Ideation with Generative AI—in Consumer Research and Beyond

JULIAN DE FREITAS
GIDEON NAVE
STEFANO PUNTONI

The use of generative AI (genAI) in consumer research is rapidly evolving, with applications including synthetic data generation, data analysis, and more. However, their role in creative ideation—a cornerstone of consumer research—remains underexplored. Drawing on the human creativity literature, we propose that ideation with genAI is facilitated by its *productivity* and *semantic breadth*, which are psychologically analogous to the dual pathways of persistence and flexibility in human ideation. Further, we distinguish between the utility of genAI as a key ideator versus humans as key ideator, conceptualized through the genAI ideation roles of *Designer* and *Writer* and of *Interviewer* and *Actor*. While genAI excels in generating incremental improvements, its potential for groundbreaking innovation could be unlocked by leveraging its ability to prompt human creativity. This article advances the theoretical and practical understanding of genAI in ideation for consumer research, offering numerous practical guidelines for integrating generative AI into research while emphasizing human–AI collaboration to achieve radical insights.

Keywords: generative AI, large language models, ideation, creativity, research, human-AI collaboration

"The real act of discovery consists not in finding new lands but in seeing with new eyes."

—Marcel Proust

Julian De Freitas (jdefreitas@hbs.edu) is assistant professor of marketing, Harvard Business School, Soldiers Field, Morgan Hall 161, Boston, MA 02163, USA. Gideon Nave (gnave@wharton.upenn.edu) is the Carlos and Rosa de la Cruz associate professor of marketing, the Wharton School, University of Pennsylvania, 749 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6304, USA. Stefano Puntoni (puntoni@wharton.upenn.edu) is the Sebastian S. Kresge professor of marketing, the Wharton School, University of Pennsylvania, 760 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, USA. Please address correspondence to Julian De Freitas.

Editor: Bernd Schmitt

This article was invited by Editor Bernd Schmitt for our special section on GenAI and Consumer Research and did not go through the journal's standard peer-review process but was subject to an expedited review and revision process.

Consumer researchers increasingly leverage generative AI (genAI) at each stage of the research process—from ideation to literature review, hypothesis generation, experimental design, data acquisition, analysis, writing, and even reviewing. This article explores the value of genAI, with a focus on large language models (LLMs), for creative ideation in consumer research. Although our discussion and examples center on consumer research, our contribution is broadly relevant across disciplines, to both researchers and practitioners seeking to utilize LLMs for ideation.

Ideation, also referred to as divergent thinking or brainstorming, is the process of generating new concepts that satisfy a specific goal (Koestler 1964). The "unit" of ideation is the "idea," whereas the goal is to create ideas that satisfy certain properties—usually, "originality" and "appropriateness." *Originality* (or novelty) is a deviation from existing ideas. Ideas that deviate only incrementally from existing ideas are called "small" and usually involve combinations of existing ideas; for example, showing that auto-correct affects consumer confidence in text-based communication refines existing research on tech and consumer confidence, rather than offering a groundbreaking

 • Vol. 52 • 2025
https://doi.org/10.1093/jcr/ucaf012

new framework. Ideas that deviate dramatically are "big" and typically involve breakthroughs that go beyond existing ideas to introduce truly novel concepts; for example, arguing that the smartphone is like a pacifier involves a surprising conjunction of concepts that necessitates a new way of conceiving smartphones (Melumad and Pham 2020). Ideas are also assessed in terms of appropriateness (or feasibility), which refers to whether the idea is practical in solving the problem (Amabile 1982; Harvey and Berry 2023). Ideas can be original but inappropriate, as exemplified by early concepts of the "metaverse" that were once dismissed as bizarre and irrelevant. Ideas can also be appropriate but unoriginal, as when a researcher replicates an existing finding in a similar population.

Originality and appropriateness in consumer research are typically assessed semantically—either objectively (e.g., by calculating the semantic similarity of the idea relative to previous ideas in consumer research in an embedding space; Berger et al. 2022) or subjectively (e.g., via human judgments). In the General Discussion, we consider broader perspectives, such as incorporating other notions like worldly "impact." For now, it is enough to emphasize that both originality and appropriateness are important to consumer researchers, who seek to produce work that excels on these dimensions, and to the field as a whole, which aims to recognize and publish such work.

As a first step, we explore the core characteristics of LLMs that make them particularly effective ideation tools. We highlight how their functionality mirrors two distinct yet equally important pathways to creativity in human psychology—persistence and flexibility (De Dreu, Baas, and Nijstad 2008)—providing several practical interventions to increase their efficacy within these pathways. Building on this foundation, we draw inspiration from the concept of "levels of automation" in human–computer interaction, to explore how LLMs can serve as both idea generators and as catalysts of human ideation. We introduce a framework of metaphorical "ideation roles" illustrating the diverse functions that LLMs can assume in ideation.

LLM IDEATION CAPABILITIES

Algorithms were originally designed to automate repetitive and routine tasks. The proliferation of generative AI has defied this expectation, leveraging advanced algorithms to create new text, images, audio, or video content. The most popular and well-studied class of these models is LLMs, which generate text outputs in response to text prompts. Training these models relies on a hybrid learning approach called self-supervised learning, which combines elements of both unsupervised learning (detecting underlying patterns in huge data corpuses without human guidance) and supervised learning (generating their own training examples from those data). To illustrate, the training algorithm might take a sentence appearing in the text corpus like "Most mornings I have coffee," and treat the first part ("Most mornings I") as input used to predict its last part ("...have coffee"), in a supervised manner. After learning from trillions of such examples, the model can generate much longer, insightful responses to human prompts.

In this section, we detail two properties of LLMs that enable creative capabilities: *productivity* and *semantic breadth* (table 1). We argue that productivity is loosely analogous to the "persistence path," and semantic breadth to the "flexibility path" to creativity in human psychology (De Dreu et al. 2008; Nijstad et al. 2010); both have long been studied by consumer research interested in creativity (Hirschman 1980). Building on these analogies, we organize recent research linking each of these factors to ideation, consider their limitations, and explore ways to expand their potential. Because the field of LLM research is young and interdisciplinary, we draw on working papers and papers published in consumer research and fields tackling related problems from different vantage points. This literature primarily investigates the influence of LLMs on ideas generated by and evaluated by laypeople—unlike in consumer research, where the human creators are typically experts (e.g., university professors), as are the evaluators (e.g., reviewers or editors).

Productivity

Productivity is the capacity of LLMs to generate a large volume of outputs over a short time. This property aligns with the concept of "persistence" in human creativity—the sustained effort to explore ideas deeply. For example, when constraints are imposed (e.g., in the idea space), the search process becomes more focused and intensive. This, in turn, increases the likelihood of discovering novel and creative solutions, as persistence helps uncover possibilities that a broader, less focused and more shallow exploration might overlook (Boyd and Goldenberg 2013; Burroughs and Mick 2004; De Dreu et al. 2008; Goldenberg, Mazursky, and Solomon 1999; Mehta and Zhu 2016). Persistence involves inhibiting irrelevant or distracting directions of thoughts, focusing on a limited number of potential ideas, and then laboriously searching for ideas in the resulting subset, which is smaller and more manageable. An example of innovation processes relying on persistence is ideation templates, also known as "thinking inside the box" (Boyd and Goldenberg 2013; Moreau and Dahl 2005), where ideators apply a set of well-defined steps. For instance, they might deliberately remove a key component from an existing product concept to see if what remains can spark a fresh, innovative design (Goldenberg et al. 1999).

