Do Large Language Models Require Prior Knowledge for Learning? A Preliminary Study

Marcus Soll; Louis Kobras

2025 IEEE Global Engineering Education Conference (EDUCON), 2025

IEEE Xplore: https://ieeexplore.ieee.org/document/11016655

DOI: 10.1109/EDUCON62633.2025.11016655

Do Large Language Models Require Prior Knowledge for Learning? A Preliminary Study

Marcus Soll
NORDAKADEMIE gAG Hochschule der Wirtschaft
25337 Elmshorn, GERMANY
https://orcid.org/0000-0002-6845-9825

Louis Kobras
NORDAKADEMIE gAG Hochschule der Wirtschaft
25337 Elmshorn, GERMANY
https://orcid.org/0000-0003-4855-2878

***Abstract*—LLMs find increasing use as an educational tool, both from the instructors' perspective as well as from the students'. This paper presents a preliminary study investigating the effects of prior knowledge concerning a given topic on the effectiveness of using an LLM (Microsoft Copilot) to study the topic. Choosing Büchi automata as an example, twelve computer science students were tasked with first giving a self-report on prior knowledge, then studying Büchi automata for 15 minutes using only an LLM as a study tool, and afterwards filling out a short topical questionnaire. Two trends could be observed: Prior knowledge of LLMs seems to increase the learning effect while prior knowledge of a topic that is related to the studied topic seems to diminish the learning effect.**

***Index Terms*—Computer Science Education, Large Language Models, Pedagogy, STEM education**

I. Introduction

The discussion about using large language models (LLMs) like ChatGPT and Microsoft Copilot for education is in full swing (see [1], [2]). The dream is that AI together with these LLMs can act as an individual tutor for students who can adapt to the prior knowledge as well as the needs of students (see [3]).

Yet, this idea is not without criticism. For example, LLMs produce wrong answers / explanations (e.g. [4]–[6]), sometimes even generating completely new facts known as *hallucinations* (e.g. [7]–[9]) or *confabulations* [10]. There are some concerns that people might depend on LLMs for reasoning, and about the inability of LLMs to generate new ideas [6]. Some criticism exists about most LLMs being controlled by corporate actors which hide crucial details about their models (cf. [11, p. 80]).

Maybe more important here, there also exists criticism about using LLMs for learning. Some first insights suggest that learning with LLMs can actually reduce the learned knowledge once students loose access to the LLM [12]. Learning with LLMs can be harmful to struggling learners and give them illusions of competence [13].

One important question is how relevant prior knowledge is for learning and whether LLMs can provide the required knowledge if a student is exploring a new subject area. Students build new knowledge based on the knowledge they already possess [14, p. 10-12]. While learning, students build connections and complex structures between different information and thus construct their individual body of knowledge [15, pp. 4-5]. This prior knowledge can have both positive effects on learning, but also can lead to bias [15, pp. 4-5]. This brings us to the theory that considering the prior knowledge is important when we want to provide good learning opportunities. Therefore, when using LLMs to learn, students with greater prior knowledge might show better results than students without prior knowledge.

In this short paper, we want to present a preliminary study addressing this issue. By better understanding the relationship between prior knowledge and learning using LLMs we hope that we can contribute to the development of teaching systems that takes the needs of students into consideration, especially regarding to the pedagogical needs of students with respect to prior knowledge.

II. Experiment Design

To determine the effect of prior knowledge on learning, we wanted to let students interact with an LLM and see how well they respond to a test afterwards. There are a few key design consideration we took into account when designing the study:

1) The topic should relate to a topic students already know (e.g., by having taken a relevant mandatory course) but at the same time they should not have learned the specific topic in said course. For this study, we choose the topic of 'Büchi automata', since students of computer science most likely have automata theory as a required course at different universities. At the same time, this specific type of automata is not discussed everywhere, especially not in introductory courses.
2) We do not want to influence the prompts students use for learning with the LLM. Unfortunately, a pre-test of knowledge might influence the students with regards to possible topics they could ask the LLM about. Therefore, pre-testing is not feasible in our study design. This means that we have to rely on self-reporting for this study.
3) We wanted students to have the ability to also generate images. Therefore, we needed an LLM which also

This research was part of the project *Flexibel kombinierbare Cross-Reality Labore in der Hochschullehre: zukunftsfähige Kompetenzentwicklung für ein Lernen und Arbeiten 4.0 (CrossLab)*, which is funded by the *Stiftung Innovation in der Hochschullehre*, Germany.

