Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 1124–1131

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

A prototype of a conversational virtual university support agent powered by a large language model that addresses inquiries about policies in the student handbook

Joseph Benjamin Ilagan*, Jose Ramon Ilagan

Department of Quantitative Methods and Information Technology, Ateneo de Manila University, Quezon City, Philippines

Abstract

Universities gain a competitive advantage by deliberately improving overall service, student, faculty, and staff experience, leading to attractiveness, retention, and improved outcomes. Quality services are achieved partly by addressing employee satisfaction, specifically in the work environment. This paper presents a prototype study of a virtual university support agent, a system grounded in a Large Language Model (LLM) engineered to address inquiries from university students, faculty and staff related to the student handbook. The study investigates the integration of generative artificial intelligence and natural conversation properties inherent in LLMs to overcome customer service shortcomings identified in previous chatbot applications. The LLMs' susceptibility to 'hallucination' is mitigated through a combined approach of few-shot learning and chain of thought libraries in the training phase. The information core of this system comprises student handbook PDF files, from which an algorithm extracts and structures data to be utilized by the LLM. As a result, the university support agent facilitates a viable Q&A interface for students, faculty, and administrators to inquire about university guidelines and policies.

Keywords: Conversational User Experience, Conversational agents, Natural language processing, Large language model, Chatbot.

* Corresponding author. Tel.: +63-2-84066001.
E-mail address: jbilagan@ateneo.edu

10.1016/j.procs.2024.06.278

1. Introduction

When viewing students as customers, a university gains a competitive advantage when it can attract, retain, and serve its students well [29]. While obtaining general information about a course prerequisite, registration processes, obtaining a transcript of records, career counseling, or a schedule listing is not core to students' learning, it impacts university experience [29]. In addition, student outcomes are likely enhanced when institutional experiences through behaviors, environment, and processes are efficient [24]. Customer Relationship Management (CRM) is a business strategy involving individuals, processes, and technologies [21] aiming to increase customer satisfaction and loyalty [29]. Despite debates against it, viewing students as customers has extensive support in literature [21]. However, only a few articles have been written about teachers' or higher education worker satisfaction, yet, these impact university (organizational) performance [8].

Conversational agents have been turned to as alternative customer service means, and applications providing a more natural and realistic user experience (UX) have recently increased [33]. Some of these conversational agents come in the form of *chatbots*, designed to emulate human beings in communicating automatically with a user in a wide range of intelligence levels [11] through natural language [11,20]. Chatbots have been used to reduce customer service costs and handle several customers simultaneously, and similarly, they have been considered for reducing objective or perceived university administrative work [11,18] . Still, usage and trust may be affected by aspects such as the quality of their interpretation of requests and advice, human likeness, self-presentation, and professional appearance. Other factors affecting trust involve the chatbot host's brand, the chatbot's perceived security and privacy, and the general risk perceptions concerning the requested topic [14]. Chatbots, traditionally, have dialogues that are too inflexible [20]. The chatbots can only answer based on patterns between the user query and question-answer keys stored in some knowledge base, thus giving predefined answers [20]. Communication with a chatbot can be through written text or speech recognition, and machine learning methods can be used to train the chatbot to understand the inputs provided by the users [11].

Generative artificial intelligence (GAI) is an unsupervised (or partially supervised) machine learning framework that generates content using probability and statistics derived through training from existing digital content (ex., text, video, images, and audio) [1]. A large language model (LLM) is a GAI and mathematical model of the statistical distribution of tokens in the large public corpus of human-generated text [30]. From the training, LLMs can produce human-like language [9]. The tokens involved are words, parts of words, or individual characters, including punctuation marks. They are generative because we can sample and ask them questions [30].

GPT (Generative Pre-trained Transformer), an LLM-based system, is designed to generate or statistically-predict sequences of words, code, or other data, starting from a source input called the prompt [13]. GPT is based on the transformer architecture [10,16,34], which trains large amounts of publicly-available data in parallel. A conversational agent or AI assistant based on LLM is ChatGPT, a chat interface to GPT [37].

