Generative AI in Entrepreneurship Research: Principles and Practical Guidance for Intelligence Augmentation

Other titles in Foundations and Trends® in Entrepreneurship

Entrepreneurs' Search for Sources of Knowledge
Albert N. Link
ISBN: 978-1-63828-296-9

The Evolution of Hidden Champions as Niche Entrepreneurs
Erik E. Lehmann and Julian Schenkenhofer
ISBN: 978-1-63828-258-7

Entrepreneurship in the Long-Run: Empirical Evidence and Historical Mechanisms
Michael Fritsch and Michael Wyrwich
ISBN: 978-1-63828-108-5

Minority Entrepreneurship 2.0
Timothy Bates
ISBN: 978-1-63828-048-4

From the Metaphor to the Concept of the Entrepreneurial Journey in Entrepreneurship Research
Tõnis Mets
ISBN: 978-1-63828-016-3

Student Entrepreneurship: Reflections and Future Avenues for Research
Bart Clarysse, Philippe Mustar and Lisa Dedeyne
ISBN: 978-1-63828-012-5

Generative AI in Entrepreneurship Research: Principles and Practical Guidance for Intelligence Augmentation

Francesco Ferrati
University of Padova
francesco.ferrati@unipd.it

Phillip H. Kim
Babson College
pkim1@babson.edu

Moreno Muffatto
University of Padova
moreno.muffatto@unipd.it

the essence of knowledge
Boston — Delft

Foundations and Trends® in Entrepreneurship

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

F. Ferrati *et al.*. *Generative AI in Entrepreneurship Research: Principles and Practical Guidance for Intelligence Augmentation*. Foundations and Trends® in Entrepreneurship, vol. 20, no. 3, pp. 245–383, 2024.

ISBN: 978-1-63828-385-0

Foundations and Trends® in Entrepreneurship
Volume 20, Issue 3, 2024

Editorial Scope

Foundations and Trends® in Entrepreneurship publishes survey and tutorial articles in the following topics:

- Nascent and start-up entrepreneurs
- Opportunity recognition
- New venture creation process
- Business formation
- Firm ownership
- Market value and firm growth
- Franchising
- Managerial characteristics and behavior of entrepreneurs
- Strategic alliances and networks
- Government programs and public policy
- Gender and ethnicity
- New business financing:
  - Business angels
  - Bank financing, debt, and trade credit
  - Venture capital and private equity capital
  - Public equity and IPOs
- Family-owned firms
- Management structure, governance and performance
- Corporate entrepreneurship
- High technology:
  - Technology-based new firms
  - High-tech clusters
- Small business and economic growth

Information for Librarians

Foundations and Trends® in Entrepreneurship, 2024, Volume 20, 8 issues. ISSN paper version 1551-3114. ISSN online version 1551-3122. Also available as a combined paper and online subscription.

Contents

Generative AI in Entrepreneurship Research: Principles and Practical Guidance for Intelligence Augmentation

Francesco Ferrati[1], Phillip H. Kim[2] and Moreno Muffatto[3]

[1]*School of Entrepreneurship (SCENT), Department of Industrial Engineering, University of Padova, Italy; francesco.ferrati@unipd.it*
[2]*Arthur M. Blank Center for Entrepreneurship, Babson College, USA; pkim1@babson.edu*
[3]*School of Entrepreneurship (SCENT), Department of Industrial Engineering, University of Padova, Italy; moreno.muffatto@unipd.it*

ABSTRACT

This monograph investigates the integration of generative artificial intelligence (AI) into the academic research process of entrepreneurship. Specifically, we explore using Large Language Models (LLMs) like ChatGPT in several research scenarios to support novice and established researchers. As a practical guide, we introduce researchers to prompt engineering – formulating instructions for the LLMs to generate a desired output. We classify different types of prompts, present various technical strategies, and suggest the design of an effective prompt formula. We illustrate the prompt engineering process with different examples for entrepreneurship research.

Francesco Ferrati, Phillip H. Kim and Moreno Muffatto (2024), "Generative AI in Entrepreneurship Research: Principles and Practical Guidance for Intelligence Augmentation", Foundations and Trends® in Entrepreneurship: Vol. 20, No. 3, pp 245–383. DOI: 10.1561/0300000121.

To assist researchers in systematically integrating LLMs into their research process, we present the "4D-Framework," which consists of four phases (Discover, Develop, Discuss, and Deliver). Each phase contains four functions accomplished through four prompts, resulting in 16 functions and 64 specific prompts. The initial stage, "Discover," involves using LLMs for project initiation tasks such as topic selection and literature review, theory exploration, conceptual or empirical puzzles, and research question identification. During the "Develop" phase, the focus shifts to operational aspects, where LLMs assist in designing methods, executing qualitative and quantitative research, and generating programming code. The third phase, "Discuss," focuses on using LLMs to analyze findings, evaluate their robustness and limitations, highlight the research contribution, and identify future research directions. Finally, the "Deliver" phase emphasizes using LLMs to draft the manuscript, craft the narrative, prepare for submission, and disseminate the findings.

We describe the application of LLMs in entrepreneurship research from a human-centric perspective, emphasizing an Intelligence Augmentation (IA) perspective for harmonizing human intelligence with AI capabilities. Given the novelty and impact of LLMs in knowledge-based areas, we also address the ethical implications of using AI in academia. We urge scholars to incorporate AI and LLMs into their research responsibly. While showcasing their potential, we also address their current limitations. We empower scholars to adopt a dynamic, AI-enhanced research approach that emphasizes the potential to unlock new insights and enhance the integrity of academic research.

Keywords: Generative AI; ChatGPT; GPT; artificial intelligence; intelligence augmentation; large language models; prompt engineering; research process; entrepreneurship.

1

Introduction

While Artificial Intelligence (AI) has been predicted for years to disrupt the nature of work, the introduction of generative AI technology has changed the ways humans interact with machines and how we conduct our work (Mollick, 2024; Hinton *et al.*, 2006; Korneeva *et al.*, 2023; Vinsel, 2023). Based on advanced deep-learning architectures, users of generative AI models can produce different types of content (e.g., texts, images, videos, code) that were not part of the explicit training dataset. Specifically, Large Language Models (LLMs) have been trained on large volumes of human language data and use this information to generate new human-like text based on the patterns learned from the training data. The model output resembles what humans generate. With appropriate training, curation, and instructions, LLMs can produce contextually relevant and coherent text in a fraction of the time and effort humans normally require (Radford *et al.*, 2018; Vaswani *et al.*, 2017). With an interactive, user-friendly interface, LLMs have become one of the most quickly adopted new technologies in history (Dale, 2021; Dalalah and Dalalah, 2023). Regardless of their AI knowledge, users of all backgrounds can interact with LLMs. As a result, generative AI technology has spread rapidly into all sectors, especially knowledge-based domains, where users continue to find new ways to integrate LLMs into their work. In this monograph, we explore the academic application of LLMs and how this technology can be incorporated into entrepreneurship research.

1.1 Generative AI and Academia

LLMs have been quickly integrated into research protocols in a wide range of scholarship. We highlight several studies in this initial wave of LLMs-powered academic research (Dwivedi *et al.*, 2023; Lund and Wang, 2023). For example, LLMs have been explored in finance as tools to support different stages in a research project (Dowling and Lucey, 2023), to investigate how financial sentiment analysis is vulnerable to adversarial attacks that alter financial texts (Leippold, 2023), and to assess LLMs as financial robo-advisors using a financial literacy test (Niszczota and Abbas, 2023). Other studies have examined how LLMs can assist economists (Korinek, 2023), legal scholars (Biswas, 2023; Liga and Robaldo, 2023), or biomedical scientists (Luo *et al.*, 2022) in their investigations. LLMs have been incorporated into education research to their impact on educators and students (AlAfnan *et al.*, 2023; Cotton *et al.*, 2023; Duong *et al.*, 2023; Lim *et al.*, 2023; Ratten and Jones, 2023; Rudolph *et al.*, 2023; Tlili *et al.*, 2023; Vecchiarini and Somià, 2023; Winkler *et al.*, 2023).

As researchers begin to apply Generative AI techniques, we highlight early efforts to integrate this technology in the entrepreneurship field. For example, LLMs have been analyzed as a tool for enhancing human-led innovation teams in new product development (Bouschery *et al.*, 2023), to create entrepreneurial content that mimicked established patterns (such as pitches in the style of prominent CEOs) (Short and Short, 2023), to enhance organizational operations and decision-making processes (Ayinde *et al.*, 2023), and to identify external enablers of entrepreneurship (Davidsson and Sufyan, 2023). Compared to other research areas, the use of AI-based research techniques in entrepreneurship is still in its early stages (Ferrati and Muffatto, 2021), despite the call for a broader integration of these tools into research methodologies and study designs (Hain and Jurowetzki, 2020; Lévesque *et al.*, 2022; Obschonka and Audretsch, 2020; Schwab and Zhang, 2019). Given the widespread access to LLM technology, we build on these early efforts and convey the same urgency to entrepreneurship researchers for understanding these techniques and evaluate their applications into their research.