The speed and scalability of LLMs enables them to generate coherent ideas in natural language with exceptional efficiency. Furthermore, unlike sourcing ideas from

TABLE 1

PROPERTIES OF LLMs FOR IDEATION

LLM property	Productivity	Semantic breadth
Psychological analogue	Persistence	Flexibility
Explanation	Thanks to their computing power, LLMs can generate a large volume of ideas in a short amount of time.	Thanks to their vast and heterogeneous training data, LLMs can generate ideas spanning diverse semantic categories.
Phenomena	Originality increases as more ideas are generated.	Originality increases as ideas connect more distant knowledge domains.
Limits	Original ideas (unique concepts) eventually plateau after a certain number of ideas are generated.	Hallucinations arise, especially when dialing up stochasticity (aka temperature parameter); negative spillover effects on collective diversity.
Practical interventions	Fine-tuning, few-shot prompting, retrieval-augmented generation.	Prompt variation, hybrid prompting, chain of thought prompting, temperature parameter.

individuals or groups of humans, one can instantaneously and repeatedly query LLMs, without exhausting them, while also evading the administrative costs involved in coordinating large groups of people (Burton et al. 2024). Demonstrating the productivity of generative AI, a study estimated that incorporating text-to-image generative AI into the workflows of visual artists increased their creative productivity by 25% (Zhou and Lee 2024).

Analogous to persistence in humans, a study found that when prompting an LLM to generate ideas for the college students market, the number of original ideas rose as the LLM generated more of them (Meincke, Mollick, and Terwiesch 2024b). At the same time, LLMs are so productive that they allow one to quantify the limit of a persistence approach to ideation (Kornish and Ulrich 2011). In the study by Meincke et al. (2024a), the number of original additions plateaued after 500 ideas, indicating that the pool of ideas eventually became exhausted—at least for the specific prompt used.

Semantic Breadth

Semantic breadth refers to the capacity of LLMs to generate ideas spanning widely different concepts. Semantic breadth is akin to "flexibility" in human creativity, where connecting disparate knowledge categories promotes originality (De Dreu et al. 2008). For example, in the product domain, a cufflink made of bike chain parts is viewed as creative because "cufflinks" and "bikes" are conceptually distant, making their unlikely combination feel original and unexpected (Caprioli, Fuchs, and Van den Bergh 2023). This process depends on stored knowledge about different categories within long-term memory (Finke, Ward, and Smith 1996; Nijstad and Stroebe 2006), which enables surveying a broad range of content categories, easily switching between categories, and harnessing associations between remote ideas rather than close ones. It also depends on attending broadly (e.g., across product categories), as opposed to focusing narrowly (e.g., within a single product category), such as by adopting a more abstract rather than concrete construal level (Förster, Friedman, and Liberman 2004; Mehta, Zhu, and Cheema 2012). Processes that rely on flexibility are thought to be psychologically accompanied by an evaluation mechanism that monitors the idea's appropriateness, thereby ensuring that the ideation process progresses toward the intended goal.

The flexibility of LLMs is rooted in their training on vast, heterogeneous datasets consisting of trillions of words sourced from diverse contents across the internet (Brown et al. 2020). This enables them to pull from numerous domains and generate ideas across broad content categories. LLMs can probabilistically draw from pieces of knowledge in disparate domains, combining these sources to complete a prompt usefully. Some studies find that LLMs are as original as humans in everyday creative tasks that require flexibility across broad categories, such as generating creative product uses (Bellemare-Pepin et al. 2024; Hubert, Awa, and Zabelina 2024).

However, semantic breadth does not guarantee diversity—the dissimilarity between ideas necessary for exploring a broad solution space (Doshi and Hauser 2024; Meincke, Nave, and Terwiesch forthcoming). While even naïve use of LLMs may increase the peak originality of an individual consumer researcher's ideas, this can come at the cost of decreasing the diversity of ideas among a group of consumer researchers using LLMs, whose ideas may become more similar to each other. For instance, the aforementioned study on artist adoption of text-to-image generative AIs found that, although the peak originality of each artist's ideas increased, average originality of ideas decreased (Zhou and Lee 2024). Studies of creative writing (Doshi and Hauser 2024) and other creative challenges, such as repurposing everyday objects (Meincke et al. forthcoming) found similar patterns. Thus, LLMs generate many ideas that are individually creative but more similar to one another than ideas generated by humans.

These side-effects likely occur because LLMs are trained to predict which tokens (e.g., words or word parts, emojis, punctuation) are most probable, following a given sequence of tokens. Since probability estimations are constrained by the training data, LLMs can only generate patterns they have already been exposed to, favoring the generation of frequently co-occurring tokens over rarer ones. This phenomenon mirrors the echo-chambers created by recommendation algorithms, which amplify homogeneity by serving similar content to similar users (Fleder and Hosanagar 2009; Lee and Hosanagar 2019; Valenzuela et al. 2024).

Practical Interventions

Productivity. LLMs can achieve greater productivity not only by exhaustively generating ideas for a single prompt, but also by adopting a more focused approach. This involves narrowing down their attention to the specific task at hand and suppressing unrelated categories of knowledge—emphasizing persistence as a blend of concentrated focus and inhibition. One way of achieving this is *fine-tuning* LLMs on specialized data, such as a corpus of publications on a specific topic (e.g., branding, motivation) to specialize them for a particular application and ensure all ideas are narrowly confined to that domain. For example, a luxury brand that fine-tuned a generative AI model to generate ideas for new t-shirt designs produced more successful designs than those produced by humans, because the generative AI designs were more faithful to the visual identity of the brand (Moreau, Prandelli, and Schreier 2023; see also De Freitas and Ofek 2024). A second approach for consumer researchers to promote persistence is focusing LLMs with *few-shot prompting*, that is, including a sample of highly relevant ideas in the prompt (Meincke et al. 2024a). For example:

> <Base prompt>
>
> Generate new research ideas for a consumer behavior researcher interested in customer journeys. The best idea will be turned into a paper submitted to the Journal of Consumer Research, where the goal is to get it published. The ideas are just ideas. The paper need not necessarily be clearly feasible. Generate 30 ideas as 30 separate paragraphs.
>
> + Here are some well received ideas for inspiration: *<Good Ideas>*

LLMs only need few exemplars to produce more specialized ideas, perhaps because similarly specialized data already exist in their knowledge database (Solaiman and Dennison 2021). Third, consumer researchers can supplement LLMs with *retrieval-augmented generation*. This technique typically utilizes an API (a structured interface that allows the AI to gather information from external sources, like Semantic Scholar) to retrieve specialized knowledge to "augment" the existing prompt, before feeding the augmented prompt into the model. With that said, one can also prompt the LLM to behave like an API, for example:

> You are an expert in consumer behavior and AI-driven recommendations. Retrieve the most recent consumer research papers on consumer trust in AI-generated product recommendations and summarize their key findings. Then, use this retrieved information to generate insights on how brands can improve consumer trust in AI-driven recommendations. Structure your response into three sections: (1) Summary of recent research, (2) Practical implications for marketers, and (3) Future research directions.

Semantic Breadth. Semantic breadth may be increased through several approaches. One effective method is *prompt variation*. For example, an LLM produced more diverse product ideas for the college market when specifically prompted to think like Steve Jobs (e.g., "You are Steve Jobs looking to generate new product ideas. <base prompt>") compared to when given the base prompt or when prompted to utilize creativity tools recommended by the Harvard Business Review (Meincke et al. 2024b). This approach, known as "persona modifiers," directs the LLM to adopt a specific perspective, often enhancing originality—though identifying the most effective persona prompt typically requires trial and error. For instance, other tactics, such as offering to tip the model, pleading with it emotionally, or threatening to shut it off, did not increase idea diversity for the case at hand.

Another approach is *hybrid prompting*, where one generates smaller idea pools using different prompts and then combines these pools (in line with the flexible path), rather than using a single prompt to generate a vast number of ideas (in line with the persistence path; Meincke et al. 2024b). This approach is akin to certain methodologies of brainstorming in humans, where the originality of groups will peak if their members first work independently and then pool ideas, rather than work as a single team (Girotra, Terwiesch, and Ulrich 2010). Analogously, hybrid prompting in LLMs likely works by increasing the number of parallel paths toward a creative solution, thereby ultimately improving the ideas (Jeppesen and Lakhani 2010; Piezunka and Dahlander 2015). For instance, one can repeat the aforementioned "base prompt" 40 times in separate LLM sessions, thereby simulating 40 participants independently generating 30 ideas each (we recommend using an LLM API to do this expediently). Next, one can "team up" sessions into 10 groups of four "people" each, yielding 120 ideas per group. Each group is tasked with whittling down to the 10 best ideas as follows:

> You are part of a team tasked with individually generating new research ideas for a consumer behavior researcher interested in customer journeys... Each team member has already generated 30 ideas. Your group consists of four members, meaning you now have 120 total ideas to work with. The following ideas were generated by your team:
>
> <list of 120 ideas>

> From these 120 ideas, select your top 10 final ideas for the group. Each idea should have a name, followed by a description of 40-80 words. Number them sequentially. The name and idea should be separated by a colon.

