# Knowledge Management of Artificial Intelligence: a framework for LLM-generated personas

Lucas Nóbrega and Luiz Felipe Martinez

Programa de Engenharia de Sistemas e Computação
COPPE - Universidade Federal do Rio de Janeiro

June 24th, 2024

**Abstract**

Recent technological advancements have boosted the text generation capabilities of Large Language Models (LLMs). Simulating personas can allow LLMs to provide answers that represent a group of people or specific well-known individuals. However, managing and validating this simulated knowledge poses significant challenges, particularly due to the impracticality of contacting a significant number of experts for direct validation. In this study, we present a framework for managing and evaluating the knowledge of personas generated by LLMs to simulate well-know experts in Delphi studies. The results indicate that LLMs can convincingly and accurately replicate expert perspectives, providing a valuable tool for more effective integration of artificial intelligence into decision-making processes based on specialized knowledge.

## 1 Introduction

In recent years, technological advancements have brought about a significant transformation in text generation capabilities, especially through the use of Large Language Models (LLMs). These models, as evidenced in recent studies, possess extensive linguistic knowledge and a great capacity for understanding problems (Brown et al., 2020; Devlin et al., 2018; OpenAI, 2023). The capabilities of LLMs have been increasingly explored in various contexts, from simple tasks like answering trivial questions to complex challenges addressing a wide variety of subjects and levels of complexity.

In this context, Nóbrega et al. (2024) have investigated the potential of LLMs for simulating expert responses in Delphi studies. The Delphi methodology, known for its effectiveness in obtaining expert insights on future issues, can significantly benefit from the ability of LLMs to emulate expert knowledge and opinions. This approach not only promises to overcome the limitations of accessibility and availability of experts but also offers an efficient way to conduct Delphi studies on a broader and more diverse scale.

However, the accuracy of responses from LLM-generated personas is a crucial aspect in contexts that require precision and quality, such as Delphi studies. In this sense, a preliminary validation was conducted with Nelson Maculan Filho, who showed high agreement with his simulated responses (Nóbrega et al., 2024). Although this result is promising, it is important to note that this isolated validation is not sufficient to ensure the absolute reliability of responses generated by LLMs. It is essential to develop validation methods that can more comprehensively guarantee the quality and reliability of the responses produced by these models.

This article proposes a framework to manage and evaluate the knowledge of personas generated by LLMs, comparing real and simulated opinions using text analysis techniques. The process includes collecting expert opinions and simulating them through LLMs, followed by validating these responses using similarity techniques to ensure their correspondence with authentic opinions. This method offers an alternative to the difficulty of accessing a sufficient number of experts, enhancing the use of LLMs in Delphi studies.

Additionally, we conducted an experiment to validate the accuracy of LLMs in simulating expert opinions in a Delphi study on the future of higher education in Brazil. The study included personalities such as Helena Nader, Cristovam Buarque, and Fernando Haddad, whose real opinions were obtained from interviews and public speeches. This process allowed us to test the ability of LLMs to convincingly replicate the knowledge and perspectives of these experts.

This work is organized as follows: Section 2 discusses related works, providing an overview of previous research that influenced our approach. Section 3 describes the methodology used, detailing the proposed framework for managing and evaluating LLMs and how the opinions were collected and simulated. In Section 4, we discuss the experiments conducted, explaining how expert opinions were used to test the effectiveness of the models. Next, Section 5 contains discussions of the results obtained, exploring them in light of related works. Finally, Section 6 summarizes the results, discusses the contributions of the study, and suggests directions for future research.

# 2 Related Work

Text generation through LLMs has been a field of intense development, resulting in significant advancements in the quality and coherence of the texts produced. Various studies have explored crucial areas for the practical application of these models, including the comparison between machine-written and human-written texts (Guo et al., 2023; Mitchell et al., 2023; Sadasivan et al., 2024), fact-checking in texts (Cheung et al., 2023; Zhang et al., 2023; Hang et al., 2024), personality analysis of personas generated by LLMs (Cheng, Piccardi, et al., 2023; Cheng, Durmus, et al., 2023; Gupta et al., 2024; Deshpande et al., 2023), and the assessment of the reliability of these personas (Xiao et al., 2023; Shao et al., 2023). Moreover, the integration of LLMs into knowledge management processes marks an important evolution, leveraging artificial intelligence to capture, store, and utilize the knowledge generated (Jarrahi et al., 2023; Sharma et al., 2023; Hu et al., 2023). Below, we will discuss these topics, providing a solid foundation for the development of the proposed framework for managing and evaluating the knowledge of personas generated by LLMs.