In this monograph, we provide practical guidance for how generative AI can be applied to entrepreneurship research and how scholars can integrate LLMs into their work. Our two objectives aim to provide researchers with guidance not only for using LLMs as a tool for data analysis and modeling but also to offer a more holistic approach to integrating AI to augment researchers' capabilities throughout the entire research process. Just as researchers need to understand AI applications in the work of entrepreneurs, we also need to reconsider how AI technologies can empower entrepreneurship researchers (Shepherd and Majchrzak, 2022). We describe several research scenarios to which our guidance directly applies.

Emily is a junior researcher in entrepreneurship starting her academic career. For someone in her situation, dealing with the extensive amount of existing literature and theories in this field can be challenging at first. Emily can use generative AI to accelerate her learning curve. The model summarizes monographs, which she skims in her first pass of potential monographs to include in a literature review. After screening for the most relevant monographs, she reads them in more detail to fully understand the arguments and findings. As she explores a topic, she asks the LLM to report the emerging trends and the unexplored areas in the research stream. She works iteratively, cycling between her reflections and the output provided by the LLM. She aims to connect the dots and formulate novel and relevant research questions for her doctoral program.

In a nearby office is José, an established researcher in entrepreneurship who is new to generative AI. He has explored how LLMs can support his traditional research workflow for a few weeks. In his research pipeline, Jose works on multiple projects simultaneously. Some are with colleagues based worldwide. Others are personal projects. He wishes to improve his research productivity and move more projects to completion. He uses the LLM in several ways. He edits an initial draft outline for a research grant due next month. He reviews the Python code he received from a colleague to increase its processing speed for a large dataset. He is also preparing a conference presentation scheduled for next week. By integrating generative AI into his workflow, Jose has found additional time to focus on his work's more complex and creative parts by delegating routine tasks to the LLM.

In another building down the street is Suwon, a computer scientist taking an interdisciplinary approach to entrepreneurship research. She wishes to find new synergies between her computer science expertise and the entrepreneurship research context. For example, she wants to explore how she can apply her machine-learning skills to analyze large entrepreneurship datasets. She looks to generative AI to help her fill different gaps in her understanding of entrepreneurship. LLMs can facilitate a personalized learning experience tailored to Suwon's specific needs and provide tips and insights that may not be initially visible to researchers outside the entrepreneurship field.

As these scenarios show, entrepreneurship researchers with different backgrounds and levels of expertise can benefit from generative AI in many ways. We write our monograph for these audiences and others who wish to use LLMs for their research under the principles of *Intelligence Augmentation*. While our focus is entrepreneurship research applications, we present our guidance in ways that can be easily adapted to other disciplines.

1.2 Artificial Intelligence and Intelligence Augmentation

Given the significant ongoing debate regarding the impact of artificial intelligence, we distinguish the concepts of Artificial Intelligence (AI) and Intelligence Augmentation (IA). AI attempts to simulate human-like reasoning and problem-solving. From self-driving cars to interactive chatbots, AI aims to develop machines that can operate autonomously, sometimes surpassing human capabilities in specific tasks. The quality of these AI machines depends on their ability to execute high-quality automation processes based on machine learning, neural networks, deep learning, and other techniques. These processes improve as they learn from repeated attempts and exposure to new data, which allows the underlying models to refine their functions over time. On the other hand, IA refers to a philosophy of cooperation between humans and machines, with the final decisions for using any machine-generated output made by humans. IA describes a human-centered design pattern that aims at enhancing, complementing, and expanding human intelligence rather than machines replacing humans and their regular functions (Skagestad, 1993; Lui and Lamb, 2018; Dellermann *et al.*, 2019; Romero-Brufau *et al.*, 2020; Ostheimer *et al.*, 2021; Van der Aalst, 2021; Vincent, 2021; Johnson *et al.*, 2022). As the label implies, IA is like having a digital assistant that whispers insights into a professional's ear, providing data-driven perspectives to inform decisions, yet always leaving the final judgment to human intuition and understanding. For example, the concept of IA has been explored as "Human-AI Collaborative Decision-Making" in organization design (Puranam, 2021) and to emphasize the paradoxical tension between automation and augmentation in management (Raisch and Krakowski, 2021). IA emphasizes the partnership between humans and machines, and the sum of their collaborative efforts can be more effective than working independently. IA equips humans with better tools to gain greater efficiency and more insight than without using them. When employed effectively and responsibly, researchers can amplify their skills through IA. For example, the labor-intensive tasks required to prepare a dataset can now be more easily automated, freeing time for researchers to tackle higher-order research activities. Researchers can generate complex programming code or synthesize published scholarly works more easily using IA than a manual process. However, even in the context of IA, human-machine cooperation can occur at different levels and according to different paradigms.[1]

In this context, Generative AI allows entrepreneurship researchers to connect to Artificial Intelligence and Intelligence Augmentation principles. When given a prompt, LLMs can independently craft responses or content to demonstrate the ability of AI to recreate the complex steps of the research process. Moreover, the essence of AI lies in the

[1]Mollick (2024) has described this as a "centaur VS cyborg" metaphor (https://www.oneusefulthing.org/p/centaurs-and-cyborgs-on-the-jagged).

machine's ability to learn from data. LLMs have acquired the ability to learn language structures, semantic codes, and other knowledge configurations through training on a large amount of text. These features offer new possibilities for entrepreneurship researchers to expand or deepen their scholarly work. At the same time, researchers are still responsible for deciding how, when, and to what extent they incorporate LLMs output into their final research products. In this way, LLMs support researchers through Intelligence Augmentation principles but do not replace them completely.

One reason for using LLMs under Intelligence Augmentation principles is their current limitations. Like other AI technologies, entrepreneurship researchers should carefully evaluate the circumstances and conditions for when and how to employ LLMs (Burtsev *et al.*, 2023). While LLMs can provide information grounded in their training, they can miss the depth, context, or understanding a human naturally brings to a conversation. For these capabilities, researchers need to adopt tools from an adjacent field of AI called Natural Language Understanding (NLU), which focuses on machine comprehension of human language and operates with a different logic than those of LLMs (Bender *et al.*, 2021). NLU allows machines to grasp the intent behind the text and the semantics, sentiment, and context in which the words are employed. Since they lack consciousness, self-awareness, and personal experiences, LLMs do not "understand" in the same way humans do. LLMs generate their output-based syntactic coherence from their training data. Depending on patterns detected in the learning phase, LLMs use previous words to predict the next word in a sequence. LLMs can produce remarkably coherent and contextually relevant text from patterns they learned during training rather than a semantic understanding of the language. For this reason, when using LLMs in the context of critical thinking, entrepreneurship researchers should consider them as brainstorming tools that can trigger and guide human judgment (Lindebaum and Fleming, 2023). They should also be prepared to take full responsibility for any LLM output that becomes integrated into the research design, results, and communication.

Another factor for researchers to consider is the *creative* capability of LLMs (Romera-Paredes *et al.*, 2023). Creativity in a human context entails original ideas, inspiration, imagination, and a specific emotional or aesthetic sensibility. LLMs, on the other hand, lack emotions, thoughts, and consciousness (Guo *et al.*, 2023). What might look like creative LLM output originated from the model's capacity to develop unique combinations based on its large training dataset and identify patterns that may not be easily discerned by human researchers. LLMs may produce "creative" content because they can accept a prompt and develop a solid, contextually appropriate output that includes details not expressly indicated in the input and unexpected by researchers. Therefore, we emphasize the difference between generative models and human creativity. Generative models produce new data points based on complex patterns learned from their training data. Human creativity involves a high level of originality and uniqueness, pushing the limits of knowledge gained from experience and learning. Moreover, human creativity often consists of intentionality and purpose, whereas generative models perform what they are instructed to do. LLM responses are based on existing patterns in the data on which they were trained rather than genuine innovative thinking. Thus, entrepreneurship researchers should not expect LLMs to fully imitate or replace their creative approaches to conduct their work.

Finally, we highlight the *emotional and empathic aspects* of LLM behavior. Like creativity, LLMs lack emotional intelligence like humans do (Kosinski, 2023). They can reproduce emotionally intelligent responses because of training on massive volumes of data that include details on human feelings and emotional interaction. In this way, they learned to generate text that mimics these expressions, creating the illusion of comprehension, empathy, or feeling. However, assuming LLMs can have unique emotional or empathic responses is a misconception since LLMs do not experience emotions. Instead, they reproduce the patterns they learned. By performing a sentiment analysis, LLMs can determine whether a text is positive, negative, or neutral and create output that matches the same attitude. Although LLMs do not experience these feelings, the output can contain information about emotional qualities in the analyzed texts. Accordingly, researchers can use LLMs to identify patterns or themes related to emotions and empathy in entrepreneurial contexts. Again, we caution researchers since LLMs may misinterpret or oversimplify complex human emotional responses, especially with complex combinations of multiple emotions, contexts, individual differences, and nonverbal communications.

