Received: 25 January 2025 | Accepted: 11 February 2025

DOI: 10.1002/jcpy.1453

METHODS DIALOGUE

Comments on "AI and the advent of the cyborg behavioral scientist"

Paul Andrew Blythe[1] | **Christopher Kulis**[1] | **A. Peter McGraw**[1] | **Michael Haenlein**[2,3] | **Kelly Hewett**[4] | **Kiwoong Yoo**[4] | **Stacy Wood**[5] | **Vicki G. Morwitz**[6] | **Joel Huber**[7]

[1]Marketing Department, Leeds School of Business, University of Colorado Boulder, Boulder, Colorado, USA

[2]Marketing Department, ESCP Business School, Paris, France

[3]University of Liverpool Management School, University of Liverpool, Liverpool, UK

[4]Haslam College of Business, University of Tennessee, Knoxville, Tennessee, USA

[5]The J. Lloyd Langdon Distinguished University Professor of Marketing, Poole College of Management, North Carolina State University, Raleigh, North Carolina, USA

[6]The Bruce Greenwald Professor of Business and Professor of Marketing, Columbia Business School, Columbia University, New York, New York, USA

[7]The Fuqua School of Business, Duke University, Durham, North Carolina, USA

Correspondence
Joel Huber, The Fuqua School of Business, Duke University, Durham, NC, USA.
Email: joel.huber@duke.edu

Abstract

Below are comments on Tomaino, Cooke, and Hoover by four teams of collaborative reviewers that helped clarify and focus its original version. Their comments on the refined version articulate how the fast-moving world of generative AI can alter authors, readers, reviewers, and consumer behavior journals. In the first comment, Blythe, Kulis, and McGraw propose that Generative AI requires substantial effort to generate research that is fast, cost-effective, and of high quality. They articulate three recommendations: to ask, to train, and to check the system. *Asking* builds on GenAI's ability to reveal its own capabilities at different stages of the research process. *Training* allows the system to be customized with relevant context, domain-specific documents, and tailored examples, enhancing its accuracy and reducing errors. *Checking* is strongly advised to validate that the outputs are both reasonable and robust. Haenlein, Hewett, and Yoo build on the capabilities of Large Language Models that go beyond the research practices central to consumer psychology. They outline strategic prompting strategies: starting broadly and gradually narrowing to specific domains, downloading information from relevant articles and data that is unlikely to be part of the current corpus, and evoking specific theories, methods, or presentation formats. They also elaborate on the ways the apparent magic of GenAI may raise learning or ethical challenges. The third comment by Stacy Wood focuses less on the capabilities of GenAI and more on how its adoption will depend on researcher feelings—in other words, how different aspects of its use may alter researchers' experiences of doing research and their identities as scholars. GenAI has the potential to both build (through increased productivity or increased accessibility) and limit (through loss of agency or faster production) pride of purpose in research. She argues that feelings from using GenAI are likely to differ across research steps, from developing novel concepts, processes, analyses, and writing of the paper. Wherever GenAI may lessen the excitement, satisfaction, motivation, and perceived status of the researcher, barriers to its use are likely to be erected. Finally, Vicki Morwitz identifies new AI capabilities beyond those explored in Tomaino et al. Those include the ability to generate synthetic data that can guide empirical experiments, a facility to create audio and visual stimuli, a capability to study group behavior, and a capacity to reliably interpret complex human statements. The comment then closes with important questions for editorial policies, raising issues about limitations on AI use by authors, its appropriate applications by review teams, and possible publishers' restrictions on uploading copyrighted articles.

KEYWORDS
artificial Inteligence, generative AI, psychological research

Accepted by David Wooten, Editor

Linked article: Tomaino, G., Cooke, A. D. J., & Hoover, J. (2025). AI and the advent of the cyborg behavioral scientist. *Journal of Consumer Psychology*, *35*, 297–315. https://doi.org/10.1002/jcpy.1452.

BREAKING THE IRON TRIANGLE IN BEHAVIORAL RESEARCH WITH GENAI

By Paul Andrew Blythe, Christopher Kulis, and A. Peter McGraw

The peer-reviewed paper is a valued currency in behavioral science research. The publication process requires deep knowledge, diverse skills, and a substantial investment of time and money. Facing single-digit acceptance rates at journals, scholars must craft novel hypotheses, design compelling studies, analyze complex data, and write clearly. These demanding requirements often exceed any single researcher's expertise, which is why impactful projects typically feature multiple authors with complementary skills (Okamura, 2019; Wuchty et al., 2007).

The rapid rise of generative artificial intelligence (GenAI) presents researchers with an opportunity to add a highly skilled, low-cost, efficient "virtual author" to their team. Unlike traditional rules-based AI models, GenAI leverages machine learning algorithms to produce outputs that mimic human reasoning (Marr, 2023). Large language models (LLMs) like ChatGPT, Claude, Elicit, and Perplexity stand out for their ability to understand and generate human language, conduct analyses, and write code.

Tomaino et al. (2025) tested GenAI's ability to conduct a research project with minimal intervention that resulted in a paper. The authors demonstrated that GenAI excels at early-stage research tasks and basic analysis. However, the models faced limitations in complex, context-specific tasks such as focused literature reviews and more advanced types of data analysis. Tomaino et al. further found that while GenAI performed many functions with impressive speed and affordability, the models were subject to "hallucinations" (i.e., fabricated, incorrect outputs; Brittain, 2023).

Breaking the iron triangle

There is a saying in business, science, and design: "Good. Cheap. Fast. Choose two." This Iron Triangle- balancing cost, time, and quality—offers a framework for assessing GenAI's benefit to research efficiency (Pollack et al., 2018). We suggest that marketing researchers, psychologists, behavioral economists, and scholars from adjacent fields can use GenAI models to reshape these trade-offs.

Speed

GenAI's ability to complete technical and procedural tasks in seconds reduces bottlenecks that slow down research progress (notably, waiting for collaborator input.) Tasks such as data analysis, programming, and experimental design, which might otherwise require weeks and the involvement of specialized collaborators, can be achieved in a fraction of the time. This efficiency gives researchers more flexibility to focus on idea generation, hypothesis testing, and data interpretation—as well as bringing important ideas to market faster.

Cost

Base versions of GenAI models are typically free—though slower and throttled compared to subscriber versions that cost about $20 per month. Paid versions offer faster processing speeds and greater responsiveness. (We suspect that subscriptions will soon be subsidized by universities in the same way email and journal access are).

