

# Generative AI, Innovation, and Trust

The Journal of Applied Behavioral Science

1–10

© The Author(s) 2024



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/00218863241285033

[journals.sagepub.com/home/jab](https://journals.sagepub.com/home/jab)

Frank T. Piller<sup>1</sup> , Mahdi Srour<sup>2</sup>,  
and Tucker J. Marion<sup>2</sup>

## Abstract

Our paper explores the integration of generative artificial intelligence (GenAI) into organizations' innovation and new product development processes, focusing on when and how to trust AI-generated outcomes in this context. We propose a framework to assess the level of trust required based on task-specific needs and the distinction between general and expert AI models. While inaccuracies in GenAI outputs can foster creativity during ideation, higher accuracy, and trust are essential for tasks requiring domain-specific expertise. The paper concludes by discussing the necessary human capabilities and organizational strategies for effectively deploying GenAI in innovation management.

## Keywords

generative AI, trust, innovation management, organizational capabilities

A defining feature of generative artificial intelligence (GenAI) is its capacity to generate novel and creative outputs, a capability that sets it apart from traditional discriminative AI (Acar et al., 2024; Bouschery et al., 2023). This has given rise to a new interest in artificial intelligence (AI) as a tool for innovation. As the majority of data pertinent to innovation projects is textual (market insights, customer needs, technological knowledge, concept descriptions, project documentation, etc.), the growing capabilities of large language models (LLMs) such as OpenAI's ChatGPT, Google's Gemini, or Meta's Llama have contributed to the increased interest in AI as a novel

<sup>1</sup>School of Business and Economics, RWTH Aachen University, Aachen, Germany

<sup>2</sup>D'Amore-McKim School of Business, Entrepreneurship & Innovation Group, Northeastern University, Boston, MA, USA

## Corresponding Author:

Frank T. Piller, RWTH Aachen University, School of Business and Economics, TIME Research Area, Templergraben 55, 52056 Aachen, Germany.

Email: [piller@time.rwth-aachen.de](mailto:piller@time.rwth-aachen.de)

innovation tool (Piller et al., 2023). While prior literature has discussed the relationship between GenAI and innovation in general (Bilgram & Laarmann, 2023; Cooper & Brem, 2024; Füller & Hutter, 2024; Roberts & Candi, 2024), we ask in this article: *When can you trust the outcomes of AI in the context of innovation?*

We propose that answering three questions can guide organizations in navigating the opportunities of AI (especially GenAI) for innovation. First, firms must structure their tasks (and desired outcomes) regarding the level of trust required when engaging a GenAI to support a development process. Second, managers need to know when general AI models are sufficient and when customized models trained for a firm's specific domain are required to generate a trustful outcome. Third, different AI models require different skills and capabilities from their users. Only when the skills match the task requirements will using GenAI lead to higher innovation performance.

## **Question 1: Do We Need Trust in the GenAI Outcome?**

One of the most prominent concerns organizations have when deploying GenAI is its inaccuracy (McKinsey, 2023), such as LLMs' infamous hallucinations of making up facts and references (Ji et al., 2023). Recent developments like combining a generative output with a web search have reduced this tendency. Still, everyone who asked ChatGPT to draft their curriculum vitae (CV) will have recognized that the algorithm makes up facts. While easily spotted in one's CV, this can lead to severe concerns when summarizing technical literature or prompting the system to identify latent customer needs. Ultimately, decisions along the innovation process must be based on accurate attributes developed from robust prediction models using the best available data. This points to the need for these systems to be developed and trained specifically for an industry or product line to derive trustworthy results from domain-specific data, leading to informed decisions.

However, there are also situations wherein inaccuracy is not inherently bad. For example, in upfront ideation and discovery tasks, organizations don't need to trust GenAI's results. Here, GenAI's hallucinations are a feature that fosters creativity. In ideation, creativity and out-of-the-box thinking are the goal. Trained models could lead to path dependencies and unwanted biases. Thus, contrary to popular belief, we suggest that the issue is not to find a way to make all AI-generated results accurate so that they are always trustworthy but rather to be able to assess the level of trust that specific use cases require and adjust the accuracy of GenAI results accordingly.

