initiating new efforts, special issues, collaborating with INFORMS and other societies and conference organizers, commissioning of papers, and so on. The EIC produces writing for a variety of avenues, including for decision letters, editorials, journal processes and guidelines (that appear on the journal website), presentations and agendas for editorial meetings and outreach activities. The EIC also consumes text by reading submissions, referee reports, and the AE and SE reports. The EIC can thus benefit from using an LLM as a *writing assistant* for communications and a *reading assistant*, for example, for summarizing incoming submissions to determine suitability and further handling. LLMs could help the EIC identify new relevant audiences (e.g., conferences) to introduce to *IJDS*. Summarization of relevant conference programs might help identify new topics and papers to pursue.

Finally, the EIC and SEs can also use LLMs as *brainstorming assistants* for identifying topics and titles for special issues, determining maturing domains with missing surveys, and identifying potential associate editors from a more diverse pool. Table 1 summarizes the different types of LLM-based assistant that can assist *IJDS* authors, reviewers, and editors.

Other *IJDS* Task Forces: Reproducibility, Media, and Management

The *IJDS* reproducibility team runs a check on each code submission to make sure it is reproducible by reviewers and future readers. Although most submissions share their code via CodeOcean, some software is not supported by CodeOcean. In such cases, the

Q:13

Table 1. Types of LLM-Based Assistants for *IJDS* Authors, Reviewers, and Editors

<i>IJDS</i> role	Type of LLM assistance
Author	Writing assistant, coding assistant, creative assistant, prereviewing assistant
Reviewer	Reading assistant, writing assistant, verification assistant, coding assistant
Editor	Reading assistant, writing assistant, integrating assistant, brainstorming assistant

reproducibility team might use LLMs as *coding assistants*, helping identify potential reproducibility problems and providing suggestions to authors.

The *IJDS* media outreach team has been sharing information on social media, producing interviews with *IJDS* editors and authors, and posting about new *IJDS* articles that appear on the journal website. Generative AI can assist the team in creating visuals, customizing interviewing scripts, and summarizing the contribution of new articles to fit social media promotion formats.

The submission management system (ScholarOne) has many complex workflows and automated emails that need constant updating. These automated emails can be improved using LLMs. Manual checks of incoming submissions could potentially also benefit from LLM *check-in assistance* that might point out undetected issues to the managing editor. Ideally, all these tasks will be significantly enhanced once such information systems have a ChatGPT-type interface.

What Else?

After the coauthors of this editorial shared their ideas, the EIC probed ChatGPT (using GPT-4) regarding how ChatGPT might be useful to an author, reviewer, and editor at a data science journal. The prompts and results are shown in Figures 1–3. We found the answers covered many of our suggestions, but not all (we humans still had some original ideas!), and some items were irrelevant to *IJDS*-type submissions. The LLM provided a useful additional team member to our brainstorming!

Dilemmas and Challenges

Since its free launch in November 2022, reports of challenges of ChatGPT have been appearing frequently in the general media and in the scientific literature. These include issues with privacy and data governance, transparency, accuracy (“AI hallucinations”), accountability, and absence of originality (“stochastic parrot”). In the context of scientific publications, journal editors have voiced a range of sentiments regarding authors’ use of LLMs: from total restrictions to strong embracement. Multiple large publishers have already listed policies regarding authors’ use of LLMs and whether an LLM can be listed as a coauthor (table 4 in Dwivedi et al. 2023).

In the following, we focus on a few challenges and dilemmas that arise when authors, reviewers, or editors use LLMs in *IJDS* submissions.

When Does Writing Assistance Become a “Coauthor”?

