Business Horizons (2024) 67, 561—570

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY
www.journals.elsevier.com/business-horizons

Innovating by prompting: How to facilitate innovation in the age of generative AI

Leif Sundberg*, **Jonny Holmström**

Swedish Center for Digital Innovation (SCDI), Department of Informatics, Umeå University, Umeå, Sweden

KEYWORDS
Prompt engineering;
Iterative prompting;
ChatGPT;
Generative AI;
AI and innovation;
Large language models

Abstract This article focuses on how recent advances in artificial intelligence (AI), particularly chatbots based on large language models (LLMs), such as ChatGPT, can be used for innovation purposes. The article begins with a brief overview of the development and characteristics of generative AI (GenAI). Elaborating on the implications of GenAI, we provide examples to demonstrate four mechanisms of LLMs: translation, summarization, classification, and amplification. These mechanisms inform a framework that highlights how LLMs enable the creation of innovative solutions for organizations through capacities in two dimensions: context awareness and content awareness. The strength of LLMs lies in the combination of capacities in both these dimensions, which enables them to comprehend and amplify content. Four managerial suggestions are presented, ranging from starting out with small-scale projects and data exploration, to scaling through integration efforts and educating prompt engineers. By presenting the framework, recommendations, and examples of use cases in various contexts, the article contributes to the emerging literature on GenAI and innovation.

1. Introduction

Recent breakthroughs in artificial intelligence (AI) involve the use of models trained on vast amounts of data with the ability to generate content, such as text, images, and code. These models fall under the umbrella term *generative AI* (GenAI; Dwivedi et al., 2023) and have been rapidly taken up since 2022, with popular platforms including ChatGPT, Midjourney, DALL-E 2, and Codepilot. This article focuses specifically on chatbots based on large language models (LLMs), such as the widely popular ChatGPT. As these technologies are still emerging, academic literature on their use in business and organizational contexts is scarce but growing (see, for example, Kanbach et al., 2023; Ritala et al., 2023). Many scholars argue that LLMs

* Corresponding author
E-mail addresses: leif.sundberg@umu.se (L. Sundberg), jonny.holmstrom@umu.se (J. Holmström)

https://doi.org/10.1016/j.bushor.2024.04.014

are double-edged swords. They can have major benefits in diverse fields, such as healthcare (Cascella et al., 2023), finance (Leippold, 2023), education (Kasneci et al., 2023), and business (Teubner et al., 2023). But they must be used with caution, as they are prone to presenting inaccurate information that may sound plausible owing to so-called hallucinations, as the models are unaware of the boundaries of their own knowledge (Azamfirei et al., 2023; Hannigan et al., 2024; Shen et al., 2023).

The debate surrounding GenAI and LLMs is also characterized by polarization, as overly optimistic hype is combined with dystopian warnings of machines increasingly taking over human functions (Hautala & Heino, 2023; Sabherwal & Grover, 2024). Moreover, owing to the emerging status of these technologies, they have been scarcely explored in previous studies on innovation (Dwivedi et al., 2023; Mariani et al., 2023). In line with previous research devoted to exploring the relationship between AI and innovation (Paschen et al., 2020), management perspectives on AI and machine learning (ML; Kietzmann & Pitt, 2020), and emerging notions of hybrid practices between humans and "machine designers" (Recker et al., 2023), we focus on how GenAI, and more specifically LLMs, can support innovation. As outlined later in this article, generating random content using GenAI is not necessarily innovative. Instead, creative and systematic approaches to generating value via prompting need to be more distinctly outlined (Robertson et al., 2024). Therefore, this article explores how GenAI can be used to stimulate innovation by engaging with academic and grey literature, and it offers practical examples of prompt engineering. Against this backdrop, the research question (RQ) addressed in this article is: How do large language models facilitate innovation?

2. The development and characteristics of GenAI and large language models

As mentioned, the discourse surrounding GenAI and LLMs is polarized, with contributors predicting both hyperbolic utopias and dystopias (Hautala & Heino, 2023; Sabherwal & Grover, 2024). We argue that these predictions—including scenarios involving the development of general, omnipotent AI systems—may hinder more productive attention to the urgent issues and possibilities associated with contemporary technologies. Clearly, even the most sophisticated AI cannot drive innovation if used inappropriately or without a clear understanding of its capabilities and limitations. Thus, in the following subsections, we outline GenAI's development and characteristics so as to provide more nuanced foundations for discussing how these technologies can be used to facilitate innovative solutions in various contexts.