Quality

GenAI provides a broad and flexible toolkit that can enhance each stage of the research process. Yet an important question remains in light of Tomaino et al.: "How good are the outputs?" The quality of GenAI outputs depend on the choice of model, its ability at the present moment, and the expertise of the researcher prompting it.

To help researchers break the Iron Triangle and simultaneously enjoy fast, low-cost, and high-quality outputs, we propose three guiding principles: "Ask it," "Train it," and "Check it."

Recommendation 1: "Ask it"

We first recommend researchers at any stage of GenAI adoption to query GenAI about its capabilities, limitations, and optimal roles. By prompting GenAI to identify its capabilities, "Ask it" helps researchers select the most appropriate model, understand its limitations, and develop strategies for optimal prompting. Researchers can even ask GenAI to suggest effective prompts about its capabilities (i.e., ask it how to "Ask it").

Before beginning their study on ethical fatigue, for example, Tomaino et al. could have prompted GenAI to outline its capabilities related to literature review, hypothesis generation, stimulus design, and data analysis. By asking, "Which tasks in consumer psychology research can you perform most effectively?" and "How might your output lack reliability?" the researchers would gain insight into the model's strengths and weaknesses. The prompts might reveal that GenAI excels in summarizing existing studies but requires additional human oversight when generating nuanced hypotheses or crafting stimuli that rely on specific emotional appeals.

"Ask it" allows researchers to assess the efficiency and quality GenAI can bring to each research stage. For instance, Tomaino et al. could have queried GenAI's utility in the experimental design phase by asking, "How can I work with you to create consistent ethical and non-ethical statements for experimental conditions?" followed by, "How reliably can you maintain the same tone across multiple stimuli?" These prompts could determine whether GenAI can generate uniform stimuli that meet the experimental goals. By clarifying GenAI's capabilities in stimulus design, Tomaino et al. would minimize the need for post-generation adjustments.

Researchers can apply "Ask it" iteratively throughout a project. In the data analysis stage, for example, Tomaino et al. could have used this approach to verify the reliability of statistical outputs by asking, "How confident are you in these statistical methods, and what potential limitations should be considered?" or "Are there additional analyses you suggest we conduct?"

Recommendation 2: "Train it"

"Train it" is akin to bringing a new research assistant up to speed on a project. Today's GenAI landscape features two distinct model types: (1) general-purpose models like ChatGPT are trained on billions of documents from a wide array of publicly available sources; (2) specialized models like Elicit are purpose-built for specific tasks such as research assistance. This distinction affects their baseline capabilities. General models offer broad knowledge but may struggle with domain-specific tasks; specialized models excel in their targeted domains but have narrower applications.

Recent advances in customization help bridge this gap. Many platforms now allow researchers to enhance model performance through parameter adjustments, behavioral instructions, and uploading documents to a specialized corpus (i.e., body of knowledge). For instance, researchers can use ChatGPT's custom GPT feature to train a model for a specific research project. In our experience, such customization particularly helps address hallucination risks which are greater for niche or novel research topics.

Before beginning a specific task, researchers can improve the GenAI's understanding by sharing relevant reference papers. For example, before starting their ethical fatigue research, Tomaino et al. could have uploaded relevant consumer psychology papers and recent work on ethical marketing into a custom model. The additions to its corpus would have helped the GenAI better understand important findings relevant to the project.

By training GenAI with real examples, researchers can improve stimulus design. For example, Tomaino et al. could have uploaded samples of actual marketing campaigns before asking GenAI to create stimuli. This approach may have improved the authenticity of experimental materials.

Recommendation 3: "Check it"

A key finding from Tomaino et al. was that GenAI can present incorrect information—from made-up papers to faulty analyses. When Tomaino et al. asked GenAI to draft manuscript sections, the model produced initial drafts that provided a useful starting point but required substantial revision.

"Check it" is akin to a senior scholar onboarding an inexperienced research assistant, verifying each of the latter's tasks. We recommend implementing verification across each stage of the research process, for example, including:

- Literature Review: Researchers should verify citations against original sources, cross-reference key findings, and treat GenAI's literature suggestions as leads to investigate rather than as facts.
- Experimental Design: Researchers should request explanations for methodological choices and ask the model to give its rationale. Akin to asking a co-author to explain their study design reasoning, Tomaino et al. could have asked GenAI to explain its stimuli choices and their connections to research goals.
- Analysis: Besides being sure to select the right model and using it to suggest analyses, we suggest checking initial outputs against statistical analyses conducted by a researcher. While GenAI can help suggest appropriate analytical approaches and write code to analyze data, Tomaino et al. found the model's actual calculations contained errors.

Another way a researcher can "Check it" involves comparing how different GenAI models respond to the same prompt. Similar to getting opinions from a co-author on a research design, the researcher might ask both ChatGPT and Claude to suggest experimental materials. Where the models agree—perhaps in how to word survey questions or structure stimuli—these suggestions may be more reliable. Where they disagree, these differences signal a need for closer examination.

Conclusion

Our three recommendations provide a practical framework for working effectively with GenAI. Think of "Ask it" like having an initial conversation with a new research assistant—taking time to understand a model's capabilities helps behavioral researchers use it more effectively. "Train it" presents a customized model's details and requirements of the project by providing context and examples to its corpus. Finally, "Check it" is a reminder to treat the output the way a principal investigator would check a research assistant's work before submission.

GenAI offers behavioral researchers an exciting new tool that helps break traditional trade-offs between speed, cost, and quality. We expect the quality of the outputs to improve as new models come to market, yet in some cases, quality remains questionable, as Tomaino et al.'s test revealed. In the meantime, a successful collaboration with GenAI requires finding the right balance between leveraging its capabilities and maintaining oversight—which is the approach we used to write this paper.

THE MAGIC, CHALLENGES, AND POTENTIALS OF LARGE LANGUAGE MODELS

Michael Haenlein, Kelly Hewett, Kiwoong Yoo

> The spirits that I summoned, I now cannot rid myself of again.

At the time of writing this comment, ChatGPT, one of the most widely known large language models (LLMs), is celebrating its 2nd anniversary, but like the Goethe quote, it will not go away. If it were a child, it would be able to talk, walk, climb, jump, and run at this age. It would also be able to sort shapes and colors and show basic interest in potty training.[1] However, few two-year-olds could act as research assistants, develop new ideas, or help design or even analyze experiments (neither, for the record, should they be tasked with such tedious work at such a tender age). Tomaino et al. (2025) provide an excellent example of the power of LLMs and how they can transform research in consumer psychology and beyond. In our comment, we build on their work and elaborate on three aspects in more detail: how to interact with LLMs (prompt engineering), their potential use in generating silicon samples, and some ideas on how they can support research in other ways.