Recent studies have investigated how texts generated by LLMs compare to those written by humans, focusing on aspects such as objectivity, formality, and communication style. Guo et al. (2023) investigated the differences between texts generated by LLMs and humans, highlighting that responses from models like ChatGPT tend to be more focused and formal, while human responses are more divergent and emotional. On the other hand, Mitchell et al. (2023) and Sadasivan et al. (2024) addressed the accuracy of methods for detecting texts generated by LLMs, focusing on distinguishing between AI-generated and human-generated content. These studies emphasize the importance of understanding how LLMs can accurately simulate the style and objectivity of human texts, which is crucial to ensuring that personas generated by LLMs are perceived as reliable and authentic in specialized knowledge contexts.

The need to efficiently validate information in LLMs is addressed by Cheung et al. (2023) and Zhang et al. (2023), both exploring advanced methods to integrate and analyze external and contextual knowledge. Cheung et al. (2023) developed an approach that combines LLMs with external evidence retrieval to improve verification accuracy, while Zhang et al. (2023) applied a hierarchical method that details the verification process in stages, reducing vulnerabilities to misinformation. Following these innovations, Hang et al. (2024) applied semantic reasoning and network analysis to effectively differentiate between rumors and verified facts, demonstrating the effectiveness of this approach in politically charged informational environments. These studies highlight the critical need to enhance fact-checking in LLMs, ensuring that the technology not only advances in capability but also in reliability and utility in contexts where accuracy is crucial.

Research on personality representation in LLMs reveals significant challenges related to stereotypes and biases. Cheng, Piccardi, et al. (2023) and Cheng, Durmus, et al. (2023) present methods to assess the representation of persona characteristics, focusing on the tendency towards caricature and the presence of stereotypes, respectively. Both studies reveal that, despite intentions for accuracy, LLMs often exaggerate or distort traits in political contexts and among marginalized groups. This issue is further explored by Gupta et al. (2024), who examine how persona attribution can introduce biases in reasoning tasks, showing a significant drop in performance when models face complex scenarios. Additionally, Deshpande et al. (2023) reveal that the toxicity of responses increases significantly depending on the simulated persona, highlighting inherent discriminatory patterns. These studies demonstrate the importance of a deep and critical understanding of the personalities modeled by LLMs, emphasizing how exaggerated or distorted characteristics can impact the validity of simulations.

Regarding the reliability of LLM simulations, Xiao et al. (2023) propose new metrics of consistency and robustness to evaluate the credibility of personas generated by LLMs, introducing a benchmark that tests the accuracy of simulations under perturbations. Furthermore, Shao et al. (2023) investigate the ability of LLMs to faithfully reproduce historical personalities, highlighting the importance of accurate and reliable simulations to ensure the authenticity of responses. These studies collectively emphasize the importance of meticulous approaches to ensure that LLM simulations are not only accurate but also robust and reliable, which is essential for the credibility of agents in practical applications.

Finally, the integration of artificial intelligence into knowledge management processes is a burgeoning field that bridges traditional management techniques with cutting-edge technology. Jarrahi et al. (2023) and Sharma et al. (2023) explore how AI can support all phases of knowledge management within organizations, from the creation and storage of knowledge to its retrieval and application. They highlight the role of AI in enhancing data analysis, generating insights, and facilitating access to organized knowledge through intelligent personal assistants. On a related note, Hu et al. (2023) delve into the specific use of ChatGPT in design knowledge management, showcasing how iterative linguistic interactions and AI-driven collaboration can significantly streamline the acquisition of targeted knowledge in design processes, although they caution about the biases and reliability of the information depending on the prompts used. These studies underscore the transformative potential of AI in optimizing knowledge management strategies, making it an essential tool for organizational development and decision-making processes.