2.1. From deep learning to GenAI

Turing (1950, p. 442) predicted that "at the end of the century…one will be able to speak of machines thinking without expecting to be contradicted." But providing a machine with the ability to understand human language was more complex than it seemed in the 1950s, as it required understanding not merely of sentence structure but also matter and context. This posed a challenge to pure behaviorist views of humans. As Russell and Norvig observed (2020, p. 31), "behaviorism discovered a lot about rats and pigeons but had less success in understanding humans."

Today's AI systems have increasing autonomy, learning capacities, and inscrutability (Berente et al., 2021), as the use of ML and of deep learning (DL) models supports innovation in various domains. The origins of DL can be traced back to the advent of regression methods in the 19th century (Cramer, 2002), and to the development of important algorithms in the second half of the 20th century (Fukushima, 1988; Goodfellow et al., 2016; LeCun et al., 1989; LeCun et al., 2015; Mackenzie, 2017; Russell & Norvig, 2020). In the 1960s, the first chatbot, Joseph Weizenbaum's ELIZA, was developed. While this early natural-language processing (NLP) invention was impressive at the time, it relied on scripted computer responses and had no understanding of context (Kietzmann & Park, 2024). So-called expert systems also appeared in the late 1960s, and some became commercially successful in the 1980s. These were the first examples of attempts to apply AI in industries, but many of these systems disappeared at the end of the 1980s and in the early 1990s owing both to "failure to meet technical performance or economic objectives" and to managerial issues (Gill, 1995, p. 51).

Today, acceleration of increases in computational processing power and access to large datasets, which have given rise to a field called big data (Günther et al., 2017; Lycett, 2013), have enabled previously invented algorithms to "learn" and have thus greatly accelerated the development of AI since the turn of the millennium

(Russell & Norvig, 2020). A key event that led to a paradigm shift was when Alex Krizhevsky and colleagues won the ImageNet Large Scale Visual Recognition Challenge in 2012 with a neural-network-driven ML model called AlexNet (Krizhevsky et al., 2017) that outperformed competitors by a large margin. This was followed by a commitment to big data and DL, not only for image data but also for music, 3D objects, and text (Denton et al., 2021). It also spurred rapid development of NLP models, such as BERT and GPT-3, as respectively described by Devlin et al. (2018) and Brown et al. (2020).

Contemporary breakthroughs in AI include advances not only in the ability to classify contents of images and text but also in the use of neural networks to create them. Generative adversarial networks allow two neural networks to generate and evaluate images, as demonstrated by Goodfellow et al. (2020). Since its development, the use of GenAI has expanded into areas such as coding, as with Codepilot, as well as to text-to-image conversion using tools such as DALL-E 2, Midjourney, and Stable Diffusion, and even to automated image generation.[1] Emerging AI techniques have also sparked controversy arising from deepfakes and copyright issues (Kietzmann et al., 2020). In 2022, an AI-generated painting won an art contest (Dataconomy, 2022), and an AI-generated imitation of a photograph won the creative open category prize in the 2023 Sony World Photography Awards (The Guardian, 2023). These events raise profound questions about the capabilities and limits of AI and prompt revisitation of Turing's ideas about thinking machines. A definitive breakthrough in GenAI occurred in 2022 with the widespread use of a chatbot capable of creating humanlike texts and conversations, namely ChatGPT.

2.2. ChatGPT

OpenAI's GPT-3 is based on a statistical model that learns by reading text. It was released as a prototype by OpenAI on November 30, 2022. Just two months after launch, the chatbot had reached 100 million users, a record for user uptake during the internet era. It took Instagram over 2 years and TikTok 9 months to reach the same number of users (Reuters, 2023).

When it receives a prompt from a user, ChatGPT outputs text based on probabilities, employing a degree of randomness ("temperature"). The model has foundational knowledge about language, enabling it to respond conversationally as it performs various generative tasks. While previous language models such as BERT relaxed the size of datasets needed to train text classifiers, new models have the potential to completely skip the training part: Their generative abilities can be used immediately. Similar LLMs include Bloom, PaLM, Chinchilla, and Llama 2.