The level of accuracy of, and therefore trust in, the GenAI depends on the kind of data and resources invested in its development and training. Organizations must first understand the factors and their interplay that make up a particular task to assess whether the GenAI's underlying models sufficiently reflect these in their ground truth (Lebovitz et al., 2023). Second, they must ensure the quality and quantity of data that the model training process requires to achieve the desired accuracy of AI-generated results. By understanding the trust dimension, organizations can manage expectations and allocate resources accordingly and effectively. For example, understanding customer sentiment on a certain product using ChatGPT will give you an

answer, but the accuracy will be low. However, using an LLM on internal customer reviews can provide highly accurate insights into customer preferences (Yuan et al., 2022). Some starting questions to explore the trust dimension further are: What are the output expectations? Will the AI's output be used as a thought starter or as the base for a decision? What type of information is required as input for the model: internal or external data? If external information is sufficient, how credible and verifiable are the underlying data and training?

## Question 2: Do We Need a General or an Expert Model?

AI models can be differentiated into general and task-specific expert models. *General models* have been trained on a large set of diverse data (e.g., corpora of information gleaned from the Internet over the last twenty years). This allows general AI platforms to be versatile and handle various tasks by ingeniously connecting information from multiple domains. Examples of General Models are the commercially available GenAI systems (e.g., OpenAI's ChatGPT, Google's Gemini, or Meta's Llama.). These powerful general models primarily constitute the current fascination with GenAI. They are easily accessible, fast, and amazingly knowledgeable about many different fields and domains (e.g., from software coding suggestions to questions on marketing). They also can connect very different bodies of knowledge innovatively, constituting their ability for lateral thinking and creativity.

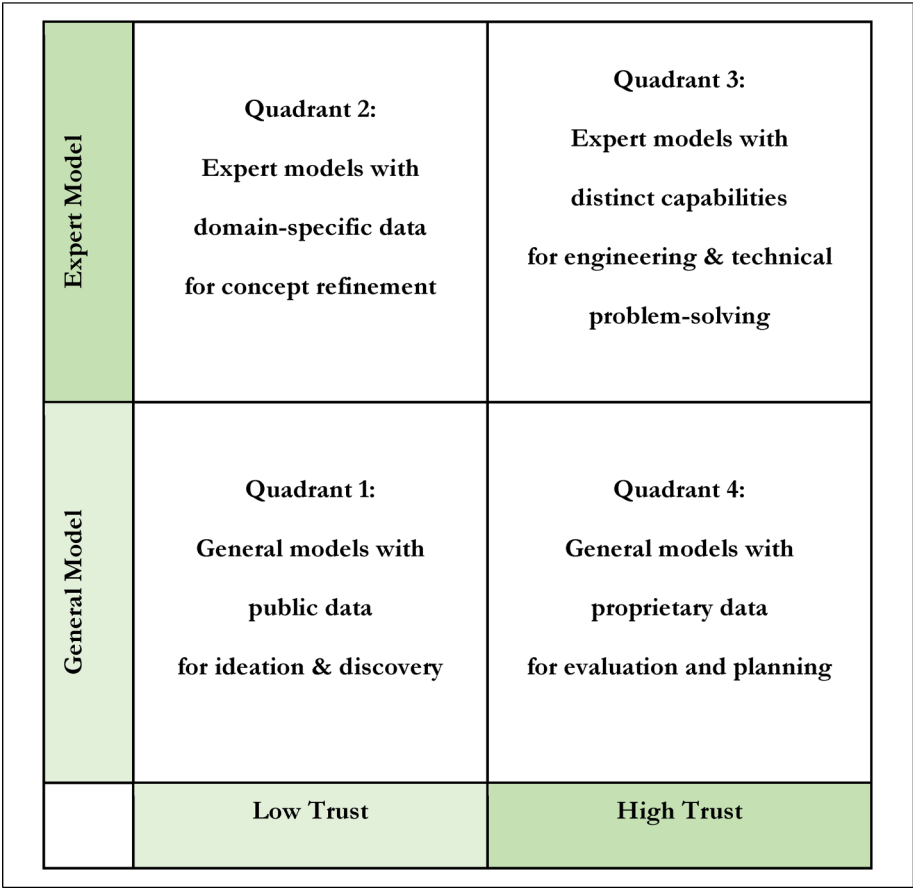
Additionally, the structure of the models and built-in fluency can help summarize and categorize information extremely efficiently. Finally, these abilities can be accessed at a very low cost. But at the same time, the knowledge capabilities of these systems cloud people's understanding and expectations of what these systems can do. Using ChatGPT for ideation can yield interesting ideas (Bouschery et al., 2024; Guzik et al., 2023; Meincke et al., 2024). However, the underlying model is not explicitly trained for a specific market. Therefore, even if the response sounds credible because the underlying language model excels at fluency, it is not credible enough to make authentic design and business decisions at advanced innovation stages that require specificity and accuracy.