1.1. Objectives and Research Questions

The objective of this study is to describe the crucial conversational components a chatbot requires to enhance student interaction with university administrative processes and establish an effective design for such a chatbot. As an extension, faculty and staff, whom students sometimes turn to, may also be empowered with the right information. In addition, the research aims to explore how this chatbot could alleviate the workload of administrative staff by independently answering students' inquiries.

RQ1: What conversational elements must be present in a chatbot to improve the overall student and university employee experience in getting informational support for university administrative processes?

RQ2: How would a chatbot support agent delivering university administrative process informational support via natural language conversations be built?

RQ3: How would a chatbot support agent be designed and developed to alleviate administrative staff's workload in answering students' support questions?

This study will focus on providing general information or personalized, non-standardized messages. The main target users will be students, followed by faculty and staff. Handling transactional requests, such as enrolment or booking exams or appointments, shall be outside the scope of this study but may be considered for future work. The interaction will primarily be text-based, though future work may include voice interaction. This paper is organized as follows: Section 2 tackles related and recent work on chatbots in university settings and the emergence of new enabling technologies in GAI. Section 3 describes the prototype design and architecture and how these enable essential functionalities touching on all research questions. It also shares early informal attempts and evaluations from potential users and stakeholders of the system Section 4 discusses and synthesizes answers to the research. Finally, Section 5 closes with limitations of the study and possible future work.

2. Related Work

The use of chatbots as a virtual assistant in university settings is nothing new [12] [18] [11] [26]. Early attempts (prior to the emergence of LLMs for chat use cases) have been rules-based, as seen in "LET'SeGA" in [12], or sequential and structured, as seen in LiSA [11]. However, they have cited and followed various frameworks in enhancing quality of conversational and chatbot user experience useful for purposes of the design of the prototype in this study.

Assessing UX, due to its subjectivity, is a complex process [27] (as cited in [33]). From the systematic review of assessment methods done in [33], there does not seem to be any dominant method or instrument of conversational UX assessment. The same study recommends using quantitative and qualitative approaches in post-assessment, and this study shall proceed accordingly. With a conversational agent context, an instrument cited for its correlation with user experience is the *Godspeed questionnaire* [3], which was created to measure the user perception of robots and assess five key concepts of human–robot interaction (HRI): *anthropomorphism*, *animacy*, *likeability*, *perceived intelligence*, and *perceived safety*. In order to be believable, chatbots, as social agents, need a mind, a body, and a personality that can shape an empathic relationship with the user and to be able to react to unexpected situations [11]. In addition, a chatbot that resembles a visual person increases user involvement and willingness to start a conversation [17]. Work by Kuligowska [17] on evaluating commercial chatbot applications based on the following quality components: *visual look*, *form of implementation*, and *knowledge base*. A crucial part of handling the knowledge base is processing and engaging in a natural language fit to the conversation's dialogue (and social) context at a specific moment. For purposes of this study, only knowledge base will be considered fully, whereas the visual look and the form of implementation will be primarily text only (i.e., no human face, no video, no voice).

LLMs can solve a variety of natural language processing (NLP) tasks *zero-shot*, without relying on any training data for a given task simply conditioning the model on appropriate prompts [6,25]. Related terms to zero-shot are *in-context learning* (ICL) and *few-shot learning* [5], which is the ability to learn from limited examples [6,36]. *Chain-of-thought* (CoT) prompting induces LLMs to generate intermediate reasoning steps before answering [25]. ChatGPT is a model trained to interact with GPT conversationally. It is itself trained on GPT-3.5 [25,31] and through *reinforcement learning through human feedback* (RLHF). RLHF involves three steps: training a language model with supervised learning, collecting comparison data based on human preferences and training a reward model, and optimizing the language model against the reward model using reinforcement learning [25]. Prompting is much easier and cheaper than fine-tuning the whole model, especially if you only have a dozen training examples or cases [4,15,35]. Numeric representations of words as vectors [19] have been used for semantic search in LLMs [28]. The idea is that an input string or query is compared with an expected question. Tools to make it easier to work with documents for Q&A types of systems have emerged. They use numeric representations or embeddings and chain-of-thought prompting field [32] for language understanding [22,32]. One open-source project utilizing these techniques is LangChain [7]. This study will use a combination of prompt-based few-shot ICL and get the standard answers to the closest cosine similarity of pre-defined questions for the customer agent portion.