These models display sophisticated combinatory properties (Holmström, 2018; Yoo et al., 2010), enabling users to generate different combinations of content through a prompt. In theory, there is no limit to the kinds of linguistic tasks these models can handle if users can formulate appropriate instructions by prompting. Thus, interest in prompting practices has rapidly increased, and several strategies to establish desired states of LLMs and to generate relevant outputs have been developed—for example, by providing a model with step-by-step instructions or examples (a "few-shot" approach), rather than trying to get it to generate something without any context (a "zero-shot" approach). Thus, there has been a surge of interest in *prompt engineering*: a refinement of human communication with GenAI models (inputs) in order to maximize the relevance of outputs (see, for example, Liu & Chilton, 2022). Thus, merely blindly generating content with these models is not likely to lead to innovative outcomes; rather, generating value from GenAI requires creative and systematic approaches (Zamfirescu-Pereira et al., 2023). To assist such approaches, researchers have provided both frameworks (Lo, 2023; Zhou et al., 2022) and examples of how to create organizational value with LLMs. For example, Bouschery et al. (2023) show how ChatGPT can not only perform various tasks during innovation processes, like text summarization and sentiment analysis, but also contribute to the generation of insights and ideas useful for an organization (see Short & Short, 2023).

At the beginning of 2023, the successor to GPT-3, GPT-4, was released. The documentation provided by its developers, OpenAI (2023), explicitly states that OpenAI will not disclose the data sources used to train the model, but other developers are more open about how their models were created. For example, the team behind GPT-SW3, a Swedish LLM that has been trained specifically on Swedish language data, has provided the sources of its training data—including academic articles, books, code, and math in various media, online forums, and Wikipedia—both at Medium.com (Sahlgren, 2022) and in more detail in an article by Öhman et al. (2023). It is important for managers to understand the types of data that an LLM is trained on so they

[1] Automated image generation is exemplified by websites such as thispersondoesnotexist.com

can be aware of any associated biases or constraints, particularly in contexts that demand transparency and accountability.

3. How large language models facilitate innovation

While LLMs often have good general knowledge in many areas, they rarely know the specific details of a given task. Thus, it is important to provide these details when prompting them to maximize the accuracy and value of their output (Lo, 2023). Prompting is an iterative activity, and various tactics can be used to improve the output of LLMs such as ChatGPT.

The ChatGPT web interface enables users to prompt the model for one-time specific tasks. In examples presented in this section, we relied upon the application programming interface (API) of ChatGPT 3.5 Turbo and a Google Colab notebook, using free OpenAI and Google accounts. Instructions for doing this and similar examples can be found via free online resources such as DeepLearning.AI (2023) and Coursera (2023). We identify four types of mechanisms, presented in Table 1 and further outlined in Sections 3.1—3.4: translation, summarization, classification, and amplification.

3.1. Translation

ChatGPT provides capacities for several basic functions, such as translation between languages, proofreading, changing formats (e.g., converting HTML to JSON, changing reference formats in academic papers), and establishing appropriate writing tones (for example, in business letters or scholarly writing). Superficially, this may not seem revolutionary since reasonably good translation tools have been available for several decades, but LLMs are more like Swiss Army knives than specialized scalpels, as they can translate a wide range of content to other formats and tones.

- Examples of use cases: content generation, making educational materials accessible to a wider audience (Baidoo-Anu & Ansah, 2023), and improving communication between doctors and patients (Khan et al., 2023).

3.2. Summarization

One of the most common applications for LLMs is summarizing large bodies of text. Typical tactics include asking a model to adhere to a specific word limit or to focus on a particular topic in the text. In the example presented below, an LLM was asked to generate a product description for an IKEA (2023) BILLY bookcase to be used on a website. The input was information from IKEA's website about the BILLY bookcase, and the output was a description for a marketing team. Delimiters such as triple quotes, backticks, and dashes are used to isolate the input from the prompt instructions (See Box 1).

Table 1. Explored mechanisms and exemplary use cases

Type	Description	Examples of use cases
Translation	Translation between languages, changing formats, and switching between tones.	Content generation, automatic translation services, making educational materials accessible to a wider audience, improving communication between doctors and patients.
Summarization	Summarizing bodies of text (i.e., making them more concise).	Cataloging and metadata generation, automated summarization of medical journals.
Classification	Categorizing and identifying different types of content within a text.	Annotating data (e.g., to train ML or DL models), detecting hate speech, identifying important features in a text (e.g., clauses in legal documents).
Amplification	Reaction, production, and combination of different kinds of content adapted to specific settings.	Enhanced and personalized customer service; integrated, smart AI assistants.