By finally aggregating the 10 best ideas from each of the 10 groups, one has 100 ideas that have been sourced in a hybrid manner.

A third path for promoting diversity is manipulating how the LLM processes the prompt. This can be achieved via *chain of thought* prompting—asking the LLM to follow distinct steps in a specific order (Wei et al. 2022). For instance, unlike when one submits only the aforementioned "base prompt," with chain of thought prompting one can influence *how* these ideas are generated, thereby helping to ensure that the ideas are of higher quality in the first place. For example, a consumer researcher can supplement the base prompt with the following step-by-step instructions:

> . . .Follow these steps. Do each step, even if you think you do not need to. First generate a list of 30 ideas (short title only). Second, go through the list and determine whether the ideas are different and bold, modify the ideas as needed to make them bolder and more different. No two ideas should be the same. This is important! Next, give the ideas a name and combine each with a paper description. The name and idea are separated by a colon and followed by a description. The idea should be expressed as a paragraph of 40-80 words. Do this step by step!

Such chain of thought prompting has been shown to improve the diversity of ideas to near-human levels, relative to using a base prompt alone (Meincke et al. 2024b).

In addition to prompting interventions, consumer researchers can dial up (or down) the degree of stochasticity of LLMs output via a *temperature parameter*, where higher values produce more diverse and unpredictable responses, and lower values result in more focused and deterministic outputs. The temperature parameter cannot be directly set through regular prompting interfaces but is a system-level setting that must be configured in the model's API or model settings before generation. With that said, consumer researchers can simulate the effects of different temperature settings indirectly through strategic prompting. For example:

> <low temperature prompt>
>
> Provide the most direct and factual answer to the following question, avoiding any unnecessary details or variations.< base prompt>
>
> <medium temperature prompt>
>
> Provide a well-balanced and somewhat creative response while ensuring clarity and coherence. <base prompt>
>
> <high temperature prompt>
>
> Give me the most creative, unexpected, and outlandish response you can think of. Feel free to be unconventional! <base prompt>

While such prompting techniques are likely to produce the desired impact on idea originality, the correlation between the literal dialing of the temperature parameter and idea originality appears to be weak (Peeperkorn et al. 2024). Further, dialing up the temperature parameter yields not just slightly more original ideas but also more factually inaccurate statements or "hallucinations" (Peeperkorn et al. 2024). While hallucinations are less of a concern in ideation—where the goal is to generate just one excellent idea, even if at the cost of much nonsense—hallucinations still induce noise that must be filtered out when evaluating which idea to choose. This is challenging for humans, who often struggle to predict which ideas will succeed (Terwiesch and Ulrich 2023).

Finally, humans face a cognitive tradeoff between engaging in persistence-driven as opposed to flexibility-driven processes, since one involves attention on a task and inhibition of remote ideas, where the other involves diffuse attention and disinhibition of remote ideas. As such, enjoying both approaches requires switching between them. Via prompting, consumer researchers can easily switch between these approaches. For instance, to take a persistence approach, a consumer researcher can prompt the LLM with, "What are 100 reasons a consumer might choose a more expensive product over a cheaper one? List only emotional factors." Alternatively, a flexible approach would instruct the LLM to probe different corners of the solution space, e.g., ". . .Include emotional, functional, and social factors" or even ". . .and then explore interactions between these factors. Be creative!"

METAPHORICAL IDEATION ROLES

Beyond the concrete prompting strategies just reviewed, how can consumer researchers use LLMs in their ideation processes? Compared to how humans have historically approached ideation, LLMs turn ideation into a co-creation process between human and machine. We draw inspiration from the concept of "levels of automation" in partially automated systems (SAE 2021), which envisions a spectrum where either the AI or the human plays the most active role in a system, depending on its configuration (Agarwal et al. 2024). Likewise, we propose that either LLMs can be key ideators (where they are the source of ideas that humans then screen) or humans can be key ideators (where LLMs "pull out" ideas from humans).

This distinction is important because current LLMs are better suited for generating "small ideas" than "big ideas," as defined earlier. In line with this notion, a recent study tasked participants with creating a toy for a 7-year-old child using three items: a paper bag, a leftover construction brick, and an unused fan (Lee and Chung 2024). Participants were randomized to perform this task in three manners: alone, with assistance from an LLM, or the LLM performed it alone. Condition and hypothesis-blind experts

TABLE 2

LLM IDEATION ROLES WITH LLM AS KEY IDEATOR

	The designer	The writer
Explanation	Increase generalizability and internal validity, by improving how diverse stimuli are selected for experimentation.	Improve how ideas are expressed, given that creativity is partially social and subjective.
Example mechanisms	Stimulus selection is easy, reproducible, and hypothesis blind. You can identify unforeseen confounds.	Ideas are more articulate, persuasive, and concrete.
Example prompt	*Please generate 5 categories of <stimulus universe> that differ in <dimension used to create categories> and provide two specific examples of <stimuli> for each.* *We are going to describe two <stimuli>, please identify 5 consequential differences between them that may impact <the dependent variable> in <the hypothesized direction>* (Simonsohn, Montealegre, and Evangelidis 2025).	*Your goal is to effectively persuade the reviewer that your proposed theory about consumer behavior is accurate. The reviewer, after reading your initial submission, has expressed skepticism about your theory, stating that <reviewer's argument>. You need to generate a response that convinces the reviewer, using their own reasoning, that your theory is indeed valid. The explanation should be clear and logical, presented in a way that is both accessible and compelling, without relying on excessive jargon* (adapted from Costello, Pennycook, and Rand (2024)).
Caveats	Large sets of superficially different stimuli are insufficient; stimuli must vary on dimensions directly related to operationalization of the latent variable of interest. The prompt must not include the hypothesis. The stimulus type (e.g., vignettes, images) should be specified, in order to be actionable, otherwise you will impractically generate too diverse stimuli.	LLMs will not necessarily prioritize accuracy. Human judgment is still needed to filter suggestions—AI has biases too, and even possesses the ability to deceive (Hagendorff 2024).

evaluated all ideas, classifying them as "small" or "big." Compared to the human-only condition, the production of big ideas was similar across LLM-assisted and LLM-alone conditions. However, both LLM conditions produced more small ideas than humans (Lee and Chung 2024). Relatedly, other studies find that exceptional human ideas still exceed those generated by today's LLMs (Koivisto and Grassini 2023). Thus, to utilize LLMs for big ideas, humans may have to assume the role of key ideator in the co-creation process.

As a practical approach, we introduce "ideation roles" as metaphors that clarify the functions that LLMs can perform (tables 2 and 3). These roles are useful to those already using LLMs for ideation, as a way of organizing what they are doing, as well as to those who do not yet utilize LLMs, as a way of activating their imaginations for what is possible. The concept of roles offers an intuitive way to understand LLM capabilities, as opposed to keeping track of the vast range of tasks LLMs can perform, which can depend on their architectures, training regimens, and databases. We do not intend these metaphors as an exhaustive list, as much as a "case study" identifying specific roles that LLMs can play. Furthermore, each ideation role can be utilized for both productivity and semantic breadth. After all, ideas are ultimately scored based on the outcome (the idea), not the process that yields it, although it may well be that some roles naturally lend themselves better to one process over the other.

LLM as Key Ideator

The Designer. Consumer researchers often seek to test hypotheses about causal relationships between underlying theoretical constructs (e.g., "identity relevance," "cognitive load," or "power"), and outcomes (e.g., "purchase intention," "choice deferral," or "decision confidence"). These hypotheses, as well as the underlying constructs, are typically expressed conceptually in natural human language (Yarkoni 2020). When designing experiments, consumer researchers seek to test these causal theories by manipulating the underlying constructs. To these ends, they randomly assign participants to conditions, and these participants receive treatments (or stimuli) differing only in the relevant dimensions. In practice, however, stimuli often differ along multiple dimensions, making it possible that differences between treatment groups might be attributed to unintended confounds, rather than the manipulation of the construct of interest. This issue undermines the internal validity of the study and its capacity to generalize results to new stimuli (Wells and Windschitl 1999).