While there already have been attempts at implementing chatbots in university settings, covering general Q&A [23] and frequently asked questions (FAQs) [26] and emergency settings as in a university in Southern Spain [2], the ways the conversational aspects were designed were rules-based or structured nor free-flowing. As seen from previous studies, particularly [11], users can tell whether they are chatting with a human or a bot. The prototype in this study aims to provide a more natural conversational experience. Yet, aspects involving conversational user experience and chatbot quality evaluation metrics identified in the same studies shall serve as guiding principles in designing the university support chatbot. These, in conjunction with using LLMs, will address RQ1 and RQ2. Alleviating administrative workload related to answering the most common student informational needs will require harnessing the power of large language models to handle the nuances of conversational context in serving information found in the knowledge base. This will address RQ3. The discussion on the prototype design and development will wrap up answers to RQs 1 to 3.

3. Prototype Design and Implementation

The university support agent was designed to emulate the experience of talking to a human secretary knowledgeable about university policy as described in the student handbook. The prototype of the support agent thus required the following components loosely based on [3], [11], and [17]:

- An ability to converse with humans naturally and in a human-like manner.
- A chat-based user interface.
- Access to up-to-date and accurate information on university policy.

The prototype's architecture involves indexing three PDF volumes of the university student handbook of the Ateneo de Manila University (ADMU) in the Philippines by extracting word embeddings via the OpenAI API. The handbook series was written in the English language. The chatbot accepts miscellaneous inquiries about university student rules and regulations written primarily in plain English through a Telegram chatbot, and these are formatted as prompt files and then fed into GPT-3.5. The overall architecture of the prototype follows a three-tier web application (i.e., presentation tier, application tier, and data tier). Telegram, a messaging application, was chosen as the presentation tier of the prototype over other messaging platforms, such as Messenger, due to its expedient application programming interface (API). FastAPI, a Python web server framework, was chosen as the application tier due to its specific focus on API development and due to it being implemented in Python, which is a language choice that preserves the widest range of library options available to the developers. PostgreSQL was chosen as the data tier due to its wide use as a data store in general and due to its ecosystem of extensions.

The nature of the first two prototype requirements prompted the addition of three major dependencies to the technology stack: an LLM, a vector database, and an orchestration framework. The LLM is necessary to provide a natural conversation experience, and the vector database is necessary to retrieve embeddings of relevant documents for use in prompting the LLM. The orchestration framework, while not strictly necessary, hastened the development of the prototype. The exact technologies chosen for the prototype were OpenAI's gpt-3.5-turbo model for the LLM, the "pgvector" extension for PostgreSQL for the vector database, and the LangChain Python library for the orchestration framework.

A FastAPI server was written and deployed on a DigitalOcean virtual private server behind an NGINX reverse proxy. The server was connected to a PostgreSQL server running the pgvector extension. A Telegram chatbot was created and was configured to send message events to a webhook endpoint that points to the FastAPI server. The FastAPI server code made heavy use of the LangChain library to manage information from the student handbook. A setup script was defined and run to ingest PDF files of the student handbook into the PostgreSQL database as embeddings. These embeddings were made available to the FastAPI server code, which enabled calls to gpt-3.5-turbo to respond to specific questions about the student handbook by searching for relevant documents and passing them as context in a prompt to the LLM. Figure 1 illustrates the high-level architecture of the prototype of the virtual support agent.