Box 1

```
# IKEA example

Billy_facts = """
[descriptive text from IKEA's website, including
dimensions, materials, care suggestions,
designer and article number]"""

prompt = f"""
Your task is to help a marketing team at IKEA
create a description for the BILLY bookshelf
based on a fact sheet from their web site.
Write a description based on the information in
the fact sheet.

Fact sheet: ```{Billy_facts}```
"""

[+ instructions to print the response]
```

This prompt led to the following output (See Box 2):

Box 2

Introducing the BILLY bookshelf - the booklover's choice that has been a timeless classic since its launch in 1979. With an impressive estimate of one BILLY bookcase sold every five seconds worldwide, it's no wonder why it's a popular choice. The adjustable shelves allow you to adapt the space between them according to your needs, making it a versatile storage solution for any room. Whether you have limited space or need a foundation for a larger storage solution, the BILLY bookshelf has got you covered. The BILLY bookshelf comes with one fixed shelf and four adjustable shelves, but you can also add extra shelves or a height extension unit for added storage space. You can even complete the look with doors available in different colours and designs. The BILLY bookshelf is made of high-quality materials, including particleboard, paper foil, melamine foil, and fibreboard. It's easy to clean with a cloth dampened in a mild cleaner. The product is also possible to recycle or use for energy recovery, making it an eco-friendly choice. With a width of 80 cm, depth of 28 cm, and height of 202 cm, the BILLY bookshelf requires a minimum ceiling height of 205 cm. Different wall materials require different types of fixing devices, so make sure to use fixing devices suitable for the walls in your home, sold separately. Designed by Gillis Lundgren, the BILLY bookshelf is a classic piece of furniture that never goes out of style. Get yours today with article number 002.638.50.

While correct, the text is not very useful, as it is simply a large paragraph that summarizes the information already provided. So we revised the prompt through several iterations, and we added further instructions, like:

- Format the output in HTML with the titles "Description,", "Materials" and "Dimensions."
- Add features such as target group (e.g., "customers").
- Add a table presenting the product's dimensions in centimeters.
- At the end of the description, present the designer's name and article number.

This resulted in the following output (See Box 3):

Box 3

Description
Introducing the BILLY bookshelf, a timeless choice for booklovers since 1979. With its adjustable shelves, you can easily adapt the space between shelves to suit your needs. Whether you have a limited space or require a larger storage solution, the BILLY bookshelf is the perfect foundation. It's no wonder that one BILLY bookcase is sold every five seconds worldwide. Add doors, extra shelves, or a height extension unit to customize and maximize your storage space. Made from high-quality materials, this bookshelf is not only functional but also environmentally friendly, as it can be recycled or used for energy recovery. Bring style and organization to your home with the BILLY bookshelf.
Materials
Basematerial: Particleboard, Paper foil, Plastic edgingSide panel: Particleboard, Paper foil, Melamine foil, Plastic edgingPlinth front: Particleboard, Paper foilBack: Fibreboard, Paint, Paper foilCare: Wipe clean with a cloth dampened in a mild cleaner.
Dimensions

Width	80 cm
Depth	28 cm
Height	202 cm

Designer: Gillis Lundgren
Article Number: 002.638.50

This is a cleaner and more targeted output, ready for publication on a website; the original output contained HTML code inappropriate for a customer-facing website.

The model's summarization function is based on changing sequences in a body of text. While it still provides a rather descriptive output, it shows how LLMs can, through iterative prompting, come to comprehend the content they are fed.

- Examples of use cases: cataloging, metadata generation (Lund & Wang, 2023), and automated summarizing of medical journals (Khan et al., 2023).

3.3. Classification

ChatGPT can be used to categorize things if prompted. For example, users can provide it with paragraphs of text and ask it to make various assessments, such as identifying positive and negative sentiments (e.g., in product reviews) and emotions (e.g., hate speech).[2] This has numerous uses, notably enabling organizations to assess rapidly what customers think about them. Returning to the example of the IKEA bookcase, information can often be obtained from a single review. We used the following review from the IKEA website as input, with instructions to classify it according to sentiment—positive or negative—and to detect whether the customer was angry or not: "I can see why the Billy bookcase has been an IKEA bestseller for decades. Great design, easy to assemble, and a good price. I would buy again."

This led to the following output (See Box 4):

Box 4

```
{ "Sentiment": "positive", "Anger": false }
```

Topics of texts can also be extracted, and news alerts created when certain topics are detected. Thus, LLMs can be used for modeling topics and for classifying tasks that previously required ML or DL modeling relying on vast amounts of domain-specific, manually labeled training data. In contrast, LLMs can make such predictions out of the box, as they have good general understanding of human language.