> The answers you get depend upon the questions you ask.

Tomaino et al. (2025) highlighted that working with LLMs requires scholars to develop prompts to use these models efficiently and effectively for the behavioral research process (i.e., prompt engineering). Marketing scholars have, until now, primarily focused on prompt engineering for marketing research (Arora et al., 2024; Goli & Singh, 2024). However, we believe these methods can be more broadly adapted for behavioral research. By establishing prompt libraries[2] or forums[3] where researchers can share best practices, the quality and precision of LLMs' outputs could be further enhanced, advancing relevant and rigorous scholarship.

As a starting point, we suggest *chain-of-thought prompting* (Wei et al., 2023), which allows the model to think through the stages of an idea, generating step-by-step insights by iteratively breaking down complex concepts. This prompting strategy can help generate ideas for theoretical frameworks or research gaps, as it allows scholars to refine the model's responses toward a topic that addresses existing literature and identifies an unexplored area (Yoo et al., 2024). For example, researchers can begin with a broad topic and then prompt the LLM to explore subtopics. They can prompt the LLM to identify major themes in the literature, summarize key studies, and identify gaps that the current study aims to address, all while directing the model toward reliable sources, potentially reducing hallucinations. Next, they can guide the model in designing an experiment or survey and proposing suitable methods and conditions.

At some stages, especially when dealing with data analysis, starting a new session with the model might be necessary to wipe out any prior memory. In data analysis, *zero-shot prompting* (Kojima et al., 2023) minimizes bias and overinterpretation by prompting the model without examples or contextual assumptions. For the extension stage, *few-shot prompting* (Sahoo et al., 2024) allows scholars to give the LLM examples of potential follow-up studies, guiding the model to generate more specific, relevant suggestions. For instance, scholars can start with a prompt such as, "Given an initial study on consumer personalization and loyalty, suggest an additional study," followed by, "Here is an example of a design for this study."

Finally, *retrieval-augmented generation* can be effective, especially in the manuscript production stage (Sahoo et al., 2024). This method allows scholars to upload research documents, datasets, or literature reviews to the model. By instructing the model to refer to specific sections of these uploaded resources, scholars can ensure that the generated text accurately reflects prior research and findings. For example, a prompt could specify, "Summarize the key findings of this study using the attached data tables and prior literature summaries." However, in this context, scholars must ensure that copyrighted materials and human respondent data are not used to train public generative AI models to avoid potential legal and ethical issues.[4]

[1]https://www.unicef.org/parenting/child-development/your-toddlers-developmental-milestones-2-years.

[2]For example: https://docs.anthropic.com/en/prompt-library/library.

[3]For example: https://community.openai.com/c/prompting/8.

[4]To our knowledge, no publicly available generative AI models (e.g., ChatGPT, Claude, Gemini) uses user-uploaded content to train and improve its models according to their Frequently Asked Questions pages.

All models are wrong, but some are useful.

Tomaino et al. (2025) underscore one limitation of their study: a lack of testing pre/pilot tests and experimental designs using synthetic datasets, also known as silicon samples (Sarstedt et al., 2024). Silicon samples, i.e., LLM-generated participants designed to mimic human-like responses, can reduce costs and time, especially during pretesting and pilot testing for qualitative and quantitative studies, by enabling scholars to refine stimuli or study designs without recruiting human participants (Li et al., 2024; Sarstedt et al., 2024). As LLMs become more sophisticated, more research is needed to understand when these models accurately mimic human attitudes and behaviors, potentially increasing methodological rigor.

Recent literature shows that LLMs can mimic certain human behaviors in survey research (Argyle et al., 2023; Dillion et al., 2023). For example, LLMs can replicate human responses in structured scenarios like economic games or moral judgment tasks (Dillion et al., 2023). This capability allows scholars to simulate human reactions to different scenarios. Researchers can potentially evaluate how different population segments respond to certain scenarios by generating synthetic responses from these segments. While qualitative assessments of LLMs are generally reliable, these models face more challenges in quantitative studies, particularly in understanding psychological mechanisms. For instance, while LLMs can simulate average responses effectively, they struggle with capturing the variability and individual differences essential to many quantitative studies, often producing "homogenized" responses (Abdurahman et al., 2024). Furthermore, LLMs may not fully grasp the underlying cognitive and emotional processes driving human behavior, which can impact the validity of quantitative findings.

Currently, these limitations make using silicon samples to generate responses to experimental designs for underlying psychological processes generally unsuccessful. They can lead to results that either do not confirm or contradict those known from human samples (Yoo et al., 2024). While silicon samples offer efficiency and scalability, they should be used alongside human samples in qualitative and quantitative research to ensure external validity and capture the full spectrum of psychological diversity. Customized LLMs, trained on a library of prior studies and their results, could lead to significant performance gains in this respect.

Automation is to time what compound interest is to money.

Research is usually a group effort that involves an initial investment in tedious tasks, from stimulus design to coding and manual annotation. In some cases, these tasks are given to junior team members (we all remember our times as PhD students), while in others, they are outsourced to research assistants, either offline or online, using platforms like Fiverr or MTurk. LLMs provide a new avenue in this context.

In stimulus design, LLMs can be useful in creating realistic-sounding brand names or visual stimuli in static (picture) or dynamic (video) formats. For example, these models can quickly generate simple and complex product packaging designs to examine consumer perceptions (Sarstedt et al., 2024). This opens possibilities for developing tailored stimuli that align closely with a study's goals.

Looking at coding tasks, LLMs can conduct sentiment analysis to understand emotional patterns in qualitative data, such as open-ended survey responses or social media content (Juroš et al., 2024). LLMs excel at detecting subtle emotions and underlying sentiments, which are often challenging to capture through traditional analysis methods. By analyzing textual data at scale, scholars can identify nuanced emotional trends and patterns, improving the accuracy and richness of insights in studies on consumer behavior. They can also be used to classify textual data into complex groups. For example, Stäbler and Haenlein (2024) use ChatGPT to classify 15,900 article titles according to their degrees of descriptiveness, complexity, and creativity.

Finally, LLMs can streamline the manual text annotation and coding processes by automating content analysis (Gilardi et al., 2023). This automation can reduce human biases and enhance consistency, making it easier to analyze large volumes of qualitative data accurately. LLMs have also been shown to either meet or outperform natural language processing models, such as Bidirectional Encoder Representations from Transformers (BERT) models, when automating text coding for typicality, that is, assessing whether content accurately represents its intended meaning (Le Mens et al., 2023). This automation allows behavioral researchers more time to focus on interpreting the results and drawing meaningful insights rather than spending extensive resources on manual coding. By applying LLM-driven text annotation, researchers can achieve reliable, consistent coding results across complex datasets, which can benefit longitudinal studies or large-scale social media analyses.