Through the review of the studies highlighted in this section, it becomes evident that there has been extensive exploration and evaluation of LLMs in various contexts, including their integration into knowledge management processes. However, the innovative proposal to compare simulated opinions by LLMs with real expert opinions using evaluation methodologies alongside effective knowledge management practices represents a relatively unexplored frontier in research. This combined approach has the potential to reveal deeper insights into the accuracy and reliability of LLM simulations, indicating a promising direction for future work at the intersection of generative AI and opinion studies.

# 3 Methodology

In this research, we developed a framework for the Knowledge Management (KM) of personas generated by LLMs, encompassing everything from the generation to the application of their knowledge. This model aims not only to measure the accuracy and reliability of simulations but also to promote efficient and systematic management of the generated knowledge. This enables continuous reuse and improvement, as well as enhancing the use of LLMs in various study and application contexts.

Knowledge management, according to Beckman (1999), is a critical process that formalizes experiences, knowledge, and expertise, making them accessible within organizations. In this context, Stollenwerk (2001) proposes a generic knowledge management model that includes essential processes such as identification, capture, selection and validation, organization and storage, sharing, application, and creation. The framework proposed in this work adapts these processes to the specific context of managing the knowledge of personas generated by LLMs, as presented in Figure 1. Below, we detail each of the stages.

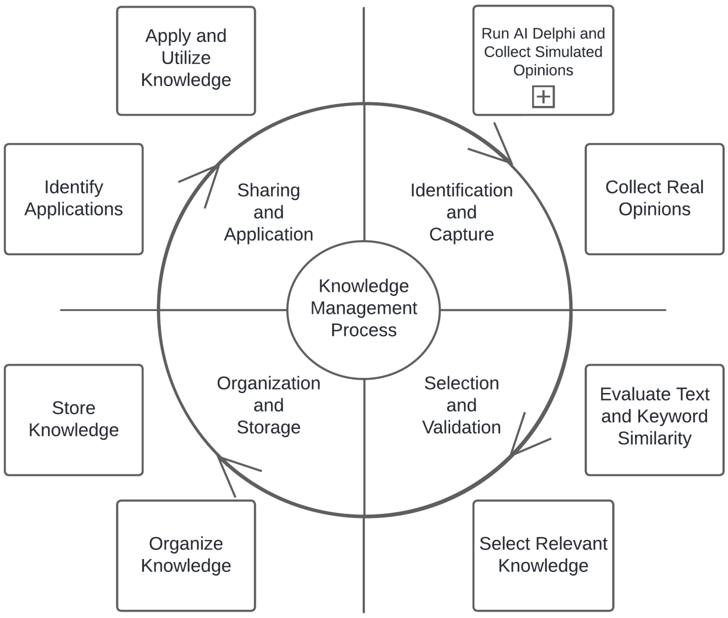


Figure 1: Knowledge Management Framework for Personas Generated by Large Language Models.

## 3.1 Identification and Capture

The first step of a knowledge management model, as described by Stollenwerk (2001), is identification, which involves determining which competencies are essential for the organization's success. The second step is capture, which refers to obtaining the necessary knowledge, skills, and experiences to develop and maintain these competencies. In the context of this study, these concepts were applied to identify the objective of capturing expert foresights on future trends, and then proceeded to collect expert opinions, both real and simulated, to create a robust and representative dataset.

Initially, to collect the simulated opinions, the AI Delphi method is used, specifically the Iconic Minds model proposed by Nóbrega et al. (2024). This model involves conducting a Delphi questionnaire, as illustrated in Figure 2, where the personas of renowned researchers represent the experts. Using LLMs configured to emulate these experts' personas, simulated opinions are generated based on prompts that reflect the themes and issues addressed in the real opinions.