We advise entrepreneurship researchers to consider these limitations when using LLMs and evaluate the output from these models. While LLMs augment researcher intelligence, researchers are ultimately responsible for evaluating this information and deciding how best to use it. Researchers who work within these limitations can still benefit from the many features available in LLMs.

1.3 Objectives and Scope

LLMs are opening new avenues of research in a variety of fields. Our monograph explores current LLM applications, their implications for entrepreneurship research, and how researchers can use IA to enhance their scholarly efforts. We address the question: "How can generative AI be integrated into the academic research process of entrepreneurship?" As we tackle this question, we also engage with a broader academic evolution regarding the role of researchers and how research is conducted. In the following sections, we present a paradigm shift in research conceptualization, implementation, and dissemination from the potential synergies between human and artificial intelligence.

We cover the following objectives. First, we take readers through the basics of constructing prompts – the building blocks for working with LLMs – and how researchers can employ prompt engineering for different purposes and stages of the research process. Second, we introduce the Discover-Develop-Discuss-Deliver (4D) framework for integrating prompts and generative AI tools into the research workflow. Third, we offer both possibilities and precautions for maximizing the strengths of LLMs responsibly and ethically.

We write this monograph for several audiences. For beginners to Generative AI and LLMs, we provide an overview of how this technology works and enable you to get started with the basics of prompt engineering in entrepreneurship research. For intermediate users with some experience with prompt engineering, we offer examples you can use to adapt to your research situations. For non-entrepreneurship researchers, we describe techniques for using Generative AI and LLMs to learn more about the entrepreneurship field and integrate this work into your current research program. These techniques can also be adapted for your primary research field. For entrepreneurship researchers curious about integrating Generative AI and LLMs into your workflow, we use the 4D framework to discuss how these tools can automate, expand, or deepen your research capabilities.

Our research framework allows entrepreneurship researchers to take advantage of the interdisciplinary application of LLMs. The entrepreneurship researcher community is multidisciplinary, with scholars from various backgrounds, including economics, psychology, sociology, and engineering. Since researchers may have different perspectives on LLMs, we aim to provide general guidance that is compatible with this diversity. With modifications, many of the proposed techniques can also be used in other research fields. However, we use entrepreneurship-specific research applications to illustrate our techniques and to minimize the researcher's transpositional effort from more general guides of LLMs. Our goal is to equip entrepreneurship researchers to leverage LLMs for their research.

Finally, given the fast-paced development of AI technology, we recommend readers use our monograph as a compass and to guide them in exploring an innovative intelligence-augmented research methodology instead of as a rigid blueprint. We anticipate new versions of LLM technology will contain advancements that improve existing capabilities. We do not claim to cover all the potential features that present and future LLMs offer for entrepreneurship research. We recommend researchers use the latest public versions of their preferred LLM application to take advantage of the most recent developments. This will usually require enrollment in a commercial (paid) or enterprise version, which offers software updates and other releases helpful to researchers. We do not recommend being solely dependent on free versions since they are unlikely to incorporate current updates and have been trained on current data. This can lead to poor outputs and discourage ongoing reliance on LLMs. We also advise readers to keep updated with current trends and insights in Generative AI by consulting resources that track the latest features and provide relevant commentary on their applications (Mollick, 2024). We write our monograph not as a complete technical reference but as an exploration of the potential of adopting a dynamic AI-powered research mindset using the principles we discuss. By becoming more familiar with generative AI tools, we aim to help scholars conduct more rigorous, relevant, and innovative research and discover new insights regarding the wide-ranging topics in entrepreneurship.

1.4 Disclaimers

Before we move into the details of applying LLMs to entrepreneurship research, we express caution to our readers. LLM adoption has occurred at an unprecedented speed. OpenAI and its ChatGPT technology hold the record for the fastest-growing consumer application in history, with one million users in just five days of its launch in November 2022[2] and over 100 million active monthly users in its first two months after launch. Other companies have raced to introduce LLM software as an alternative to ChatGPT. While many sectors have quickly integrated these tools into their regular operations, academia will likely be much slower in adopting LLMs and other IA tools into its practices. Before proceeding to the next section, we highlight three key disclaimers: the varying LLM-use policies of journal publishers, the privacy of the input data provided to LLMs, and the accuracy of the output generated by LLMs.

1.4.1 On Journal Policies

Since the introduction of LLMs and other AI-powered research tools, the academic community has neither fully defined nor accepted a common policy for integrating these tools into the research process. This is especially relevant for how researchers use LLMs to prepare manuscripts for publication. Researchers and their organizations, journal editors and reviewers, and journal publishers may have different views on what can be done. The evolution of these norms will likely occur at varying speeds

[2] https://twitter.com/gdb/status/1599683104142430208?lang=en.

across and within each stakeholder group. Moreover, each discipline may have unique LLM-policy features that differ from other disciplines. Academic norms and ethical guidelines for new practices such as IA will likely take time to emerge. Until then, researchers may encounter unclear and inconsistent boundaries for what is appropriate and what violates generally accepted norms for conducting and communicating research.

For example, according to Grimes *et al.* (2023), researchers must wrestle with the uncertainties associated with two conditions – societal regulation of AI and AI systems transparency and the four possible scenarios these conditions produce. In scenario 1 (low regulation and transparency), the questionable credibility of AI poses a threat to academic professional integrity, leading to an increased emphasis on rewarding authentic human expertise. In scenario 2 (low regulation and high transparency), making AI systems more transparent toward scholarly work can foster greater acceptance of AI in academia. However, the inevitably rapid increase in knowledge production challenges the integrity of the current reward system in the academic profession. In scenario 3 (high regulation and low transparency), rigorous regulations discourage the widespread adoption of Al in scholarship, and the academic profession cautiously allows and rewards intelligence-augmented research. Finally, in scenario 4 (high regulation and transparency), academic AI adoption starts slowly and gains momentum exponentially over time, leading the profession to reward this form of knowledge production with verification and impact. The relevance of these scenarios to individual researchers depends on how local and general norms evolve.

At the time of writing this monograph, publishers of entrepreneurship journals have begun to outline requirements for submitting research manuscripts. Policies regarding "Generative-AI" use can be found in their Guides for Authors. According to a BMJ study (Ganjavi *et al.*, 2024), 24 of the world's 100 major publishers - responsible for more than 28,000 journals - have specific policies on generative AI. Typically, journals with generative AI policies allow authors some level of use of LLMs if they are acknowledged transparently. On the other hand, reviewers are not usually granted permission to use LLMs to complete their work. To evaluate the range of policies for entrepreneurship journals, we read the policies of major academic publishers of entrepreneurship research. In general, these guidelines included being transparent in acknowledging any use of AI tools during the research process, including preparing the manuscript, banning LLMs as co-authors (Teubner *et al.*, 2023) and being entirely responsible for the content of their manuscript, including the pieces produced by an AI tool. Specifically, some publishers referred to the Committee on Publication Ethics (COPE)[3] position statement on AI tools. For this monograph, we have deliberately chosen not to provide specific policies for different publishers because of the rapidly changing nature of this issue and to avoid conveying outdated and misleading information to readers. However, we advise researchers to review these policies carefully before integrating these techniques into their research process, remembering that what applies to one journal may be different for another. Over time, we anticipate guidelines will become more specific as the academic research community converges on ethical practices, norms, and boundaries for Generative AI. Until these details are established, we encourage researchers to use AI tools sensibly and responsibly and avoid taking shortcuts to produce quick results.

Throughout this monograph, we will offer different recommendations for how researchers can use our techniques to accomplish this. For example, we advise researchers to enable future replication of any data analysis by verifying the results through internal replication. Researchers will be responsible for the integrity of their work, and we counsel Generative AI users to take appropriate steps to ensure this.

We also wish to convey that the contents of this monograph should in no way be interpreted as legal advice or legitimizing the use of generative AI in research. The study of the legal or ethical aspects

[3]Many leading publishers and journals have adopted the positions offered by the Committee on Publication Ethics (COPE) regarding ethical research practices, which has recommended at the time of this writing that "Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of the paper, or in the collection and analysis of data, must be transparent in disclosing in the Materials and Methods (or similar section) of the paper how the AI tool was used and which tool was used. Authors are fully responsible for the content of their manuscript, even those parts produced by an AI tool, and are thus liable for any breach of publication ethics." https://publicationethics.org/cope-position-statements/ai-author.

of using this technology in research is a rapidly emerging area but outside the scope of our monograph. Instead, we offer principles that could guide entrepreneurship researchers interested in using augmented intelligence to support their work and inform future examinations into policies on the appropriate use of LLMs in academic research, especially in entrepreneurship.

1.4.2 On Data Privacy

When using Generative AI tools, we urge researchers to exercise care regarding data privacy and confidentiality. If you plan to conduct any analysis (for both qualitative and quantitative data) with AI tools, please carefully review the data privacy policies for the software provider. Free or paid individual access plans may incorporate any information submitted in prompts (e.g., datasets, interview notes, unpublished manuscripts) into future training LLMs. Enterprise plans may provide better privacy protections. On-premise LLMs offer the best privacy protections since the software is installed locally within your organization, and the information submitted to the local LLM does not leave the "premise" and provides researchers complete control over the information during the analysis. Researchers should evaluate the tradeoffs regarding LLMs' processing speed, accuracy, costs, and other issues when determining their data privacy and confidentiality arrangements. University ethics boards and institutional review boards (IRBs) may also impose additional guidelines for researchers to anticipate and integrate into their research designs.