Any sufficiently advanced technology is indistinguishable from magic.

The examples given above and the work of Tomaino et al. (2025) are just starting points for integrating LLMs, like ChatGPT, into the research process. In our experience, the best way to think about these tools is to compare them to smart undergraduate students interested in research but with limited experience in the field. Sometimes, one may be surprised by their performance or level of creativity, while in other cases tasks considered simple can take a surprisingly high amount of effort to get them right. As with real-life research assistants, identifying potential areas of support requires trial and error and differs from person to person. One of the authors of this comment regularly uses the voice feature in the ChatGPT app to structure new research projects, get feedback, and refine ideas before discussing them with co-authors – similar to having a phone call with a patient and senior colleague. Another one has failed miserably with the simple task of getting the same tool to reformat a reference list in the style of a specific journal. The point is that the magic of LLMs is part of our reality and is here to stay. Like the introduction of computers that made writing manuscripts and running analyses more efficient (some may still remember when manuscripts were written on typewriters and submitted as paper printouts), LLMs will have a profound and lasting impact on our work. O brave new world that has such tools in it!

HOW DOES IT FEEL TO BE A CYBORG BEHAVIORAL SCIENTIST?

Stacy Wood

Acknowledgments

I'd like to thank Joel Huber and Vicki Morwitz for their good counsel and contributions to this work. I also thank my fellow editors at *JCR*, Bernd Schmitt, June Cotte, Andrew Stephen, and Markus Giesler, for our many conversations on the role of AI in research, which helped shape my thinking.

Introduction

What I really wanted to know from Tomaino et al. (2025) and their experience as cyborg behavioral scientists is what it all *felt* like. Of course, their purpose was to use AI as much as possible to conduct a viable consumer research project and objectively assess the outcome—how good was the work AI produced? But I wonder if the more pressing question—the question that has more insight into our future—is what the experience of off-loading as many tasks as possible to generative AI was like for Tomaino, Cooke, and Hoover as scholars? What did it *feel* like?

Now, an alternative important commentary on this paper would be to examine how the work did or did not fully utilize currently available AI tools and to consider how the specific path these researchers took might be very different from another project, thus limiting what we can learn about AI's current abilities now. And to consider how journals could or should regulate the use of AI in marketing research through norms, practices, rules, and restrictions. Luckily for all of us. Vicki Morwitz takes this approach in the current issue.

But, for me, I am stuck on the feelings. Tomaino et al. (2025) diligently and clinically report each destination on their AI journey; but they offer an emotionally cool travelogue largely lacking personal impressions. I see this as the primary and critical missing piece of their report. Like going to Italy and having a room with a view but offering no lavish, revelatory descriptions. Completely unsatisfying. Yet, some may ask, why do the feelings matter?

Feelings matter because all utilitarian discussions of what AI can do right now will be largely meaningless a year from now as the development of AI abilities far outpaces our imaginings. What will not change is how the researcher's personal experience of using AI to do certain aspects of research will create, as I see it, *three early movements* that will shape future adoption of AI in research: (1) norms and ethics, (2) scholastic identity, and (3) status and compensation. An expanded view of Tomaino et al. and their test case could have offered early insights into the likely direction of those movements. In this commentary, I outline these three movements and offer some examples of researcher experience and what they might mean for the future of adoption of AI in the social sciences.

A movement of norms and ethics

There are many researchers grappling with the ethics of using AI in research from different philosophical vantage points. However, from a pragmatic approach, it may be valuable to first look to what norms develop around the early use of AI. While it is not my aim to open a debate between ethical relativism (where cultural norms drive ethical stances) and ethical objectivism (where universal ethical truths are held), ethics often develop in research by scientists trying new things, developing normative behaviors, and then elevating and codifying some norms as ethically grounded processes. Moreover, germane to the dialogue here, many norms begin with early experiences and the simple ways that first users feel that one path is better or worse than the other.

For example, consider the use of AI in generating hypotheses. To ask generative AI to suggest hypotheses for consumer psychology requires the researcher to choose a corpus or the source material. Right now, free access realistically means that the researcher uses something like the ChatGPT corpus, a mammoth body of information from websites, books, articles, forums, and other unspecified material sourced from the internet. Although carefully crafted prompts can help narrow the material sourced, it's still a wild landscape. But, with time or the ability to pay for it, one could customize one's source material to use only academic papers from open-source consumer-related disciplines, marketing industry publications, and certain consumer forums with the aim of developing a "better" set of hypotheses. Publishers might someday make and sell access to corpora crafted from their copyright-protected material. The hypotheses with a custom corpus might be seen as better weighted on cutting-edge consumer psychology, as less likely to be contaminated by bias from non-filtered internet sources, or as more specific to the kinds of questions that consumer psychologists ask. Or could they be seen as *worse* because they are wrought from a small genetic pool, iterating an ever-increasing minutia? What feels right? Given uncertainty in utilitarian value (knowing what's right), many consumer psychologists will rely on their feelings to make the choice—as many consumer psychologists have shown! If scholars feel the dirty chaos of the unfiltered internet to be toxic, then this feeling may spur the use of a customized corpus. If scholars feel the use of a small conceptual pool evokes distaste at inbreeding and navel-gazing, then this feeling may spur the use of a free, unfiltered corpus. Tomaino et al. chose to use an unfiltered corpus—how did that feel? Worrisome? Exhilarating? If we knew, we might have some sense of how norms around selecting a corpus for hypothesis development might emerge and the downstream effect on the adoption of generative AI (specifically free versions) for this task. Or, worse yet, did the use of generative AI for hypothesis generation evoke guilt or detachment? Yikes! If using artificial intelligence to do the one hardest job of research—to ask a good question—feels artificial to scholars, we might give it a pass altogether.

Hypothesis generation is just one example. For each step on the journey Tomaino et al. took with AI, we might ask about feelings or try to infer them. Was using AI to guide data analysis a real disappointment? Was it disappointing only when it was wrong? Was using AI to write the paper a relief or embarrassing? Would we feel angry or ambivalent if a reviewer used AI to generate critiques of our paper if the critiques were justified? In each case, feelings commonly experienced throughout a community will emerge as norms. Norms that seem important will likely lead to ethical views and subsequent rules. Scientific discovery of what *can* be done will always precede scientific consideration of what *should* be done. Through feelings in early practice, we can glean insights into the direction that AI use might first take in research and how it might evolve, perhaps even independent of subsequent utilitarian advances in the accuracy of the technology.