[2] This type of content should not be removed from the training data, as explained by the team involved in the training of the previously mentioned GPT-SW3 (Sahlgren, 2022).

- Examples of use cases: annotating data, detecting hate speech (Huang et al., 2023), and analyzing legal documents (e.g., checking them for presence or absence of certain legally binding clauses; Gandhi & Talwar, 2023).

3.4. Amplification

If translation is an example of the basic functionality of LLMs, their amplifying capacity is one of the key reasons that we refer to these models as generative. They can produce and combine content of different kinds, and hence they can explore novel—as well as more familiar—combinations of content. This is where the true strength of the technology lies. Amplification occurs when an LLM is provided with a snippet of text that a user wants to elaborate upon. In the next example, the same BILLY review was used to automate a response to the customer who wrote it.

First, ChatGPT was told its own role: We told it to act as a customer-service AI assistant at IKEA, and to keep a professional tone. Then we asked ChatGPT to adapt its response to the sentiment of the review. This led to the following output (See Box 5):

Box 5

Dear valued customer,

Thank you for taking the time to leave a review about your experience with the Billy bookcase. We are thrilled to hear that you found the design great, easy to assemble, and at a good price. Your feedback is important to us and we appreciate your kind words.

We are glad to hear that you would buy again and we hope that you continue to enjoy your purchase. If you have any further questions or concerns, please do not hesitate to contact us.

Thank you again for your review and for choosing IKEA.

Best regards,
IKEA's AI customer assistant

As with before, several settings can be tuned here. In this example, ChatGPT was instructed to keep a professional tone toward the customer. In other settings, depending on the task and context, it may be instructed to provide a tone that is more casual or that is aligned with a certain domain or business. Meanwhile, the temperature setting

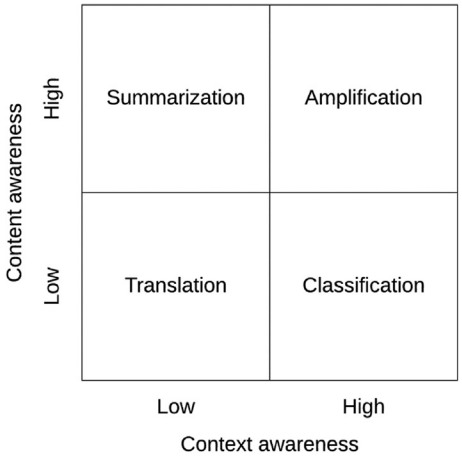
determines the randomness of the model. If the temperature is set at 0, the model will always choose the most likely of the possible answers it considers. Increasing the value allows the model to generate less likely and more divergent responses, which can assistant in producing more creative responses, albeit at the cost of predictability.

- Examples of use cases: enhancing customer service (Ferraro et al., 2024; Kanbach et al., 2023), creating smart AI assistants (e.g., web browsers and search engines, productivity tools, and educational support systems).

3.5. LLMs and innovation: A framework

Despite the evident potential of LLMs, managers must understand and contextualize the ways in which they facilitate innovation. To that end, we propose a framework for guiding LLM's strategic development and use. Specifically, we introduce a two-by-two matrix that distinguishes four mechanisms through which LLMs may facilitate innovation along two axes: context awareness and content awareness (Figure 1). Context awareness refers to a model's ability to activate knowledge based on its training data to make inferences regarding new data. Content awareness refers to a model's ability to change and repackage data it is fed into appropriate formats for the task(s) at hand. These dimensions are further explained below. It should be noted that the word "awareness" is not here an indicator of consciousness. LLMs are not conscious beings but statistical models that generate content based on probabilities. Their strength lies in their ability to mimic awareness by comprehending human language. LLMs do not understand concepts in the way that humans do.

Figure 1. The LLM innovation framework



As shown in Figure 1, translation (lower left square) does not require high levels of either dimensions, since the primary task is simply to change the language or format of given content. Summarization (upper left square) requires high content awareness, since LLMs must generate relevant output based on content given during prompting. For example, in the IKEA BILLY case, relevant content had to be generated by repackaging the input product information. These two mechanisms distinguish between two elements of what Ritala et al. (2023) refer to as ChatGPT's enhancement of knowledge work through reorganizing content (e.g., "iteratively working with users to refine textual materials for style, content, and format and to summarize, reformulate and even translate textual materials"). In classifying activities (lower right square), LLMs require high context awareness, as they must use their understanding of language to add meaning to a prompt, as exemplified by the classification of reviews expressing either positive or negative sentiments. Content awareness is not as crucial here, as classification demands a general understanding of language rather than understanding of a specific domain, such as furniture. Finally, amplification (upper right square) requires high levels of both content and context awareness, as a given input must be elaborated on and fleshed out based on specific parameters. Amplification is therefore the most potent mechanism for fostering disruptive innovation. For instance, LLMs can be used to create an AI customer assistant that adheres to the tone and system of a specific business, thereby illustrating the true strengths of GenAI in comparison to previous systems.