Wie schätzen Sie ihr Vorwissen bezüglich KIs wie **Microsoft Copilot / ChatGPT / Large Language Models** ein? (How would you rate your prior knowledge regarding AIs like **Microsoft Copilot / ChatGPT / Large Language Models**?)

Anfänger (beginner knowledge) ☐ ☐ ☐ ☐ ☐ ☐ ☐ Experte (expert knowledge)

Wie schätzen Sie ihr Vorwissen bezüglich **Automatentheorie** ein? (How would you rate your prior knowledge regarding **automata theory**?)

Anfänger (beginner knowledge) ☐ ☐ ☐ ☐ ☐ ☐ ☐ Experte (expert knowledge)

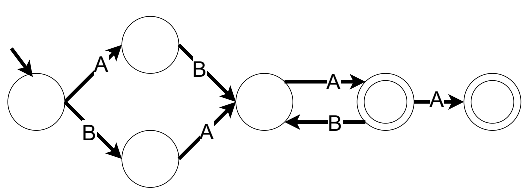
Wie schätzen Sie ihr Vorwissen bezüglich **Büchi-Automaten** ein? (How would you rate your prior knowledge regarding **Büchi automaton**?)

Anfänger (beginner knowledge) ☐ ☐ ☐ ☐ ☐ ☐ ☐ Experte (expert knowledge)

Aufgabe (**Exercise**): Nutzen Sie sie nächsten 15 Minuten, um sich in das Thema Büchi-Automaten einzuarbeiten. Nutzen Sie dafür ausschließlich Microsoft Copilot sowie ihre Vorerfahrung. (Please take 15 minutes to familiarise yourself with Büchi automata. Only use Microsoft Copilot and your prior knowledge.)

Fig. 1: Questions (original German question and translated) used for the pre-test to determine students' prior knowledge

1. Was ist ein Büchi-Automat? Was zeichnet ihn aus? (What is a Büchi automaton? What makes it special?)
   - 0 points: (Partially) wrong answer
   - 1 point: A single aspect is named
   - 2 points: Two or more aspects are named
2. Betrachten Sie folgenden Büchi-Automaten. Welche Sprache akzeptiert er? (Consider the following Büchi automaton. What language does it accept?)
   - 0 points: Wrong language
   - 1 point: Language is correct, but formal writing is not correct (e..g, no omega, wrong characters)
   - 2 points: Language is correct



3. Nennen Sie zwei verschiedene Arten von Wörtern, die ein Büchi-Automat nicht akzeptieren kann. Begründen Sie kurz Ihre Antwort. (Name to kinds of words which can not be accepted by Büchi automata. Justify your answer.)
   - 0 points: (Partly) wrong answer
   - 1 point: One correct word or multiple words of the same kind
   - 2 points: Two or more words

Fig. 2: Questions (original German question and translated) and scoring schema used for the post-test to determine students' learned knowledge

allowed this. This brought us to using Microsoft Copilot for this study.

Therefore, the study flow for students is as following: The study was conducted outside of teaching context. It begins with a self-reporting (see Fig. 1) on prior knowledge relevant to the topic and on their LLM knowledge using a 7-point Likert scale (for the reason see item 2), with a 1 representing *beginner knowledge* and a 7 representing *expert knowledge*. After that, students are asked to interact with Microsoft Copilot (see item 3) to learn about Büchi automata (see item 1). Students interacted with Microsoft Copilot directly through the standard web interface for 15 minutes. Finally, they have to answer a short quiz in German (without access to Microsoft Copilot) which consists of three questions aimed at gauging fundamental subject knowledge and understanding to determine what they have learned from their session.

The quiz is manually rated by the researchers following a strict scoring guide (see Fig. 2). The points students score on the quiz is then tested for correlation with their self-reported prior knowledge using Kendall rank correlation coefficient.