Simultaneously, the real opinions of the experts simulated by LLMs are collected through published interviews, public speeches, social media posts, and academic publications. This process ensures the inclusion of a wide

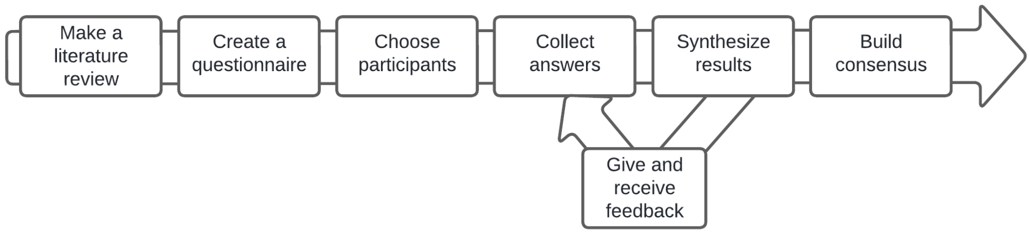


Figure 2: Delphi Process Steps. Adapted from: Argôlo et al. (2022)

range of expert perspectives, providing a solid foundation of authentic and representative knowledge.

## 3.2 Selection and Validation

After the identification and capture of knowledge, Stollenwerk (2001) addresses the selection and validation stage, which involves filtering knowledge, assessing its quality, and synthesizing it for future application. In the context of this study, these concepts were applied to evaluate whether the knowledge of personas simulated by LLMs is similar to the real opinions of experts. This stage is crucial to ensuring that the simulations generated by LLMs are accurate and reliable, guaranteeing the integrity and applicability of the generated knowledge.

To conduct the evaluation, a combination of text similarity analysis and keyword analysis is used. Text similarity analysis allows us to investigate how similar one document is to another, examining the similarity between meanings, ideas, and contexts in the texts (Gomaa et al., 2013). This method is fundamental to ensuring that the simulated opinions accurately reflect the perspectives and knowledge of real experts.

On the other hand, keyword similarity analysis is essential in identifying and comparing the keywords present in both real and simulated opinions, obtaining the key terms that best summarize a document (Witten et al., 1999; Hasan et al., 2010). If the main topics of the simulated opinions are similar to or related to those of the real opinions, this may indicate shared opinions and the accuracy of the simulations. This approach ensures that not only the semantic structure but also the specific content and central themes are faithfully replicated in the simulated opinions.

Based on the results of these analyses, the relevant knowledge from the simulated opinions is filtered, retaining only those that show high similarity and accuracy compared to the real opinions. This filtering process ensures that only high-quality and relevant knowledge is considered for the subsequent phases of organization and application.

## 3.3 Organization and Storage

The next stage of a knowledge management model involves the organization and storage of knowledge. According to Stollenwerk (2001), these phases involve ensuring the quick, easy, and accurate retrieval of knowledge through the use of effective storage systems. In the context of this study, this translates to organizing and storing the validated knowledge of simulated personas in an efficient and accessible manner.

After the selection and validation of the personas generated by LLMs, it is essential that this information be organized and stored properly, including archiving both the generated responses and real data collected in a database. Utilizing AI technologies, such as Natural Language Processing for content classification and machine learning for taxonomy generation, enhances the organization and retrieval of knowledge (Sharma et al., 2023). This approach allows for its practical and effective future use, ensuring that validated knowledge is available for applications in subsequent research and studies.

## 3.4 Sharing and Application

The final stages of a knowledge management model refer to the sharing and application of knowledge. Sharing involves ensuring that knowledge is widely formalized and organized to facilitate access, according to Stollenwerk (2001). Application, in turn, ensures that knowledge, experiences, and information are used in real situations, producing tangible benefits. In the context of this study, this means making the validated knowledge of personas generated by LLMs accessible and practical.

First, it is essential to identify the possible applications in which the knowledge generated by the personas will be used. This involves analyzing different contexts where the personas' knowledge can be useful, such as in decision-making processes. Identifying specific applications allows for the effective direction of knowledge to areas where it can have the greatest impact, ensuring that the information is used beneficially. Thus, the knowledge of personas generated by LLMs can finally be applied and utilized practically.

# 4 Experiments

In this section, we detail the experiments conducted to evaluate the accuracy of personas generated by LLMs in simulating the knowledge of real experts[1]. This study primarily focuses on developing a KM framework specifically tailored for LLM-generated personas. Although the processes of identification, capture, organization, storage, sharing, and application are well-established in traditional knowledge management, validating machine-generated knowledge presents unique challenges. Therefore, our experiments specifically target the validation aspect, employing the AI Delphi model to generate simulated opinions and various techniques to compare these with real expert opinions. We will also discuss the methodologies used to analyze the similarities between simulated and real opinions and present the key findings.