These examples indicate the potential disruptions that GenAI and LLMs may cause, particularly by facilitating service innovation. Thus, we can expect significant changes in ways that organizations interact with their stakeholders and customers in the near future, as LLMs are likely to disrupt business models in domains such as marketing and advertising. But the literature also demonstrates emerging use of GenAI in domains such as law, education, and medicine, where LLMs are changing how professionals conduct, communicate, and document their work. The next section presents suggestions for managers based on the framework and examples discussed in this section.

4. Suggestions for managers

Based on the outline of how LLMs facilitate innovation (Section 3), we propose four managerial

suggestions. These can be performed in phases, to advance through the four types proposed in the previous section.

First, start out small. In line with other AI tools, described for example by Sundberg and Holmström (2023), GenAI and LLMs are emerging technologies; they are rapidly changing, and we do not know how they will develop (Berthon et al., 2024). New models can process multimodal content such as images, text, and tabular data, and they can also perform various analytical tasks. Moreover, they can create excellent—albeit inconsistent—user experiences, but they are also challenging to develop, as their failures often occur silently, from the perspective of the organization. Thus, while GenAI models add value through their ability to comprehend content and add context, they need to be aligned with the organization's norms and values (Kasirzadeh & Gabriel, 2023). Nevertheless, they are rapidly becoming commodities through integration into diverse types of software, such as office suites, search engines, and AI chatbots. Thus, organizations that want to reap the benefits of using these technologies should start out small and apply them for clearly delimited purposes—for example, translating informational content (e.g., websites, code). At this stage, sandboxing environments are also important to test AI systems before deploying them.

Second, use LLMs to explore organizational data. All organizations possess textual datasets in one form or another, and both the nature and format of the datasets should be considered to optimize the future use of LLMs. Thus, an important part of engaging with GenAI is to identify and explore valuable datasets that can be used to tailor future AI assistants to given organizational contexts. This step also enables organizations to learn how to manage and adjust LLMs' outputs, which is vital to ensuring safe and relevant use of these technologies, especially in sensitive settings where silent failures may have harmful consequences (e.g., when LLMs are used to summarize medical information). Here, we see opportunities associated with the summarizing mechanism to repackage and refine textual data (e.g., internal documentation, intranet information, or Slack data, depending on the organizational structures) for further use.

Third, scale through integration. When an organization has acquired the necessary capabilities and then successfully applied GenAI for a few purposes, the systems can be further scaled through integration efforts. The versatility and flexibility of LLMs enable their application in myriads of tasks, as illustrated in notes on general-purpose technologies by Bresnahan and Trajtenberg (1995), and they can be scaled easily via APIs. These key characteristics allow patient organizations to integrate them in core business functions and thus to maximize their value-creating capacities. An important feature here is LLMs' ability to classify text data—for example, through creating labels via sentiment analysis. Via these labels, LLMs can be used to generate training data for ML and DL models for deployment in the organization's AI ecosystem, both internally and externally. During this stage, questions concerning who is responsible for the AI system's functioning become increasingly important. For example, what are service provider's responsibilities for their system's behavior, and what are its liabilities if the system should malfunction or fail to meet expectations? Clearly, organizations need to find ways to safeguard themselves against failures and potential legal issues.

Fourth, educate prompt engineers. Effective prompting requires not only good domain knowledge (Zhou et al., 2022) but also AI expertise (Zamfirescu-Pereira et al., 2023). While prompters do not need to be engineers, they do need to understand the business side of the organization and to possess technological skills. Thus, our fourth recommendation, to educate people in prompting, should be conducted in parallel with the other steps. Several online resources are available for this purpose, all of which emphasize language proficiency and the ability to communicate concisely. For organizations that want to enjoy the benefits of LLMs and to carefully monitor the relevance of their outputs, it is important to create diverse teams with data scientists and domain experts collaborating to set relevant contexts and tones for the models and to delimit their application to designated areas. At this stage, organizations are likely to experiment with their first attempts to develop a chatbot powered by GenAI. In line with our recommendation to start with small-scale projects, developing an internal chatbot for a modest designated process and well-specified domain is an appropriate first application. Summarized and classified organizational datasets, processed in line with the previous recommendations, will facilitate the creation of such a chatbot and its tailoring to the specific domain and organization. This application may serve as a stepping stone to extend the use of the chatbot to encompass enhancing external customer encounters and harnessing the power of amplification.