*Expert models* can bridge the gap between sounding credible and being credible. They are designed to handle the additional complexity introduced by domain- and organization-specific factors. Expert applications built out of expert models excel in their domain of expertise and deliver tailored results compared to general models. To transition to expert models by integrating specific information into LLMs, an expert model can either be created from scratch through fine-tuning existing LLMs on domain-specific data (e.g., using OpenAI's GPT builder to create custom ChatGPTs) or prompt-tuning commercially available LLMs (e.g., input expertise into OpenAI's ChatGPT using prompts without changing its underlying model) so that the AI-generated outcomes reflect experts' real-world knowledge processes and thus the complexity that certain use cases ask for. A growing trend in developing specialized expert models is models that serve as "AI agents" (Xi et al., 2023). These agents generate specialized information and can collaborate with other models,

fostering a more connected and efficient AI ecosystem. This approach leverages the strengths of individual expert models while allowing them to work together to provide comprehensive solutions. This trend toward collaboration among AI agents can enhance accuracy and specificity, making them valuable assets in complex decision-making processes that require detailed and precise information (Chen et al., 2024).

**Navigating the Possibilities of GenAI for Innovation Management**

A synthesis of these two questions and decision flows reveals a landscape that can facilitate navigation of the potential applications of GenAI in innovation management,



**Figure 1.** Matching GenAI applications to specific innovation tasks.

as shown in Figure 1. The horizontal axis asks about the level of trust and confidence a task requires. The vertical axis shows the type of GenAI, either a general model or an expert system. The two left quadrants (#1 and #2) correspond to tasks where trust in the outcome is unimportant, as there will be other checks and evaluations of the results. These are generally questions in the problem space of an innovation project to develop compelling concept alternatives. The quadrants on the right (#3 and #4) address the solution space, where, for example, the best concept is transformed into a technical solution and launched into the market. Hence, these situations require high trust in the AI's prediction. In the following, we explore these four quadrants in more detail.

*Quadrant 1: General models with public data for ideation & discovery:* This quadrant represents tasks characterized by a relatively low level of trust required in the AI's output and the applicability of general models. Typical tasks in this quadrant include ideating novel ideas and brainstorming, discovering initial customer needs and market requirements, and exploring new topics to gain a basic understanding of various domains. General models in the form of the popular commercial LLMs can support these. This class of GenAI enhances lateral thinking capabilities by providing diverse perspectives and facilitating connections across different knowledge areas, which can lead to more innovative and creative solutions. Innovation teams get quick access to more knowledge from many domains that might otherwise be outside the team's expertise.

Additionally, these models assist in navigating larger general knowledge fields by quickly assimilating and presenting information that would otherwise require extensive manual research, thereby enabling innovation teams to focus on applying this knowledge to develop and refine their concepts. In addition to knowledge extraction capabilities, text-to-image generators DALL-E, Stable Diffusion, or Midjourney are robust general models that can generate basic mock-ups to make "crazy" ideas more tangible. Much of the prior literature on GenAI for innovation can be placed in this field.

*Quadrant 2: Expert models with domain-specific data for concept refinement:* In this cluster, tasks are still related to exploration and creativity but require more specific knowledge about a domain. Typical tasks include concept refinement, latent need identification and validation, and experimentation and prototyping. An example could be further elaborating concepts or ideas from the first quadrant or generating ideas in highly specialized areas not adequately covered by general models. Evaluating the market fit of a concrete idea or initial validations that concepts are technically, financially, and operationally viable and aligned with an organization's overall strategy and goals requires access to specific, often in-public information. Still, results don't have to be completely accurate and 100% trustworthy to be useful at this stage. However, such tasks require GenAI outputs reflecting solution information covered by engineering expertise to produce acceptable results. Thus, the GenAI must be trained on domain-specific data to understand the required information and become an expert system (Bouschery et al., 2023; Piller et al., 2023). In addition to critical technical requirements, the GenAI output should also reflect domain- and organization-specific knowledge, such as evaluating the viability of the concepts using customer sentiment

data culled from customer reviews, which consider the context of the specific industry. This requires training on the product's particular attributes and its industry vertical. The result is a more targeted and precise approach to developing ideas, ensuring that they are not only innovative but also viable within the specific context of the industry. Furthermore, by leveraging expert models, organizations can refine concepts with a higher degree of relevance and accuracy, thus increasing the likelihood of successful market entry.