A movement of scholastic identity

One emotional genre that I was particularly eager to hear about in the experience of Tomaino et al. was their satisfaction. Did their project feel worthwhile? I do not mean as an experiment into the process and outcome of using AI, but in the actual research they produced. In other words, if they were constrained to use this method of producing research for the rest of their careers, would they find that satisfying? Would they still want to be consumer researchers?

Obviously, I am waving my hands around a precise conceptual definition of satisfaction. In many ways, what I am trying to assess is how the experience of using AI fits with one's identity as a consumer researcher. Some may argue that AI does not touch on identity at all—it's just a tool, so how could it matter as long as it is accurate? Others will say it changes the whole ballgame.

Researchers have been adopting more and more modern tools in research since time immemorial, or at least since the advent of scientific journals. Did researchers feel more like scholars when they had to punch cards to analyze data or find their way to the musty lower levels of the university library to track down a reference in a *book*? Did we feel differently about our jobs and ourselves when "doing some initial research" didn't mean Googling but was instead dropping by the office of whichever colleague had the most encyclopedic memory?

AI is probably not a monolithic hammer falling on the scholarly identity. It would be interesting to know if using AI for different tasks feels more (or less) fun, engaging, intriguing, free, creative, thorough, discriminating, relevant, or impactful, characteristics that often define our identities as researchers. If early users like Tomaino et al. spoke to their experiences of such feelings while engaging in any of the tasks where they ceded control to AI, we might learn where AI-enabled tasks run counter to the academic raisond'être. Interestingly, as I typed that last sentence, I can see where my own presumptive feelings lie—the use of "ceded control" speaks to an individual who sees independence as a critical facet of identity.

I suspect many will feel that these identity-shaping tasks are just a subset of what AI can do for the researcher, and I agree. Perhaps it even differs by researcher! Some of us might feel that our core task is in coming up with research questions. Others may see their critical function as designing the perfect experiment, whatever the hypothesis. In our recent *JCR* editorial (Schmitt et al., 2024), my fellow editors and I thought about a future in which each task in the research and review process might be a continuum where reliance on AI shifts from light assistance to complete responsibility. Only as researchers increasingly use AI will we see what tasks feel too central to our purpose to be offloaded to "an assistant," be that assistant human or not. As Tomaino et al. note, the robots currently need human direction. Perhaps those scholars with a long experience of "directing" research by using their highly capable research assistants, pre- and post-docs, and doctoral students to originate and execute every step of a project will find it least upsetting to hand it all over to AI.

A movement of status and compensation

The last sentence of the last paragraph was not purely snark. It was a snarky way to say that the feelings created by AI use will have great influence on how scholars build status and earn compensation in an AI-enabled future. In the future some may see AI use as a golden tool bequeathed by a well-funded university to its most illustrious faculty. Wherever AI tools are costly or restricted, their use will convey aspects of status. I wonder if in these cases we will see more value in both the AI tool and the researcher who uses it. An expensive tool suggests an expert craftsman.

It is interesting to know how early researchers using free tools or the simplest form of AI tools feel. Are they slightly ashamed that they do not have the resources (in terms of either money, time, or ability) to use something more sophisticated? Or is any use of AI currently cutting-edge in the minds of the users and the reviewers? This may be affected by our position in the academy as university-affiliated researchers. In the minds of academics, is AI associated with high-tech pedagogy or cheating undergrads? Is the primary underlying feeling enthusiasm or skepticism?

Yet, AI has the promise to be a great equalizer in academia by helping scholars in places where knowledgeable colleagues, big libraries, well-kitted labs, and copious research participants are not available. It can help scholars around the world for whom the language of the top journals (typically English now) is not a first language or even a spoken language. For these scholars, will the creation of research through nontraditional methods give rise to a new type of bias? Will their work be prejudged as 'factory-made' or 'inauthentic' because of the use of AI? Will these biases be exacerbated in communities where existing scholars have the most to lose from a new population of scholars who can produce high-quality research faster? Won't we have to argue that doing faster or easier research *must* be lower quality or be forced into a new gauge of productivity ourselves? In our recent editorial (Schmitt et al., 2024), we argued that a tiered system like factory farms versus boutique/family farms could emerge where "hand-made" research was seen as slower, more authentic, and more expensive, though not necessarily better than "mass-produced" or "automated" research.

How researchers feel about themselves and others who use AI at different levels will impact the likelihood of articles being accepted and subsequently our compensation. Now we labor for a fairly ambiguous standard of research performance. It's hard to know exactly what number of and type of articles we must publish to be in good standing. What number must we publish to be elite? What is the quality of an article and how is it judged? Should we be working to impress the academics in our department or must we also impress the university stakeholders who are increasingly skeptical about the value of what we produce? Can we simply publish at the same rate, but use AI to speed along our research and create more work-life balance? Can we use AI to focus on the important parts of research and avoid the drudgery? If anyone with AI can do decent research, will status and compensation be created by those who can best translate and advocate for their findings? These are the discussions that will shape our future in a real way.

Here, we can again use feelings to see movements emerge. As scholars, do we feel fear of being replaced in any of the current AI capabilities? Or do we feel the opposite—an increase in confidence and hope? Wherever we feel fear, there is likely a hurdle that we will throw up to slow down adoption. Wherever we feel a release from fear or the surge of hope, there is likely a place where we will work to facilitate adoption. I wonder how Tomaino et al. felt when they finished and looked at what they had created—fear for their own personal future or hope? I would like to sit down with them, order a whiskey sour, and hear the real story.

Conclusion

The experience documented by Tomaino et al. (2025) was interesting and will surely generate much discussion. I hope one discussion, as more scholars use more AI tools, is how it felt to do research differently and what those feelings might mean for the pattern of adoption of AI in social science research. We can look to books like *The Last Human Job* by Allison Pugh (2024) or investigate domains where the robotic has been claimed as human, like The Alternative Limb Project (www.alternativelimbproject.org). With honest self-reflection of how we feel and what it might mean, we will test, adapt, and adopt innovations that best improve our work and ourselves.

CYBORG SCHOLARS AND CONSUMER RESEARCH

Vicki G. Morwitz

Tomaino et al. (2025) provide an intriguing illustration of how artificial intelligence (AI), and particularly generative AI, can support experimental consumer behavior researchers in developing research ideas, theorizing, designing, executing, and analyzing studies, as well as writing for scholarly journals like the *Journal of Consumer Psychology*. While they offer insights into how generative AI might assist scholars, they mostly caution against over-reliance, citing limitations in AI capabilities and access to scholarly content, at least at this moment in time. They also warn editors and others involved in our academic journals about potential concerns surrounding AI use by reviewers. The accompanying comments address other ways that AI may influence scholarship that may have a fundamental impact on our field and as Stacy

Wood emphasizes, it affects scholars' feelings about, and perceptions of, their accomplishment.