5. Concluding remarks

While contemporary breakthroughs in AI offer remarkable opportunities for innovation, realizing this potential requires mindful management. It

demands human oversight and a focus on human—AI collaboration. With these considerations in place, organizations and societies can leverage AI to drive truly meaningful and responsible innovation.

Against this backdrop, the aim of this study was to answer the RQ: How do large language models facilitate innovation? To answer the RQ, we described the emergence of GenAI and LLMs in relation to the broader development of AI, which is important for understanding the innovation processes involved. These distinctions were then illuminated by practical examples intended to show how LLMs facilitate innovation through four mechanisms—translation, summarization, classification, and amplification—that can be classified into two dimensions: content awareness and context awareness. Finally, the examples and framework were applied to construct four straightforward recommendations for managers: start out small, explore organizational data, scale through integration, and educate prompters.

The evolution of AI, particularly in the field of generative models, has shown great potential in facilitating the creation of innovative solutions for organizations. By understanding AI capabilities in terms of context awareness and content awareness, we can harness the power of the amplification mechanism to stimulate innovative thinking, accelerate the innovation process, and democratize access to these tools.

Facilitating innovation by using LLMs demands a mindful and strategic approach to managing and integrating AI into organizational and societal processes. The future undoubtedly holds both promises and surprises as we continue exploring this fascinating frontier.

References

Azamfirei, R., Kudchadkar, S. R., & Fackler, J. (2023). Large language models and the perils of their hallucinations. *Critical Care, 27*, 120.

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52—62.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433—1450.

Berthon, P., Yalcin, T., Pehlivan, E., & Rabinovich, T. (2024). Trajectories of AI technologies: Insights for managers. *Business Horizons, 67*(5), 461—470.

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management, 40*(2), 139—153.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics, 65*(1), 83—108.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877—1901.

Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *Journal of Medical Systems, 47*(1), 33.

Coursera. (2023). *Prompt engineering for ChatGPT*. Available at https://www.coursera.org/learn/prompt-engineering

Cramer, J. S. (2002, December). The origins of logistic regression. *SSRN*. Available at https://ssrn.com/abstract=360300

Dataconomy. (2022, September 26). *AI artwork wins art competition*. Available at https://dataconomy.com/2022/09/26/ai-artwork-wins-art-competition/

DeepLearning.AI. (2023). *ChatGPT prompt engineering for developers*. Available at https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/

Denton, E., Hanna, A., Amironesei, R., Smart, A., & Nicole, H. (2021). On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data and Society, 8*(2), 20539517211035955.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. Available at https://doi.org/10.48550/arXiv.1810.04805

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., … Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management, 71*, 102642.

Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. *Business Horizons, 67*(5), 549—559.

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks, 1*(2), 119—130.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191—209.

Gandhi, P. A., & Talwar, V. (2023). Artificial intelligence and ChatGPT in the legal context. *Indian Journal of Medical Sciences, 75*(1), 1—2.

Gill, T. G. (1995). Early expert systems: Where are they now? *MIS Quarterly, 19*(1), 51—81.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139—144.

Hannigan, T. R., McCarthy, I. P., & Spicer, A. (2024). Beware of botshit: How to manage the epistemic risks of generative chatbots. *Business Horizons, 67*(5), 471—486.

Hautala, J., & Heino, H. (2023). Spectrum of AI futures imaginaries by AI practitioners in Finland and Singapore: The unimagined speed of AI progress. *Futures, 153*, 103247.

Holmström, J. (2018). Recombination in digital innovation: Challenges, opportunities, and the importance of a theoretical framework. *Information and Organization, 28*(2), 107—110.

Huang, F., Kwak, H., & An, J. (2023). Is ChatGPT better than human annotators? Potential and limitations of ChatGPT in

explaining implicit hate speech. *arXiv*. Available at https://doi.org/10.48550/arXiv.2302.07736

IKEA. (2023). *BILLY bookcase*. Available at https://www.ikea.com/gb/en/p/billy-bookcase-white-00263850/

Kanbach, D. K., Heiduk, L., Blueher, G., Schreiter, M., & Lahmann, A. (2023). The GenAI is out of the bottle: Generative artificial intelligence from a business model innovation perspective. *Review of Managerial Science, 18*, 1189–1220.