*Quadrant 3: Expert models with distinct capabilities for engineering & technical problem-solving:* This quadrant represents tasks that cover a critical area where the highest level of trust and confidence in AI outputs is required. In this quadrant, tasks such as technical or scientific problem-solving, engineering and design, and managing regulatory and compliance procedures are central. For years, computer-aided engineering (CAE) and advanced simulation systems have defined this space. These expert systems must incorporate knowledge of scientific laws, industry standards, safety regulations, and any applicable legal requirements that designers and engineers must recognize when developing a product's features or technical solutions to defined problems. This information constitutes the expert model's "ground truth"—the annotated data that the model is trained on—and thus significantly influences the trustworthiness of the AI-generated outcomes.

The promise of GenAI in this stage is to navigate the overall complexity that comes with the traditional tools and find optimal solutions more quickly. At the same time, engineers and designers often focus too much on the technical solutions they already know and are comfortable with. GenAI can help explore a more extensive solution space and overcome path dependencies. We see this combination of established (but also complex to operate) expert systems with GenAI capabilities as a significant opportunity (Davenport et al., 2023; Salvador & Sting, 2022). Marion et al. (2024) discuss the case of *Siemens Simcenter*. This established simulation package allows engineers to simulate the physical behavior of products or processes, replacing physical prototypes and testbeds with digital ones. While very powerful, Simcenter traditionally required long setup times and could only be used after extensive training. Interpreting the results of a simulation also traditionally demands deep expertise. To overcome these challenges, Siemens combined the established expert system with a GenAI-based frontend (user interface), creating an expert application called *HiSimcenter*. It can handle tasks ranging from simple queries, such as selecting the proper CAE tool for a given task, to a fully automated design capability that takes product requirements (in natural language) as input and directly generates a compliant design. This expert application not only makes the use of CAE tools more accessible but also makes their results more understandable to a broader set of users who have domain expertise but are not simulation specialists.

*Quadrant 4: General models with proprietary data for evaluation and planning:* In the final stage of the innovation process, general models can be used to increase the efficiency of these tasks if the innovation team can provide data about the solution to be evaluated and planned to build confidence in the outcome. In this phase, tasks encompass summarization and categorizing diverse content, such as text for user

manuals or training documentation. The scope of tasks extends to launch support, including creating media announcements or instructions for service representatives who provide user support. Additionally, activities can involve code exploration and debugging, ensuring the efficiency of software development processes. All of these tasks have to be executed so that the user can fully trust the outcome of the AI. This quadrant has recently seen much progress, with the ability to quickly train a general model with context-specific information, such as creating one's own ChatGPT. Normally installed on a secure and private service instance, organizations here upload corporate documentation, like product descriptions or engineering files, to provide context and industry knowledge to the model. Still, the main efficiency gain of using GenAI in this cluster stems from building on powerful general models that can be enhanced easily by context-specific information to establish trust in the results.

### **Question 3: Do We Have the Right AI Capabilities Within the Innovation Process?**

Our discussion of the first two questions proposes that organizations navigating the deployment of GenAI for innovation face a nuanced challenge, balancing concerns about inaccuracies, especially in critical tasks, with the benefits of GenAI's creativity during ideation. The trustworthiness of GenAI hinges on understanding the specific requirements of different innovation tasks and adjusting its accuracy accordingly. This involves discerning between versatile general models, exemplified by widely used systems like ChatGPT, and task-specific expert models tailored to handle domain-specific complexities. Firms need to ensure GenAI aligns effectively with the intricacies of the specific use cases on the task level within their innovation process.