Tomaino et al. took a unique approach to assess the value of generative AI for experimental consumer research scholars – they decided to delegate almost all tasks in the development and writing of a single paper to AI. I have not seen this done before. The approach was an interesting choice that provided a comprehensive look at how a scholar might consider using AI throughout the research process for a given project. One benefit of this approach was that it allows us to read and evaluate an embedded largely AI-generated manuscript on an interesting topic. It is also helpful for identifying areas where, in its current state, experimental consumer researchers can benefit from AI's use and where caution is needed.

However, this approach may underestimate AI's value for consumer researchers because, of course, not every paper uses all aspects and tools of scholarship. Given the topic selected by generative AI together with the Tomaino et al. team, some of the more rapidly evolving and potentially promising uses of generative AI were not evaluated as they were not relevant for the selected research topic.

An alternative approach for evaluating the use of generative AI in consumer research is to assess what we know from existing scholarship regarding how well it performs at different tasks that are employed in related fields, based on multiple tests involving those tasks. To do this, we can turn to evaluations already conducted by other scholars regarding its use for research in psychology (Abdurahman et al., 2024), marketing (Peres et al., 2023), and more broadly social science (Bail, 2024). These papers identify several additional ways that generative AI may be beneficial to consumer researchers beyond those examined in Tomaino et al. and identify additional areas where caution is warranted. I summarize relevant problems below and conclude with some implications for journals.

AI-related opportunities for consumer researchers

Stimuli development

Tomaino et al. used generative AI to generate ethical and non-ethical brand positioning statements as experimental stimuli. They concluded that this was one of the areas where generative AI performed well, and they noted this was consistent with prior findings (Sarstedt et al., 2024). Given the nature of the research topic they explored, the stimuli that generative AI created were relatively simple and were textual in nature.

As discussed in Bail (2024), generative AI can be used by consumer researchers to develop a wider range of stimuli. These stimuli include textual statements but also images, music, or even complex multimedia elements that operationalize psychological constructs. Consumer researchers can use generative AI to create nuanced, contextually relevant stimuli, including vignettes, images, music, and videos provided while avoiding potential copyright issues. AI-generated images also help avoid the use of real individuals' images and related privacy and ethical concerns. Currently, such stimuli are only accessible to researchers with the skills and budget needed to create them, but AI would open up these options to a broader set of scholars, including scholars who want to develop stimuli related to countries and cultures different from their own.

Studying group behavior

Tomaino et al. (2025), like most, but not all, consumer research papers, focused on individual reactions to marketing stimuli. However, much consumption happens in small groups like dyads, peers, and households and sometimes in larger groups (segments, communities, tribes, organizations) (MacInnis et al., 2020). Yet, group consumption decisions have received limited study in consumer research (for an example of an exception, see Dzhogleva & Lamberton, 2014).

Bail (2024) discusses how generative AI could be used to study group-based decision-making processes in social science contexts. He notes that it can be challenging and costly to recruit and assemble groups of participants and that generative AI might be able to simulate group members and help approximate group dynamics, processes, and behavior.

For example, generative AI could be used to examine how consumers interact and make consumption-related decisions when others dynamically express views that agree or disagree with their own. Currently, such research would involve human confederates assigned by the researcher to play out specific roles, but this can become challenging when researchers want to study larger groups or longer interactions. In cases like these, AI agents could be created and trained to interact with human participants in controlled and realistic consumption-related environments to investigate how group dynamics shape consumer decision-making and behavior in complex social interactions. For example, research could examine how human participants react dynamically when needing to make a group decision, where members of the group have different preferences and constraints. One such example could involve asking respondents to imagine attending a conference dinner where the job is to choose dinners and discussion topics from a set of options. The research could reveal how participants navigate (a) larger group dynamics: the groups could be created to include people from different backgrounds and who have different taste preferences; (b) potential disagreement: group members may have conflicting dietary restrictions and budget considerations; and (c) resolution: how group influence and majority and minority preferences and constraints influence the dynamics of how decisions are made.

Contexts that are difficult or unethical to study with human participants

Bail (2024) also discusses how generative AI can be used to study topics that are difficult to study in real life. AI might be helpful when researchers aim to study topics such as high-risk consumption situations and activities, the consumer behavior of illegal substances and activities, and interacting with others in illegal market activities (Grossmann et al., 2023).

For example, studying the dynamics of emotions and behavior in gambling is particularly challenging in a laboratory setting due to the complexity, the stochastic elements, and the longitudinal nature of real-world gambling environments. Researchers could use generative AI to create dynamic, interactive simulations that mimic gambling contexts. For example, a generative AI-created e-sport betting platform could simulate realistic betting scenarios where participants decide whether to place bets, experience wins and losses, and read and respond to commentaries written by other simulated users on the platform. This approach would allow researchers to observe how the nuanced interplay of emotions and behaviors evolves in real time to influence subsequent gambling decisions, in ways that would be difficult to replicate in static lab experiments and in ways that avoid the ethical and privacy concerns of using real-world gambling contexts.

Analyzing and coding text

Although AI is frequently used for text analysis, Tomaino et al. note that they did not employ it because it was not relevant for their research topic. This is a major area of AI application, with our quantitative colleagues in marketing increasingly adopting it (e.g., Netzer et al., 2019), and where there is significant potential for expanded use by consumer researchers. For example, consumer researchers could use generative AI to analyze social media posts, consumer reviews, and open-ended survey responses. It provides an efficient and quick way to analyze textual content, even for large data sets.

When analyzing text from research participants or secondary data, consumer researchers typically use human coders to code and categorize topics mentioned. While Tomaino et al. caution that generative AI may have limitations for coding certain constructs, such as creativity, evidence from multiple contexts shows that it can perform at similar levels to, and in some cases surpass, human coders. For example, Gilardi et al. (2023) found that generative AI outperformed mTurk coders at classifying tweets by topic, sentiment, and narrative framing. Mellon et al. (2024) found that generative AI was as accurate as highly trained human coders when analyzing statements about British elections. Overall, while there is cause for concern about accuracy, at least for some constructs, AI would allow for the coding of text at unprecedented scale and speed.