III. Results

The study was conducted during November 2024 and included a total of 12 computer science students from three universities. While the self-reported scores on prior knowledge concerning LLMs and general automata theory are somewhat spread out (with answers ranging from 3 to 7 and 2 to 5, respectively; lower / higher values were not reported by students), no students reported prior knowledge concerning Büchi automata with every student ranking themselves at *beginner knowledge* (1).

In general, students managed to mostly list one or two correct properties of Büchi automata for question 1 while

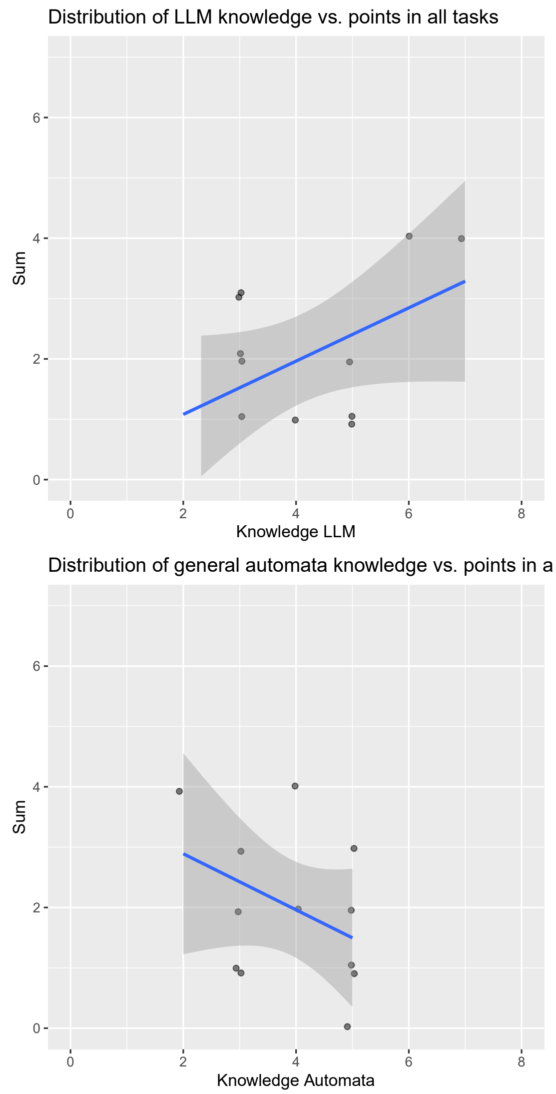


Fig. 3: Comparing the distribution of the self-report questions from the pre-test concerning LLMs and general automata theory against the response scores from the post-test.

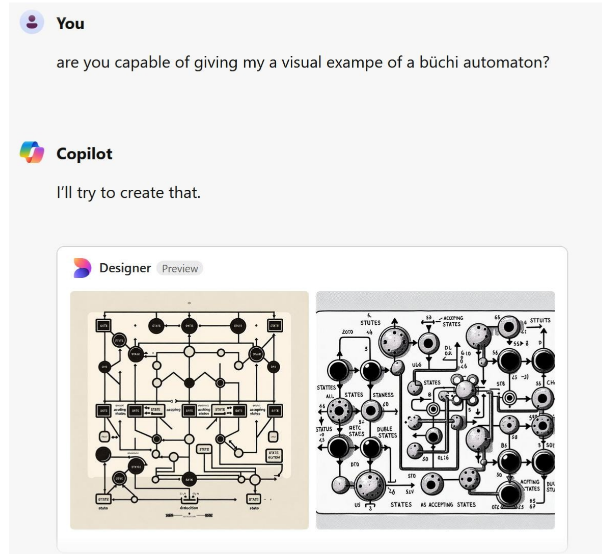


Fig. 4: When prompted, Copilot *tries* to draw a Büchi automaton – with less than stellar results. Repeated requests to properly draw an automaton were met with similar artistic renderings before the student gave up on this line of questioning.

struggling more with question 3, where only one student managed to achieve a score of two points. However, ten of the twelve students did not score any points for question 2 regarding the accepted language of a given sample Büchi automaton, with the remaining two students correctly identifying the language but writing it down with an incorrect syntax, thus both receiving a score of one for this question.