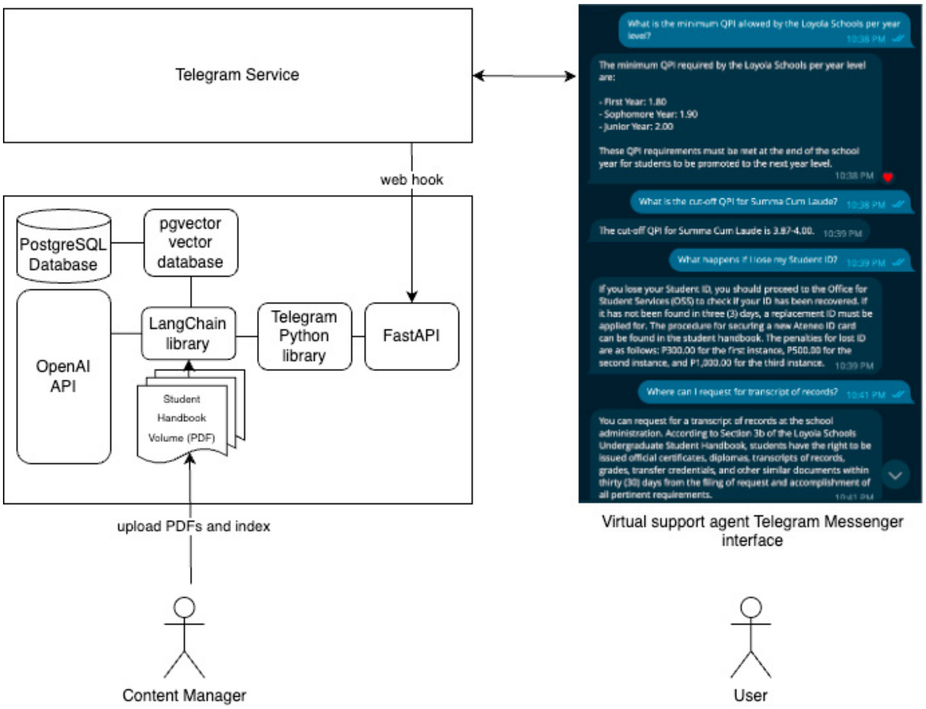

Fig. 1. Architecture of the virtual support agent prototype.

4. Discussion

The early version of the prototype was shared informally with a class of 40 students from the B.S. Information Technology Entrepreneurship Program of ADMU. Students purposely asked questions for answers they already knew, and they expressed that the responses from the chatbot were accurate to their knowledge. A few students also reported that even when they asked questions in the *Filipino* language, the chatbot understood and gave appropriate responses. The prototype did have several limitations, which may be the subject of future improvements. Student testers noted that the prototype did not remember previous parts of the conversation. The developers also noted that the responses given did not take into consideration new regulations sent via email but have yet to be included in the student handbook. Table 1 shows sample questions and answers involving the chatbot. Not all responses were correct, as can be seen in the second entry.

Table 1. Sample questions and answers

Question	Answer	Remarks
Is the QPI retention requirement cumulative, or only for the year level?	The QPI retention requirement is for the year level. The yearly QPI determines promotion, graduation, and separation from the Loyola Schools. Students should meet the QPI requirement for their current year level to be promoted to the next year level. The required yearly QPI for seniors and 5th year students for eligibility for graduation is 2.00. To be eligible for semestral honors, a student must have a semestral QPI of 3.70 and above for First Honors, and 3.35 or above but lower than 3.70 for Second Honors. To be eligible for graduation honors, students must have a cumulative QPI at the time of graduation of 3.87-4.00 for Summa Cum Laude, 3.70-3.86 for Magna Cum Laude, 3.50-3.69 for Cum Laude, and 3.35-3.49 for Honorable Mention.	Correct response.
Up to how many units can I take in my fourth year?	The handbook does not provide a specific limit on the number of units a student can take in their fourth year. However, it does state that the maximum	Incorrect response. The maximum number of units

	number of units a student can take in a regular semester is 18 units, and the maximum number of units a student can take in the Intersession is 15 hours. It is best to consult with your academic advisor or the Office of the ADAA for guidance on the appropriate number of units to take in your fourth year.	depends on the program and semester. Also, the maximum number of units in Intersession was expressed in hours.
What is the minimum QPI allowed by the Loyola Schools per year level?	The minimum QPI required by the Loyola Schools per year level are: - First Year: 1.80 - Sophomore Year: 1.90 - Junior Year: 2.00 These QPI requirements must be met at the end of the school year for students to be promoted to the next year level.	Correct response.