There is a third question organizations have to ask next to understanding the required trust on a task level and the difference between general and expert models: How to assess and develop the human capabilities of their innovation teams (Gama & Magistretti, 2023; Igna & Venturini, 2023; Kemp, 2024). Given the speed with which these technologies are evolving, the question of integration within the organization becomes more critical. As Siemens demonstrates, a top-down approach of strategically investing in internal development and strategic partnerships is one approach. However, these initiatives take time, and the latency of implementation may result in technology deployment that is already dated. Hence, bottom-up, more democratized approaches are required, too, where teams and individuals select, use, and build tools as they see fit. While this comes with challenges within an organization, such as control and security, this "citizen development" approach fosters speed and agility—and is primarily supported by the abilities of GenAI (Davenport et al., 2023).

*Merck*, the German healthcare and electronics company, provides an example in this regard. A long-time user of conventional AI for laboratory automation, Merck has developed more than 300 predictive models of compound properties through past experimentation with various AI systems. One of their primary goals in adding GenAI (LLMs) to their AI portfolio was to create an interface to these specialized

discovery systems, similar to Siemens' HiSimcenter, in a very top-down, strategic approach. In parallel, Merck started a bottom-up exploration with GenAI by giving all 64,000 employees early access to ChatGPT (in a secure private cloud environment). Providing broad access was seen as the best way to increase GenAI literacy. To get people excited about the possibilities of GenAI, access was phased in at a rate of a few thousand users per week, and in return, users were asked to participate in local communities of practice and share their GenAI experiments. By fostering peer learning and collaboration, the company expects to increase R&D productivity by 50% over the next few years through more effective use of traditional AI systems. We suggest combining both approaches to allow companies to strategically build better data and trusted solutions while enabling dynamic experimentation. However, much future research is needed to explore this proposition and many of the contingencies of building trust in an AI output in the context of innovation.

## **Conclusions and Outlook**

Incorporating GenAI into the innovation process presents a unique set of opportunities and challenges. As previously stated, the most effective means of leveraging GenAI is understanding the subtle nuances of trust, the distinction between general and expert models, and the alignment of human capabilities with AI tools. The decision-making process regarding the degree and manner of trust to be accorded to AI outputs is not merely a technical consideration; instead, it is a strategic one that can influence the trajectory of innovation within an organization. Our investigation underscores the significance of contextual factors in determining the optimal degree of trust in AI-generated outcomes. In the ideation and discovery phases, where creativity and unconventional thinking are paramount, the inaccuracies of GenAI can be utilized as a distinctive feature rather than a shortcoming. Conversely, in the later stages of product development, where precision and domain-specific knowledge are paramount, reliance on expert models becomes indispensable.

This dual approach to AI deployment suggests several avenues for future research. It is incumbent upon scholars and practitioners alike to investigate the manner in which disparate industries and organizational contexts shape the equilibrium between creativity and accuracy in AI-generated outcomes. Furthermore, the ever-changing landscape of AI capabilities, like the current emergence of AI agents, necessitates further investigation to ascertain the optimal ways of utilizing these tools to enhance decision-making and innovation. In this context, Hagendorff et al. (2023) suggest studying LLMs' behavior and reasoning abilities. They propose to engage LLMs in behavioral experiments that have traditionally been aimed at understanding human cognition and behavior to better understand their "reasoning" and factors that generate a specific response. Our findings highlight the necessity of a strategic and multifaceted approach to AI integration for practitioners. It is imperative that organizations not only invest in the requisite technological infrastructure but also cultivate the human expertise necessary to interpret and leverage AI outputs effectively—as well as the human expertise to know the nature of the task in the first place. The potential of GenAI to democratize



innovation processes through “citizen development” initiatives, as well as the risks associated with inadequate training and oversight, require careful consideration.

In conclusion, as GenAI continues to reshape the innovation landscape, the critical challenge for researchers and practitioners is to develop frameworks that balance the inherent trade-offs between creativity and accuracy, generalization and specialization, and automation and human judgment. By addressing these challenges head-on, we may come closer to unlocking the full potential of GenAI for innovation.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: This paper is based on research supported by a National Science Foundation grant (#2050052) and a grant by Deutsche Forschungsgemeinschaft (DFG) (#390621612).