1.4.3 On Output Accuracy

Our experience with computers and software has taught us to expect accurate and precise outcomes. Given the complexities of such processes, we rarely challenge the machine's result. The use of LLMs challenges this approach. One of the most critical issues confronting LLM systems now is the phenomenon of AI hallucination (Bender *et al.*, 2021). The expression "hallucination" refers to circumstances in which LLMs provide output that appears convincing but is inaccurate or not founded in reality. When conducting research, a hallucination might be a fictitious response, an inaccurate definition, a reference to a nonexistent academic publication, or a nonfunctional programming language code. As explained in the Appendix, and as anticipated earlier in talking about NLU, this behavior is due to how these models work. At the time of this writing, LLMs are not designed to understand the content like humans do. Instead, they detect and replicate text patterns that are statistically significant in the training data without understanding what the words represent. For this reason, all outputs produced by LLMs, whether in textual or numerical form, must be properly verified, especially in research and knowledge production. In fact, according to the logic of augmented research, LLMs should support and not substitute the researcher's activity. In exploring the approach presented in this work, we urge the reader to always critically evaluate the results produced by the LLMs since the author is responsible for the published content.

Appendix

A

Technical Notes

We provide a basic primer in this Appendix for how LLMs work technically. We include these details to help researchers grasp the essentials of the technical foundations of LLMs. These explanations are not intended to replace more advanced explanations available elsewhere. We encourage researchers to investigate these sources for specific questions or details not covered in this primer. This may be necessary when evaluating methodology design choices, data privacy guidelines, or how to convey study parameters.

A.1 Discriminative AI and Generative AI

To fully appreciate the disruptive innovation introduced by generative AI, we start with a comparative overview of discriminative AI, which has accounted for almost all the application to date (Haenlein and Kaplan, 2019).

Discriminative AI models distinguish between distinct types of outputs (i.e., groups, classes, labels) given specific inputs (i.e., variables, features). They are trained to respond to questions such as "Given this new element that you have never seen before, does it belong to class A or class B?" For example, an email spam filter is trained to respond to the question "Given this new email, is it spam or a regular message?" These types of tasks represent examples of *classification* problems. In machine learning, a model can be training in a supervised mode to learn the optimal decision boundaries between several categories of data (Obschonka and Audretsch, 2020). In a simple classification, models are trained to determine if a case should be labeled as category 1 (yes) or category 0 (no). Logistic regression, support vector machines, random forest and most deep neural networks are examples of discriminative models (Mahesh, 2020). Discriminative AI models are typically built to perform very specific tasks, e.g., detecting objects in pictures, or forecasting market prices. In this regard, they can be considered as task-specific models. i.e., a model is developed to optimally perform on a single specific task (e.g., recognizing objects in images). These models would not work other contexts (e.g., forecasting market prices) since they are not designed or trained for these classifications. Since AI models learn from data, high-performing classification models need to be trained on data that is closely relevant to the task that it is designed to execute. While this approach has been used to power many successful AI applications, it has also significant limitations. Since the models are task specific, each new task requires collecting and labeling new training data, building new models, and training them on the new specific task.

To overcome these limitations, researchers have introduced *foundation* models as a new paradigm for building AI systems. The expression was popularized by the Stanford Institute for Human-Centered Artificial Intelligence's (HAI) Center for Research on Foundation Models (CRFM) (Bommasani *et al.*, 2021). Instead of developing unique models each specific to a single task, foundation models are pre-trained initially on a large corpus of data (to capture a broad understanding of human language or vision or audio, etc.) and then fine-tuned with a smaller, task-specific dataset to perform a particular task more accurately without extensive task-specific data or architecture changes. In this way, building a foundation model is like establishing the foundation for a building. The foundation is created first and is general-purpose to support a variety of structures built on top of it. In the same way, a foundation model is pre-trained on an extensive dataset as the base for

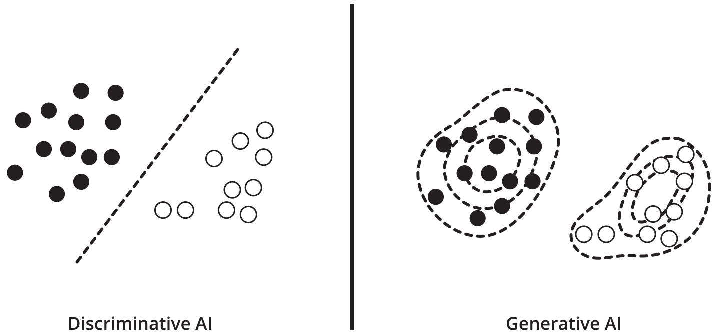

Figure A.1: Discriminative AI and generative AI.

various potential task-specific models. From this perspective, discriminative AI models are specialized use cases of foundation models.

A properly fine-tuned foundation model can complete extremely specific tasks effectively but requires less computation since the foundation model leverages transfer learning – i.e., skills acquired for one task can improve the model's performance on other different tasks. As a result, foundation models can be more effective, flexible, and practical for users. After pre-trained on a vast amount of diverse data, foundation models can detect previously unknown patterns underlying the data and can use this knowledge to generate new, similar data. In this case, a foundation model can also be considered a generative AI model (i.e., generative foundation model) when it produces new, original, and coherent content such as text, images, or audio. It is worth emphasizing that many foundation models are generative AI models, but not all generative AI models are foundation models. For example, some generative AI models may be specifically developed and trained for a single task, rather than being pre-trained on a large corpus of data and then fine-tuned for specific tasks. As represented in Figure A.1, while a discriminative AI model learns the optimal decision boundaries between two or more classes of data during training, a generative AI model learns the true data distribution of the training set to generate new data points with some variations.

A.2 Large Language Models – Architecture

Generative foundation models can be pre-trained on large amounts of data of many forms, such as text, images, or audio. When trained using textual data, a generative foundation model is called a Large Language Model (LLM).[1] LLMs are state-of-the-art generative AI systems capable of processing and comprehending massive amounts of human language data. These models can generate contextually relevant and coherent text, similar to the text on which they have been trained. LLMs typically use a type of neural network architecture called a transformer. To better understand the working mechanisms of LLMs, we describe the main techniques and models used in Natural Language Processing (NLP) that laid the technical foundations for LLMs.

A.2.1 Natural Language Processing Before Large Language Models

NLP draws on computer science, artificial intelligence, and linguistics principles that enable computers to understand, interpret, and generate human language meaningfully and practically (Chowdhary and Chowdhary, 2020). Speech recognition, smart assistants, sentiment analysis, machine translation, autocomplete, and autocorrect are examples of NLP applications widely used in practice. In entrepreneurship and organizational research, NLP techniques have been employed to analyze archival texts and other large corpora of textual data (e.g., Croidieu and Kim, 2018).

NLP relies on finding an appropriate numerical representation of textual data. This conversion can be performed using different strategies considering the text's semantic attributes. The Bag-of-Words (BoW) is one of the most basic techniques in NLP (Harris, 1954). A text (such as a sentence or an entire document) is represented as a bag (multiset) of its words. The numerical representation focuses on the

[1]We note all LLMs are foundation models (since they are trained on large amounts of data and may be fine-tuned for a variety of tasks), but not all foundation models are LLMs (as some may be learned on non-text data). On the other hand, LLMs are a subset of generative AI models that are specifically designed and trained to generate "new data points" in the form of sequences of words or sentences.

frequency of words, while grammar and word order are ignored in the BoW. Another conversion technique called Term Frequency-Inverse Document Frequency (TF-IDF) represents text as vectors where each word's value is replaced by a score that considers a word's frequency and how unique the word is to a particular document (Robertson, 2004). This numerical statistic reflects how important (i.e., uncommon) a word is to a document in a collection of documents (or corpus).

While both NLP techniques extract features from texts (for later use, for example, with machine learning algorithms), they do not track word order. Researchers have developed the word embedding technique for semantic text analysis to interpret contextual details. Words are represented by dense vectors (word embeddings) in a high-dimensional space so that words with similar meanings are represented by vectors that are close to each other, while dissimilar words are represented by vectors that are far away. Word embeddings have the advantage of capturing the context of a word in a document, semantic and syntactic similarity, and even relationships with other words. Models learn word-embedding patterns from large corpora of texts (e.g., all Wikipedia monographs). These embedding can then be created using popular methods such as Word2Vec (developed by researchers at Google) (Mikolov *et al.*, 2013a,b), GloVe (developed by researchers at Stanford) (Pennington *et al.*, 2014), and FastText (developed by researchers at Facebook) (Bojanowski *et al.*, 2017). These word vectors capture semantic meanings and relationships between words. For example, Word2Vec can resolve analogies, such as "the king is to the queen as man is to woman," by performing simple vector arithmetic: "king" − "man" + "woman" ≈ "queen."