For example, consider a consumer researcher testing the hypothesis that identity-relevant ads are more persuasive than identity-neutral ones. They randomize participants to view one of two ads: Identity-Relevant Ad ("For athletes like you, who never back down," with images of diverse professional athletes that emphasize performance and grit) or Identity-Neutral Ad ("Shoes that get the job done," with images of people jogging in a park that emphasize comfort

TABLE 3

LLM IDEATION ROLES WITH HUMAN AS KEY IDEATOR

	The interviewer	The actor
Explanation	Prompt the human with thought-provoking questions to progress toward ideation goals.	Get ideas from "interviewing" LLMs that roleplay consumers.
Example mechanisms	Reverse-interview; consider different perspectives of the "inner crowd."	Get ideas for consumer interviews, new consumer samples, or entire projects.
Example prompt	*I want to conduct a research project, to be published in the Journal of Consumer Research, on a new topic related to <topic>. Help me come up with an interesting and original premise. I'd like all the ideas to come from me, but I want your help eliciting them. First, provide me with 5 questions to: (i) Inspire my creativity and imagination (ii) Prompt me to juxtapose disparate concepts or settings to create novel ideas (iii) Recall meaningful memories from my own consumption and life experiences. Then, ask me each question one at a time. For each response, ask two follow-up questions, one at a time, before moving on to the next question.* (adapted from (OpenAI 2024)	*You are a respondent in an in-depth interview. Today is November 21, 2019. I am Ally. I will be guiding you through an online discussion. You have been selected with a handful of others across the country to share your thoughts and opinions in this research discussion, and I look forward to hearing what you have to say! You have been chosen to be a part of this discussion because you previously mentioned you will either be hosting or attending a Friends-giving this year! Your name is Scott. You are a 32-year-old Caucasian Male. You are a Host of the Friendsgiving party. For the remainder of this discussion, we are going to be talking about Friendsgiving. I would love to understand your opinions and thoughts on this! Answer all the questions using as much detail as possible. There are no wrong answers!* (Arora, Chakraborty, and Nishimura 2025)
Caveats	Only works if the LLM is prompted to take a Socratic approach, where it asks questions rather than provides answers. The latter is less likely to stimulate creativity and learning of the creative process.	Given the tendency to anthropomorphize LLMs, it is tempting to see them as synthetic participants occupying a simulated world. This would be a mistake.

and durability). Suppose participants exposed to the identity-relevant ad show greater purchase intentions. Is this a successful test of the theory? This conclusion is true only if the observed difference between the two groups cannot be attributed to any factor other than identity relevance (Yarkoni 2020). Unfortunately, the stimuli differ on multiple dimensions other than just identity relevance, including emotional tone (e.g., "never back down" is more inspirational and energizing than "get the job done"), and visual appeal (e.g., professional athletes might attract more attention than generic joggers).

An often-touted solution is to sample diverse stimuli in the experiment, ensuring that idiosyncratic confounds between individual ads are balanced out (Monin and Oppenheimer 2014; Wells and Windschitl 1999). Nonetheless, sampling diverse stimuli might simply introduce random variation (Simonsohn et al. 2025). Returning to the example above, the consumer researcher could add more "replicates" to the design of their study. For instance, for identity-relevant ads, they could add a makeup ad with the tagline "Empower your natural beauty, your way" and a gaming console ad with the tagline "For players who dominate the game," while for identity-neutral ads they could add a makeup ad with the tagline "Quality that lasts" and a gaming console ad with the tagline "Entertainment for everyone." While these ads introduce variation, they do not necessarily mitigate the key confounds present in the original two stimuli. Even with diverse stimuli, the identity-relevant ads may still, on average, use more emotionally charged language (e.g., "empower," "dominate") compared to the more generic language of the neutral ads. Thus, to overcome this, consumer researchers must deliberately sample stimuli that vary along dimensions that might explain the observed effect.

But that is not all. Even when deliberate in their sampling, experimenters are biased due to a conflict of interest: they have a vested interest in their study producing the desired result. Because experimenters can imagine how participants might respond to the stimuli, they might still unconsciously tweak the stimuli to align with their expectations, inadvertently introducing additional confounds (Strickland and Suben 2012). Compounding concerns, consumer researchers often pre-test lots of stimuli but only report the ones that yield favorable results, substantially increasing the chance that their findings are merely artifacts of the stimuli chosen.

To overcome these challenges, we need a process for generating new stimuli that is deliberate, reproducible, hypothesis blind and, ideally, straightforward. LLMs, with their high productivity and disinterestedness in research outcomes, are well-suited to the task (see table 2 and Simonsohn et al. 2025 for detailed examples and prompts). Using carefully constructed prompts, Simonsohn et al. (2025) propose a structured approach where LLMs help to (i) define the experimental paradigm to be used, (ii) identify the universe of possible stimuli to be used, then (iii) systematically sample from these stimuli in a stratified manner (for a related approach, see Tomaino et al. 2025). Furthermore, consumer researchers can prompt an LLM to identify any other factors that might have a potentially confounding effect on the outcome variable, both before and after stimuli generation. Notably, this approach combines productivity and semantic breadth. Concerning productivity, it involves narrowing the "universe of stimuli" to a manageable subset that the LLM must sample from, ensuring the task remains tractable. As for semantic breadth, the approach relies on LLMs to identify potential confounds, leveraging their capacity to identify relationships between distant categories that might otherwise be missed.

The Writer. Consumer researchers seek to produce work that is not only objectively creative but also recognized as such by others. Although academic research and science in general are often perceived as purely objective endeavors, they are inherently social constructs, shaped by human researchers who decide what topics to study and what research to submit for publication. Then, human editors and reviewers evaluate these works for their rigor and originality, and human readers decide whether to read them, share them, or cite them. Thus, creativity assessments of scientific research are not a purely objective matter, but rather an inherently communicative affair, akin to how humans evaluate art, literature, and music. Given their communication capacities, we propose that LLMs can help the ideation process by how they write, thereby enhancing perceived creativity in the research process. LLMs can do that in various ways such as refining the articulation of ideas, enhancing their persuasiveness, and making them simpler.

LLMs can increase perceived idea originality by augmenting the persuasive communication of ideas. In the aforementioned study by Lee and Chung (2024), participants were randomized to use either web-search or an LLM when coming up with ideas for a novel dining table. A separate group of judges rated the ideas' creativity. Ideas generated with LLM assistance were rated as more creative than ones produced using web-search, and the effect was mediated by how articulate the expression of the idea was. Furthermore, recent studies found that LLMs were more persuasive than humans in domains where swaying opinions is challenging: politics (Hackenburg and Margetts 2024) and belief in conspiracy theories (Costello et al. 2024). Another study revealed that LLMs did not merely produce more complex grammatical and lexical structures, but also utilized more expansive moral foundations than humans—making their arguments particularly appealing to care-related virtues, fairness, authority virtues, and sanctity virtues (Carrasco-Farre 2024). LLMs are also capable of personalizing persuasive messages to the characteristics of the recipient (Matz et al. 2024). Finally, another study found that LLM-generated summaries of scientific papers written for a general audience were clearer, less complex, more understandable, and better comprehended than the same types of summaries written by humans (Markowitz 2024)—an effect likely driven not just by the communicative abilities of LLMs but also the tendency of academics to communicate in an overly abstract manner (Pinker 2015). Simplifying one's work with LLMs is straightforward. For example, one can simplify an abstract as follows (Markowitz 2024):

> The following text is an academic abstract from the Journal of Consumer Research. Based on this abstract, create a new abstract that provides enough context for the paper's implications to be clear to readers. The statement should not contain references and should avoid numbers, measurements, and acronyms unless necessary. It should explain the significance of the research at a level understandable to an undergraduate-educated scientists outside their field of specialty. Finally, it should include no more than 120 words. Write the abstract here:

Simple, concrete language may increase processing fluency and the ability to visualize what is being described (Jessen et al. 2000; Markowitz 2024; Paivio 2014)—all of which may enhance perceptions of the idea itself.

