BIRMINGHAM CITY University

CMP6200
Individual Undergraduate Project
2024 - 2025

A1 - Proposal

University Artifically Intelligent Assistant

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

Contents

Introduction

1.1 Background and Rationale

With artificial intelligence (AI) becoming increasingly more powerful and useful in recent years, in some cases even surpassing humans in areas like language and image recognition (Giattino et al., 2023). It is therefore essential that higher education institutions take advantage of it and ensure to keep up-to-date with its developments in the interest of academic integrity, which was also stated in the Higher Education Policy Report dated 28 March 2024 - "Higher education will have to adopt, adapt, collaborate and lead to take advantage of AI while managing the risks" (HEPI, 2024).

This project aims to leverage recent technical developments in natural language processing (NLP) to create a digital assistant for university life that students can use to gain information on topics such as university policies and locations on campus. This is because attending university for the first time is a daunting and stressful experience for many, often due to it being a new and unfamiliar environment where students have full independence unlike their previous educational settings, which could possibly lead to declines in academic performance and social activity. Some students may not wish to speak with newer people who they don't know about university topics for fear of ridicule or embarrassment, and would benefit from a digital companion to help them to become acquainted with their new environment without the perceived risk of social judgement.

Chatbots are also a significant tool across the tech sphere, with 73% of surveyed web users expecting companies to have chatbots for convenient interactions (Cherniak, 2024). By giving students a quick and easy tool to get the information they need, they can instead shift their focus to their studies instead of being distracted by smaller issues.

1.2 Key Themes/Topics

This project undertakes the following key themes:

- Natural Language Processing (NLP)
  - As the backbone of this project, extensive research into this topic will be necessary to ensure users have a smooth experience.
- Large Language Models (LLMs)
  - The final product will be utilising a trained LLM for its backend calculations for what to display to the user. It is therefore essential that considerable study is taken into how LLMs function from both a developer and user perspective.
- Human-computer interactions
  - The interactions between users and chatbots will require meticulous study to ensure my project can perform as expected.
- Ethics in AI
  - Given that the chatbot will be engaging in two-way conversation with end-users, it will be important that an ethical standard is maintained.

Aims and Objectives

2.1 Project Aim

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.

2.2 Project Objectives

- Develop a chatbot capable of accurately answering user queries related to university buildings, policies, and clubs with a 95% accuracy rate.
- Conduct a thorough literature review on AI and Natural Language Processing.
- Document 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.

Project Planning

3.1 Initial Project Plan

1. Research
   (a) Conduct heavy research into machine learning and natural language processing to bolster my knowledge of the topics to assist in the development of the chatbot.
   (b) Identify similar projects that already exist to understand where challenges may arise in development and how to differentiate my work to make it stand out and provide unique value.
2. Data collection
   (a) To present users with information from the university, I must first collect this information for myself from sources such as the university website and the Student Union.
   (b) It will also be useful to survey various students for their perspectives on this project, and how useful it would be to them, as they are the target audience.
3. Development of a prototype
   (a) A prototype of the chatbot could be very useful in demonstrating the potential of this project, even before its completion.
   (b) This prototype could be distributed as part of a survey of users.
4. Evaluation
   (a) It will be important to identify how smooth the process of development was, and how the final product compared to its original expectations. Within the evaluation, any challenges faced along the way, as well as how they were overcome during development, should be documented in the interests of further research.
5. Finalisation and submission
   (a) At the very end of the process, when the chatbot and documentation are in acceptable states, final changes will be made based on supervisor and user feedback before making the final submission.

Below is a Gantt chart documenting an early concept of the development timeline throughout the year.