Although language editing has been well recognized as a standard presubmission task for most authors, there are different views on the extent of writing improvements that would make the editing provider become a coauthor. This ethical question quickly became a serious issue as early users of ChatGPT published papers with ChatGPT as a coauthor (King and ChatGPT 2023). To date, the consensus across many publishers is that an LLM cannot be listed as an author or coauthor, because authors are accountable for their papers. At the time of writing, INFORMS journals do not have a single code of practice regarding LLMs, yet at the *IJDS*, we currently take the position that an LLM should be considered an assistant and not a coauthor. Our reasoning is that LLMs create a real authorship challenge in fields where the writing itself is the contribution. Yet, at the *IJDS*, language is just an intermediary for conveying and communicating about a data science methodology, data, and its decision-making context. For this reason, data science journals have typically encouraged authors to hire professional editing services or use tools such as AI-based *Grammarly*.¹ In fact, some journals even offer authors paid editing services.²

Will Improved Writing Cover up Methodological Flaws?

A possible concern relates to how reviewers and editors might be tempted to skip critical reading and deep thinking when a manuscript was made to “look” good by using an LLM. In other words, the packaging would hide the content from proper review. From our current experience at the *IJDS*, this is not likely to occur because well-written manuscripts make it easier to identify important issues, thereby requiring fewer review rounds to a final decision than difficult-to-read manuscripts. Moreover, well-written articles give an advantage to native English speakers who are good writers and to authors with funding for writing assistance. The use of LLMs as writing assistants can democratize writing skills by making them more broadly accessible. Similar advantages arise in the generation of “nice looking” diagrams that no longer requires artistic talent or funding for outsourcing the task.

LLM Availability

We note that a current practical challenge of using tools such as ChatGPT is their availability. ChatGPT is currently

Q:17

Figure 1. (Color online) ChatGPT's Response for How It Might Be Useful for Authoring a Data Science Paper



how can ChatGPT be useful for authoring a paper for a data science journal?



ChatGPT can be a valuable tool for authoring a paper for a data science journal by assisting you in various aspects of the research and writing process. Here's how it can help:

1. **Idea generation:** ChatGPT can help you brainstorm research topics, data science techniques, and potential applications. Provide a broad subject area or problem, and the AI can suggest specific ideas to investigate.
2. **Outlining:** Create a structured outline for your paper by specifying the sections and their content. ChatGPT can help you generate a logical flow, ensuring that your paper covers all relevant aspects of the data science research.
3. **Background research:** The AI can help you gather and summarize relevant background information, such as previous research, related algorithms, and key concepts in data science.
4. **Data collection and preprocessing:** ChatGPT can provide guidance on data collection methods, data sources, and data preprocessing techniques, such as handling missing data, outliers, or data transformation.
5. **Algorithm selection and implementation:** The AI can assist you in selecting suitable algorithms and models for your research, as well as help you understand their advantages, disadvantages, and implementation details.
6. **Experiment design:** ChatGPT can provide suggestions for designing your experiments, including validation techniques, performance metrics, and ways to minimize biases or overfitting.
7. **Results and analysis:** The AI can aid in presenting and summarizing your results, offering insights into the data and identifying trends or patterns. It can also help you interpret the performance of your models and their implications.
8. **Discussion and interpretation:** ChatGPT can help you discuss your findings, interpret your results, and identify the implications and limitations of your study in the context of data science.
9. **Writing assistance:** ChatGPT can help improve your writing by providing suggestions for better phrasing, grammar, or syntax. Additionally, it can help ensure your paper follows the appropriate style guide and formatting rules for a data science journal.
10. **Citation and referencing:** The AI can provide guidance on citing sources and creating a bibliography according to the required citation style.
11. **Abstract and title:** ChatGPT can help you craft a concise and informative abstract and suggest potential titles that accurately represent your data science research.
12. **Editing and proofreading:** The AI can help you review your paper, identify inconsistencies or errors, and suggest improvements for clarity and readability.