Synthetic data

Tomaino et al. opted to obtain data from humans and avoided AI-generated synthetic data. They pointed to literature that suggests synthetic data may be appropriate for piloting research, but that data from humans should be used for actual tests (Abdurahman et al., 2024).

There is some evidence that generative AI can closely capture responses from nationally representative survey respondents, including respondents from a range of different demographic groups (Argyle et al., 2023), including groups that might otherwise be difficult to obtain. There is also evidence in consumer contexts that AI can generate realistic survey results (Grossmann et al., 2023). Bail (2024) discusses how AI-generated samples can represent diverse groups, more diverse than our commonly used student samples and online panels. AI also offers the advantage of administering longer studies, since AI does not get tired, but humans do (Grossmann et al., 2023).

At the same time, caution is warranted (for a full discussion of challenges and opportunities see Sarstedt et al., 2024). While AI samples can match mean human estimates, they have lower variance, sometimes exaggerate extreme responses (Bisbee et al., 2023), and even display affirmation biases in binary yes/no questions (Dentella et al., 2023). GenAI modules are also trained on Western, WEIRD samples (Henrich et al., 2010), so they will not reflect other, more diverse samples. For these reasons, Tomanio et al., as well as Bail (2024), Grossmann et al. (2023), and Sarstedt et al. (2024) conclude that although synthetic samples may not yet be sufficient in themselves to test research hypotheses, they could be useful in pilots and pretests, particularly for participant groups that are difficult to recruit in real life, and findings from synthetic data could later be confirmed with human participants.

Additional cautions for consumer researchers regarding the use of generative AI in their research

Tomaino et al.(2025) did a good job identifying many areas where consumer researchers need to be cautious when using AI in conducting research. However, the literature has mentioned a few other areas where caution is warranted, which I summarize below.

Biases in LLM models

Since AI is primarily trained on data created by humans, AI responses often reflect human biases. This is likely especially true for generative AI since these models are usually trained on data from the internet, where biases have proliferated (Peres et al., 2023). For example, research has shown that LLMs may exhibit biases against women and racial minorities (Kotek et al., 2024). For consumer researchers, these biases may pose challenges as they inadvertently lead to the creation of biased stimuli, questions, or data.

Reproducibility, ethical, and privacy-related concerns

Bail (2024) discusses concerns related to reproducibility. Since generative AI models operate probabilistically, not deterministically, their outputs can vary from one occasion to the next, even given the same input and prompts. In addition, many generative AI models are closed source and are updated frequently. Minor differences in the wording used for prompts can also produce different outputs, creating additional challenges. Altogether, this means that when AI is used in the research process, researchers cannot rely on previous versions to replicate studies over time. These issues pose serious challenges for establishing reproducible findings.

Bail (2024) also discusses concerns related to ethics and privacy. For example, if AI is used to impersonate other people in a study, do participants need to be informed that they may be interacting with or responding to AI rather than a human? Bail (2024) suggests that consent documents disclose that participants may interact with AI. Research has shown that most participants given such consent were uncertain if they had interacted with a human or with AI (Allamong et al., 2023).

Additionally, there are privacy issues when researchers use AI to analyze participant data, such as interview transcripts or open-ended responses. Entering participant data into AI could violate rules of institutional review boards, confidentiality agreements, and legal requirements. These ethical and privacy risks underscore the need for careful consideration and clear guidelines from universities and journals regarding ways in which generative AI may be incorporated into scholarly research.

Implications for journals

As the previous editor-in-chief of the *Journal of the Association for Consumer Research*, I have observed frequent AI-related questions at meet the editor sessions at academic conferences. Potential authors seek clarity on allowable AI use and worry about reviewers' reliance on it.

As generative AI rapidly evolves, so must our journal policies, which are generally set by journal policy boards. The current AI policies at many academic journals state that AI cannot be a co-author, authors are responsible for all content of their papers, and some journals require that the use of AI in the research process should be declared, with exceptions sometimes provided for tasks such as copy editing. Some journal policies go further and explicitly limit AI use for the writing of manuscripts and the creation of scholarly content. As generative AI becomes increasingly integrated into consumer research, academic journal policies must evolve to ensure transparency, rigor, and ethical standards. Policies will need to require clear disclosure of when and how AI was used in the scholarship, specifying its role in idea generation and development, stimuli development, data generation, analysis, and manuscript preparation. Journal policies will need to address issues of originality and authorship, defining how contributions involving AI align with intellectual accountability. Guidelines might also require that authors are responsible for AI outputs. Additionally, journal policies, as well as institutional review boards at universities, must consider ethical implications, such as data privacy when using AI and informed consent when human participants interact with AI agents.

The use of generative AI by journal reviewers introduces additional challenges for review teams. While generative AI can assist reviewers who seek help with improving their writing exposition or by seeking to better understand a method used by the authors, its use by reviewers also raises serious concerns, beginning with but not limited to copyright and intellectual property issues associated with uploading a manuscript one is reviewing into an AI system. Reviewers' reliance on AI may lead to superficial and homogeneous evaluations compared with human peer teams. Journal review guidelines should emphasize that AI should augment, not replace, human judgment and encourage reviewers to critically validate any allowable AI-generated insights.

By addressing these considerations regarding the use of AI by authors and reviewers, journals can support the responsible use of generative AI while preserving the integrity of academic scholarship and the intellectual integrity of peer evaluation.

Conclusion

In conclusion, I agree with Tomaino et al. that while generative AI offers a powerful new tool set for consumer researchers, it also raises critical concerns that must be carefully managed. In addition to the benefits and costs already discussed by Tomaino et al. and in the comments that follow, there are larger economic, cultural, and societal costs that should also be considered. Economically, reliance on AI tools may exacerbate inequalities, as only well-funded researchers may have access to the most advanced systems, leaving scholars with fewer resources at a disadvantage. As Wood discusses in her comment, AI may change cultural aspects of the research process, including how scholars feel about their own research contributions, and how scholars are judged by the research community depending on their use of AI. Finally, consumer researchers should be mindful of the environmental impact associated with AI tools. Bail (2024) states that a single large language model can generate as much carbon dioxide as the lifetime emissions of five cars due to the immense computational power required. As sustainability becomes increasingly important, researchers should weigh the benefits of AI against its environmental impact.

