BIRMINGHAM CITY University

CMP6200
Individual Undergraduate Project
2024 - 2025

A2 - Literature Review and Methods

University Artifically Intelligent Assistant

Course: Computer & Data Science
Student Name: Lewis Higgins
Student Number: 22133848
Supervisor Name: Dr. Atif Azad

Contents

Report Introduction

1.1 Aims and Objectives

This project aims to leverage to aid new and existing students alike while they are attending university with helpful information about university itself, such as university societies, locations/campuses, and policies through the medium of a digital chatbot companion to converse with. The project's objectives are to:

- Conduct a thorough literature review on the surrounding topics, namely AI, LLMs and NLP.
- Create effective documentation for all stages of development, highlighting challenges faced during the process.
- Leverage Retrieval-Augmented Generation alongside a cloud-based LLM to query a vector database of university-related data.
- Develop a chatbot capable of accurately answering user queries related to university buildings, policies, and societies with a minimum 95% accuracy rate.
- Evaluate the effectiveness of an AI assistant on university student acclimatization.

1.2 Literature Search Methodology

My literature search will be performed using multiple reputable databases for academic papers, including:

- IEEE Xplore
- Scopus / Elsevier
- Google Scholar
- arXiv
- BCU Online Library

By using multiple different databases to source my information from, I can ensure that any potentially relevant literature will be found. Figure 1.1 depicts how in a search for 1685 articles about employee retention strategies and turnover, only 582 (25.7%) appeared in multiple databases (Wanyama et al., 2022), meaning that the remaining 74.3% of articles were exclusive to the single database in which they were found, emphasising the importance of searching multiple databases.

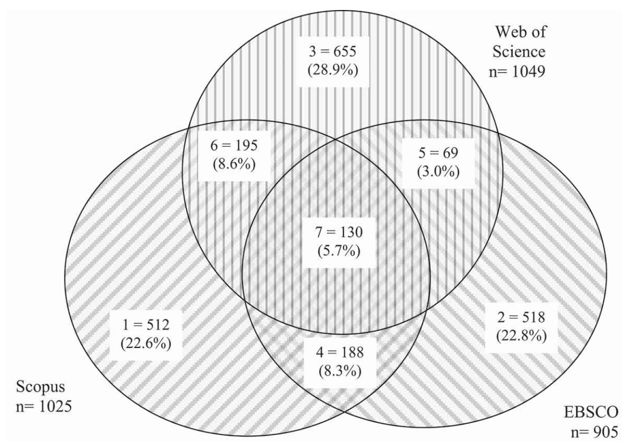


Figure 1.1: Distribution of searched articles across databases. (Wanyama et al., 2022)

All searches performed for recent literature will have a heavy preference to more recent literature, due to the constantly evolving fields my project is based on. The search terms I will use to retrieve the data I will be studying are:

- Artifical Intelligence / AI
- Natural Language Processing / NLP
- Large Language Models / LLMs
- Chatbots / Conversational Agents
- Retrieval-Augmented Generation / RAG

- User Experience / UX
    - Human-Computer Interaction

By using these specific terms that are directly relevant to the core themes of my project, I will be ensuring that I only retrieve literature that will be of crucial use in its development.

Literature Review

2.1 Themes

To develop the artefact and conduct thorough background research on relevant literature to further my knowledge of the subject areas, key general themes of the project were identified. From these themes, further keywords to be used in the literature search were derived to ensure that retrieved literature is directly relevant to my research and development of the final artefact. Due to the constantly evolving fields the project focuses on, it will be necessary to limit the results to primarily those written in recent years as there are frequent new developments in the subject areas.

Theme	Description	Keywords
AI	A field of computing dedicated to allowing computers to simulate human learning by training them on large amounts of data so that they can recognise patterns to classify or predict unknown data. AI can only be as good as the data it is trained upon, and can develop biases if it is fed too much data of a certain type.	Generative AI, Human-Centred AI, Explaianble AI, AI Ethics, AI Bias
Natural Language Processing	NLP refers to the use of machine learning to encode and process text to understand it in a similar way to humans, which can be used to allow direct two-way conversation between users and computers.	Deep learning, Vectorisation Sentiment analysis, Entity linking
LLMs	Large Language Models are a type of machine learning model dedicated to the recognition and generation of text. As suggested by their name, they are trained on enormous amounts of text data, which allows them to have active conversations with users. There are many different LLMs, and as their size and complexity increases, so too does the necessary processing power.	Retrieval augmented generation (RAG), Fine-tuning, Prompt engineering, Impact on industry, GPT4o, LLaMA, Gemini, Evaluation
Chatbot Conversational Agent	Software that simulates a natural conversation between the computer and end user. Many chatbots, including the one to be produced in this project, utilise recent developments such as Generative AI and natural language processing (NLP) to interpret and respond to user queries. (IBM, 2024c)	NLP, Digital assistant, ChatGPT, Risks, Impact on industry
Retrieval-Augmented Generation (RAG)	The optimisation of the generated text output of an LLM, incorporating an external data source to enhance its contextual knowledge and enhancing the subject relevancy of outputs.	Embedding, Vector databases, Document retrieval, Prompt engineering
User Experience (UX)	The end user's overall experience of using a system, such as its ease of use and whether it is enjoyable to use (Cambridge Dictionary, 2024). In the context of this project, it will refer to the user's ability to smoothly converse with the chatbot and how human-like it is.	Conversational design, usability, market research, human-computer interaction

2.2 Review of Literature

2.2.1 Artificial Intelligence (AI)

Researchers have always wanted to harness the processing power of computers to act in a similar manner indistinguishable from that of humans, most notably from as long ago as 1950, where the question was posed 'Can machines think?' (Turing, 1950). Ever since, constant innovations were made in computer intelligence and machine learning, from playing games of checkers at a better level than human players (Samuel, 1959) to classifying the contents of millions of images using convolutional neural networks (Krizhevsky et al., 2012). Recently, AI is used across many disciplines and for different purposes to complete tasks faster than, and in some cases better than, human workers. Wirtz et al. (2018) write that 'service robots' [1] can complete a variety of tangible or intangible actions, such as reading and sending text as a chatbot, seen in Figure 2.1.