Fig. 3 compares the scores of the pre-test self-reports against the scores of the post-test quiz. The general trend seems to go up with more prior knowledge concerning LLMs and going down with more prior knowledge concerning general automata theory. As all participants gave the same response with regards to familiarity with Büchi automata, no trend can be observed. The pre- and post-scores correlate with $p = 0.2055$ for LLMs and $p = 0.2265$ for general automata theory. To avoid finding significance because of data-analysis decisions (see [16], [17]), we only performed statistical evaluations that were pre-registered. We were not able to compute correlations containing prior knowledge of Büchi automata since all participants responded with *beginner knowledge* (1) on the Likert scale here.

IV. Discussion

Having a look at the trends displayed in Fig. 3, the positive correlation between prior knowledge concerning LLMs and the post-test score may not be all too surprising. It might be assumed that experience on how to formulate prompts might help elicit more helpful answers from the LLM. Cowan, Watanobe and Shirafuji [18], for example, show an increase in helpful conversations with an LLM in the context of programming education after applying structured pre-processing to prompts.

The second trend, however, warrants more scrutiny: How could it be that students who possess *more* knowledge perform *worse* in the post-test? One possible interpretation, which also seems to correspond to the answers students gave in post-test question 2, lies in the fact that Büchi automata share a visual language with another type of automaton, the (non-)deterministic finite automaton. When asked to note the formal language the example automaton accepts, answers more often than not yielded a formal language for an NFA, not a Büchi automaton – finite words or words that end in the "trap" accepting state on the right end. Here, prior knowledge may actually be a hurdle to overcome as students conflate the types of automaton based on the shared visual language. This issue is further reinforced by the fact that Copilot seems to be unable to properly create visual representations of Büchi automata; when asked for a visual examples, the result is anything but, see also Fig. 4 for a possible result.

With a significance niveau of $p \approx 0.2$ each, however, both trends hardly seem to be significant. Nonetheless, we hope that our results might also influence future curricula development in the light of LLMs and required prior knowledge.

A. Possible Reasons

Learning strategies which are more successful facilitate different strategies like the retrieval, summarisation, and explanation of information as well as applying learned content [15, p. 5]. It is likely that this kind of activity is not present when students simply interact with LLM, instead it merely generates textual answers to their questions.

In addition, facts are rarely learned by experiencing them a single time [15, p. 49]. When they are learned at a single time, the learner generally already had a vast background of knowledge in the area [15, p. 49].

Another issue that arose is found in Copilot blatantly inventing things when it couldn't interpret a question. While the concept of hallucinations / confabulation as made-up outputs for AI tools is not new [19], it was nonetheless interesting to see that Copilot vehemently tried to teach a student about book vending machines after the student accidentally asked for information regarding "Büchli automata"[1] – a term that as far as the authors can tell any web search engine corrects to "Büchi automata".

B. Limitations of the Study

One limitation of the study is that the test was directly after the learning phase. Humans have two kind of memories, working memory and long-term memory, with the latter being subdivided even further [15, pp. 77-82]. Testing knowledge directly after learning does not allow us to test all kinds of memory equally, which limits the generalisation of the study. However, we still believe that this is a good insight into learning which could be extended by other studies.

12 participants is quite a low turnout, especially considering that a control group of participants familiar with Büchi automata is missing. This puts into question the representativity of the conducted study, especially since it is below the number of participants we aimed for in our pre-registration. Unfortunately, despite repeated attempts, the authors were unable to recruit more students to participate. However, most incentives (e.g., payment for participation) would have warranted a thorough validation of the study by the local ethics board. Additionally, rewarding students with bonuses for their studies if they were recruited directly from a course, e.g. with bonus points for the exam, was not an option as to not endanger plausible voluntariness and to prevent any concern regarding conflict of interest from arising. Be that as it may, as this study was only intended as a preliminary study from the beginning, first insights indicating a certain direction – that using LLMs does not help with learning – suffice to justify further efforts.

[1]With *Buch* being the German word for *book* and *Büchli* ostensibly being a diminutive of the same word. Microsoft Copilot seems to have used this interpretation despite both prompt and answer being in English.