Another key task performed in NLP is sequence modeling, which predicts or generates a subsequent data point (or sequence of data points) based on a provided sequence. Because these are data sequences, the word order and their relationship are critical to the accuracy of the output. N-grams capture the local structure of sentences by considering not just individual words, but also how words appear together in pairs, triples, etc. (Brants *et al.*, 2007; Moore and Lewis, 2010). N-grams are contiguous sequences of n items (e.g., words) from a given text sample.

Based on the previous N-1 words, N-gram models anticipate the next word in a sequence so that the prediction comes with some context.

Although very useful in many language tasks, N-grams are limited in capturing long-range dependencies between words, can suffer from data sparsity, and are computationally inefficient as N increases. Hidden Markov Models (HMMs) address these limitations (Rabiner, 1989). An HMM statistical model assumes a system as a Markov process with unobserved (hidden) states. HMMs are frequently used in NLP for part-of-speech tagging and other sequence prediction applications. HMMs determine the likelihood of a sequence of words or tags by considering both the likelihood of individual words or tags and the likelihood of transitions between them. Unlike N-grams, HMMs have states and transitions that allow them to record more complicated dependencies. HMMs assume that the current state depends only on the previous state (the Markov property) and not on the sequence of states that preceded it, which can limit their application for certain tasks. Moreover, HMMs rely on strong statistical assumptions to model the probability distributions of sequences of observable data and hidden states.

To address the limitations of HMM, Recurrent Neural Networks (RNNs) and their advanced variants like Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Units (GRU) (Cho *et al.*, 2014) provide an alternative approach. RNN is a type of neural network, and as such, it does not start from any statistical assumptions. Instead, RNNs learn to map sequences of inputs to sequences of outputs based on patterns in the training data. This gives them the flexibility to model complex relations and dependencies over time. RNNs have internal states that pass from one step in the sequence to the next, allowing them to capture temporal dependencies and maintain an internal memory. In fact, due to their recurrent structure, RNNs could theoretically remember all past information, although in practice they often struggle with long-term dependencies due to issues like vanishing or exploding gradients (vanishing gradients occur when the updates needed for learning shrink and learning becomes more difficult. Exploding gradients happen when these updates become excessively large, leading to unstable learning).

A.2.2 The Transformer Architecture

Transfomer-based architectures are the next advancements in machine learning to overcome the limitations of their predecessor sequence-to-sequence models like RNNs and LSTMs (Vaswani *et al.*, 2017). This innovative neural network architecture represents a critical breakthrough in the field of NLP and serves as the cornerstone of LLMs (Qiu *et al.*, 2020). The transformer architecture contains two main elements, an encoder and a decoder (in fact, the encoder-decoder structure represents the heart of the transformer architecture), both of which consist of multiple identical layers, each with two primary components: a multi-head self-attention mechanism and a position-wise fully connected feed-forward network.

The model receives a text as input. The input text is given to an *input embedding* layer which converts the input tokens[2] into vectors. These vectors are high-dimensional representations learned throughout the training phase and serve as the foundation for the model's knowledge of word relationships. Since the transformer model does not inherently capture the order of tokens in a sequence (unlike RNNs or LSTMs do), it employs *positional encodings* to provide relative or absolute location of tokens in the context. In this way, the input embeddings and the positional encodings are incorporated to form a single input vector for each token.

[2]In the field of NLP, a **token** represents a single unit of data or an instance of a sequence in a specific text. *Tokenization* is the process of breaking down text into individual tokens which allows computers to better interpret and analyze text. The specific definition of a "token" varies depending on the considered level of granularity. In its most intuitive case, one token corresponds to a single word. However, tokens are not always one word. For example, Byte Pair Encoding (BPE), is a technique of sub-word tokenization (also used by OpenAI's GPT models). With BPE, a single English word might represent many tokens for the model. From a technical perspective, BPE replaces the most common pair of consecutive bytes in the data repeatedly and statistically. It begins by tokenizing at the character level. Then, it combines frequent pairings of symbols iteratively, building a vocabulary of larger and longer sequences of characters. Since it can break them down into recognized sub-words, BPE can help with the problem of out-of-vocabulary (OOV) words, which are words that were not observed in the training corpus and may not be in the model's vocabulary. This makes the language model more flexible as well as capable of comprehending and generating a broader range of words.

The embedded inputs enter a set of identical layers known as *encoders*. Each *encoder* consists of two sublayers: a multi-headed *self-attention mechanism* and a position-wise fully connected *feed-forward neural network*. A residual connection surrounds each of the two sub-layers, followed by layer normalization.

In the *self-attention mechanism*, the model computes a score to evaluate the relevance of various input tokens in the context of other tokens in the sequence while generating an output token (i.e., how they affect one another). Specifically, within the multi-head attention mechanism, the input is transformed into three vectors called *query*, *key*, and *value* vectors. The compatibility of each token with every other token is calculated by executing a dot product on the *query* and *key* vectors, followed by a softmax operation to guarantee the weights add up to 1. To obtain a weighted representation, these weights are then multiplied with *value* vectors and summed to produce the output of the *self-attention* sub-layer. The procedure is repeated several times (multi-head) with different learned linear input projections. This technique essentially lets the model decide on which parts of the input sequence to "pay attention" when generating an output, hence the name "attention."

Following the *self-attention* sub-layer, the outputs move through a position-wise fully connected *feed-forward network* composed of two linear transformations (a kind of simple type of neural network), with a ReLU (Rectified Linear Unit) activation function in between. The role of this module is to refine the representations produced by the self-attention layers before passing it on to succeeding layers (or to the output layer in the case of the final layer in the stack). This allows to add an extra level of abstraction to model complex patterns within the input data. The same feed-forward network is applied to each position independently (position-wise). Despite being a fully connected network, it does not share parameters between positions in the sequence.

Finally, each sub-layer (self-attention and feed-forward) is surrounded by a residual connection before layer normalization. This contributes to the stabilization of the learning process and reduces training time.

The *decoder* is also made up of identical layers and has a similar structure as the *encoder* but with an additional third sub-layer. In fact, in addition to the two sub-layers found in the encoder (i.e., self-attention and feed-forward), *decoders* also include a third sub-layer called *masked multi-headed self-attention mechanism.* In this case, the self-attention mechanism is "masked" to maintain the auto-regressive property. This means that the attention scores are calculated so that each position is restricted to paying attention to earlier positions in the output sequence, excluding future positions from consideration. This masking is critical to preventing the model from "cheating" by making predictions based on future tokens in the sequence. The decoder's last layer is a linear layer followed by a softmax function to build a probability distribution over the target vocabulary for each incoming token. As the output for that time step, the token with the highest probability is chosen.

During the training phase, the transformer uses a specific practice called "teacher forcing," specifically concerning the input to the decoder. The model is trained to predict each token (word, for example) in the output sequence by observing the tokens that come before it. During training, the correct output sequence is given to the decoder, but it is "shifted right" by one position, i.e., each token is used to predict the next one. This approach allows the model to learn how to create the token at each position by only looking at the tokens that came before it. For example, consider the sequence *"I am writing an academic paper."* When provided into the decoder as input during training, the sequence would be shifted right, like so: *"<start> I am writing an academic."* and the model would be trained to predict the output *"I am writing an academic paper."* On the one hand, it should be noted that this "shifted right" technique is only used during the training phase. On the other hand, when the model generates new text, the decoder generates one token at a time and provides its previous outputs as input to the next step.

A.3 Large Language Models – How They are Developed

Since the transformer architecture was introduced, several tech giants have developed their LLMs in a race to establish the top performers and to introduce them as commercially viable products. Current examples of transformer-based LLMs include GPT by OpenAI, BERT by Google, LLaMA by Meta, Megatron by Nvidia, and Claude by Anthropic. In this section, we present the sequence of activities required to develop an LLM (Naveed *et al.*, 2023; Zhao *et al.*, 2023)

A.3.1 Data Collection and/or Selection

Building LLMs is a sophisticated process that starts with the essential foundation of any machine-learning project: large amounts of high-quality data. LLMs must be trained on an immense amount of textual data to learn statistical patterns of human language. Over the years, several organizations have conducted efforts to automatically collect online textual data into massive datasets to advance NLP technology. Many of these datasets are publicly available. A flexible model capable of performing multiple language tasks (e.g., answering factual inquiries, summarizing monographs, translating texts) requires various sources with different types of content and complexity. To appreciate the volume and complexity of these information sources, we list some of the datasets used to train current LLMs: CommonCrawl, WebText, WebText2, C4 (Colossal Clean Crawled Corpus), Book Corpus, Books1, Books2, Pile - Books3, Wikipedia, SQuAD 1.1 (Stanford Question Answering Datasets 1.1: Q&A), SQuAD 2.0 (Stanford Question Answering Datasets 2.0: Reading Comprehension), and SWAG (Situations with Adversarial Generations).