It is essential to remember that while ChatGPT can be a helpful tool for authoring a data science paper, it should not replace your own expertise, critical thinking, or domain knowledge. Always verify the information provided by the AI and use it as a starting point or supplement to your own research and writing efforts.

banned in some countries (Italy) and unavailable in others³ (e.g., China, Egypt, Iran, Russia, Syria), and to some authors, the software will be unavailable if they are not able to create a free account. Already, OpenAI's

freemium model for ChatGPT creates inequalities (as of writing, the \$20/month ChatGPT Plus version gives access to GPT-4, whereas the free version uses the less advanced GPT-3.5).

Figure 2. (Color online) ChatGPT's Response for How It Might Be Useful for Reviewing a Submission to a Data Science Journal

how can ChatGPT be useful for reviewing a submission to a scientific journal?



ChatGPT can be useful for reviewing a submission to a scientific journal in several ways, helping to improve the review process and provide valuable feedback to authors. Here are some ways the AI can assist:

1. **Pre-screening:** ChatGPT can help you quickly evaluate if the submission adheres to the journal's guidelines, including formatting, word count, and citation style. It can also help identify any potential ethical concerns, such as plagiarism or conflicts of interest.
2. **Understanding the content:** The AI can help summarize the main points, objectives, methods, and findings of the paper, enabling you to grasp the core aspects of the research.
3. **Assessing novelty and relevance:** ChatGPT can provide insights into the novelty of the research, its significance, and how it relates to existing literature. It can also help you determine if the research aligns with the scope and aims of the journal.
4. **Evaluating the methodology:** ChatGPT can assist in analyzing the methods used in the research, identifying any potential flaws or biases, and suggesting improvements or alternative approaches.
5. **Analyzing results and data:** The AI can help you scrutinize the results and data presented in the paper, identifying inconsistencies, statistical errors, or any unsupported conclusions.
6. **Assessing the discussion and interpretation:** ChatGPT can provide feedback on the clarity and validity of the conclusions drawn, as well as the implications, limitations, and potential future research directions suggested by the authors.
7. **Reviewing language and writing quality:** The AI can help you identify issues with grammar, syntax, and clarity in the paper, offering suggestions for improvements.
8. **Providing constructive feedback:** ChatGPT can help you formulate specific, actionable feedback and recommendations for the authors, ensuring that the review is constructive and helpful.
9. **Streamlining the review process:** By using ChatGPT to assist in the review process, you may save time and increase your efficiency as a reviewer.

It is essential to remember that ChatGPT should not replace your own expertise and critical thinking when reviewing a scientific paper. The AI can be a valuable tool to augment your review process, but your own experience, knowledge, and judgment remain crucial to providing a thorough and accurate evaluation of the submission.

Will Use of LLM by Reviewers Lead to Inferior Referee Reports?

Given the *IJDS* structured reviewer form, it is possible that reviewers will try to use LLMs to answer the five questions. Reviewers are expected to use LLMs to supplement their own knowledge and to trigger deeper investigations (as described in the previous section). Dealing with low-quality reviews is not new to the reviewing system, which is why the AE's role in identifying reviewers and evaluating their reports is so important. Low-quality reviews reflect on reviewers

and affect their reputation in the community (although reviewers are anonymous to the authors, they are known to the editors). Hence, we do not foresee a decline in referee reports. However, *IJDS* editors will need to be more vigilant.

A more likely challenge related to reviews is the following: Human reviewers typically start their report with a brief summary of the paper and its contribution, so authors can determine whether their main message was perceived correctly by the reviewer. However, if a reviewer uses an LLM tool to write this summary, this

Figure 3. (Color online) ChatGPT’s Response for How It Might Be Useful to an Editor of a Data Science Journal



how can an editor of a data science journal use ChatGPT?



