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

> **Draft notice**
>
> This is a very early draft of this literature review and will be subject to major change over the next month. I've marked sections that I've taken from the proposal, or ones that I'm currently uncertain of, with boxes similar to this one.

1.1 Aims and Objectives

> **Copied from proposal**
>
> These are still subject to change pending the feedback from my proposal.

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

- Develop a chatbot capable of accurately answering user queries related to university buildings, policies, and societies with a minimum 95% accuracy rate.

- 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.

- Manage time effectively to ensure all project milestones are met on a consistent and regular timeframe.

- Evaluate the effectiveness of an AI assistant on university student acclimatization.

1.2　Literature Search Methodology

> **Uncertainty - Unfinished section**
>
> I assumed that I should include the exact search terms including special characters like asterisks as well as AND + OR queries, but I'm not sure how to best display that in this document. I'm also aware that a few of these are parts of each other (Deep learning and NLP for example) though I would surely need to research both anyway?

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

- IEEE Xplore

- Scopus / Elsevier

- Google Scholar

- 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, McQuaid, and Kittler, 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.
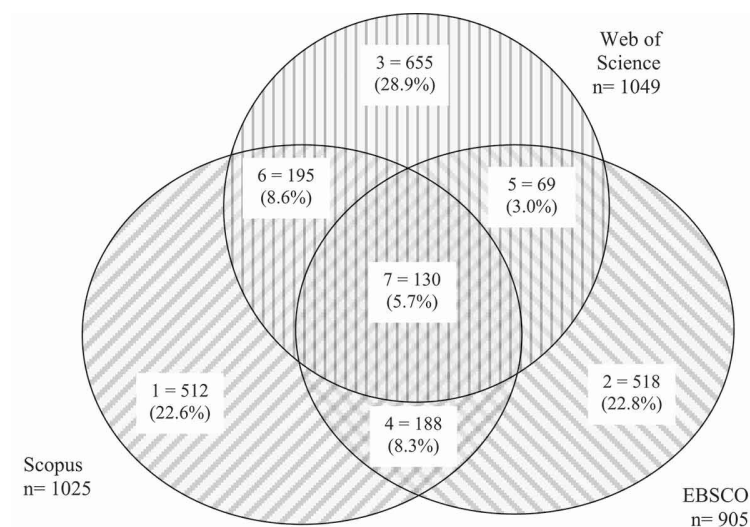

Figure 1.1: Distribution of searched articles across databases. (Wanyama, McQuaid, and Kittler, 2022)

All searches performed will be on papers published during or after 2020, 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

  - Ethics

    * Bias
    * Fairness
    * Privacy

- Chatbots / Digital Assistants

- Generative AI

- Large Language Models / LLMs

  - Fine tuning

- Natural Language Processing / NLP

  - Intent recognition
  - Entity extraction

- Deep learning

- User Experience / UX

  - Design

- Information retrieval

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

My project has many underlying themes, including:

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, AI Ethics, AI Bias
Generative AI	AI dedicated to the generation of content rather than prediction or classification. It is possible for generative AI to produce text, images and more recently, even video and sound.	LLMs, Tokens, Embedding
Chatbot Digital Assistant	Software that simulates a natural conversation between the computer and end user. Many chatbots, including the one I intend to develop, utilise recent developments such as Generative AI and NLP to interpret and respond to user queries. (IBM, 2024b)	User Experience, Conversational Design, Microsoft Bot Framework, Watson Assistant, ChatGPT
LLM	Large Language Models are a type of AI 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.	GPT4o, LLaMA, Gemini, Claude

2.2 Review of Literature

2.2.1 Review

Check the template document because I'm not too sure if it's formatted this way.

2.2.2 Theory

Check the template document because I'm not too sure if it's formatted this way.

2.3 Summary
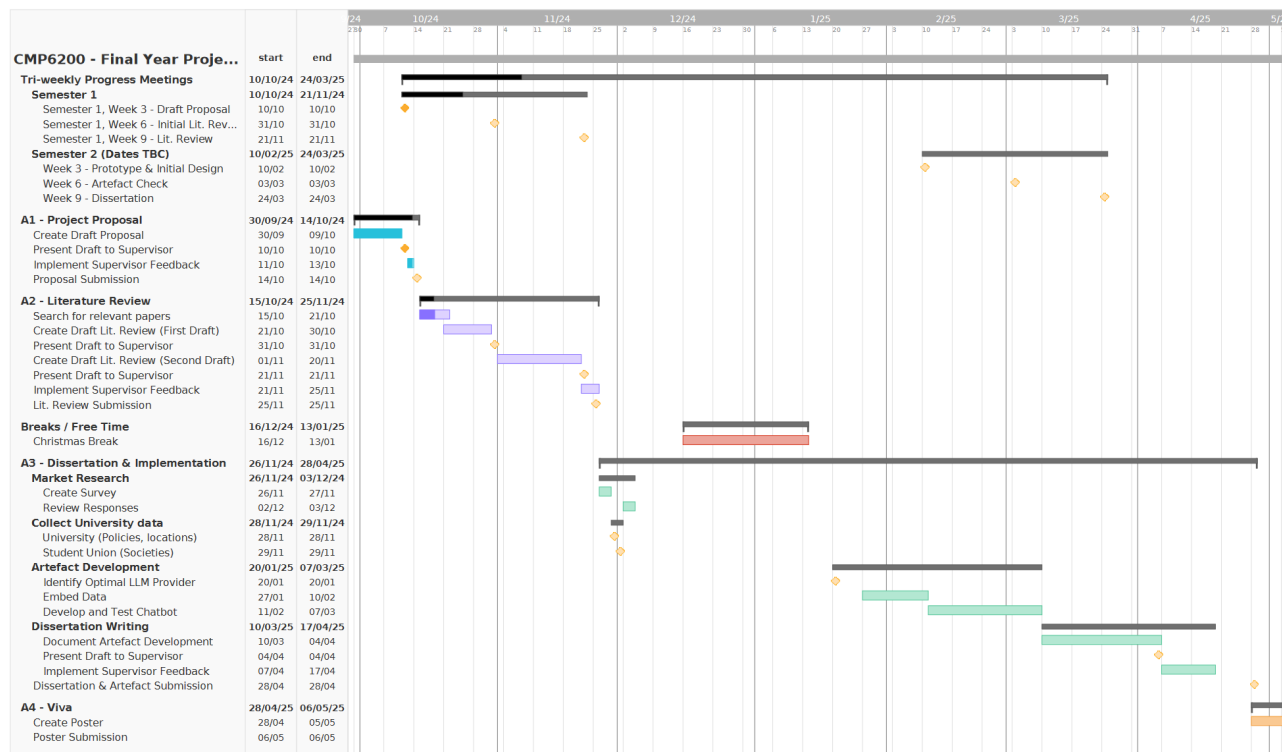
Appendix

3.1 Gantt Chart

Figure 3.1: A conceptual Gantt chart of a development timeline.

References and bibliographies aren't the same thing. Your references are your directly cited sources, whereas your bibliography is everything you consulted for information, even if you didn't directly cite it.

References

Wanyama, Seperia B., Ronald W. McQuaid, and Markus Kittler (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. DOI: 10.1080/13645579.2021.1892378.

Bibliography

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

IBM (2024c). *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).

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