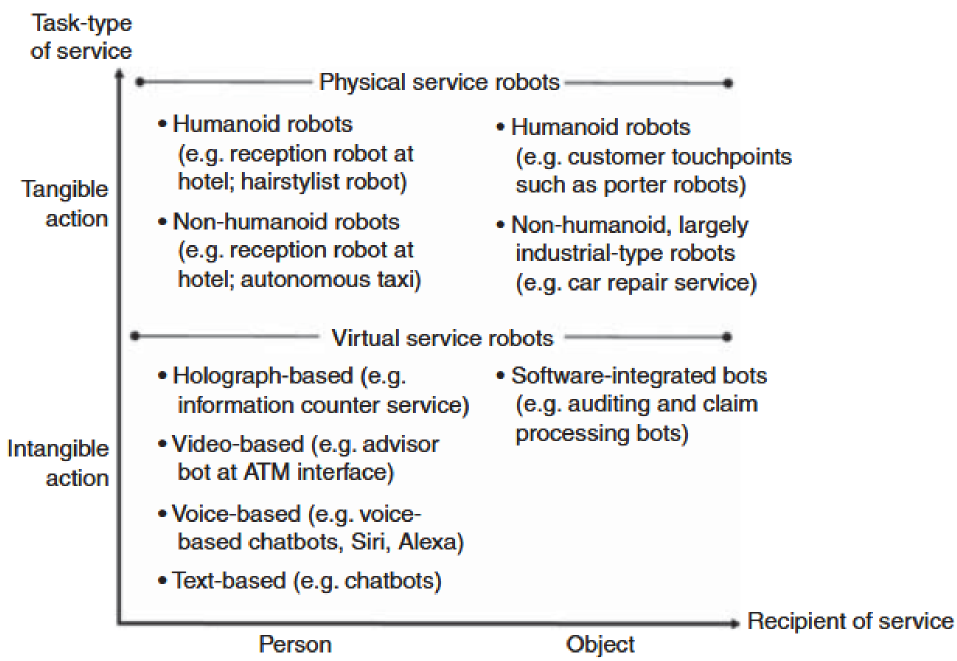


Figure 2.1: Service robots categorization by task-type and recipient of service (Wirtz et al., 2018).

Today, AI is still a constantly evolving field that is seeing bleeding-edge developments on a highly frequent basis, and more recently, is becoming instrumental in many people's work and private lives with the introduction of large language models (LLMs) (Maedche et al., 2019). However, when developing a project that utilises AI, it is important that they are ethical and human-centred in the development process, which is known as Human-Centred AI (HCAI). Another issue is the "black-box problem" - the inability to know an

[1] Defined as "system-based autonomous and adaptable interfaces that interact, communicate and deliver service to an organization's customers" (Wirtz et al., 2018, p.909)

AI's reasoning, meaning that eXplainable AI (XAI) is a growing necessity (Miró-Nicolau et al., 2025). In focusing on HCAI and XAI, the focus shifts from the machine executing the algorithms, and instead to the user and their experience using the AI (Shneiderman, 2020). In his article, Shneiderman strongly advocates for the promotion of HCAI for the benefit of both companies and their users, which is a commonly accepted idea due to the ethical risks of using AI. Because AI calculates outcomes from its training data rather than understanding social norms and perspectives, the use of it in sociotechnical systems poses serious risks due to the 'traps' it can fall into, because it cannot account for every possibility such as the personal tendencies and biases of its users (Selbst et al., 2019), and therefore developers require a shift in focus - from the final product at the end of development to the development process itself and end users, which also echoes Shneiderman's views.

2.2.2 Natural language processing (NLP)

The ability for a computer to interpret and understand human language greatly enhances the scale of their capabilities. This was recognised during the 1950s, where machine translation from Russian to English was demonstrated for the first time, albeit in a basic form (Jones, 1994). NLP has persistently been a key topic in computing, and even more so become in recent years, with its applications becoming very wide in scope with modern processing power. performing tasks such as sentiment analysis, which is the classification of the intent of a sentence, whether positive or negative for example, using recent developments in AI such as recurrent neural networks (RNNs) via libraries like TensorFlow (Abadi et al., 2016). Another use of NLP, as previously mentioned in a rudimentary form in the 1950s, is language translation. Back then, there were very limited technical options compared to those that exist today, and with today's AI, translation can be extremely accurate, albeit more computationally intensive. Both of these applications use RNNs, which are neural networks that are often superior to their alternatives such as convolutional (CNNs) and feedforward neural networks (FNNs) when analysing text due to the fact that they can retain information in their internal memory, which can allow them to recall context, allowing them to determine the linguistic relations between sentences within a document (Tang et al., 2015), which is especially useful in conversational interfaces where the user may say "it" to contextually refer to a previous noun from their last prompt. However, an even better option is a long short-term memory (LSTM), an updated form of an RNN with an even greater memory capacity that allows it to solve problems with long-term dependencies (Hochreiter and Schmidhuber, 1997) that often produce excellent results (Sherstinsky, 2020).