A.3.2 Model Design

Language models can be created using different deep-learning architectures. As previously discussed, transformer-based models that employ attention processes are the preferred architectures because they have been shown to be especially effective for language modeling tasks. Once the transformer architecture has been selected, users need to decide the number of layers, the number of attention heads, and the model's size (in terms of parameters), which are critical choices when creating language models. These judgments are frequently based on empirical findings from earlier studies, and model size (number of parameters) is usually as large as the available computational resources allow. The term "large" in LLMs points to two key aspects: the size of the massive training datasets and the number of parameters in these datasets. Parameters are the weights in the model's various layers defined through training and considered internally adjustable components that enable the model to learn from the data and generate human-like responses when optimally set. LLMs often include tens to hundreds of billions of parameters, which allow them to capture complex patterns in the data and perform sophisticated text generation tasks. The number of parameters varies significantly throughout models, with the more advanced models having more parameters than their early predecessors. The BERT model, for instance, uses 340 million parameters, whereas the GPT-3 model has 175 billion parameters, and the GLAM model 1.2 trillion parameters. Although increasing the number of parameters may lead to improved performance (although this is not always the case), it also leads to an increase in model complexity, resulting in longer computation time and higher energy consumption. As a result, model performance may not always improve with more parameters. This tradeoff should be carefully evaluated at the model design stage.

A.3.3 Pre-Training

Model development involves multiple steps. In machine learning, model development generally involves training, cross-validation, and test phases. LLMs require another preliminary phase called pre-training as their first step. Pre-training is a technique that is applied across many machine-learning domains, including but not limited to LLMs and generative AI. The core concept behind pre-training is to use large datasets to learn generic patterns that can later be used to start learning more specific tasks (through a subsequent fine-tuning process). Given their extensive scale, pre-training datasets are often unlabeled or only semi-structured. At the pre-training stage, the model learns from a vast quantity of text data without specific tasks as unsupervised learning or, more specifically, self-supervised learning. Unsupervised learning is a subset of machine learning that uses unlabeled input to enable a model to independently explore patterns and structures in data. Within this field, the pre-training phase of LLMs is generally considered an instance of self-supervised learning. Self-supervised learning[3] is a type of unsupervised learning in which a model learns representations from data by predicting some portions of the data based on other parts. In this scenario, the "labels" are produced automatically from the data. For example, in the case of LLMs, the models learn to predict the next word in a sentence based on the previous ones. The task of predicting the next word (the "label") given the previous words (the "input") is generated by the data itself, thus the expression "self-supervised." The pre-training approach helps build versatile models to understand and generate coherent human-like text. The pre-trained model can further learn to perform specific tasks effectively through the subsequent fine-tuning process. A corpus of billions or even trillions of sentences is used extensively to enable the model to learn a good "initial" representation of human language, including grammar, syntax, common phrases, and some general information about the world. The model learns to effectively predict the next word in a sentence, given the previous ones.

A.3.4 Fine-Tuning

In the LLM development process, fine-tuning follows the initial pre-training stage to refine the pre-trained model using a narrower, task-specific dataset. During fine-tuning, at least one internal model parameter of the pre-trained model is specifically trained. This iterative process aligns the general-purpose model's behavior with the desired output, such as performing translations, generating stories, or completing other tasks. The main advantage of fine-tuning is achieving better model performance, while only requiring a smaller number of manually labeled examples than models relying on supervised training.

There are three common approaches to fine-tune a pre-trained model: self-supervised, supervised, and reinforcement learning. Specifically, Reinforcement Learning from Human Feedback (RLHF) is a particularly effective fine-tuning technique. It combines elements of traditional reinforcement learning and supervised learning but can be tailored to

[3] While all self-supervised learning is a form of unsupervised learning, not all unsupervised learning is self-supervised.

suit the challenges and needs of language models. Traditionally, reinforcement learning involves an agent taking actions in an environment to maximize a cumulative reward. However, in the context of LLMs, determining the exact "reward" for a given piece of text can be extremely difficult. To overcome this challenge, humans are brought into the training process (human-in-the-loop). Instead of relying on a preset metric, humans provide feedback that serves as the reward signal. The pre-trained model generates a series of responses to a set of prompts. Human evaluators review and rank these model-generated responses based on quality or accuracy. For example, raters can rank two or more responses to the same prompt. The human feedback forms the basis of a reward model. This model can predict the reward (or human preference score) for new model-generated outputs. With the reward model in place, techniques like Proximal Policy Optimization (PPO) can be used to fine-tune the language model. The model is incentivized to generate outputs that would score higher based on the reward model. The model can be fine-tuned more precisely with multiple iterations of new human feedback.

A.3.5 Performance Evaluation

After fine-tuning, the next phase of LLM development is performance evaluation (Chang *et al.*, 2023). This phase ensures LLMs operate as intended and meet required standards for accuracy, coherence, and applicability in real-world scenarios. Evaluation includes several methods and metrics, each tailored to test different aspects of LLMs' capabilities.

The first approach is automated evaluation, which employs predefined metrics to measure LLMs' predictive ability, fluency, and grammatical correctness. Key metrics in this category include Perplexity, BLEU, ROUGE, and the F1 Score. Perplexity measures how well an LLM can predict the next word based on the prior context, with lower scores indicating better performance. BLEU (Bilingual Evaluation Understudy) was originally used for machine translation quality assessment but is now also applied to assess the fluency and grammatical correctness of generated text. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is primarily used in summarization tasks to evaluate the overlap between model-generated text and reference text. The F1 Score balances precision and recall, which is particularly important in scenarios where both false positives and false negatives are significant.

In addition to automated evaluation, the second approach is human evaluation. This involves having evaluators assess LLM-generated text based on criteria such as relevance to the given prompt or context, logical coherence, topic consistency, grammatical correctness, stylistic consistency, and vocabulary diversity.

The third approach occurs when testing in real-world scenarios. This involves deploying an LLM in practical applications and monitoring its performance in actual user interactions and specific settings. This type of testing is essential for gathering feedback on the LLM's effectiveness and appropriateness in real-world situations. The evaluation also reveals the model's efficacy in tasks it is specifically designed for to complete.

As these evaluations are conducted, LLMs are checked for their ethical uses, bias or fairness in their results, and their security robustness. The goal is to ensure the model operates impartially, dissuasively, and resistant to malicious inputs of any kind, particularly regarding sensitive issues. Finally, the scale and effectiveness of LLMs are also evaluated. When assessing the performance of LLMs, it is necessary to test them with a variety of input data sizes and measure the computing power required for optimum operation.

A.4 Barriers to Entrance

Since LLMs require tremendous resources to build, the most notable developments have occurred in well-funded corporations. This raises complex questions about the future trajectory of Generative AI regarding control, oversight, and applications.

Training LLMs is computationally intensive, necessitating advanced hardware like Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). While these devices excel in parallel processing tasks essential for neural network computations, they come at a significant cost. State-of-the-art LLMs often require clusters of these units, multiplying the expense. A cluster is essentially a collection of interconnected GPUs or TPUs that work in tandem, distributing the computational load amongst themselves. This distributed approach allows for the concurrent processing of different portions of a task, vastly accelerating the training process. While this might sound like a straightforward solution, it requires significant financial resources to acquire cutting-edge GPUs or TPUs. A single unit can cost thousands of dollars. LLM training might involve hundreds of these, so the costs can quickly escalate into the millions. Such financial implications can be prohibitive for startups with limited capital, highlighting challenges in democratizing access to advanced AI training.

The creation of a competitive LLM is not just a matter of financial and computational resources; it's also a significant investment of time. These models, by their very nature, are vast neural networks with millions, if not billions, of parameters. Each parameter must be adjusted during the training process to minimize prediction errors, and doing so requires processing vast amounts of data, sometimes multiple times in repeated epochs. With model's intricate architecture and the enormous datasets, conducting an iterative refining process requires time. Also, post-training, the model undergoes further rigorous evaluation, fine-tuning, and potential retraining phases based on its performance on unseen data. Each of these stages adds to the overall time commitment for a robust and efficient LLM. This temporal barrier becomes even more apparent when considering the pace of advancements in the AI field. The state-of-the-art today might become obsolete in a matter of months, meaning that a model trained over a long period might face stiff competition from newer, more advanced models even prior to its deployment.

LLMs has brought with it not only technological advancements but also heightened environmental concerns. As the computational demands of these models have grown, so too has their energy consumption, placing them squarely in the crosshairs of environmental debates (Henderson *et al.*, 2020). In fact, when used in clusters for extended periods, GPUs and TPUs can consume electricity equivalent to that of small towns. For example, the training of one BERT base model was calculated to require as much energy as a trans-American flight even without considering hyperparameter tuning (Bender *et al.*, 2021).

This energy consumption, especially when sourced from non-renewable resources, results in a significant carbon footprint from LLM providers. Environmentalists and concerned researchers have raised alarms about the tradeoff spawned by these advanced technologies that on the one hand, promise a better future and, on the other, negatively impact the environment amid concerns over global climate change and to reduce carbon emissions (Strubell *et al.*, 2019).