An editor of a data science journal can utilize ChatGPT to streamline various tasks and improve the overall publication process. Here are some ways ChatGPT can be helpful:

1. **Initial screening:** ChatGPT can help with the initial assessment of submitted manuscripts by checking their adherence to submission guidelines, including formatting, word count, citation style, and potential ethical concerns.
2. **Assigning reviewers:** The AI can assist in identifying potential reviewers by analyzing the content of a manuscript and matching it with the expertise of researchers in the data science field.
3. **Managing reviewer feedback:** ChatGPT can help summarize reviewer feedback, identify common themes, and suggest possible revisions or clarifications for the authors. This can streamline the communication between authors, reviewers, and the editorial team.
4. **Editing and proofreading:** ChatGPT can help identify inconsistencies or errors in the text, suggest improvements for clarity and readability, and ensure that the manuscript follows the journal's style guide and formatting rules.
5. **Decision letters:** The AI can assist in drafting decision letters, summarizing the reviewers' feedback, and providing clear instructions to authors regarding revisions, resubmissions, or publication decisions.
6. **Content curation:** ChatGPT can help you identify trending topics and emerging areas of interest in data science to guide the journal's focus and scope, ensuring it remains relevant and cutting-edge.
7. **Promotion and outreach:** The AI can assist in creating engaging summaries, press releases, or social media content to promote the journal's articles and attract a wider readership.
8. **Streamlining communication:** ChatGPT can help draft emails, respond to queries, and facilitate communication between authors, reviewers, and the editorial team, saving time and maintaining a high level of professionalism.
9. **Monitoring publication metrics:** ChatGPT can help you keep track of important metrics, such as citation counts, download statistics, and impact factors, to evaluate the journal's performance and identify areas for improvement.

It is important to remember that ChatGPT should be used as a tool to support and enhance the editor's work, rather than replace their expertise and judgment. Human oversight and intervention are crucial to ensure the quality and integrity of the publication process.

human-to-human signaling mechanism becomes moot. Moreover, an incorrect summary can cause the reviewer to misunderstand the paper's contribution. We therefore strongly recommend reviewers to manually create the initial draft of this important summary and then use LLMs to polish the language (if needed).

Accuracy

Current LLMs have been shown to produce incorrect (and even “imaginary”) output, stated with as much certainty as correct output (Ji et al. 2023). This applies to

answers to questions, text summarization, software code, results of computations, and more. Although LLMs promise to be more efficient than search engines or Google Scholar for citing references and identifying missing literature, currently, some LLMs such as ChatGPT suffer from “AI hallucinations,” which include inventing non-existent references.

Because these accuracy issues are a known problem, several have already been improved in recent versions of ChatGPT (Bubeck et al. 2023). Although we anticipate these issues will be solved in future versions, until

then, we strongly recommend authors and reviewers to meticulously examine and double-check any output by LLMs.

Sensitive Information and Privacy

Authors who insert confidential ideas, methods, or code into a tool such as ChatGPT might effectively be giving away the information, as recently occurred in a case involving Samsung.⁴

Authors should take special care not to insert personal or sensitive data about data subjects without making sure their data are not used by the LLM. For example, OpenAI's policy states they use data inserted by ChatGPT users but not from API users.⁵ Not only might entered information be available to the hosting company (e.g., OpenAI), but it can potentially also be probed by other users.

Gaming the System

We have assumed thus far that the human counterparts are acting in good faith, using LLMs to improve their performance and output. However, malintent and abuse are certainly possible. One such possibility is authors who learn to strategically use LLMs to optimize their submission's chance for acceptance not based on true contribution but rather by identifying language, structure, and other factors that are more likely to get accepted. In other words, identifying "dark patterns" that might trick the review team. Although such tactics might work in some areas, we believe it is more difficult to do so for *IJDS*-type submissions. With multiple sets of independent, diverse, and expert eyes on each submission, including a reproducibility check of submitted code, it is currently difficult to imagine "dark patterns" that would escape all these eyes.

Accountability

A major challenge with the use of LLMs is accountability: Is the LLM to be blamed for an incorrect summary of reviewers' feedback? Can we blame the LLM's training data for a discriminatory editorial letter? Can the authors blame OpenAI for fictitious references? In a recent article about AI liability in *The Hill*,⁶ Vasant Dhar from New York University wrote the following:

"There's a simple rule from the physical world that seems useful in controlling the risks of AI, namely, a credible threat. If you knowingly release a product that is risky for society, like a toxic effluent into the environment, you are liable for damages. The same must apply to AI."