Initially, simulated opinions were generated using AI Delphi, specifically the Iconic Minds model, which is used to simulate responses from well-known personalities. The Delphi process on the future of higher education in Brazil applied in the study by Nóbrega et al. (2024) was replicated, involving the same experts and the same set of questions. These questions assessed aspects such as the likelihood and desirability of various future scenarios in higher education, along with additional events that could influence these outcomes. The experts were a mix of individuals; some were suggested by GPT based on their relevance to the theme and the questionnaire, while others were manually selected, focusing on those recognized by the GPT. In parallel, real opinions[2] from the same experts were manually collected from authentic sources, such as interviews, public speeches, and academic publications, to establish a comparative set.

After generating the simulated opinions and collecting the real expert opinions about the future of higher education in Brazil, we proceeded with detailed similarity analysis to quantify the accuracy of the generated personas. We use two main approaches: text similarity and keyword similarity.

For Text Similarity, each document was initially segmented into sentences, with each sentence converted into a vector using the OpenAI (*OpenAI* 2024) embedding model "text-embedding-3-small". To obtain a representative vector for each complete document, we applied the simple centroid technique,

[1] The repository with the codes and data used in the experiments is available at https://github.com/LucasONobrega/COS738_CPS831_KnowledgeManagementOfArtificialIntelligence.

[2] The real opinions collected are available at https://docs.google.com/spreadsheets/d/1hqo3PhVr9GFzmsuuMu7lLU4QoV5PgqbiMNLoE9nmL5g/edit?usp=sharing.

which combines the embeddings of each sentence. Subsequently, the cosine similarity between the document vectors of the simulated and real opinions was calculated, providing a quantitative measure of the semantic similarity between the opinions.

On the other hand, for a more focused analysis, Keyword Similarity was explored. Initially, the most significant keywords were extracted from each document using the KeyBERT (*KeyBERT* 2024) model. Each keyword was converted into a vector using the same embedding model. Similarly, the vectors were aggregated using the centroid technique for each set of keywords. Thus, the cosine similarity between the keyword sets of the simulated and real opinions was calculated, highlighting the agreement in terms of key themes and topics.

The results obtained from the text and keyword similarity analyzes directly reflect the accuracy of personas generated by LLMs in replicating expert knowledge. For text similarity, the values ranged from 0.702 for Vladimir Pinheiro-Safatle to 0.868 for Claudia Costin, with a mean of 0.796 and a standard deviation of 0.055. In contrast, keyword similarity recorded values from 0.728 for Helena Nader to 0.922 for Claudia Costin, with a mean of 0.842 and a standard deviation of 0.050. Table 1 presents the complete results of the similarity values for each of the simulated personas, while Table 2 provides additional descriptive statistics, including means, standard deviations, percentiles, minimums, and maximums for both metrics. These results demonstrate a significant ability of the simulated personas to reflect expert opinions, showing considerable alignment between the simulations and the real opinions.

Table 1: Similarity Results

| Name | Text Similarity | Keyword Similarity |
|---|---|---|
| Claudia Costin | 0.868 | 0.922 |
| Cristovam Buarque | 0.830 | 0.841 |
| Daniel Cara | 0.844 | 0.890 |
| Denise Pires de Carvalho | 0.865 | 0.857 |
| Fernando Haddad | 0.733 | 0.838 |
| Helena Nader | 0.741 | 0.728 |
| Lino de Macedo | 0.820 | 0.831 |
| Miguel Nicolelis | 0.749 | 0.886 |
| Mozart Neves Ramos | 0.864 | 0.854 |
| Muniz Sodré | 0.772 | 0.839 |
| Nelson Maculan Filho | 0.790 | 0.794 |
| Renato Janine Ribeiro | 0.820 | 0.884 |
| Vladimir Pinheiro-Safatle | 0.702 | 0.845 |
| Yvonne Maggie | 0.752 | 0.777 |