Finally, another element that prevents new entrants is the availability of talented human resources. The development of LLMs requires specialized knowledge in deep learning, neural networks, and NLP. The surge in AI popularity has produced a notable talent shortage in these specialized domains. This mismatch between demand and supply has driven up compensation for AI experts, often sidelining smaller entities or startups that can't match the offers of well-funded tech giants. Consequently, new advancements in AI technologies may become more concentrated within a few resource-rich organizations at the expense of the broader AI industry.

References

AlAfnan, M. A., S. Dishari, M. Jovic, and K. Lomidze (2023). "Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses". *Journal of Artificial Intelligence and Technology*. 3(2): 60–68.

Aldrich, H. E. and P. H. Kim (2007). "Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search". *Strategic Entrepreneurship Journal*. 1(1–2): 147–165.

Anderson, N., D. L. Belavy, S. M. Perle, S. Hendricks, L. Hespanhol, E. Verhagen, and A. R. Memon (2023). "AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise medicine manuscript generation". *BMJ Open Sport & Exercise Medicine*. 9(1): e001568.

Ayinde, L., M. P. Wibowo, B. Ravuri, and F. B. Emdad (2023). "ChatGPT as an important tool in organizational management: A review of the literature". *Business Information Review*. 40(3): 137–149.

Baker, T. and F. Welter (2018). "Contextual entrepreneurship: An interdisciplinary perspective". *Foundations and Trends® in Entrepreneurship*. 14(4): 357–426.

Bansal, P. and K. Corley (2011). "The coming of age for qualitative research: Embracing the diversity of qualitative methods". *Academy of Management Journal*. 54(2): 233–237.

Basta, C., M. R. Costa-Jussà, and N. Casas (2019). "Evaluating the underlying gender bias in contextualized word embeddings". arXiv preprint arXiv:1904.08783.

Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell (2021). "On the dangers of stochastic parrots: Can language models be too big?" In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 610–623.

Biswas, S. S. (2023). "Potential use of Chat GPT in global warming". *Annals of Biomedical Engineering*. 51(6): 1126–1127.

Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov (2017). "Enriching word vectors with subword information". *Transactions of the Association for Computational Linguistics*. 5: 135–146.

Bommasani, R., D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, . . ., and P. Liang (2021). "On the opportunities and risks of foundation models". arXiv preprint arXiv:2108.07258.

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

Bradshaw, T. (2022). "Is becoming a prompt engineer the way to save your job from AI?" URL: https://www.ft.com/content/0deda1e7-4fbf-46bc-8eee-c2049d783259.

Brants, T., A. C. Popat, P. Xu, F. J. Och, and J. Dean (2007). "Large language models in machine translation". In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. Association for Computational Linguistics, Prague, Czech Republic. 858–867.

Buchanan, J., S. Hill, and O. Shapoval (2024). "ChatGPT hallucinates non-existent citations: Evidence from economics". *The American Economist*. 69(1): 80–87.

Burtsev, M., M. Reeves, and A. Job (2023). "The working limitations of large language models". *MIT Sloan Management Review*.

Chang, Y., X. Wang, J. Wang, Y. Wu, K. Zhu, H. Chen, . . ., and X. Xie (2023). "A survey on evaluation of large language models". arXiv preprint arXiv:2307.03109.

Cho, K., B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation". arXiv preprint arXiv:1406.1078.

Chowdhary, K. and K. R. Chowdhary (2020). "Natural language processing". In: *Fundamentals of Artificial Intelligence*. 603–649.

Cotton, D. R., P. A. Cotton, and J. R. Shipway (2023). "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT". In: *Innovations in Education and Teaching International*. 1–12.

Croidieu, G. and P. H. Kim (2018). "Labor of love: Amateurs and lay-expertise legitimation in the early US radio field". *Administrative Science Quarterly*. 63(1): 1–42.

Dalalah, D. and O. M. Dalalah (2023). "The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT". *The International Journal of Management Education*. 21(2): 100822.

Dale, R. (2021). "GPT-3: What's it good for?" *Natural Language Engineering*. 27(1): 113–118.

Davidsson, P. and M. Sufyan (2023). "What does AI think of AI as an external enabler (EE) of entrepreneurship? An assessment through and of the EE framework". *Journal of Business Venturing Insights*. 20: e00413.

Dellermann, D., P. Ebel, M. Söllner, and J. M. Leimeister (2019). "Hybrid intelligence". *Business & Information Systems Engineering*. 61: 637–643.

Dowling, M. and B. Lucey (2023). "ChatGPT for (finance) research: The bananarama conjecture". *Finance Research Letters*. 53: 103662.

Du, Y. and P. H. Kim (2021). "One size does not fit all: Strategy configurations, complex environments, and new venture performance in emerging economies". *Journal of Business Research*. 124: 272–285.

Duong, C. D., T. N. Vu, and T. V. N. Ngo (2023). "Applying a modified technology acceptance model to explain higher education students' usage of ChatGPT: A serial multiple mediation model with knowledge sharing as a moderator". *The International Journal of Management Education*. 21(3): 100883.

Dwivedi, Y. K., N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, I. K. Kar, ..., and R. Wright (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.

Ferrati, F. and M. Muffatto (2021). "Entrepreneurial finance: Emerging approaches using machine learning and big data". *Foundations and Trends® in Entrepreneurship*. 17(3): 232–329.

Ganjavi, C., M. B. Eppler, A. Pekcan, B. Biedermann, A. Abreu, G. S. Collins, ..., and G. E. Cacciamani (2024). "Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: Bibliometric analysis". *Bmj*: 384.

Grant, A. M. and T. G. Pollock (2011). "Publishing in AMJ—Part 3: Setting the hook". *Academy of Management Journal*. 54(5): 873–879.

Grimes, M., G. von Krogh, S. Feuerriegel, F. Rink, and M. Gruber (2023). "From scarcity to abundance: Scholars and scholarship in an age of generative artificial intelligence". *Academy of Management Journal*. 66(6): 1617–1624.

Grudin, J. (2022). *From Tool to Partner: The Evolution of Human-Computer Interaction*. Springer Nature.

Guo, C., Y. Lu, Y. Dou, and F. Y. Wang (2023). "Can ChatGPT boost artistic creation: The need of imaginative intelligence for parallel art". *IEEE/CAA Journal of Automatica Sinica*. 10(4): 835–838.

Haenlein, M. and A. Kaplan (2019). "A brief history of artificial intelligence: On the past, present, and future of artificial intelligence". *California Management Review*. 61(4): 5–14.

Hain, D. S. and R. Jurowetzki (2020). "The promises of machine learning and big data in entrepreneurship research". In: *Handbook of Quantitative Research Methods in Entrepreneurship*. Edward Elgar Publishing. 176–220.

Hansson, S. O. (2007). "What is technological science?" *Studies in History and Philosophy of Science Part A*. 38(3): 523–527.

Harris, Z. S. (1954). "Distributional structure". *Word*. 10(2–3): 146–162.

Henderson, P., J. Hu, J. Romoff, E. Brunskill, D. Jurafsky, and J. Pineau (2020). "Towards the systematic reporting of the energy and carbon footprints of machine learning". *The Journal of Machine Learning Research*. 21(1): 10039–10081.

Hinton, G. E., S. Osindero, and Y. W. Teh (2006). "A fast learning algorithm for deep belief nets". *Neural Computation*. 18(7): 1527–1155.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". *Neural Computation*. 9(8): 1735–1780.

Hutchinson, B., V. Prabhakaran, E. Denton, K. Webster, Y. Zhong, and S. Denuyl (2020). "Social biases in NLP models as barriers for persons with disabilities". arXiv preprint arXiv:2005.00813.

Johnson, M., A. Albizri, A. Harfouche, and S. Fosso-Wamba (2022). "Integrating human knowledge into artificial intelligence for complex and ill-structured problems: Informed artificial intelligence". *International Journal of Information Management*. 64: 102479.

Kim, P. H. and H. E. Aldrich (2005). "Social capital and entrepreneurship". *Foundations and Trends® in Entrepreneurship*. 1(2): 55–104.

Kim, P. H. and M. Li (2014). "Seeking assurances when taking action: Legal systems, social trust, and starting businesses in emerging economies". *Organization Studies*. 35(3): 359–391.

Kim, P. H., K. Wennberg, and G. Croidieu (2016). "Untapped riches of meso-level applications in multilevel entrepreneurship mechanisms". *Academy of Management Perspectives*. 30(3): 273–291.

Kojima, T., S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa (2022). "Large language models are zero-shot reasoners". *Advances in Neural Information Processing Systems*. 35: 22199–22213.

Korinek, A. (2023). "Generative AI for economic research: Use cases and implications for economists". *Journal of Economic Literature*. 61(4): 1281–1317.

Korneeva, E., T. O. Salge, T. Teubner, and D. Antons (2023). "Tracing the legitimacy of artificial intelligence: A longitudinal analysis of media discourse". *Technological Forecasting and Social Change*. 192: 122467.

Kosinski, M. (2023). "Theory of mind may have spontaneously emerged in large language models". arXiv preprint arXiv: 2302.02083.