Kasirzadeh, A., & Gabriel, I. (2023). In Conversation with Artificial Intelligence: Aligning language Models with Human Values. *Philosophy and Technology, 36*(2). https://doi.org/10.1007/s13347-023-00606-x

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274.

Khan, R. A., Jawaid, M., Khan, A. R., & Sajjad, M. (2023). ChatGPT-Reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences, 39*(2), 605–607.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons, 63*(2), 135–146.

Kietzmann, J., & Park, A. (2024). Written by ChatGPT: Large language models, conversational chatbots, and their place in society and business. *Business Horizons, 67*(5), 453–459.

Kietzmann, J., & Pitt, L. F. (2020). Artificial intelligence and machine learning: What managers need to know. *Business Horizons, 63*(2), 131–133.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. In D. S. Touretzky (ed.), *Advances in neural information processing systems 2* (pp. 396–404). San Francisco, CA: Morgan Kaufman.

Leippold, M. (2023). Thus spoke GPT-3: Interviewing a large-language model on climate finance. *Finance Research Letters, 53*, 103617.

Liu, V., & Chilton, L. B. (2022, April). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery.

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*(4), 102720.

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News, 40*(3), 26–29.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems, 22*(4), 381–386.

Mackenzie, A. (2017). *Machine learners: Archaeology of a data practice*. Cambridge, MA: MIT Press.

Mariani, M. M., Machado, I., & Nambisan, S. (2023). Types of innovation and artificial intelligence: A systematic quantitative literature review and research agenda. *Journal of Business Research, 155*, 113364.

Öhman, J., Verlinden, S., Ekgren, A., Gyllensten, A. C., Isbister, T., Gogoulou, E., Carlsson, F., & Sahlgren, M. (2023). The nordic pile: A 1.2 TB nordic dataset for language modeling. *arXiv*. Available at https://doi.org/10.48550/arXiv.2303.17183

OpenAI. (2023). GPT-4 technical report. *arXiv*. Available at https://arxiv.org/abs/2303.08774

Paschen, U., Pitt, C., & Kietzmann, J. (2020). Artificial intelligence: Building blocks and an innovation typology. *Business Horizons, 63*(2), 147–155.

Recker, J., von Briel, F., Yoo, Y., Nagaraj, V., & McManus, M. (2023). Orchestrating human-machine designer ensembles during product innovation. *California Management Review, 65*(3), 27–47.

Reuters. (2023, February 2). *ChatGPT sets record for fastest-growing user base — Analyst note*. Available at https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Ritala, P., Ruokonen, M., & Ramaul, L. (2023). Transforming boundaries: How does ChatGPT change knowledge work? *Journal of Business Strategy, 45*(3), 214–220.

Robertson, J., Ferreira, C., Botha, E., & Oosthuizen, K. (2024). Game changers: A generative AI prompt protocol to enhance human-AI knowledge co-construction. *Business Horizons, 67*(5), 499–510.

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach*. London, UK: Pearson Education.

Sabherwal, R., & Grover, V. (2024). The societal impacts of generative artificial intelligence: A balanced perspective. *Journal of the Association for Information Systems, 25*(1), 13–22.

Sahlgren, M. (2022, September 22). The Nordic pile. *Medium*. Available at https://medium.com/ai-sweden/the-nordic-pile-a8d5aaf3db60

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology, 307*(2), e230163.

Short, C. E., & Short, J. C. (2023). The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. *Journal of Business Venturing Insights, 19*, e00388.

Sundberg, L., & Holmström, J. (2023). Democratizing artificial intelligence: How no-code AI can leverage machine learning operations. *Business Horizons, 66*(6), 777–788.

Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of ChatGPT et al.: The prospects of large language models. *Business and Information Systems Engineering, 65*(2), 95–101. https://doi.org/10.1007/s12599-023-00795-x

The Guardian. (2023, April 17). *Photographer admits prize-winning image was AI-generated*. Available at https://www.theguardian.com/technology/2023/apr/17/photographer-admits-prize-winning-image-was-ai-generated

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 49*, 433–460. Available at https://www.jstor.org/stable/2251299

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research, 21*(4), 724–735.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023, April). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery.

Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. *International Journal of Computer Vision, 130*(9), 2337–2348.