2.2.3 Large language models

LLMs are colossal machine learning models that leverage NLP to generate text. To do so, the training data required is immense, reaching 45 terabytes of pure text data for ChatGPT in 2023 (Dwivedi et al., 2023). This data is harvested from the web (Dubey et al., 2024) and social media due to it being one of the largest repositories of opinionated text data (Z. Wang et al., 2016), such as posts on platforms like Facebook and X. However, meticulous care is taken into the specific sources used to remove Personally Identifiable Information (PII) to minimise privacy and ethical concerns (Dubey et al., 2024). They are currently used widely across an assortment of industries in place of technical support and human resources systems, and can be supplied with text prompts from users which will cause the LLM to generate a response. Vaswani et al. (2017) proposed the Transformer architecture, which became a staple in LLMs due to the major reduction in necessary processing power to produce higher-quality results. The architecture they proposed underpins many massive LLMs today, including ChatGPT (Brown et al., 2020). Even with this revelation, however, LLMs are still extremely performance-intensive, requiring more than 8 top-range server-grade GPUs to run some of the most powerful high-parameter open-source models of today like LLaMA 3.1's 405 billion parameter model (Dubey et al., 2024). However, it is also important to note that the amount of parameters in a model does not entirely account for the quality of its responses, as studied by Ouyang et al. (2022) in Figure 2.2 wherein their surveys revealed their fine-tuned LLM with over 100x less parameters than a 175 billion parameter GPT3 model would often give answers preferred by its human assessors, which reveals that the fine-tuning of an LLM is of vital importance to the quality of its responses, even moreso than the amount of parameters.

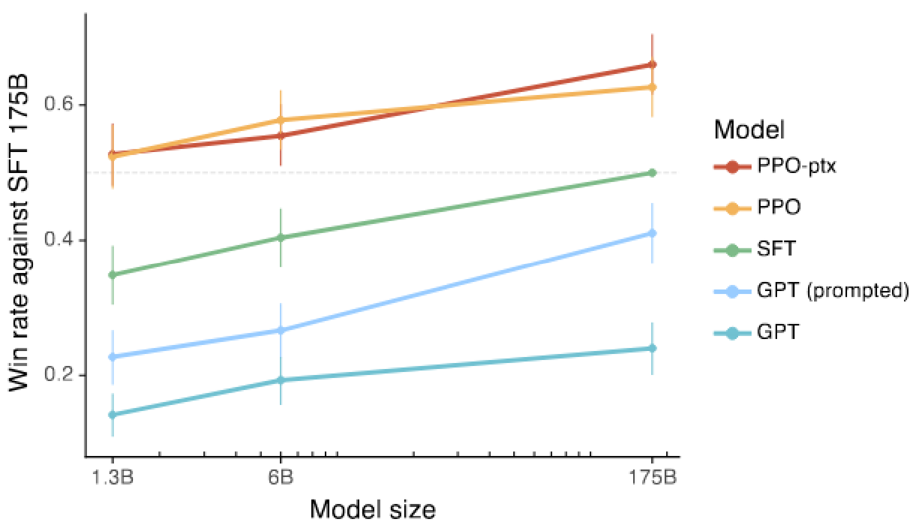


Figure 2.2: Human evaluations of the GPT models produced by Ouyang et al. (2022). PPO and PPO-ptx are their models.

2.2.4 Chatbots / Conversational Agents

Conversational agents, better known as chatbots, leverage natural language processing in order to simulate a conversational flow between a user and machine, and have become mainstream products in recent years (Liao et al., 2018), though have existed as far back as 1966 with the creation of "ELIZA" for the IBM-7094 (Weizenbaum, 1966). As time has passed, advancements in chatbots have occurred in "waves", where each new wave has brought a major innovation (Schöbel et al., 2024). As a product of the considerable developments in the field, chatbots are now widely used across industries such as education (Kuhail et al., 2023). However, the use of the latest wave of chatbots based on LLMs, especially in educational settings, poses significant risks as studied by Neumann et al. (2024) due to the risk of hallucinations being interpreted as absolute fact, although Shuster et al. (2021) argued that this risk can be greatly reduced through introducing RAG to the backend LLM, which is further backed by the RAG-based chatbot created by Ge et al. (2023), which they found to also give superior answers in their medical field of study to those of a general-purpose chatbot without RAG.