Kotha, R., P. Crama, and P. H. Kim (2018). "Experience and signaling value in technology licensing contract payment structures". *Academy of Management Journal*. 61(4): 1307–1342.

Lakatos, I. (1970). "History of science and its rational reconstructions". In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*. Vol. 1970. Cambridge University Press. 91–136.

Leippold, M. (2023). "Sentiment spin: Attacking financial sentiment with GPT-3". *Finance Research Letters*. 103957.

Lévesque, M., M. Obschonka, and S. Nambisan (2022). "Pursuing impactful entrepreneurship research using artificial intelligence". *Entrepreneurship Theory and Practice*. 46(4): 803–832.

Liga, D. and L. Robaldo (2023). "Fine-tuning GPT-3 for legal rule classification". *Computer Law & Security Review*. 51: 105864.

Lim, W. M., A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina (2023). "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators". *The International Journal of Management Education*. 21(2): 100790.

Lindebaum, D. and P. Fleming (2023). "ChatGPT undermines human reflexivity, scientific responsibility and responsible management research". *British Journal of Management*.

Locke, K. and K. Golden-Biddle (1997). "Constructing opportunities for contribution: Structuring intertextual coherence and 'problematizing' in organizational studies". *Academy of Management Journal*. 40(5): 1023–1062.

Lui, A. and G. W. Lamb (2018). "Artificial intelligence and augmented intelligence collaboration: Regaining trust and confidence in the financial sector". *Information & Communications Technology Law*. 27(3): 267–283.

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

Lund, B. D., T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang (2023). "ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing". *Journal of the Association for Information Science and Technology*. 74(5): 570–581.

Luo, R., L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T. Y. Liu (2022). "BioGPT: Generative pre-trained transformer for biomedical text generation and mining". *Briefings in Bioinformatics*. 23(6): bbac409.

Mahesh, B. (2020). "Machine learning algorithms-a review". *International Journal of Science and Research (IJSR)*. 9(1): 381–386.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013a). "Efficient estimation of word representations in vector space". arXiv preprint arXiv:1301.3781.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013b). "Distributed representations of words and phrases and their compositionality". *Advances in Neural Information Processing Systems*. 26.

Mollick, E. (2024). *Co-Intelligence: Living and Working with AI*. Random House.

Moore, R. C. and W. Lewis (2010). "Intelligent selection of language model training data". In: *Proceedings of the ACL, 2010 Conference Short Papers*. 220–224.

Munafò, M. R., B. A. Nosek, D. V. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, . . ., and J. Ioannidis (2017). "A manifesto for reproducible science". *Nature Human Behaviour*. 1(1): 1–9.

Naveed, H., A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, . . ., and A. Mian (2023). "A comprehensive overview of large language models". arXiv preprint arXiv:2307.06435.

Niszczota, P. and S. Abbas (2023). "GPT has become financially literate: Insights from financial literacy tests of GPT and a preliminary test of how people use it as a source of advice". *Finance Research Letters*. 58: 104333.

Nosek, B. A., T. E. Hardwicke, H. Moshontz, A. Allard, K. S. Corker, A. Dreber, . . ., and S. Vazire (2022). "Replicability, robustness, and reproducibility in psychological science". *Annual Review of Psychology*. 73: 719–748.

Obschonka, M. and D. B. Audretsch (2020). "Artificial intelligence and big data in entrepreneurship: A new era has begun". *Small Business Economics*. DOI: 10.1007/s11187-019-00202-4.

Open Science Collaboration (2015). "Estimating the reproducibility of psychological science". *Science*. 349(6251): aac4716.

Ostheimer, J., S. Chowdhury, and S. Iqbal (2021). "An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles". *Technology in Society*. 66: 101647.

Pennington, J., R. Socher, and C. D. Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.

Puranam, P. (2021). "Human-AI collaborative decision-making as an organization design problem". *Journal of Organization Design*. 10(2): 75–80.

Qiu, X., T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang (2020). "Pre-trained models for natural language processing: A survey". *Science China Technological Sciences*. 63(10): 1872–1897.

Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE*. 77(2): 257–286.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever (2018). "Improving language understanding by generative pre-training".

Raisch, S. and S. Krakowski (2021). "Artificial intelligence and management: The automation-augmentation paradox". *Academy of Management Review*. 46(1): 192–210.

Ratten, V. and P. Jones (2023). "Generative artificial intelligence (ChatGPT): Implications for management educators". *The International Journal of Management Education*. 21(3): 100857.

Robertson, S. (2004). "Understanding inverse document frequency: On theoretical arguments for IDF". *Journal of Documentation*. 60(5): 503–520.

Robledo, S., A. M. Grisales Aguirre, M. Hughes, and F. Eggers (2023). "'Hasta la vista, baby'-will machine learning terminate human literature reviews in entrepreneurship?" *Journal of Small Business Management*. 61(3): 1314–1343.

Romera-Paredes, B., M. Barekatain, A. Novikov, *et al.* (2023). "Mathematical discoveries from program search with large language models". *Nature*. DOI: 10.1038/s41586-023-06924-6.

Romero-Brufau, S., K. D. Wyatt, P. Boyum, M. Mickelson, M. Moore, and C. Cognetta-Rieke (2020). "What's in a name? A comparison of attitudes towards artificial intelligence (AI) versus augmented human intelligence (AHI)". *BMC Medical Informatics and Decision Making*. 20(1): 1–5.

Rudolph, J., S. Tan, and S. Tan (2023). "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching*. 6(1).

Schönberger, D. (2018). *Deep Copyright: Up- and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)*. SSRN.

Schwab, A. and Z. Zhang (2019). "A new methodological frontier in entrepreneurship research: Big data studies". *Entrepreneurship Theory and Practice*. 43(5): 843–854.

Shepherd, D. A. and A. Majchrzak (2022). "Machines augmenting entrepreneurs: Opportunities (and threats) at the Nexus of artificial intelligence and entrepreneurship". *Journal of Business Venturing*. 37(4): 106227.

Shinn, N., F. Cassano, A. Gopinath, K. R. Narasimhan, and S. Yao (2023). "Reflexion: Language agents with verbal reinforcement learning". In: *Thirty-Seventh Conference on Neural Information Processing Systems*.

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

Skagestad, P. (1993). "Thinking with machines: Intelligence augmentation, evolutionary epistemology, and semiotic". *Journal of Social and Evolutionary Systems*. 16(2): 157–180.

Strubell, E., A. Ganesh, and A. McCallum (2019). “Energy and policy considerations for deep learning in NLP”. arXiv preprint arXiv:1906.02243.

Tan, Y. C. and L. E. Celis (2019). “Assessing social and intersectional biases in contextualized word representations”. *Advances in Neural Information Processing Systems*. 32.

Teubner, T., C. M. Flath, C. Weinhardt, W. van der Aalst, and O. Hinz (2023). “Welcome to the era of chatgpt et al. the prospects of large language models”. *Business & Information Systems Engineering*. 65(2): 95–101.

Thorp, H. H. (2023). “ChatGPT is fun, but not an author”. *Science*. 379(6630): 313–313.

Tlili, A., B. Shehata, M. A. Adarkwah, A. Bozkurt, D. T. Hickey, R. Huang, and B. Agyemang (2023). “What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education”. *Smart Learning Environments*. 10(1): 15.

Van der Aalst, W. M. (2021). “Hybrid Intelligence: To automate or not to automate, that is the question”. *International Journal of Information Systems and Project Management*. 9(2): 5–20.

Van Dis, E. A., J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting (2023). “ChatGPT: Five priorities for research”. *Nature*. 614(7947): 224–226.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . . ., and I. Polosukhin (2017). “Attention is all you need”. *Advances in Neural Information Processing Systems*. 30.

Vecchiarini, M. and T. Somià (2023). “Redefining entrepreneurship education in the age of artificial intelligence: An explorative analysis”. *The International Journal of Management Education*. 21(3): 100879.

Vedula, S. and P. H. Kim (2019). “Gimme shelter or fade away: The impact of regional entrepreneurial ecosystem quality on venture survival”. *Industrial and Corporate Change*. 28(4): 827–854.

Vincent, V. U. (2021). “Integrating intuition and artificial intelligence in organizational decision-making”. *Business Horizons*. 64(4): 425–438.

Vinsel, L. (2023). “Don’t get distracted by the hype around generative AI”. *MIT Sloan Management Review*. 64(3): 1–3.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, ..., and D. Zhou (2022). "Chain-of-thought prompting elicits reasoning in large language models". *Advances in Neural Information Processing Systems*. 35: 24824–24837.

Winkler, C., B. Hammoda, E. Noyes, and M. Van Gelderen (2023). "Entrepreneurship education at the dawn of generative artificial intelligence". *Entrepreneurship Education and Pedagogy*. 6(4): 579–589.

Zech, H. (2021). "Liability for AI: Public policy considerations". *ERA Forum*. 22: 147–158. DOI: 10.1007/s12027-020-00648-0.

Zhao, W. X., K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, ..., and J. R. Wen (2023). "A survey of large language models". arXiv preprint arXiv:2303.18223.