Human as Key Ideator

The Interviewer. Consumer researchers often seek creative "inspiration," where they move beyond habitual lines of thought to grasp fundamentally new avenues. Helpfully, LLMs can prompt consumer researchers themselves to engage in flexible or persistent thinking toward this end. In practice, we suggest using the LLM as a "Socratic interviewer," by prompting it to ask the consumer researcher a series of probing questions that are likely to draw out new insights from them (table 3). Inspired by research in education, the key is for the LLM to not simply surrender an answer, but to guide and support the human's thinking process. This distinction is significant, as individuals using LLMs as a tool for thinking and reflection achieve better learning outcomes than those who treat them as an answering machine (Bastani et al. 2024; Kumar et al. 2023; Lehmann, Cornelius, and Sting 2024). LLMs can thus be prompted to ask the right, thought-provoking questions at the right time, in the name of stimulating ideation. For instance, an LLM can ask provocative questions that tap into relevant experience, challenge assumptions, or force one to juxtapose disparate concepts (e.g., "How would you blend insights from childhood nostalgia with digital personalization?"), pushing consumer researchers beyond their usual thought patterns.

As an interviewer, the LLM can also be probed to make consumer researchers think of the problem from different perspectives. Indeed, research on the "inner crowd" finds that people are more accurate when "internally sampled" multiple times and their answers averaged, than when simply asked once (Herzog and Hertwig 2014). For example, after providing a first answer, the LLM can be prompted as follows:

> Start by assuming your initial idea might be flawed. Acknowledge that your first concept may not fully hit the mark. Identify potential reasons for this. Reflect on what assumptions or considerations might have led to gaps or weaknesses in your initial idea.
>
> Explore the implications of these new insights. Ask yourself: Do these considerations suggest that your idea was too ambitious, too simplistic, or off-target in another way? Develop an alternative perspective. Using this fresh understanding, refine or reframe your original idea, creating a second, improved version.

The Actor. Consumer researchers are interested in ensuring that their theory development and experimental design are informed by an empathetic grasp of consumer psychology and behavior. To these ends, LLMs can also serve as "actors," imitating a consumer that you interview for ideas, without treating them as an actual simulation of a consumer (Dengel et al. 2023). When prompting LLMs to behave as consumers with certain characteristics, like an American who often attends Friendsgiving parties (see table 3 for a prompt example), the model's most probable response conforms to the description in its prompt. One reason to expect LLMs to respond in a way that is valuable for ideation purposes is that the text corpus that LLMs are trained on contains enormous information about consumer behavior and decision-making (e.g., social media posts about Friendsgiving parties). Based on these interviews, a consumer researcher might, for instance, find new consumer samples worth incorporating, correct errors in a study before initiating it with real human participants (Sarstedt et al. 2024), or gain ideas for specific questions to include in a qualitative interview with real consumers (Arora et al. 2025). More useful still, the hope is that engaging with LLMs in this way may offer consumer researchers insights about new hypotheses and phenomena to investigate.

Nonetheless, consumer researchers must be careful to avoid assuming they are interacting with a true simulation of a consumer within a fully simulated world (Arora et al. 2025; Brand, Israeli, and Ngwe 2023; Shanahan, McDonell, and Reynolds 2023). LLMs simply generate the best response to a prompt, continuously adapting to the conversational context rather than to a constant simulated reality. Indeed, challenging the validity of synthetic samples, studies have shown that a common assumption—that LLMs can experimentally manipulate a single variable in a simulated world while holding all else constant—is routinely violated by these models (Gui and Toubia 2023). For instance, asking an LLM to act like an average customer and express its willingness to pay for Coca-Cola at various price levels would likely result in the model assuming that competitors' prices have also fluctuated. Even when instructed to keep competitor prices constant, the LLM would make other assumptions that introduced further "confounds" (Gui and Toubia 2023).

While LLM actors do not represent actual people with relevant histories, experiences, and preferences, letting go of this misconception is challenging, in large part because of the tendency for people, including consumer researchers, to anthropomorphize LLMs (De Freitas and Cohen 2025). Another (current) barrier to using these models as actors is that today's LLMs over-represent the views of Western, wealthy, liberal individuals as compared to other demographic groups, with some groups (e.g., elderly, widows), being highly under-represented (Santurkar et al. 2023). It is challenging to predict such biases (Saumure, De Freitas, and Puntoni 2025), given that LLMs are shaped by myriad sources: internet users providing data, crowd workers annotating the data based on guidelines provided by a company, and software engineers who make and tweak the models. Thus, special care should be taken when studying underrepresented groups, brands, or newer events that are unlikely to be represented in existing datasets.

LOOKING AHEAD

Consumer research, or any research field for that matter, will profoundly change as researchers use LLMs for ideation and other tasks. The literature accumulated to date suggests that if consumer researchers use LLMs as is, they will benefit individually by increasing their creativity, but our ideas as a field might become more homogenous. This would create a type of prisoner's dilemma (Doshi and Hauser 2024; Meincke et al. forthcoming), where each consumer researcher decides whether to favor themselves or the collective. Further, if average originality increases, then ideas we consider "big" before the widespread adoption of LLMs might seem "small" or incremental after adoption of LLMs. In short, these developments might backfire, posing challenges for the field.

Or not. Even if none of the recommendations for increasing the creativity of LLMs advocated in this article are utilized, we believe that peer review will provide a natural selection force that weeds out homogenous ideas, pushing

TABLE 4

TEN GUIDELINES FOR UTILIZING LLMs IN IDEATION

Guideline	Explanation
1. Increase productivity until originality plateaus	Generate more ideas within a narrow domain (productivity) by utilizing few-shot prompting (including a sample of highly relevant ideas in the prompt), retrieval-augmented generation (an API that fetches specialized data to augment the prompt), or fine-tuning an LLM on specialized data. Realize that the number of original ideas generated through this approach will eventually plateau.
2. Increase semantic breadth, and beware of negative spillover effects on collective diversity	Generate ideas spanning more diverse semantic categories (semantic breadth) by using prompt variation (varying prompts to enhance originality, as via persona modifiers), hybrid prompting (generating smaller idea pools using different prompts and then combining these pools), chain of thought prompting (asking the LLM to follow distinct, ordered steps in generating, expanding, and revising an idea), or increasing the temperature parameter (dialing up the stochasticity of the ideas to produce more diverse and unpredictable responses). These approaches help ensure an increase in originality without the cost of decreasing overall diversity of ideas.
3. Utilize the best of both the productivity and semantic breadth approaches to creativity	Get the best of the productivity and semantic breadth approaches by using prompting to switch between them, such as by instructing the LLM to "list only emotional factors" (productivity) or to "explore interactions between factors. Be creative!" (semantic breadth).
4. Beware of small ideas by considering different co-creation roles	Since current LLMs are better suited for "small ideas" than "big ideas," for big ideas treat the human as a key ideator (where LLMs "pull out" ideas from the human) rather than LLMs as key ideators (where they are the source of ideas that humans then screen).
5. Employ LLMs as a "Designer"	Use LLMs for generalizability and internal validity, by improving how diverse stimuli are selected for experimentation: Stimulus selection is easy, reproducible, and hypothesis blind, and you can identify unforeseen confounds.
6. Leverage LLMs as a "Writer"	Use LLMs to improve how ideas are expressed, given that creativity is partially social and subjective: Ideas are more articulate, persuasive, and concrete.
7. Prompt the LLM to "Interview" you	Prompt humans with thought-provoking questions to progress toward ideation goals: Reverse-interview the human, and prompt them to consider different perspectives of the "inner crowd."
8. Cast the LLM as an "Actor"	"Interview" LLMs that roleplay consumers: Get ideas for consumer interviews, new consumer samples, or entire projects.
9. Beware of inaccuracies, de-skilling, and anthropomorphizing	Filter all LLM suggestions, since they do not necessarily prioritize accuracy. Use Socratic approaches to stimulate learning and skill-building, rather than always using LLMs as answering machines. Resist the tendency to view them as synthetic participants occupying a simulated world.
10. Consider impact	Consider optimizing for impact and relevance—not just originality.

consumer researchers to consider new ways of answering the ever-relevant question: how can I differentiate my ideas from the competition and what has been said before? Furthermore, continued innovations in LLMs will likely alleviate some of their existing shortcomings.