## ORCID iD

Frank T. Piller  <https://orcid.org/0000-0003-2532-4020>

## References

- Acar, O. A., Tuncdogan, A., van Knippenberg, D., & Lakhani, K. R. (2024). Collective creativity and innovation. *Journal of Management*, 50(6), 2119-2151. <https://doi.org/10.1177/01492063231212416>
- Bilgram, V., & Laarmann, F. (2023). Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Engineering Management Review*, 51(2), 18-25. <https://doi.org/10.1109/EMR.2023.3272799>
- Bouschery, S., Blazevic, V., & Piller, F. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management*, 40(2), 139-153. <https://doi.org/10.1111/jpim.12656>
- Bouschery, S. G., Blazevic, V., & Piller, F. T. (2024). Artificial Intelligence-augmented brainstorming: how humans and AI beat humans alone. Available at SSRN 4724068.
- Chen, W., You, Z., Li, R., Guan, Y., Qian, C., Zhao, C., & Sun, M. (2024). Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence. arXiv:2407.07061.
- Cooper, R. G., & Brem, A. M. (2024). The adoption of AI in new product development: Results of a multi-firm study in the US and Europe. *Research-Technology Management*, 67(3), 44-53. <https://doi.org/10.1080/08956308.2024.2324241>
- Davenport, T. H., Barkin, I., & Tomak, K. (2023). We're all programmers now. *Harvard Business Review*, 101(5), 98-107.
- Füller, J., & Hutter, K. (2024). Add correct reference to Füller article in the same JABS issue.
- Gama, F., & Magistretti, S. (2023). Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. *Journal of Product Innovation Management*, 136, 103081. <https://doi.org/10.1111/jpim.12698>

- Guzik, E. E., Byrge, C., & Gilde, C. (2023). The originality of machines: AI takes the Torrance test. *Journal of Creativity*, 33(3), 100065. <https://doi.org/10.1016/j.yjoc.2023.100065>
- Hagendorff, T., Dasgupta, I., Binz, M., Chan, S., Lampinen, A., Wang, J., Akata, Z., & Schulz, E. (2023). Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods. arXiv:2303.13988.
- Igna, I., & Venturini, F. (2023). The determinants of AI innovation across European firms. *Research Policy*, 52(2), 104661. <https://doi.org/10.1016/j.respol.2022.104661>
- Ji, Z., Yu, T., Xu, Y., Lee, N., Ishii, E., & Fung, P. (2023). Towards mitigating LLM hallucination via self reflection. In Findings of the Association for Computational Linguistics (EMNLP 2023). 1827–1843.
- Kemp, A. (2024). Competitive advantage through artificial intelligence. *Academy of Management Review*, 49(3), 618-635. <https://doi.org/10.5465/amr.2020.0205>
- Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2023). The No. 1 question to ask when evaluating AI tools. *MIT Sloan Management Review*, 64(3), 27-30.
- Marion, T. J., Moghaddam, M., Ciuccarelli, P., & Wang, L. (2023). AI For user-centered new product development: From large-scale need elicitation to generative design. In *The PDMA handbook of innovation and new product development* (4th ed., pp. 425-442). Wiley.
- Marion, T., Srour, M., & Piller, F. (2024). When Generative AI meets product development. *MIT Sloan Management Review*, 66(1).
- McKinsey (2023). *The state of AI in 2023: Generative AI's breakout year*. McKinsey & Company.
- Meincke, L., Mollick, E. R., & Terwiesch, C. (2024). Prompting diverse ideas: Increasing AI idea variance. arXiv:2402.01727.
- Piller, F., Bouschery, S., & Blazevic, V. (2023). Hybrid intelligence for innovation: Augmenting NPD teams with artificial intelligence and machine learning. In *The PDMA handbook of innovation and new product development* (4th ed., pp. 407-425). Wiley.
- Roberts, D. L., & Candi, M. (2024). Artificial intelligence and innovation management: Charting the evolving landscape. *Technovation*, 136, 103081. <https://doi.org/10.1016/j.technovation.2024.103081>
- Salvador, F., & Sting, S. (2022). How your company can encourage innovation from all employees. *Harvard Business Review Online*, September 19, 2022, <https://hbr.org/2022/09/how-your-company-can-encourage-innovation-from-all-employees>.
- Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., & Gui, T. (2023). The rise and potential of large language model based agents: A survey. arXiv:2309.07864.
- Yuan, C., Marion, T., & Moghaddam, M. (2022). Leveraging end-user data for enhanced design concept evaluation: A multimodal deep regression model. *Journal of Mechanical Design*, 144(2), 021403.