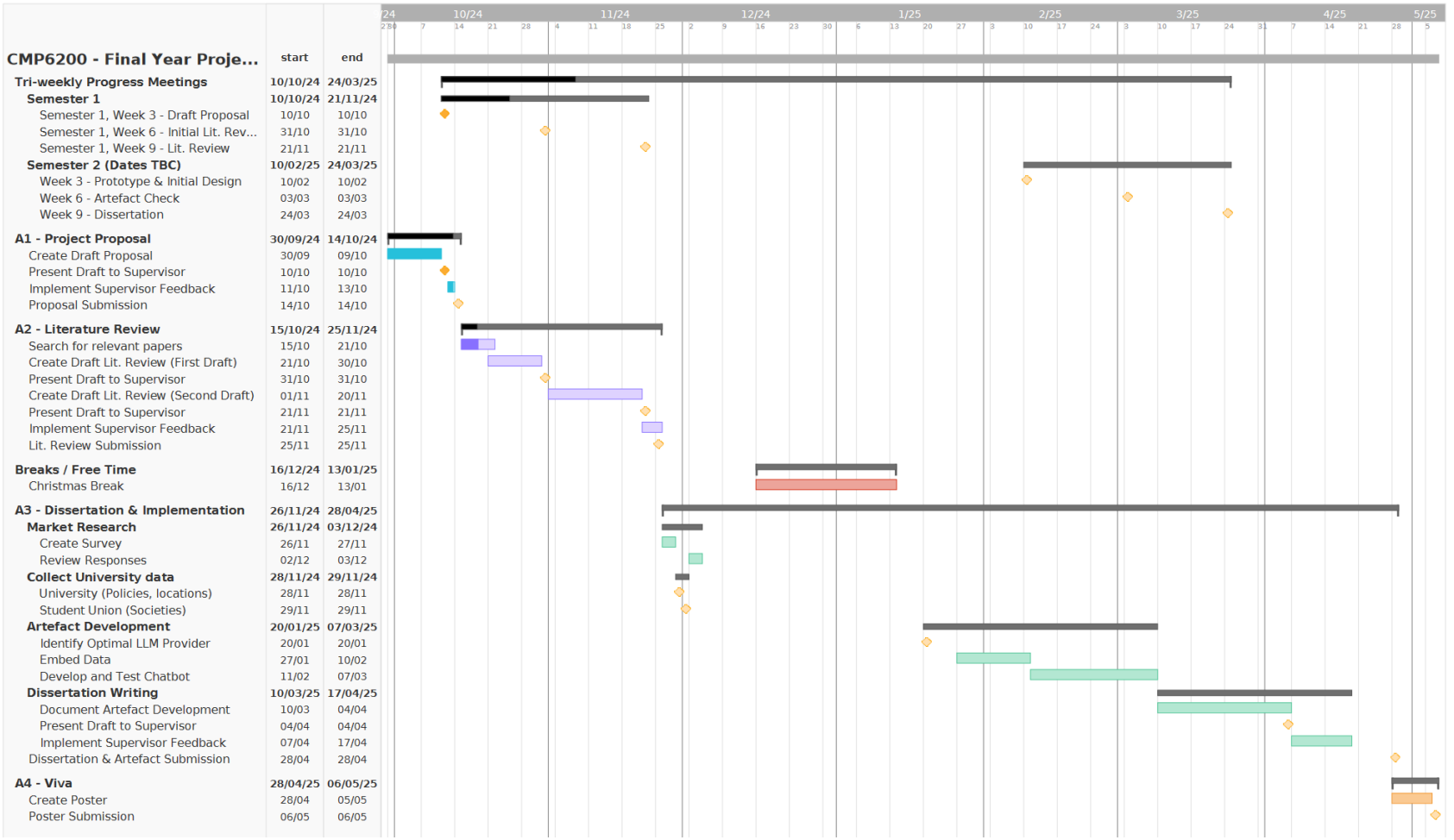

Figure 3.1: A conceptual Gantt chart of a development timeline.

3.2 Resources

- An integrated development environment (IDE) for Python
  - Visual Studio Code is a lightweight editor that supports most programming languages, including Python.
  - An alternative could be JetBrains' PyCharm Professional, which I can access at no charge due to being a student.
- Version control software
  - I will use Git for version control, and will store my versions on the cloud via Github.
  - In doing so, I can ensure that my work will be backed up in a secure location, and that the procedural changes I make throughout development can be tracked.
- Access to a pre-trained LLM
  - It is not likely to be feasible to train my own LLM in the time that I have for the development of this project, and especially unlikely that my own trained model could measure in comparison to those that already exist.
  - Examples include:
    * IBM Watson Discovery
    * OpenAI's GPT-4o mini
    * Google's Gemini 1.5 Flash
    * Microsoft's AI Azure Bot Service
- A platform for the chatbot.
  - Many messaging services allow developers to add bots, such as:
    * Facebook Messenger (used universally)
    * WhatsApp (used universally)
    * Telegram (less common, used universally)
    * Microsoft Teams (used for university by students)
    * Discord (used by students, not necessarily for university)

3.3 Risk Assessments

Risk	Likelihood	Severity	Overall	Mitigation
Time constraints could potentially cause rushed and poor-quality work if development is not to a consistent and regular schedule balanced with other university modules.	3	4	Medium-High	Ensure that work is completed at regular intervals, even if the amount at each interval is small. In doing so, it will be much easier to balance three simultaneous deadlines.
Personal circumstances could cause progress to slow or halt if something unexpected were to occur during the year.	3	4	Medium-High	Try to be ahead of deadlines where possible to ensure that there is free time that could be used in the event of a sudden break becoming necessary.
The devices used to write code and documentation could encounter hard drive corruption, potentially losing considerable amounts of work.	2	5	Medium	Ensure that all work is regularly backed up to the cloud and/or a secondary device. Github will be used as a version control system for the project to keep an audit trail of all changes.
Loss of internet connection would leave the bot inoperable until it is restored.	2	5	Medium	Where possible, use an Ethernet connection for higher network stability. A temporary alternative solution could be to use a mobile device as a hotspot.
Budget constraints could be an issue during development.	2	2	Low	Use open-source or lower cost resources during development, and create a budget to adhere to.