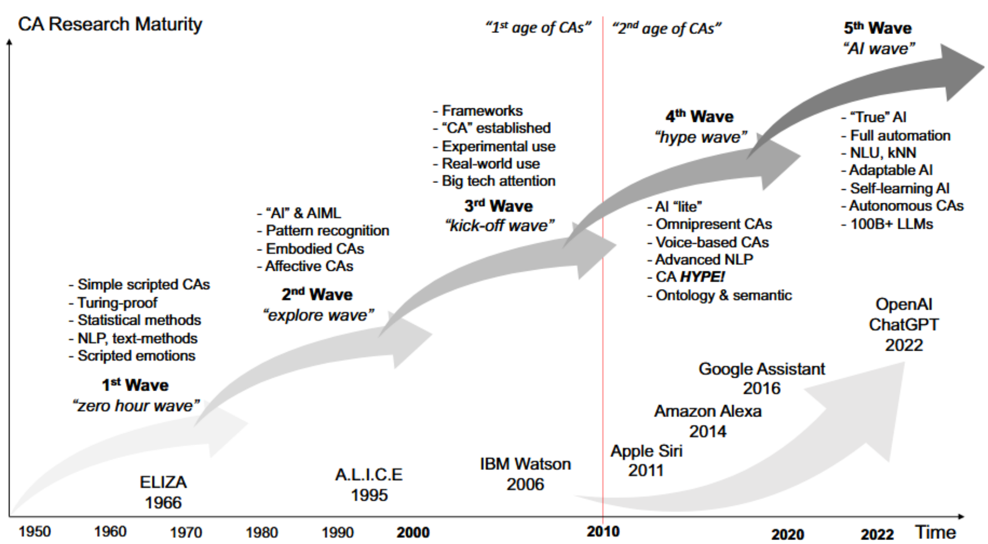


Figure 2.3: The five waves of conversational agent research (Schöbel et al., 2024).

Many platforms exist to aid chatbot development

2.2.5 Retrieval-Augmented Generation

While LLMs are highly useful tools across many industries, they are not without limitations. The most notable of these limitations are hallucinations (P. Lewis et al., 2021), where the LLM will fabricate information that conflicts with user input, earlier conversation context or true facts (Zhang et al., 2023). This occurs as a direct result of the LLM's parametric memory[2] being overfitted or biased, which can be counteracted through introducing an external knowledge source, known as non-parametric memory (Komeili et al. (2022), Siriwardhana et al. (2023)).

[2]Knowledge that the LLM has from its training data (Siriwardhana et al., 2023).

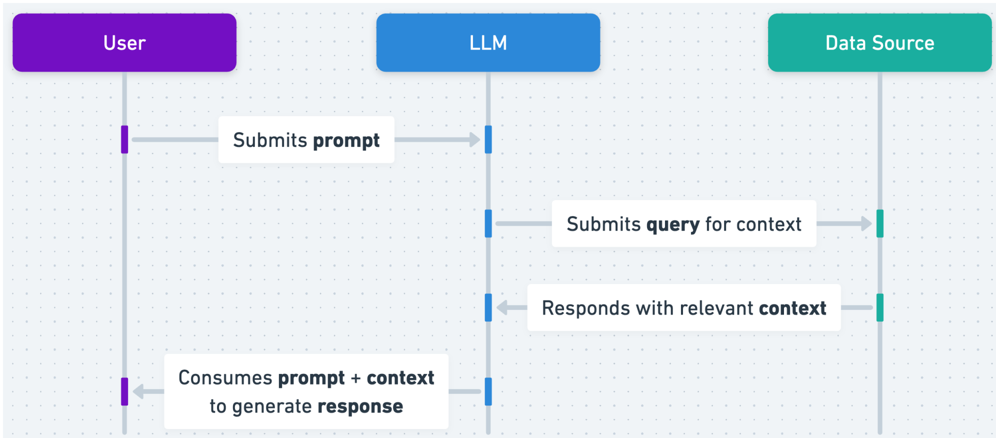


Figure 2.4: A basic overview of a RAG workflow (OpenAI, 2024).

Siriwardhana et al. (2023) expanded upon the earlier works of Karpukhin et al. (2020) and M. Lewis et al. (2020) by creating "RAG-end2end", an open-source model that allowed for dynamic updates to its attached knowledge base during training and also performed better than older RAG models in their evaluations.

RAG is often dependent upon vector databases, which store and process the vectors produced by embedding models for non-parametric memory (Li, 2023), which will make them an essential part of the backend of a RAG-enabled chatbot as studied by Odede and Frommholz (2024). Many options exist for the creation of vector databases, such as Milvus (J. Wang et al., 2021), Pinecone and Chroma. A study by Xie et al. (2023) compared these three, and cited Pinecone's 'robust distributed computing capabilities and scalability'. Pinecone was also used in Odede and Frommholz (2024)'s RAG chatbot, showcasing its potential as a vector database solution.

2.2.6 User experience and Human-Computer Interaction

The way people interact with their devices has drastically evolved over the years, from early MS-DOS command-line interfaces (CLIs) to mouse-based graphical user interfaces (GUIs), to touch screens (Kotian et al., 2024), greatly broadening the userbase of computers as a whole from exclusively those heavily invested in tech to a vast majority of the world's population. As such, inclusive and accessible design is increasingly important to maximise the audience of any software, especially considering the growing disabled population (Putnam et al., 2012). As well as being inclusive, the design should also be user-centred, meaning it should be an iterative process that is constantly taking user feedback into account (Chammas et al., 2015). However, there are some barriers in this process when developing chatbots, as studied by Clark et al. (2019) in their survey of university students who stated that they would always view a chatbot as a tool, and would not converse with them in the same way as they would a person, which would limit their potential use and hinder the overall design process. In this same context, it is also important to understand that users may struggle to get the chatbot to respond with information they want, as their prompts may be poorly understood due to issues like overgeneralisation (Zamfirescu-Pereira et al., 2023), and that users can quickly grow impatient after around 2 to 6 failed attempts, often branding the product as poor if this occurs (Luger and Sellen, 2016).