In an even better scenario, consumer researchers will strategically adopt the practical guidelines we have provided to proactively increase the diversity of their ideas—summarized in table 4. They may even invent new roles, such as using LLMs to uncover overlooked hypotheses from data of experiments that have already been conducted (Batista and Ross 2024; Yang et al. 2024). To generate not just small ideas but big ones, consumer researchers will leverage LLMs in ways that preserve the role of humans as key ideators. This approach will be complemented by interventions that augment the productivity and semantic breadth of LLMs, enabling more original and diverse outcomes (e.g., via fine-tuning and hybrid prompting; table 1). They will also be mindful to maintain an active role, even when they are using LLMs as key ideators.

One open question is whether LLMs can help not just in generating ideas, but in whittling down a list of ideas to the best idea. Traditionally, this process involves first narrowing down options as much as possible, such as through voting, and then executing "minimal studies"—akin to "minimal viable products" (Terwiesch and Ulrich 2023)—to test which ideas have the most practical promise. But could LLMs streamline this process even further? Because the only way to know a paper's impact is to "iterate on the world," LLMs may need to be trained on actual outcomes to be able to predict a paper's likely impact before it is executed. Luo et al. (2025), for example, demonstrate that it is possible to train an LLM to beat neuroscientists at predicting the results of neuroscience experiments. Such efforts hint at the possibility to move beyond a merely semantic definition of appropriateness to one that is defined in terms of metrics like "success" (i.e., collecting real data that make a paper publishable) and real-world "relevance" and "impact" (i.e., other consumer researchers or external stakeholders care about the findings and find them useful; Pham 2013; Schmitt et al. 2022). Solving this problem is a cutting-edge frontier for academia and industry alike.

Even once an idea has succeeded and been submitted for publication, LLM-generated "impact scores" could be useful to editors seeking to predict whether a paper is likely to have an impact, potentially reducing power dynamics in the field and encouraging more diverse and impactful submissions (Chawla 2024). Some commentators have noted that, for the last 10 years, around 70% of consumer research articles garner hardly any citations at all (Pham 2013). This is alarming for the field, especially when considering that it ostensibly deals with relevant marketplace consumption phenomena, suggesting that in practice consumer researchers are not embracing this aspect of the field as much as they should (MacInnis et al. 2020; Schmitt et al. 2022). LLMs could help identify articles that are likely to make an impact, incentivizing researchers to move beyond this disappointing status quo.

However, such automated approaches require agreeing upon informative metrics for quantifying impact, which itself requires thought leadership. For instance, some have argued why citation counts alone, and even impact factors, are inappropriate as a yardstick of impact, recommending that the field instead track the percentile of the paper relative to other papers published in the same journal within the same year (Pham, Wu, and Wang 2024). On the other hand, some may worry that this approach penalizes highly original research, which, due to its distance from the status quo, could take longer to be picked up by the field. For example, a highly original paper on "Relational spending at funerals" (Whitley et al. 2022) recently won the "Early Contribution Award" from the *Journal of Consumer Psychology*, yet has just six citations to date, likely because few consumer researchers study funeral spending. Given the personal, social, and economic importance of bereavement, this lack of attention is an indictment of the field, not of the authors. Others will push back on the use of LLM-generated impact scores altogether, under the argument that readers are the best judges of the paper's value.

A final related question is whether LLMs will be able to provoke new ways of seeing ideas that, as a research field, we would typically deem too radical. Because of cognitive fixedness and our cultural predilection for following what is popular, scientific fields always risk getting stuck in incremental "paradigms." LLMs may help us avoid these constraints on our thinking. For example, Shin et al. (2023, e2214840120) examine the impact that the release of the algorithm AlphaGo had on the decision making of professional Go players. They conclude that the arrival of superhuman AI led players to "break away from traditional strategies and induced them to explore novel moves". As it might do for other stubborn societal problems like road fatalities and loneliness (Agarwal et al. 2024; De Freitas et al. 2024), AI might help solve systemic ideation problems in the field that we have been unable to solve ourselves.

REFERENCES

Agarwal, Stuti, Julian De Freitas, Anya Ragnhildstveit, and Carey K. Morewedge (2024), "Acceptance of Automated Vehicles Is Lower for Self Than Others," *Journal of the Association for Consumer Research*, 9 (3), 269–81.

Amabile, Teresa M. (1982), "Social Psychology of Creativity: A Consensual Assessment Technique," *Journal of Personality and Social Psychology*, 43 (5), 997–1013.

Arora, Neeraj, Ishita Chakraborty, and Yohei Nishimura (2025), "Express: AI-Human Hybrids for Marketing Research: Leveraging LLMs as Collaborators," *Journal of Marketing*, 89 (2), 43–70.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Ozge Kabakcı, and Rei Mariman (2024), "Generative AI Can Harm Learning," preprint SSRN 4895486. http://dx.doi.org/10.2139/ssrn.4895486.

Batista, Rafael M. and James Ross (2024), "Words That Work: Using Language to Generate Hypotheses," preprint. http://dx.doi.org/10.2139/ssrn.4926398.

Bellemare-Pepin, Antoine, François Lespinasse, Philipp Thölke, Yann Harel, Kory Mathewson, Jay A. Olson, Yoshua Bengio, and Karim Jerbi (2024), "Divergent Creativity in Humans and Large Language Models," arXiv, arXiv:2405.13012, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2405.13012.

Berger, Jonah, Grant Packard, Reihane Boghrati, Ming Hsu, Ashlee Humphreys, Andrea Luangrath, Sarah Moore, Gideon Nave, Christopher Olivola, and Matthew Rocklage (2022), "Marketing Insights from Text Analysis," *Marketing Letters*, 33 (3), 365–77.

Boyd, Drew and Jacob Goldenberg (2013), *Inside the Box*. New York: Simon & Shuster.

Brand, James, Ayelet Israeli, and Donald Ngwe (2023), "Using LLMs for Market Research," Working Paper (23–062), Harvard Business School Marketing Unit, Boston, MA 02163.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, and Amanda Askell (2020), "Language Models Are Few-Shot Learners," *Advances in Neural Information Processing Systems*, 33, 1877–901.

Burroughs, James E. and David Glen Mick (2004), "Exploring Antecedents and Consequences of Consumer Creativity in a Problem-Solving Context," *Journal of Consumer Research*, 31 (2), 402–11.

Burton, Jason W., Ezequiel Lopez-Lopez, Shahar Hechtlinger, Zoe Rahwan, Samuel Aeschbach, Michiel A. Bakker, Joshua A. Becker, Aleks Berditchevskaia, Julian Berger, Levin Brinkmann, Lucie Flek, Stefan M. Herzog, Saffron Huang, Sayash Kapoor, Arvind Narayanan, Anne-Marie Nussberger, Taha Yasseri, Pietro Nickl, Abdullah Almaatouq, Ulrike Hahn, Ralf H. J. M. Kurvers, Susan Leavy, Iyad Rahwan, Divya Siddarth, Alice Siu, Anita W. Woolley, Dirk U. Wulff, and Ralph Hertwig (2024), "How Large Language Models Can Reshape Collective Intelligence," *Nature Human Behaviour*, 8 (9), 1643–55.

Caprioli, Sara, Christoph Fuchs, and Bram Van den Bergh (2023), "On Breaking Functional Fixedness: How the Aha! Moment Enhances Perceived Product Creativity and Product Appeal," *Journal of Consumer Research*, 50 (1), 48–69.

Carrasco-Farre, Carlos (2024), "Large Language Models Are as Persuasive as Humans, but How? About the Cognitive Effort and Moral-Emotional Language of LLM Arguments," arXiv, arXiv:2404.09329, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2404.09329.