Our approach at the *IJDS* is that the output of these tools, when included in an *IJDS* submission/evaluation, should be subject to careful human scrutiny. **The author, reviewer, editor, and journal administrator are all still accountable for any materials and communications**

arriving from them. In other words, we treat these new applications as decision-support systems (with currently many faults) while maintaining our role as human decision makers. All the uses mentioned in the previous section take this stance. This approach also means that acknowledging the use of, say ChatGPT, is not needed if an author uses it for improving grammar or if an editor uses it to polish the language in the report. Even without explicitly using a tool such as ChatGPT, LLMs are being integrated into many productivity applications such as Microsoft Word and Google docs and are also being used to write software programs and produce models. These applications will be widely used by authors to write papers and conduct research and become part of routine practice and unlikely to raise red flags. In conclusion, our position is that no attribution is needed as long as the human scrutinizes the LLM's result and as long as the LLM itself did not *propose* the new data science methodology (Prompt: "create a new data science methodology that addresses an important decision making problem in business, evaluate its performance on real data, and assess its practical and ethical implications").

Conclusions

At the *IJDS*, we are enthusiastic about the possibilities of new LLMs for improving the quality of submissions and making the review process more effective and efficient. We mentioned uses of an LLM as a *reading assistant*, *writing assistant*, *coding assistant*, *creative assistant*, *integration assistant*, *prereviewing assistant*, *verification assistant*, and *brainstorming assistant*. Other useful assistant roles might also appear. Yet, we are aware that "unknown unknowns" will likely surprise us (as they will other data science journals), and our editorial board is prepared to discuss and tackle issues that appear.

Although effective use of LLMs for journal submissions, reviewing, and management will take time to evolve, early adopters will likely already experiment with different uses. Recent research has shown that "presenting humans with arguments for two competing answer options, where one is correct and the other is incorrect, allows human judges to perform more accurately, even when one of the arguments is unreliable and deceptive" (Parrish et al. 2022). Although we encourage experimentation, we remind authors, reviewers, and editors that current LLMs often fail without showing a trace of uncertainty. The choice of prompt has already shown to distinguish between correct and incorrect answers, prompting the new field of "prompt engineering."

As mentioned, multiple publishers have already come up with restrictions for using/attributing LLMs, and at this time of writing, INFORMS journals do not yet have such a code of practice. We anticipate that such a code of practice will appear and evolve as our journals start experiencing and identifying relevant

challenges. For *IJDS* papers, which focus on novel data science methodology for decision making, we expect the humans involved will use generative AI mostly for communication, evaluation, and dissemination of knowledge. This can improve equalizing effort and submission barriers related to language proficiency, lower costs and speed up editing and creating visuals, enrich authors' and reviewers' perspectives, improve communications, make the reviewing process more effective and efficient, increase the visibility and readership of *IJDS* publications, and hopefully free up our most precious and limited resource: time. Whether and how LLMs will help directly accelerate the development of novel data science methodology for decision making is an open question.

Although our focus has been on using LLMs for authoring and reviewing *IJDS* manuscripts, we of course anticipate a spike in LLM-based research. We encourage submissions that look into the design, development, and use of such tools, as well as ethical considerations and implications. From this research perspective, editors and reviewers will need to update their knowledge and ability to handle submissions of LLM-based research artifacts and respond appropriately via *IJDS*-hosted workshops and training sessions. Submissions to *IJDS* may use LLMs as objects of study as data generating systems, examining face validity, provenance, statistical properties, and so on, of the resulting data. LLMs may also be applied to study consequential sectors and domains such as healthcare where researchers want to study efficiency, effectiveness, safety, and quality of healthcare systems and how workflows get changed when such solutions are deployed in real life settings. Safe, responsible, and meaningful ways to use this technology are currently being debated, and *IJDS* policies will evolve with them.