C. Implications for LLM-based educational tools

Although this was a preliminary study, we can conclude that future development of LLM-based educational tools has to take prior knowledge into consideration, for example by using adaptive learning systems (see [20]) which respect prior knowledge during the adaptive process. However, we should also be aware that LLMs might have harming effects on learning compared to traditional learning, e.g. shown in [12], [13], and that the problems found in this paper add to that. Therefore, it might be more beneficial to students to use traditional teaching methods compared to LLM-based learning.

V. Conclusion

This paper presents a preliminary study that aims to investigate the effects of prior knowledge on a topic on how well students can utilize LLMs for their own learning efforts. The participants, all Computer science students, were first asked to fill out a self-assessment pre-test concerning their familiarity with LLMs, general automata theory, and Büchi automata. They were then given 15 minutes with an LLM to learn about Büchi automata, after which they had to fill out a post-test questionnaire.

Two effects could be observed comparing the pre-test and post-test scores:

- Participants with more familiarity with LLMs achieved better scores in the post test.
- Participants with more familiarity with general automata theory achieved worse scores in the post test.

Unfortunately, no participants reported any prior knowledge regarding Büchi automata, preventing any comparisons in that regard. A possible interpretation for the first effect might be that certain prompts elicit better responses from an LLM and students with more familiarity with an LLM are able to formulate more fitting prompts. However, further evaluation of the prompts used by students was not possible since those prompts were deleted after the study due to data protection. A possible interpretation for the second observation may be that students conflate Büchi automata with their knowledge on other automata models, some evidence for which can be found when qualitatively investigating the students' answers in the post-test; e.g. when asked to produce the accepted formal language of a given Büchi automata, most participants responded with the formal language for a finite automaton. This issue is compounded by the fact that Büchi automata and finite automata share the same visual notation.

Further issues students need to deal with when using an LLM to learn stem from the inability of an LLM to know what it is doing. When a student asked for a visual representation of a Büchi automaton they were represented with some artistic rendering (see Fig. 4) that had nothing to do with the standardised visual notation of a Büchi automaton. Another student made a typo in their initial prompt, asking for a "Büchli" automaton, after which the LLM spent the entire time teaching the student about book vending machines.

As only 12 participants could be won, the data set is quite small. As such, the representativity of the results is questionable. With the results being as they are, while some effects between pre- and post-test could be observed (see Fig. 3), they are not statistically significant ($p > 0.2$ for both), indicating a need to gather more data before conclusive insights can be gained.

ACKNOWLEDGEMENTS

This study was pre-registered at https://aspredicted.org/88nj-dtzx.pdf. All participants gave their informed consent. A basic ethics evaluation was done by the authors which concluded that no votum of the local ethics committee was needed.

REFERENCES

[1] S. Filippi and B. Motyl, "Large language models (llms) in engineering education: A systematic review and suggestions for practical adoption," *Information*, vol. 15, no. 6, 2024. [Online]. Available: https://www.mdpi.com/2078-2489/15/6/345

[2] O. E. Chinonso, A. M.-E. Theresa, and T. C. Aduke, "Chatgpt for teaching, learning and research: Prospects and challenges," *Global Academic Journal of Humanities and Social Sciences*, vol. 5, no. 02, p. 33–40, Mar. 2023. [Online]. Available: https://www.gajrc.com/media/articles/GAJHSS_52_33-40.pdf

[3] B. D. Nye, D. Mee, and M. G. Core, "Generative large language models for dialog-based tutoring: An early consideration of opportunities and concerns," in *Proceedings of the Workshop on Empowering Education with LLMs - the Next-Gen Interface and Content Generation 2023*, ser. CEUR Workshop Proceedings, S. Moore, J. Stamper, R. Tong, C. Cao, Z. Liu, X. Hu, Y. Lu, J. Liang, H. Khosravi, P. Denny, A. Singh, and C. Brooks, Eds., vol. 3487. Tokyo, Japan: CEUR, Jul. 2023, p. 78–88. [Online]. Available: https://ceur-ws.org/Vol-3487/#paper4

[4] R. Balse, V. Kumar, P. Prasad, and J. M. Warriem, "Evaluating the quality of llm-generated explanations for logical errors in cs1 student programs," in *Proceedings of the 16th Annual ACM India Compute Conference*, ser. COMPUTE '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 49–54. [Online]. Available: https://doi.org/10.1145/3627217.3627233

[5] H. A. Nguyen, H. Stec, X. Hou, S. Di, and B. M. McLaren, "Evaluating chatgpt's decimal skills and feedback generation in a digital learning game," in *Responsive and Sustainable Educational Futures*, O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, and T. Papathoma, Eds. Cham: Springer Nature Switzerland, 2023, pp. 278–293.