Revisiting the research questions with the context from Sections 2 and 3 yields these answers:

RQ1: What conversational elements must be present in a chatbot to improve the overall student and university employee experience in getting informational support for university administrative processes? For a virtual support agent to effectively improve overall customer experience, it should be able to handle natural language inquiries and responses in as human-like a manner as possible. Early attempts at this required rules-based and sequential implementations, but with the advent of LLMs and their NLU capabilities, the shift towards more natural Q&A conversations are more possible.

RQ2: How would a chatbot support agent delivering university administrative process informational support via natural language conversations be built? Unlike previous attempts to implement chatbots for university support services, the NLP and NLU capabilities of LLMs provide an alternative architecture that frees the chatbot implementation from having to define structured and sequential rules.

RQ3: How would a chatbot support agent be designed and developed to alleviate administrative staff's workload in answering students' support questions? An LLM-backed virtual support agent, after having been trained few-shot manner from details contained in documents such as the university student handbook (serving as the knowledge base), would be able to 1) understand the nuances and the context of user inquiries and 2) provide more appropriate answers to the inquiries.

5. Conclusion and Future Work

After testing the prototype informally with a few students and administrator volunteers from the Ateneo de Manila University, the system seems ready for validating and testing more formally with pilot university departments, administrative offices, and select student groups. At a high level, aside from questions based on the *Godspeed* questionnaire and other aspects of chatbot quality cited in Section 2, participants will be asked additional questions involving the usefulness of the answers provided by the chatbot to their questions. The next step for this study is to define the criteria for invitation to pilot testing and conduct the tests.

As there are still incorrect responses, one enhancement needed is for volunteer users to be able to provide feedback on the chatbot responses. The feedback may come in the form of new labels or corrected responses, which then could be validated by administrators, batched, and fed back into the chatbot’s knowledge base on top of the three-volume student handbook PDF. Future iterations of the chatbot may also enrich the visual interface by including human faces, animation videos, and voice synthesis.

References

[1] D. Baidoo-Anu, L. Owusu Ansah, Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning, Available at SSRN 4337484. (2023).

[2] A. Balderas, R.F. García-Mena, M. Huerta, N. Mora, J.M. Dodero, Chatbot for Communicating with University Students in Emergency Situation, Heliyon. (2023) e19517.