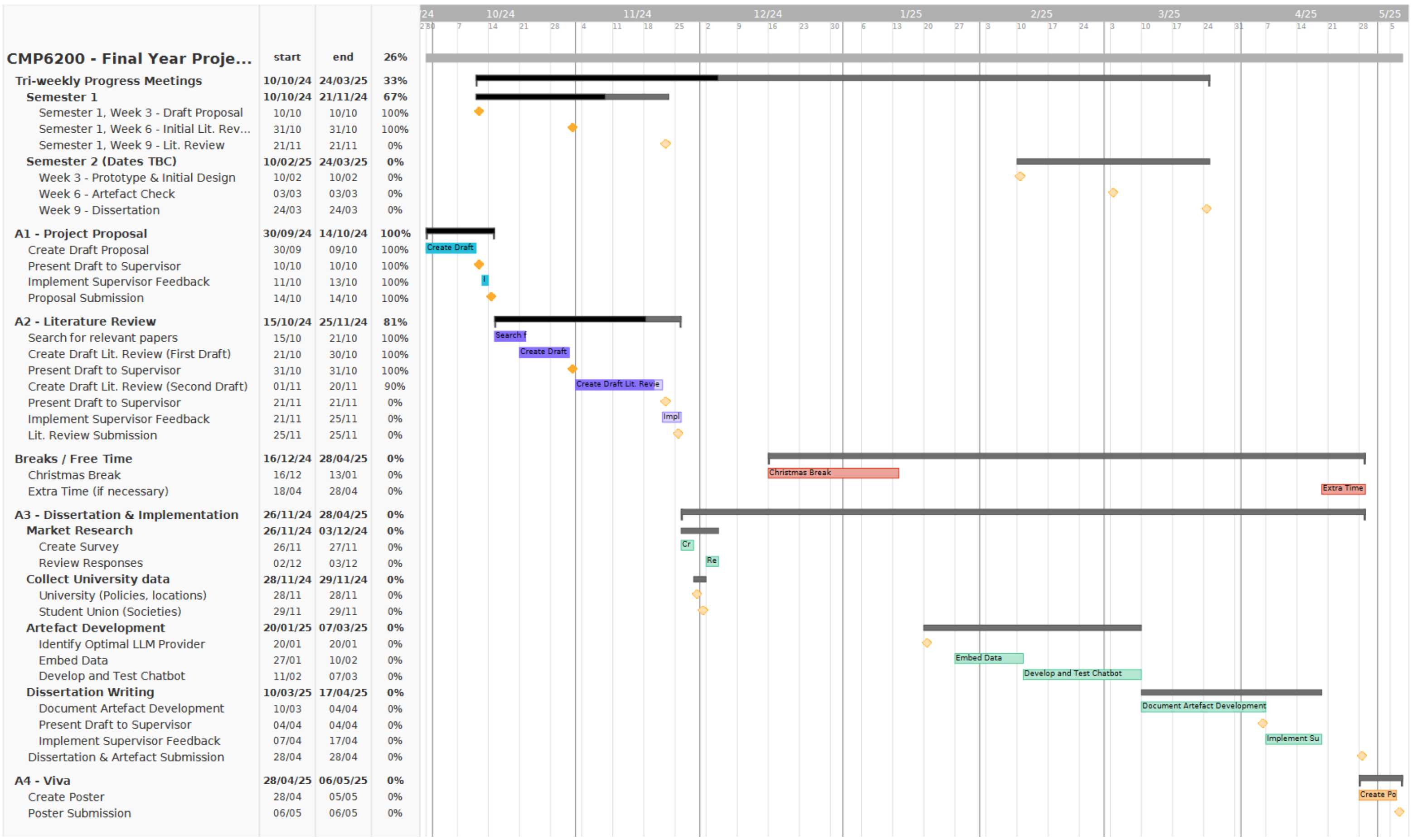
2.3 Summary

In conclusion, this literature review has revealed multiple key areas of focus for the development of the chatbot. The overall design of the chatbot must be iterative and human-centred, and user feedback should be obtained at every possible opportunity to ensure the resultant product is high quality. A deep exploration into AI, specifically in its applications in NLP and LLMs, has revealed that the best option for the chatbot will be to leverage a pre-existing cloud-based LLM's RAG capabilities to attach it to a database of university information, as to do so on a local machine would require an infeasible amount of processing power. In doing so, the LLM-based chatbot will be able to give good general answers as well as accurate university-specific answers regarding policies and societies.

Additionally, the discovery of many issues in the development of chatbots will greatly influence the design process, such as the need for users to be able to access the information they want in as few queries as possible to ensure user retention. Doing so will require prompt engineering to ensure that the LLM backend is generating specific information relevant to university rather than generic information that it may have been pre-trained with that would not be directly relevant.