Chawla, Dalmeet Singh (2024), "Can Novelty Scores on Papers Shift the Power Dynamics in Scientific Publishing?," *Nature Index*. https://www.nature.com/articles/d41586-024-04021-w.

Costello, Thomas H., Gordon Pennycook, and David G. Rand (2024), "Durably Reducing Conspiracy Beliefs through Dialogues with AI," *Science (New York, N.Y.)*, 385 (6714), eadq1814.

De Dreu, Carsten K. W., Matthijs Baas, and Bernard A. Nijstad (2008), "Hedonic Tone and Activation Level in the Mood-Creativity Link: Toward a Dual Pathway to Creativity Model," *Journal of Personality and Social Psychology*, 94 (5), 739–56.

Freitas De and Julian I. Glenn Cohen (2025), "Disclosure, Humanizing, and Contextual Vulnerability of Generative AI Chatbots," *New England Journal of Medicine AI*, 2 (2), AIpc2400464.

Freitas De and Julian Elie Ofek (2024), "How AI Can Power Brand Management," *Harvard Business Review*, 103 (9), 108–4.

De Freitas, Julian, Ahmet K. Uğuralp, Zeliha O. Uğuralp, and Stefano Puntoni (2024), "AI Companions Reduce Loneliness," Working Paper, 24–078, Harvard Business School, Boston, MA 02163.

Dengel, Andreas, Rupert Gehrlein, David Fernes, Sebastian Görlich, Jonas Maurer, Hai Hoang Pham, Gabriel Großmann, and Niklas Dietrich Genannt Eisermann (2023), "Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and Bard on Computer Science Education," *Informatics*, 10 (4), 78.

Doshi, Anil R. and Oliver P. Hauser (2024), "Generative AI Enhances Individual Creativity but Reduces the Collective Diversity of Novel Content," *Science Advances*, 10 (28), eadn5290.

Finke, Ronald A., Thomas B. Ward, and Steven M. Smith (1996), *Creative Cognition: Theory, Research, and Applications*. Cambridge, MA: MIT Press.

Fleder, Daniel and Kartik Hosanagar (2009), "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," *Management Science*, 55 (5), 697–712.

Förster, Jens, Ronald S. Friedman, and Nira Liberman (2004), "Temporal Construal Effects on Abstract and Concrete Thinking: Consequences for Insight and Creative Cognition," *Journal of Personality and Social Psychology*, 87 (2), 177–89.

Girotra, Karan, Christian Terwiesch, and Karl T. Ulrich (2010), "Idea Generation and the Quality of the Best Idea," *Management Science*, 56 (4), 591–605.

Goldenberg, Jacob, David Mazursky, and Sorin Solomon (1999), "Toward Identifying the Inventive Templates of New Products: A Channeled Ideation Approach," *Journal of Marketing Research*, 36 (2), 200–10.

Gui, George and Olivier Toubia (2023), "The Challenge of Using Llms to Simulate Human Behavior: A Causal Inference Perspective," arXiv, arXiv:2312.15524, preprint: not peer reviewed.

Hackenburg, Kobi and Helen Margetts (2024), "Evaluating the Persuasive Influence of Political Microtargeting with Large Language Models," *Proceedings of the National Academy of Sciences of the United States of America*, 121 (24), e2403116121.

Hagendorff, Thilo (2024), "Deception Abilities Emerged in Large Language Models," *Proceedings of the National Academy of Sciences of the United States of America*, 121 (24), e2317967121.

Harvey, Sarah and James W. Berry (2023), "Toward a Meta-Theory of Creativity Forms: How Novelty and Usefulness Shape Creativity," *Academy of Management Review*, 48 (3), 504–29.

Herzog, Stefan M. and Ralph Hertwig (2014), "Harnessing the Wisdom of the Inner Crowd," *Trends in Cognitive Sciences*, 18 (10), 504–6.

Hirschman, Elizabeth C. (1980), "Innovativeness, Novelty Seeking, and Consumer Creativity," *Journal of Consumer Research*, 7 (3), 283–95.

Hubert, Kent F., Kim N. Awa, and Darya L. Zabelina (2024), "The Current State of Artificial Intelligence Generative Language Models Is More Creative Than Humans on Divergent Thinking Tasks," *Scientific Reports*, 14 (1), 3440.

Jeppesen, Lars Bo and Karim R. Lakhani (2010), "Marginality and Problem-Solving Effectiveness in Broadcast Search," *Organization Science*, 21 (5), 1016–33.

Jessen, Frank, Reinhard Heun, Michael Erb, D.-O. Granath, Uwe Klose, Andreas Papassotiropoulos, and Wolfgang Grodd (2000), "The Concreteness Effect: Evidence for Dual Coding and Context Availability," *Brain and Language*, 74 (1), 103–12.

Koestler, Arthur (1964), *The Act of Creation*, London: Hutchinson.

Koivisto, Mika and Simone Grassini (2023), "Best Humans Still Outperform Artificial Intelligence in a Creative Divergent Thinking Task," *Scientific Reports*, 13 (1), 13601.

Kornish, Laura J. and Karl T. Ulrich (2011), "Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas," *Management Science*, 57 (1), 107–28.

Kumar, Harsh, David M. Rothschild, Daniel G. Goldstein, and Jake M. Hofman (2023), "Math Education with Large Language Models: Peril or Promise?," preprint. http://dx.doi.org/10.2139/ssrn.4641653.

Lee, Byung Cheol and Jaeyeon Chung (2024), "An Empirical Investigation of the Impact of ChatGPT on Creativity," *Nature Human Behaviour*, 8 (10), 1906–14.

Lee, Dokyun and Kartik Hosanagar (2019), "How Do Recommender Systems Affect Sales Diversity? A Cross-

Category Investigation Via Randomized Field Experiment," *Information Systems Research*, 30 (1), 239–59.

Lehmann, Matthias, Philipp B. Cornelius, and Fabian J. Sting (2024), "AI Meets the Classroom: When Does ChatGPT Harm Learning?," arXiv, arXiv:2409.09047, preprint: not peer reviewed.

Luo, Xiaoliang, Akilles Rechardt, Guangzhi Sun, Kevin K Nejad, Felipe Yáñez, Bati Yilmaz, Kangjoo Lee, Alexandra O. Cohen, Valentina Borghesani, Anton Pashkov, Daniele Marinazzo, Jonathan Nicholas, Alessandro Salatiello, Ilia Sucholutsky, Pasquale Minervini, Sepehr Razavi, Roberta Rocca, Elkhan Yusifov, Tereza Okalova, Nianlong Gu, Martin Ferianc, Mikail Khona, Kaustubh R. Patil, Pui-Shee Lee, Rui Mata, Nicholas E. Myers, Jennifer K. Bizley, Sebastian Musslick, Isil Poyraz Bilgin, Guiomar Niso, Justin M. Ales, Michael Gaebler, N. Apurva Ratan Murty, Leyla Loued-Khenissi, Anna Behler, Chloe M. Hall, Jessica Dafflon, Sherry Dongqi Bao, and Bradley C. Love (2025), "Large Language Models Surpass Human Experts in Predicting Neuroscience Results," *Nature Human Behaviour*, 9 (2), 305–15.

MacInnis, Deborah J., Vicki G. Morwitz, Simona Botti, Donna L. Hoffman, Robert V. Kozinets, Donald R. Lehmann, John G. Lynch Jr, and Cornelia Pechmann (2020), "Creating Boundary-Breaking, Marketing-Relevant Consumer Research," *Journal of Marketing*, 84 (2), 1–23.

Markowitz, David M. (2024), "From Complexity to Clarity: How AI Enhances Perceptions of Scientists and the Public's Understanding of Science," *PNAS Nexus*, 3 (9), pgae387.

Matz, Sandra C., Jacob D. Teeny, Sumer S. Vaid, Heinrich Peters, Gabriella M. Harari, and Moran Cerf (2024), "The Potential of Generative AI for Personalized Persuasion at Scale," *Scientific Reports*, 14 (1), 4692.