[6] H. C. Lucas, J. S. Upperman, and J. R. Robinson, "A systematic review of large language models and their implications in medical education," *Medical Education*, vol. 58, no. 11, pp. 1276–1285, 2024. [Online]. Available: https://doi.org/10.1111/medu.15402

[7] H. Alkaissi and S. I. McFarlane, "Artificial hallucinations in chatgpt: Implications in scientific writing," *Cureus*, Feb. 2023. [Online]. Available: doi.org/10.7759/cureus.35179

[8] Z. Ji, T. Yu, Y. Xu, N. Lee, E. Ishii, and P. Fung, "Towards mitigating LLM hallucination via self reflection," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 1827–1843. [Online]. Available: https://aclanthology.org/2023.findings-emnlp.123

[9] J. Gravel, M. D'Amours-Gravel, and E. Osmanlliu, "Learning to fake it: Limited responses and fabricated references provided by chatgpt for medical questions," *Mayo Clinic Proceedings: Digital Health*, vol. 1, no. 3, pp. 226–234, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949761223000366

[10] A. L. Smith, F. Greaves, and T. Panch, "Hallucination or confabulation? neuroanatomy as metaphor in large language models," *PLOS Digital Health*, vol. 2, no. 11, p. e0000388, Nov. 2023. [Online]. Available: https://dx.plos.org/10.1371/journal.pdig.0000388

[11] R. Klemke and H. Jardozka, *Exploring New Horizons: Generative Artificial Intelligence and Teacher Education*. Association for the Advancement of Computing in Education (AACE), 2024, p. 76–87. [Online]. Available: https://www.learntechlib.org/primary/p/223928/

[12] H. Bastani, O. Bastani, A. Sungu, G. Haosen, K. Ozge, and R. Mariman, "Generative ai can harm learning," preprint. [Online]. Available: https://hamsabastani.github.io/education_llm.pdf

[13] J. Prather, B. N. Reeves, J. Leinonen, S. MacNeil, A. S. Randrianasolo, B. A. Becker, B. Kimmel, J. Wright, and B. Briggs, "The widening gap: The benefits and harms of generative ai for novice programmers," in *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1*, ser. ICER '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 469–486. [Online]. Available: https://doi.org/10.1145/3632620.3671116

[14] N. R. Council, *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: The National Academies Press, 2000. [Online]. Available: https://nap.nationalacademies.org/catalog/9853/how-people-learn-brain-mind-experience-and-school-expanded-edition

[15] National Academies of Sciences, Engineering, and Medicine, *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: The National Academies Press, 2018. [Online]. Available: https://nap.nationalacademies.org/catalog/24783/how-people-learn-ii-learners-contexts-and-cultures

[16] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant," *Psychological Science*, vol. 22, no. 11, p. 1359–1366, Nov. 2011. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0956797611417632

[17] A. Gelman and L. Eric, "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time," 2013. [Online]. Available: http://www.stat.columbia.edu/~gelman/research/unpublished/forking.pdf

[18] B. Cowan, Y. Watanobe, and A. Shirafuji, "Enhancing programming learning with llms: Prompt engineering and flipped interaction," in *Proceedings of the 2023 4th Asia Service Sciences and Software Engineering Conference*, ser. ASSE '23. New York, NY, USA: Association for Computing Machinery, 2024, p. 10–16. [Online]. Available: https://doi.org/10.1145/3634814.3634816

[19] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Comput. Surv.*, vol. 55, no. 12, Mar. 2023. [Online]. Available: https://doi.org/10.1145/3571730

[20] S. Ennouamani and Z. Mahani, "An overview of adaptive e-learning systems," in *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2017, pp. 342–347.