Appendix

3.1 Gantt Chart



CMP6200 - Final Year Proje...	start	end	26%
Tri-weekly Progress Meetings	10/10/24	24/03/25	33%
Semester 1	10/10/24	21/11/24	67%
Semester 1, Week 3 - Draft Proposal	10/10	10/10	100%
Semester 1, Week 6 - Initial Lit. Rev...	31/10	31/10	100%
Semester 1, Week 9 - Lit. Review	21/11	21/11	0%
Semester 2 (Dates TBC)	10/02/25	24/03/25	0%
Week 3 - Prototype & Initial Design	10/02	10/02	0%
Week 6 - Artefact Check	03/03	03/03	0%
Week 9 - Dissertation	24/03	24/03	0%
A1 - Project Proposal	30/09/24	14/10/24	100%
Create Draft Proposal	30/09	09/10	100%
Present Draft to Supervisor	10/10	10/10	100%
Implement Supervisor Feedback	11/10	13/10	100%
Proposal Submission	14/10	14/10	100%
A2 - Literature Review	15/10/24	25/11/24	81%
Search for relevant papers	15/10	21/10	100%
Create Draft Lit. Review (First Draft)	21/10	30/10	100%
Present Draft to Supervisor	31/10	31/10	100%
Create Draft Lit. Review (Second Draft)	01/11	20/11	90%
Present Draft to Supervisor	21/11	21/11	0%
Implement Supervisor Feedback	21/11	25/11	0%
Lit. Review Submission	25/11	25/11	0%
Breaks / Free Time	16/12/24	28/04/25	0%
Christmas Break	16/12	13/01	0%
Extra Time (if necessary)	18/04	28/04	0%
A3 - Dissertation & Implementation	26/11/24	28/04/25	0%
Market Research	26/11/24	03/12/24	0%
Create Survey	26/11	27/11	0%
Review Responses	02/12	03/12	0%
Collect University data	28/11/24	29/11/24	0%
University (Policies, locations)	28/11	28/11	0%
Student Union (Societies)	29/11	29/11	0%
Artefact Development	20/01/25	07/03/25	0%
Identify Optimal LLM Provider	20/01	20/01	0%
Embed Data	27/01	10/02	0%
Develop and Test Chatbot	11/02	07/03	0%
Dissertation Writing	10/03/25	17/04/25	0%
Document Artefact Development	10/03	04/04	0%
Present Draft to Supervisor	04/04	04/04	0%
Implement Supervisor Feedback	07/04	17/04	0%
Dissertation & Artefact Submission	28/04	28/04	0%
A4 - Viva	28/04/25	06/05/25	0%
Create Poster	28/04	05/05	0%
Poster Submission	06/05	06/05	0%

Figure 3.1: The updated Gantt Chart for the development timeline.

References

Chammas, Adriana, Manuela Quaresma, and Cláudia Mont'Alvão (Jan. 1, 2015). "A Closer Look on the User Centred Design". In: *Procedia Manufacturing*. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015 3, pp. 5397–5404. ISSN: 2351-9789. DOI: 10.1016/j.promfg.2015.07.656.

Clark, Leigh, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan (May 2, 2019). "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. New York, NY, USA: Association for Computing Machinery, pp. 1–12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300705.

Dubey, Abhimanyu et al. (Aug. 15, 2024). *The Llama 3 Herd of Models*. DOI: 10.48550/arXiv.2407.21783. arXiv: 2407.21783.

Dwivedi, Y.K., N. Kshetri, L. Hughes, E.L. Slade, A. Jeyaraj, A.K. Kar, A.M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, H. Albanna, M.A. Albashrawi, A.S. Al-Busaidi, J. Balakrishnan, Y. Barlette, S. Basu, I. Bose, L. Brooks, D. Buhalis, L. Carter, S. Chowdhury, T. Crick, S.W. Cunningham, G.H. Davies, R.M. Davison, R. Dé, D. Dennehy, Y. Duan, R. Dubey, R. Dwivedi, J.S. Edwards, C. Flavián, R. Gauld, V. Grover, M.-C. Hu, M. Janssen, P. Jones, I. Junglas, S. Khorana, S. Kraus, K.R. Larsen, P. Latreille, S. Laumer, F.T. Malik, A. Mardani, M. Mariani, S. Mithas, E. Mogaji, J.H. Nord, S. O'Connor, F. Okumus, M. Pagani, N. Pandey, S. Papagiannidis, I.O. Pappas, N. Pathak, J. Pries-Heje, R. Raman, N.P. Rana, S.-V. Rehm, S. Ribeiro-Navarrete, A. Richter, F. Rowe, S. Sarker, B.C. Stahl, M.K. Tiwari, W. van der Aalst, V. Venkatesh, G. Viglia, M. Wade, P. Walton, J. Wirtz, 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". In: *International Journal of Information Management* 71. DOI: 10.1016/j.ijinfomgt.2023.102642.

Ge, Jin, Steve Sun, Joseph Owens, Victor Galvez, Oksana Gologorskaya, Jennifer C Lai, Mark J Pletcher, and Ki Lai (Nov. 1, 2023). "Development of a Liver Disease-Specific Large Language Model Chat Interface using Retrieval Augmented Generation". In: *medRxiv*, p. 2023.11.10.23298364. DOI: 10.1101/2023.11.10.23298364.

Hochreiter, Sepp and Jürgen Schmidhuber (Nov. 1, 1997). "Long Short-Term Memory". In: *Neural Comput.* 9 (8), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

Karpukhin, Vladimir, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih (Nov. 2020). "Dense Passage Retrieval for Open-Domain Question Answering". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. EMNLP 2020. Ed. by Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu. Online: Association for Computational Linguistics, pp. 6769–6781. DOI: 10.18653/v1/2020.emnlp-main.550.

Komeili, Mojtaba, Kurt Shuster, and Jason Weston (May 2022). "Internet-Augmented Dialogue Generation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL 2022. Ed. by Smaranda Muresan, Preslav Nakov, and Aline Villavicencio. Dublin, Ireland: Association for Computational Linguistics, pp. 8460–8478. DOI: 10.18653/v1/2022.acl-long.579.