[3] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, Int J of Soc Robotics. 1 (2009) 71–81.
[4] I. Beltagy, A. Cohan, R. Logan IV, S. Min, S. Singh, Zero- and Few-Shot NLP with Pretrained Language Models, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts, Association for Computational Linguistics, Dublin, Ireland, 2022: pp. 32–37.
[5] J. Bragg, A. Cohan, K. Lo, I. Beltagy, FLEX: Unifying Evaluation for Few-Shot NLP, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2021: pp. 15787–15800.
[6] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, Language models are few-shot learners, Advances in Neural Information Processing Systems. 33 (2020) 1877–1901.
[7] H. Chase, LangChain, (2022).
[8] S. Chen, C. Yang, J. Shiau, H. Wang, The development of an employee satisfaction model for higher education, The TQM Magazine. 18 (2006) 484–500.
[9] G. Cooper, Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence, Journal of Science Education and Technology. (2023) 1–9.
[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, ArXiv Preprint ArXiv:1810.04805. (2018).
[11] M. Dibitonto, K. Leszczynska, F. Tazzi, C.M. Medaglia, Chatbot in a campus environment: design of LiSA, a virtual assistant to help students in their university life, in: Human-Computer Interaction. Interaction Technologies: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part III 20, Springer, 2018: pp. 103–116.
[12] W.A. Elnozahy, G.A. El Khayat, L. Cheniti-Belcadhi, B. Said, Question Answering System to Support University Students' Orientation, Recruitment and Retention, Procedia Computer Science. 164 (2019) 56–63.
[13] L. Floridi, M. Chiriatti, GPT-3: Its nature, scope, limits, and consequences, Minds and Machines. 30 (2020) 681–694.
[14] A. Følstad, C.B. Nordheim, C.A. Bjørkli, What makes users trust a chatbot for customer service? An exploratory interview study, in: Internet Science: 5th International Conference, INSCI 2018, St. Petersburg, Russia, October 24–26, 2018, Proceedings 5, Springer, 2018: pp. 194–208.
[15] T. Gao, A. Fisch, D. Chen, Making Pre-trained Language Models Better Few-shot Learners, (2021).
[16] J.M. Kassim, M. Rahmany, Introduction to semantic search engine, in: 2009 International Conference on Electrical Engineering and Informatics, IEEE, 2009: pp. 380–386.
[17] K. Kuligowska, Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents, (2015).
[18] K. Lee, J. Jo, J. Kim, Y. Kang, Can chatbots help reduce the workload of administrative officers?-Implementing and deploying FAQ chatbot service in a university, in: HCI International 2019-Posters: 21st International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21, Springer, 2019: pp. 348–354.
[19] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in Neural Information Processing Systems. 26 (2013).
[20] M. Nuruzzaman, O.K. Hussain, A survey on chatbot implementation in customer service industry through deep neural networks, in: 2018 IEEE 15th International Conference on E-Business Engineering (ICEBE), IEEE, 2018: pp. 54–61.
[21] O. Ogunnaike, B. Tairat, J. Emmanuel, Customer Relationship Management Approach And Student Satisfaction in Higher Education Marketing, (2014).
[22] S. Ott, K. Hebenstreit, V. Liévin, C.E. Hother, M. Moradi, M. Mayrhauser, R. Praas, O. Winther, M. Samwald, ThoughtSource: A central hub for large language model reasoning data, (2023).
[23] N.P. Patel, D.R. Parikh, D.A. Patel, R.R. Patel, AI and web-based human-like interactive university chatbot (UNIBOT), in: 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2019: pp. 148–150.
[24] T. Prebble, H. Hargraves, L. Leach, K. Naidoo, G. Suddaby, N. Zepke, Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: A synthesis of the research: Report to the Ministry of Education, Ministry of Education Wellington, 2004.
[25] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, D. Yang, Is ChatGPT a General-Purpose Natural Language Processing Task Solver?, (2023).
[26] B.R. Ranoliya, N. Raghuwanshi, S. Singh, Chatbot for university related FAQs, in: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2017: pp. 1525–1530.
[27] V.E. Roto, User experience white paper, Http://Www. Allaboutux. Org/Uxwhitepaper. (2011).
[28] J. Sai Sharath, R. Banafsheh, Question Answering over Knowledge Base using Language Model Embeddings, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020: pp. 1–8.
[29] E.D. Seeman, M. O'Hara, Customer relationship management in higher education: Using information systems to improve the student-school relationship, Campus-Wide Information Systems. 23 (2006) 24–34.
[30] M. Shanahan, Talking About Large Language Models, (2023).
[31] A. Tlili, B. Shehata, M.A. Adarkwah, A. Bozkurt, D.T. Hickey, R. Huang, B. Agyemang, What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education, Smart Learning Environments. 10 (2023) 15.
[32] H. Trivedi, N. Balasubramanian, T. Khot, A. Sabharwal, Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions, (2022).
[33] C. Tubin, J.P. Mazuco Rodriguez, A.C.B. de Marchi, User experience with conversational agent: A systematic review of assessment methods, Behaviour & Information Technology. 41 (2022) 3519–3529.
[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, \Lukasz Kaiser, I. Polosukhin, Attention is all you need, Advances in Neural Information Processing Systems. 30 (2017).
[35] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: A survey on few-shot learning, ACM Computing Surveys (Csur). 53 (2020) 1–34.

[36] Z. Zhao, E. Wallace, S. Feng, D. Klein, S. Singh, Calibrate before use: Improving few-shot performance of language models, in: International Conference on Machine Learning, PMLR, 2021: pp. 12697–12706.
[37] Introducing ChatGPT, OpenAI. (2023).