Table 2: Descriptive statistics

| Descriptive statistics | Text Similarity | Keyword Similarity |
|---|---|---|
| Mean | 0.796 | 0.842 |
| Standard deviation | 0.055 | 0.050 |
| Minimum | 0.702 | 0.728 |
| 25% | 0.750 | 0.833 |
| 50% | 0.805 | 0.843 |
| 75% | 0.841 | 0.877 |
| Maximum | 0.868 | 0.922 |

# 5 Discussion

The results of the experiments in this study directly reflect the capabilities and challenges faced by LLMs in replicating expert knowledge, an area of growing interest in artificial intelligence literature. The similarity found between simulated and real opinions illustrates both the potential and limitations of LLMs in generating authentic text and emulating specific personalities.

As evidenced by Guo et al. (2023), Mitchell et al. (2023), and Sadasivan et al. (2024), LLMs are capable of generating responses that convincingly mimic the style and formality of human texts. Our study adds to this narrative by demonstrating that LLMs can, to a certain extent, simulate complex expert opinions with reasonable accuracy. However, as highlighted by Cheung et al. (2023), Zhang et al. (2023), and Hang et al. (2024), accuracy in fact-checking and the ability to integrate contextual knowledge are crucial for the reliability of these simulations. The need for robust methods to validate and verify the information generated by LLMs is a recurring theme that our results also support.

Studies by Cheng, Piccardi, et al. (2023), Cheng, Durmus, et al. (2023), Gupta et al. (2024), and Deshpande et al. (2023) reveal that LLM-generated personalities often include exaggerations or distortions, especially in politicized or culturally sensitive contexts. This highlights that, while our study indicates that LLMs can capture key elements of expert opinions, complete accuracy and the absence of distortions are still areas that require ongoing attention, particularly when the simulation involves complex cultural and contextual nuances.

Additionally, the research by Xiao et al. (2023) and Shao et al. (2023) on the reliability of LLM simulations suggests that consistency and robustness are vital metrics for assessing the credibility of generated personas. Our results endorse this view and suggest that, while LLMs can effectively replicate general aspects of expert knowledge, there are still significant challenges in ensuring that simulations are consistently accurate in all cases.

Furthermore, the integration of artificial intelligence in knowledge management practices highlighted in this study aligns with the insights provided by Jarrahi et al. (2023), Sharma et al. (2023), and Hu et al. (2023), who demonstrate how AI can facilitate the entire spectrum of knowledge management activities. Our findings extend these observations by illustrating how LLMs, when embedded within a structured knowledge management framework, can not only replicate but also enhance the accessibility and applicability of expert knowledge in many contexts.

Finally, the contribution of this study to the existing literature is twofold: it reinforces the viability of LLMs as tools for simulating specialized knowledge and highlights the imperative need for improved methodologies to assess the fidelity and reliability of these simulations. This outlines a promising path for future research focused on enhancing the accuracy of generative AI technologies in critical knowledge contexts.

# 6 Conclusion

LLM models have significantly advanced in recent years in the task of text generation, achieving levels of complexity and thematic variety that can mimic human fluency. However, validating the accuracy and relevance of the responses generated by these models represents a major challenge, especially in contexts where accuracy is vital, such as in Delphi studies. The difficulty of accessing experts for direct validations due to their limited availability necessitates the search for alternative methods that ensure the reliability of LLM simulations.

In this work, we propose to develop and test a framework that not only facilitates the generation and comparison of simulated and real expert opinions but also provides a systematic way to manage the obtained knowledge. The experiments focused on simulating expert opinions on the future of higher education in Brazil using the AI Delphi model, and comparing these simulations with real opinions to validate the accuracy of the generated personas.

The main contribution of this work is the development and application of an innovative framework for the knowledge management of personas generated by LLMs. This framework not only evaluates the accuracy of personas in simulating the knowledge of real experts but also provides a structure for efficiently organizing, storing, and reusing this knowledge. Through the integration of similarity analysis techniques, the study demonstrates how LLM simulations can be rigorously validated, providing a valuable tool for environments that rely on the accuracy and depth of specialized knowledge. Additionally, by establishing a systematic procedure for knowledge validation in LLMs, this work expands the potential application of these models in critical contexts where the reliability of information is essential, representing a significant step towards more effective integration of artificial intelligence in decision-making processes based on specialized knowledge.