Kotian, Abhijith L, Reshna Nandipi, Ushag M, Usha Rani S, VARSHAUK, and Veena G T (Jan. 2024). "A Systematic Review on Human and Computer Interaction". In: *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 1214–1218. DOI: 10.1109/IDCIoT59759.2024.10467622.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc. DOI: 10.1145/3065386.

Kuhail, M.A., N. Alturki, S. Alramlawi, and K. Alhejori (2023). "Interacting with educational chatbots: A systematic review". In: *Education and Information Technologies* 28 (1), pp. 973–1018. DOI: 10.1007/s10639-022-11177-3.

Lewis, Mike, Marjan Ghazvininejad, Gargi Ghosh, Armen Aghajanyan, Sida Wang, and Luke Zettlemoyer (June 26, 2020). *Pre-training via Paraphrasing*. DOI: 10.48550/arXiv.2006.15020. arXiv: 2006.15020.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela (Apr. 12, 2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. DOI: 10.48550/arXiv.2005.11401. arXiv: 2005.11401.

Li, Feifei (Aug. 1, 2023). "Modernization of Databases in the Cloud Era: Building Databases that Run Like Legos". In: *Proc. VLDB Endow.* 16 (12), pp. 4140–4151. ISSN: 2150-8097. DOI: 10.14778/3611540.3611639.

Liao, Q. Vera, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N. Sadat Shami, and Werner Geyer (Apr. 19, 2018). "All Work and No Play?" In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, pp. 1–13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173577.

Luger, Ewa and Abigail Sellen (May 7, 2016). ""Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. New York, NY, USA: Association for Computing Machinery, pp. 5286–5297. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858288.

Maedche, Alexander, Christine Legner, Alexander Benlian, Benedikt Berger, Henner Gimpel, Thomas Hess, Oliver Hinz, Stefan Morana, and Matthias Söllner (Aug. 1, 2019). "AI-Based Digital Assistants". In: *Business & Information Systems Engineering* 61 (4), pp. 535–544. ISSN: 1867-0202. DOI: 10.1007/s12599-019-00600-8.

Miró-Nicolau, M., A. Jaume-i-Capó, and G. Moyà-Alcover (2025). "A comprehensive study on fidelity metrics for XAI". In: *Information Processing and Management* 62 (1). DOI: 10.1016/j.ipm.2024.103900.

Neumann, Alexander Tobias, Yue Yin, Sulayman Sowe, Stefan Decker, and Matthias Jarke (2024). "An LLM-Driven Chatbot in Higher Education for Databases and Information

Systems". In: *IEEE Transactions on Education*. Conference Name: IEEE Transactions on Education, pp. 1–14. ISSN: 1557-9638. DOI: 10.1109/TE.2024.3467912.

Odede, Julius and Ingo Frommholz (Mar. 10, 2024). "JayBot – Aiding University Students and Admission with an LLM-based Chatbot". In: *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval*. CHIIR '24. New York, NY, USA: Association for Computing Machinery, pp. 391–395. ISBN: 9798400704345. DOI: 10.1145/3627508.3638293.

OpenAI (2024). *Retrieval Augmented Generation (RAG) and Semantic Search for GPTs | OpenAI Help Center*. URL: https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts (visited on 11/23/2024).

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe (Mar. 4, 2022). *Training language models to follow instructions with human feedback*. DOI: 10.48550/arXiv.2203.02155. arXiv: 2203.02155.

Putnam, Cynthia, Kathryn Wozniak, Mary Jo Zefeldt, Jinghui Cheng, Morgan Caputo, and Carl Duffield (Oct. 22, 2012). "How do professionals who create computing technologies consider accessibility?" In: *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ASSETS '12. New York, NY, USA: Association for Computing Machinery, pp. 87–94. ISBN: 978-1-4503-1321-6. DOI: 10.1145/2384916.2384932.

Samuel, A. L. (July 1959). "Some Studies in Machine Learning Using the Game of Checkers". In: *IBM Journal of Research and Development* 3 (3). Conference Name: IBM Journal of Research and Development, pp. 210–229. ISSN: 0018-8646. DOI: 10.1147/rd.33.0210.

Schöbel, Sofia, Anuschka Schmitt, Dennis Benner, Mohammed Saqr, Andreas Janson, and Jan Marco Leimeister (Apr. 1, 2024). "Charting the Evolution and Future of Conversational Agents: A Research Agenda Along Five Waves and New Frontiers". In: *Information Systems Frontiers* 26 (2), pp. 729–754. ISSN: 1572-9419. DOI: 10.1007/s10796-023-10375-9.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi (Jan. 29, 2019). "Fairness and Abstraction in Sociotechnical Systems". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. FAT* '19. New York, NY, USA: Association for Computing Machinery, pp. 59–68. ISBN: 978-1-4503-6125-5. DOI: 10.1145/3287560.3287598.

Sherstinsky, Alex (Mar. 1, 2020). "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network". In: *Physica D: Nonlinear Phenomena* 404, p. 132306. ISSN: 0167-2789. DOI: 10.1016/j.physd.2019.132306.

Shneiderman, Ben (Oct. 16, 2020). "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems". In: *ACM Trans. Interact. Intell. Syst.* 10 (4), 26:1–26:31. ISSN: 2160-6455. DOI: 10.1145/3419764.

Shuster, Kurt, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston (Nov. 2021). "Retrieval Augmentation Reduces Hallucination in Conversation". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Findings 2021. Ed. by Marie-

Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih. Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 3784–3803. DOI: 10.18653/v1/2021.findings-emnlp.320.