Mehta, Ravi and Meng Zhu (2016), "Creating When You Have Less: The Impact of Resource Scarcity on Product Use Creativity," *Journal of Consumer Research*, 42 (5), 767–82.

Mehta, Ravi, Rui Zhu, and Amar Cheema (2012), "Is Noise Always Bad? Exploring the Effects of Ambient Noise on Creative Cognition," *Journal of Consumer Research*, 39 (4), 784–99.

Meincke, Lennart, Karan Girotra, Gideon Nave, Christian Terwiesch, and Karl T. Ulrich (2024a), "Using Large Language Models for Idea Generation in Innovation," Working Paper. http://dx.doi.org/10.2139/ssrn.4526071.

Meincke, Lennart, Ethan R. Mollick, and Christian Terwiesch (2024b), "Prompting Diverse Ideas: Increasing Ai Idea Variance," arXiv, arXiv:2402.01727, preprint: peer not reviewed.

Meincke, Lennart, Gideon Nave, and Christian Terwiesch (forthcoming), "ChatGPT Enhances Individual Creativity but Undermines Collective Diversity in Brainstorming," *Nature Human Behaviour*.

Melumad, Shiri and Michel Tuan Pham (2020), "The Smartphone as a Pacifying Technology," *Journal of Consumer Research*, 47 (2), 237–55.

Monin, Benoît and Daniel M. Oppenheimer (2014), "The Limits of Direct Replications and the Virtues of Stimulus Sampling," *Social Psychology*, 45 (4), 299–300.

Moreau, C. Page and Darren W. Dahl (2005), "Designing the Solution: The Impact of Constraints on Consumers' Creativity," *Journal of Consumer Research*, 32 (1), 13–22.

Moreau, Page, Emanuela Prandelli, and Martin Schreier (2023), "Generative Artificial Intelligence and Design Co-Creation in Luxury New Product Development: The Power of Discarded Ideas," preprint. http://dx.doi.org/10.2139/ssrn.4630856.

Nijstad, Bernard A., Carsten K. W. De Dreu, Eric F. Rietzschel, and Matthijs Baas (2010), "The Dual Pathway to Creativity Model: Creative Ideation as a Function of Flexibility and Persistence," *European Review of Social Psychology*, 21 (1), 34–77.

Nijstad, Bernard A. and Wolfgang Stroebe (2006), "How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups," *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 10 (3), 186–213.

OpenAI (2024), "Writing with AI," Last Accessed January 5, 2025. https://openai.com/chatgpt/use-cases/writing-with-ai/.

Paivio, Allan (2014), "Intelligence, Dual Coding Theory, and the Brain," *Intelligence*, 47, 141–58.

Peeperkorn, Max, Tom Kouwenhoven, Dan Brown, and Anna Jordanous (2024), "Is Temperature the Creativity Parameter of Large Language Models?," arXiv, arXiv:2405.00492, preprint: not peer reviewed.

Pham, Michel Tuan (2013), "The Seven Sins of Consumer Psychology," *Journal of Consumer Psychology*, 23 (4), 411–23.

Pham, Michel Tuan, Alisa Yinghao Wu, and Danqi Wang (2024), "Benchmarking Scholarship in Consumer Research: The P-Index of Thought Leadership," *Journal of Consumer Research*, 51 (1), 191–203.

Piezunka, Henning and Linus Dahlander (2015), "Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing," *Academy of Management Journal*, 58 (3), 856–80.

Pinker, S. (2015), *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*, New York, NY: Penguin Books.

SAE (2021), "Sae Levels of Driving Automation™ Refined for Clarity and International Audience," Last Accessed January 5, 2025. https://www.sae.org/blog/sae-j3016-update.

Saumure, Roger, Julian De Freitas, and Stefano Puntoni (2025), "Humor as a Window into Generative AI Bias," *Scientific Reports*, 15 (1), 1326.

Santurkar, Shibani, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto (2023), "Whose Opinions Do Language Models Reflect?," in *International Conference on Machine Learning*, Vol. 202, PMLR, 29971–30004.

Sarstedt, Marko, Susanne J. Adler, Lea Rau, and Bernd Schmitt (2024), "Using Large Language Models to Generate Silicon Samples in Consumer and Marketing Research: Challenges, Opportunities, and Guidelines," *Psychology & Marketing*, 41 (6), 1254–70.

Schmitt, Bernd H., June Cotte, Markus Giesler, Andrew T. Stephen, and Stacy Wood (2022), "Relevance—Reloaded and Recoded," *Journal of Consumer Research*, 48 (5), 753–5.

Shanahan, Murray, Kyle McDonell, and Laria Reynolds (2023), "Role Play with Large Language Models," *Nature*, 623 (7987), 493–8.

Shin, Minkyu, Jin Kim, Bas Van Opheusden, and Thomas L. Griffiths (2023), "Superhuman Artificial Intelligence Can Improve Human Decision-Making by Increasing Novelty," *Proceedings of the National Academy of Sciences of the United States of America*, 120 (12), e2214840120.

Simonsohn, Uri, Andres Montealegre, and Ioannis Evangelidis (2025), "Stimulus Sampling Reimagined: Designing Experiments With Mix-and-Match, Analyzing Results with

Stimulus Plots," preprint. http://dx.doi.org/10.2139/ssrn.4716832.

Solaiman, Irene and Christy Dennison (2021), "Process for Adapting Language Models to Society (Palms) With Values-Targeted Datasets," *Advances in Neural Information Processing Systems*, 34, 5861–73.

Strickland, Brent and Aysu Suben (2012), "Experimenter Philosophy: The Problem of Experimenter Bias in Experimental Philosophy," *Review of Philosophy and Psychology*, 3 (3), 457–67.

Terwiesch, Christian and Karl T. Ulrich (2023), *The Innovation Tournament Handbook: A Step-by-Step Guide to Finding Exceptional Solutions to Any Challenge*. Philadelphia, PA: University of Pennsylvania Press.

Tomaino, Geoff, Asaf Mazar, Ziv Carmon, and Klaus Wertenbroch (2025), "A Simple Method for Improving Generalizability in Behavioral Science: Scope Testing with AI-Generated Stimuli (Stags)," *Consumer Psychology Review*, 8 (1), 87–97.

Valenzuela, Ana, Stefano Puntoni, Donna Hoffman, Noah Castelo, Julian De Freitas, Berkeley Dietvorst, Christian Hildebrand, Young Eun Huh, Robert Meyer, Miriam E Sweeney, Sanaz Talaifar, Geoff Tomaino, and Klaus Wertenbroch (2024), "How Artificial Intelligence Constrains the Human Experience," *Journal of the Association for Consumer Research*, 9 (3), 241–56.

Wei, Jason, Wang Xuezhi, Schuurmans Dale, Bosma Maarten, Xia Fei, Chi Ed, Le Quoc V., and Zhou Denny (2022), "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," *Advances in Neural Information Processing Systems*, 35, 24824–37.

Wells, Gary L. and Paul D. Windschitl (1999), "Stimulus Sampling and Social Psychological Experimentation," *Personality and Social Psychology Bulletin*, 25 (9), 1115–25.

Whitley, Sarah C., Ximena Garcia-Rada, Fleura Bardhi, Dan Ariely, and Carey K. Morewedge (2022), "Relational Spending in Funerals: Caring for Others Loved and Lost," *Journal of Consumer Psychology*, 32 (2), 211–31.

Yang, Minwen, Wang Qiurun, Gülden Ulkümen, Claire Tsai, and Carey K. Morewedge (2024), "Large Language Models Improve Hypothesis Generation by Reducing Effort," arXiv:2404.04326v3, preprint: not peer reviewed.

Yarkoni, Tal (2020), "The Generalizability Crisis," *The Behavioral and Brain Sciences*, 45, e1.

Zhou, Eric and Dokyun Lee (2024), "Generative Artificial Intelligence, Human Creativity, and Art," *PNAS Nexus*, 3 (3), pgae052.

Journal of Consumer Research, 2025, 52, 18–31
https://doi.org/10.1093/jcr/ucaf012
Gen AI and Consumer Research