Regarding the limitations and risks to the validity of the research, we recognize that the similarity-based approach, while providing a quantitative measure of the similarity between simulated and real opinions, may not fully capture the depth and richness of expert opinions. Specifically, the simple centroid technique employed to calculate the representative vectors of complete documents may not reflect all the nuances contained in the opinions, resulting in a possible underestimation of the diversity of perspectives and details present in the original content. This simplification may limit the ability to accurately assess the fidelity with which LLM-generated personas replicate specialized knowledge. While the proposed framework advances the knowledge management of LLM-generated personas, the practical application of these techniques requires careful consideration of these limitations to ensure more robust and comprehensive validations in future implementations.

For future work, it is essential to consider enhancing similarity analysis techniques by exploring advanced algorithms that can more accurately capture the complex nuances in expert opinions. Investigating alternative validation methods that integrate quantitative and qualitative assessments is also crucial, as this would allow for a deeper understanding of the accuracy and richness of LLM-generated simulations. Additionally, developing adaptations of the framework to handle different data modalities and types of specialized knowledge can offer new opportunities to validate the reliability and applicability of simulations in more varied contexts.

In conclusion, our research reinforced the utility of LLMs as powerful tools for simulating specialized knowledge. The proposed framework is essential for evaluating and managing this knowledge, addressing significant challenges related to the accuracy and reliability of simulations. This study highlights the ongoing need to develop and refine methodologies that ensure the integrity of LLM-generated simulations, underscoring the crucial role of such technologies in contexts where expertise is critical.

# References

Argôlo, Matheus, Mauricio Miranda, Rodrigo Pagliusi, Yuri Lima, Herbert Santos, Carlos Eduardo Barbosa, Alan Lyra, and Jano Souza (Aug. 2022). "Managing Knowledge About Future of Work: A Model for Higher Education Institutions". In: *European Conference on Knowledge Management* 23, pp. 28–36. DOI: 10.34190/eckm.23.1.319 (cit. on p. 7).

Beckman, T. J. (1999). "The Current State of Knowledge Management". In: *Knowledge Management Handbook*. Ed. by Liebowitz. CRC Press (cit. on p. 5).

Brown, Tom B, Benjamin Mann, Nick Ryder, Melvin Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. (2020). "Language Models are Few-Shot Learners".

In: *arXiv preprint arXiv:2005.14165*. URL: https://arxiv.org/abs/2005.14165 (cit. on p. 1).
Cheng, Myra, Esin Durmus, and Dan Jurafsky (2023). *Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models*. arXiv: 2305.18189 [cs.CL] (cit. on pp. 3, 4, 12).
Cheng, Myra, Tiziano Piccardi, and Diyi Yang (2023). *CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations*. arXiv: 2310.11501 [cs.CL] (cit. on pp. 3, 4, 12).
Cheung, Tsun-Hin and Kin-Man Lam (2023). *FactLLaMA: Optimizing Instruction-Following Language Models with External Knowledge for Automated Fact-Checking*. arXiv: 2309.00240 [cs.CL] (cit. on pp. 3, 12).
Deshpande, Ameet, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan (2023). *Toxicity in ChatGPT: Analyzing Persona-assigned Language Models*. arXiv: 2304.05335 [cs.CL] (cit. on pp. 3, 4, 12).
Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018). "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (cit. on p. 1).
Gomaa, Wael and Aly Fahmy (Apr. 2013). "A Survey of Text Similarity Approaches". In: *international journal of Computer Applications* 68. DOI: 10.5120/11638-7118 (cit. on p. 7).
Guo, Biyang, Xin Zhang, Ziyuan Wang, Minqi Jiang, Jinran Nie, Yuxuan Ding, Jianwei Yue, and Yupeng Wu (2023). *How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection*. arXiv: 2301.07597 [cs.CL] (cit. on pp. 3, 12).
Gupta, Shashank, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot (2024). *Bias Runs Deep: Implicit Reasoning Biases in Persona-Assigned LLMs*. arXiv: 2311.04892 [cs.CL] (cit. on pp. 3, 4, 12).
Hang, Ching Nam, Pei-Duo Yu, and Chee Wei Tan (2024). "TrumorGPT: Query Optimization and Semantic Reasoning over Networks for Automated Fact-Checking". In: *2024 58th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. DOI: 10.1109/CISS59072.2024.10480162 (cit. on pp. 3, 4, 12).
Hasan, Kazi and Vincent Ng (Jan. 2010). "Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-Art." In: vol. 2, pp. 365–373 (cit. on p. 7).
Hu, Xin, Yu Tian, Keisuke Nagato, Masayuki Nakao, and Ang Liu (2023). "Opportunities and challenges of ChatGPT for design knowledge management". In: *Procedia CIRP* 119. The 33rd CIRP Design Conference, pp. 21–28. ISSN: 2212-8271. DOI: https://doi.org/10.1016/j.procir.2023.