Siriwardhana, Shamane, Rivindu Weerasekera, Elliott Wen, Tharindu Kaluarachchi, Rajib Rana, and Suranga Nanayakkara (Jan. 12, 2023). "Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering". In: *Transactions of the Association for Computational Linguistics* 11, pp. 1–17. ISSN: 2307-387X. DOI: 10.1162/tacl_a_00530.

Tang, Duyu, Bing Qin, and Ting Liu (Sept. 2015). "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2015. Ed. by Lluís Màrquez, Chris Callison-Burch, and Jian Su. Lisbon, Portugal: Association for Computational Linguistics, pp. 1422–1432. DOI: 10.18653/v1/D15-1167.

Turing, A. M. (Oct. 1, 1950). "I.—COMPUTING MACHINERY AND INTELLIGENCE". In: *Mind* LIX (236), pp. 433–460. ISSN: 1460-2113, 0026-4423. DOI: 10.1093/mind/LIX.236.433.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin (Dec. 4, 2017). "Attention is all you need". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., pp. 6000–6010. ISBN: 978-1-5108-6096-4. DOI: 10.48550/arXiv.1706.03762.

Wang, Jianguo, Xiaomeng Yi, Rentong Guo, Hai Jin, Peng Xu, Shengjun Li, Xiangyu Wang, Xiangzhou Guo, Chengming Li, Xiaohai Xu, Kun Yu, Yuxing Yuan, Yinghao Zou, Jiquan Long, Yudong Cai, Zhenxiang Li, Zhifeng Zhang, Yihua Mo, Jun Gu, Ruiyi Jiang, Yi Wei, and Charles Xie (June 18, 2021). "Milvus: A Purpose-Built Vector Data Management System". In: *Proceedings of the 2021 International Conference on Management of Data*. SIGMOD '21. New York, NY, USA: Association for Computing Machinery, pp. 2614–2627. ISBN: 978-1-4503-8343-1. DOI: 10.1145/3448016.3457550.

Wang, Zhaoxia, Chee Seng Chong, Landy Lan, Yinping Yang, Seng Beng Ho, and Joo Chuan Tong (Dec. 2016). "Fine-grained sentiment analysis of social media with emotion sensing". In: *2016 Future Technologies Conference (FTC)*. 2016 Future Technologies Conference (FTC), pp. 1361–1364. DOI: 10.1109/FTC.2016.7821783.

Wanyama, Seperia B., Ronald W. McQuaid, and Markus Kittler (May 4, 2022). "Where you search determines what you find: the effects of bibliographic databases on systematic reviews". In: *International Journal of Social Research Methodology* 25 (3), pp. 409–422. ISSN: 1364-5579. DOI: 10.1080/13645579.2021.1892378.

Weizenbaum, Joseph (Jan. 1, 1966). "ELIZA—a computer program for the study of natural language communication between man and machine". In: *Commun. ACM* 9 (1), pp. 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168.

Wirtz, J., P.G. Patterson, W.H. Kunz, T. Gruber, V.N. Lu, S. Paluch, and A. Martins (2018). "Brave new world: service robots in the frontline". In: *Journal of Service Management* 29 (5), pp. 907–931. DOI: 10.1108/JOSM-04-2018-0119.

Xie, Xingrui, Han Liu, Wenzhe Hou, and Hongbin Huang (Dec. 2023). "A Brief Survey of Vector Databases". In: *2023 9th International Conference on Big Data and Information Analytics (BigDIA)*. 2023 9th International Conference on Big Data and Information

Analytics (BigDIA). ISSN: 2771-6902, pp. 364–371. DOI: 10.1109/BigDIA60676.2023.10429609.

Zamfirescu-Pereira, J.D., Richmond Y. Wong, Bjoern Hartmann, and Qian Yang (Apr. 19, 2023). "Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI '23. New York, NY, USA: Association for Computing Machinery, pp. 1–21. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581388.

Zhang, Yue, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, Longyue Wang, Anh Tuan Luu, Wei Bi, Freda Shi, and Shuming Shi (Sept. 24, 2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. DOI: 10.48550/arXiv.2309.01219. arXiv: 2309.01219.

Bibliography

Sources consulted but not directly cited

Cambridge Dictionary (2024). *Meaning of user experience in English.* URL: https://dictionary.cambridge.org/dictionary/english/user-experience (visited on 10/28/2024).

Cloudflare (2024). *What is a large language model (LLM)?* URL: https://www.cloudflare.com/en-gb/learning/ai/what-is-large-language-model/ (visited on 10/28/2024).

IBM (Aug. 16, 2024a). *What is AI?* URL: https://www.ibm.com/topics/artificial-intelligence (visited on 10/28/2024).

IBM (Aug. 11, 2024b). *What Is NLP (Natural Language Processing)? | IBM.* URL: https://www.ibm.com/topics/natural-language-processing (visited on 11/04/2024).

IBM (2024c). *What is a chatbot?* URL: https://www.ibm.com/topics/chatbots (visited on 10/28/2024).

IBM (2024d). *What is generative AI?* URL: https://research.ibm.com/blog/what-is-generative-AI (visited on 10/28/2024).

ICO (2024). *Definitions.* URL: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/ (visited on 10/28/2024).

MIT (Nov. 9, 2023). *Explained: Generative AI.* URL: https://news.mit.edu/2023/explained-generative-ai-1109 (visited on 10/28/2024).