As Tomaino et al. and others have illustrated, generative AI can aid researchers in many aspects of the research process, from developing realistic experimental stimuli to analyzing vast amounts of text with unprecedented speed and scale. However, as AI technology evolves, consumer researchers must remain vigilant, balancing AI's potential benefits with a commitment to ethical standards, privacy protections, and environmental considerations. Guidelines from journals, universities, and ethics boards will be essential in navigating these complexities, ensuring that the use of generative AI contributes responsibly to the advancement of consumer research.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Paul Andrew Blythe https://orcid.org/0009-0008-7685-4447
Christopher Kulis https://orcid.org/0009-0001-0287-9616
A. Peter McGraw https://orcid.org/0000-0002-0373-1314
Stacy Wood https://orcid.org/0000-0001-5256-1772

REFERENCES

Abdurahman, S., Atari, M., Karimi-Malekabadi, F., Xue, M. J., Trager, J., Park, P. S., Golazizian, P., Omrani, A., & Dehghani, M. (2024). Perils and opportunities in using large language models in psychological research. *PNAS Nexus*, *3*(7), 1–14. https://doi.org/10.1093/pnasnexus/pgae245

Allamong, M. B., Trexler, A., Alqabandi, F., Tucker, T., Bail, C., Hillygus, D. S., & Volfovsky, A. (2023). Outnumbered online: An experiment on Partisan imbalance in a dynamic social media environment. OSF [Preprints], 10.

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Arora, N., Chakraborty, I., & Nishimura, Y. (2024). EXPRESS: AI-human hybrids for marketing research: Leveraging LLMs as collaborators. *Journal of Marketing*, *20*, 6529. https://doi.org/10.1177/00222429241276529

Bail, C. A. (2024). Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, *121*(21), e2314021121.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2023). Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*, *20*, 1–16.

Brittain, S. (2023). New York lawyers sanctioned for using fake ChatGPT cases in legal brief. *Reuters*. https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

Dentella, V., Günther, F., & Leivada, E. (2023). Systematic testing of three language models reveals low language accuracy, absence of response stability, and a yes-response bias. *Proceedings of the National Academy of Sciences*, *120*(51), e2309583120.

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

Dzhogleva, H., & Lamberton, C. P. (2014). Should birds of a feather flock together? Understanding self-control decisions in dyads. *Journal of Consumer Research*, *41*(2), 361–380.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30), e2305016120. https://doi.org/10.1073/pnas.2305016120

Goli, A., & Singh, A. (2024). Frontiers: Can large language models capture human preferences? *Marketing Science*, *43*(4), 709–722. https://doi.org/10.1287/mksc.2023.0306

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, *380*(6650), 1108–1109.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Juroš, J., Majer, L., & Šnajder, J. (2024). LLMs for targeted sentiment in news headlines: Exploring different levels of prompt prescriptiveness. *arXiv*. 10.48550/arXiv.2403.00418.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2023). Large language models are zero-shot Reasoners. *arXiv* https://doi.org/10.48550/arXiv.2205.11916

Kotek, H., Sun, D. Q., Xiu, Z., Bowler, M., & Klein, C. (2024). Protected group bias and stereotypes in large language models. arXiv preprint, arXiv:2403.14727.

Le Mens, G., Kovács, B., Hannan, M. T., & Pros, G. (2023). Uncovering the semantics of concepts using GPT-4. *Proceedings of the National Academy of Sciences*, *120*(49), e2309350120. https://doi.org/10.1073/pnas.2309350120

Li, P., Castelo, N., Katona, Z., & Sarvary, M. (2024). Frontiers: Determining the validity of large language models for automated perceptual analysis. *Marketing Science*, *43*(2), 254–266. https://doi.org/10.1287/mksc.2023.0454

MacInnis, D. J., Morwitz, V. G., Botti, S., Hoffman, D. L., Kozinets, R. V., Lehmann, D. R., Lynch, J. G., Jr., & Pechmann, C. (2020). Creating boundary-breaking, marketing-relevant consumer research. *Journal of Marketing*, *84*(2), 1–23.

Marr, B. (2023). The difference between generative AI and traditional AI - an easy explanation for anyone. *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/07/24/the-difference-between-generative-ai-and-traditional-ai-an-easy-explanation-for-anyone/

Mellon, J., Bailey, J., Scott, R., Breckwoldt, J., Miori, M., & Schmedeman, P. (2024). Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale. *Research & Politics*, *11*(1), 20531680241231468.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, *56*(6), 960–980.

Okamura, K. (2019). Interdisciplinarity revisited: Evidence for research impact and dynamism. *Palgrave Communications*, *5*, 141. https://doi.org/10.1057/s41599-019-0352-4

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGpt and beyond: How generative artificial intelligence my affect research, teaching, and practice. *International Journal of Research in Marketing*, *40*(2), 269–275.

Pollack, J., Helm, J., & Adler, D. (2018). What is the iron triangle, and how has it changed? *International Journal of Managing Projects in Business*, *11*, 527–547. https://doi.org/10.1108/IJMPB-09-2017-0107

Pugh, A. J. (2024). *The last human job: The work of connecting in a disconnected world*. Princeton University Press.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models. *arXiv* https://doi.org/10.48550/arXiv.2402.07927

Sarstedt, M., Adler, S. J., Rau, L., & Schmitt, B. (2024). Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines. *Psychology & Marketing*, *41*, 1254–1270. https://doi.org/10.1002/mar.21982

Schmitt, B., Cotte, J., & Giesler, M. (2024). Andrew T Stephen, Stacy wood, will we Be the last human editors of JCR? *Journal of Consumer Research*, *51*(3), 451–454. https://doi.org/10.1093/jcr/ucae053

Stäbler, S., & Haenlein, M. (2024). The unheard voice of marketing research: Breaking through to news and social media. *Journal of the Academy of Marketing Science*, *53*(1), 105–128. https://doi.org/10.1007/s11747-024-01038-5

Tomaino, G., Cooke, A. D. J., & Hoover, J. (2025). AI and the advent of the cyborg behavioral scientist. *Journal of Consumer Psychology*, *35*, 297–315. https://doi.org/10.1002/jcpy.1452

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models. *arXiv* https://doi.org/10.48550/arXiv.2201.11903

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*, 1036–1039. https://doi.org/10.1126/science.1136099

Yoo, K., Haenlein, M., & Hewett, K. (2024). A whole new world: charting unexplored territories in consumer research with generative artificial intelligence. (Marketing Science Institute Working Paper Series No. 24-123). https://thearf-org-unified-admin.s3.amazonaws.com/MSI_Report_24-123.pdf

How to cite this article: Blythe, P. A., Kulis, C., McGraw, A. P., Haenlein, M., Hewett, K., Yoo, K., Wood, S., Morwitz, V. G., & Huber, J. (2025). Comments on "AI and the advent of the cyborg behavioral scientist". *Journal of Consumer Psychology*, *35*, 316–328. https://doi.org/10.1002/jcpy.1453