05.001. URL: https://www.sciencedirect.com/science/article/pii/S2212827123004274 (cit. on pp. 3, 4, 12).

Jarrahi, Mohammad Hossein, David Askay, Ali Eshraghi, and Preston Smith (2023). "Artificial intelligence and knowledge management: A partnership between human and AI". In: *Business Horizons* 66.1, pp. 87–99. ISSN: 0007-6813. DOI: https://doi.org/10.1016/j.bushor.2022.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0007681322000222 (cit. on pp. 3, 4, 12).

*KeyBERT* (2024). https://github.com/MaartenGr/KeyBERT (cit. on p. 10).

Mitchell, Eric, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn (2023). *DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature*. arXiv: 2301.11305 [cs.CL] (cit. on pp. 3, 12).

Nóbrega, Lucas, Luiz Felipe Martinez, Luísa Marschhausen, Yuri Lima, Marcos Almeida, Alan Lyra, Carlos Eduardo Barbosa, and Jano Moreira de Souza (2024). "AI-Delphi: Emulating Personas Toward Machine-Machine Collaboration". In: (cit. on pp. 2, 6, 9).

*OpenAI* (2024). https://openai.com/ (cit. on p. 9).

OpenAI (2023). *GPT-4 Technical Report*. Tech. rep. OpenAI. URL: https://arxiv.org/pdf/2303.08774v3.pdf (cit. on p. 1).

Sadasivan, Vinu Sankar, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil Feizi (2024). *Can AI-Generated Text be Reliably Detected?* arXiv: 2303.11156 [cs.CL] (cit. on pp. 3, 12).

Shao, Yunfan, Linyang Li, Junqi Dai, and Xipeng Qiu (2023). *Character-LLM: A Trainable Agent for Role-Playing*. arXiv: 2310.10158 [cs.CL] (cit. on pp. 3, 4, 12).

Sharma, Ruchi, Girish Lakhera, and Manu Sharma (2023). "Artificial Intelligence in KM Processes: Emerging Trends and Challenges". In: *2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI)*. Vol. 1, pp. 1–6. DOI: 10.1109/IDICAIEI58380.2023.10406876 (cit. on pp. 3, 4, 8, 12).

Stollenwerk, M. F. L. (2001). "Gestão do conhecimento: conceitos e modelos". In: *Inteligência organizacional e competitiva*. Ed. by K. Tarapanoff. Brasília: UNB, pp. 143–163 (cit. on pp. 5–8).

Witten, Ian H., Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig G. Nevill-Manning (1999). *KEA: Practical Automatic Keyphrase Extraction*. arXiv: cs/9902007 (cit. on p. 7).

Xiao, Yang, Yi Cheng, Jinlan Fu, Jiashuo Wang, Wenjie Li, and Pengfei Liu (2023). *How Far Are We from Believable AI Agents? A Framework*

*for Evaluating the Believability of Human Behavior Simulation*. arXiv: 2312.17115 [cs.CL] (cit. on pp. 3, 4, 12).
Zhang, Xuan and Wei Gao (2023). *Towards LLM-based Fact Verification on News Claims with a Hierarchical Step-by-Step Prompting Method*. arXiv: 2310.00305 [cs.CL] (cit. on pp. 3, 12).