We end this editorial by inviting readers to share their thoughts and feedback: Do you have further ideas

of how authors, reviewers, editors, and the *IJDS* could benefit from using generative AI? What do you think about the suggested uses? Please feel free to post on social media or email the EIC.

Endnotes

¹ See <https://www.grammarly.com>.

² Examples include Taylor & Francis and Wiley: <https://authorservices.taylorandfrancis.com/publishing-your-research/writing-your-paper/editing-services-improve-your-manuscript/> and <https://onlinelibrary.wiley.com/page/journal/15405915/homepage/forauthors.html> (accessed April 21, 2023).

³ See <https://platform.openai.com/docs/supported-countries> (accessed April 21, 2023).

⁴ See <https://mashable.com/article/samsung-chatgpt-leak-details> (accessed April 21, 2023).

⁵ See <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance> (accessed April 21, 2023).

⁶ See <https://thehill.com/opinion/cybersecurity/3938554-why-its-time-to-safeguard-against-ai-liability/> (accessed April 21, 2023).

References

- Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al. (2023) Sparks of artificial general intelligence: Early experiments with gpt-4. Preprint, submitted XX, <https://arxiv.org/abs/2303.12712>.
- Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. (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* 71:102642.
- Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. (2023) Survey of hallucination in natural language generation. *ACM Comput. Survey* 55(12):1–38.
- King MR, ChatGPT (2023) A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular Molecular Bioengr.* 16(1):1–2.
- Parrish A, Trivedi H, Nangia N, Phang J, Padmakumar V, Saimbhi AS, Bowman SR (2022) Two-turn debate does not help humans answer hard reading comprehension questions. Preprint, submitted XX, <https://arxiv.org/abs/2210.10860>.

Q:15
Q:16

Proof Only

AUTHOR QUERIES

DATE 5/10/2023

JOB NAME IJDS

ARTICLE 20230007

QUERIES FOR AUTHOR Shmueli et al.

THIS QUERY FORM MUST BE RETURNED WITH ALL PROOFS FOR CORRECTIONS

- Q: 1_Please provide short title for running head. Please note that entire running head, including author names, must be <81 characters/spaces.
- Q: 2_Please provide all funding information, if applicable, including the names of the institutions as well as any grant numbers associated with the funding, or confirm that no funding was received.
- Q: 3_Your color figures appear in color in your page proof to simulate their presentation in color online, but they will be converted to grayscale in the print version of your article because you have not requested color processing for print. Please (a) check your figures to confirm that they will be sufficiently clear in grayscale presentation in the print version of your article and (b) modify your figure legends as necessary to remove any references to specific colors in the figure.
- Q: 4_Please confirm that the article title, author names, affiliations, and email addresses are set correctly. If applicable, please provide author ORCID numbers.
- Q: 5_Please confirm that corresponding author designation in author line is correct.
- Q: 6_Please provide postal codes for all affiliations and spell out IROM in affiliation e.
- Q: 7_Please spell out GPT at first mention in text.
- Q: 8_Please update the missing senior editor name in XXX.
- Q: 9_Please confirm that keywords are correct as set.
- Q: 10_Please clarify use of “my” with many authors.
- Q: 11_Per journal style, throughout text, “while” has been changed to “whereas” or “although” to clarify a causal rather than a temporal relationship. Please confirm.
- Q: 12_Please confirm that heading levels are correct as set.
- Q: 13_Per journal style, throughout text, “since” has been changed to “because” to clarify a causal rather than a temporal relationship. Please confirm.
- Q: 14_Please spell out API at only mention in text.
- Q: 15_For Bubeck et al., Dwivedi et al., and Ji et al., please provide 7 authors before et al. per INFORMS style.
- Q: 16_For ALL arXiv references, please provide submission date and update if possible.

Q: 17_Please check that all tables and figures, including their titles and notes (which may contain edits) are set correctly.