Project Review and Methodology

4.1 Critique of Past Similar Projects

While researching similar projects, I identified three useful final year projects by BCU students. The first was a 2019 project by Sanah Mehreen Hussain, which was the creation of a chatbot to assist students with their module content. As another chatbot project, the processes they undertook in their development will be heavily interlinked with my own, which made it imperative to review. Their report was highly detailed, providing a clear and strong explanation of each step of their development process. They had created many diagrams to clearly explain their processes, making it a powerful learning resource. Their chatbot was placed within Moodle itself, which is used by BCU students to access all of their work, which would massively improve its engagement. Sanah also conducted surveys for user feedback which proved useful in their evaluation.

Another useful project to review was Ali Akbar Rashid's 2022 chatbot for BCU IT support. This was another highly detailed exploration of the chatbot development process. Ali used the Agile project management methodology in the development of their project, citing its fast approach and lower cost in relation to other methodologies such as Waterfall, as well as frequent opportunities for feedback. Even still, Ali mentioned time limitations due to the balancing of his project with his other university modules, which suggests that I must take particular care to ensure I am setting and meeting frequent goals as mentioned in Section 3.3. Ali's evaluation and research was also heavily detailed, with a strictly defined methodology for both.

The third useful project was "KURA", a chatbot to assist users with information about the 2022 World Cup, developed in 2023 by Stanley Eweniwe Osuozah. While it is not as directly linked to a university chatbot like the others, it is functionally similar in that it is also retrieving information and displaying it to users, though simply of a different topic. This project was not documented quite as well as the other two, though it did still have some useful information to be learned. Their user survey was somewhat confusing to interpret, with questions with different implications such as "the KURA chatbot is welcoming" followed immediately by "the KURA chatbot is not clear about what it is meant for." This shift from a positive statement to a negative one within the survey could confuse participants and give potentially misleading results due to their misinterpretation of what they are being asked. The dataset they published to Kaggle was also very limited, and did not yield much useful information. From this project in particular, I have gathered that I must convey my intentions clearly and in significant detail to ensure that my work could be used by others in future to assist the development of their own projects.

4.2 Literature Search Methodology

There will be a considerable amount of research required to develop this project to a good standard; therefore, I will make use of online resources such as but not limited to the BCU library portal, IEEE Xplore, Elsevier and Google Scholar to find high-quality documents to base my project upon. To do so, I will need to use curated search terms to yield the specific information that I will need. Therefore, I will use the following terms:

- Artificial intelligence
- Natural Language Processing
- Chatbots
- Machine learning
- Deep learning
- Ethical AI
- Information retrieval
- User experience

My search will also be limited to papers from at least 2020, as AI and NLP are cutting-edge fields changing constantly, and information beyond that point is almost certain to no longer be relevant.

4.3 Initial Literature Search Results

In my initial search of these topics, I identified the following papers that I believe will be instrumental in the development of this project.

- I. Lauriola, A. Lavelli and F. Aiolli (2022). 'An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools'. In: *Neurocomputing* 470, pp. 443–456. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2021.05.103

  - This article from the Neurocomputing journal, a peer-reviewed journal covering AI, machine learning and neural networks with an impact factor of 6.0, gives a wide range of knowledge on NLP and the processes surrounding it such as one-hot encoding and vectorisation, which will be essential for me when supplying my chatbot with university related information. It shows detailed diagrams and explanations of the aforementioned processes as well as explaining the software that can be used, such as Python's SpaCy library. While my project won't be directly training a model of its own through one of these libraries, it is still essential to have the relevant background knowledge to create a product of equal or superior quality to others.

- E. Adamopoulou and L. Moussiades (2020). 'An Overview of Chatbot Technology'. In: *Artificial Intelligence Applications and Innovations*. Ed. by I. Maglogiannis, L. Iliadis and E. Pimenidis. Cham: Springer International Publishing, pp. 373–383. DOI: 10.1007/978-3-030-49186-4_31

  - This conference paper elaborates on what a chatbot precisely is as well as their applications. It also elaborates on their history and advancements made through time. More importantly, it discusses the architecture of a chatbot as well as key terms such as entities and contexts, and the design and development processes of chatbots, which will be absolutely integral to my project.

- S. Du and C. Xie (2021). 'Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities'. In: *Journal of Business Research* 129, pp. 961–974. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2020.08.024

  - This article from the Journal of Business Research, a journal with an impact rating of 10.5, describes the ethical implications of AI, which will be an important topic in relation to my project due to the fact that my chatbot will be interacting with students. Additionally, this article, due to being part of a business journal, refers to the widespread application of AI in large businesses such as Amazon and Netflix, and furthermore, the growth of digital personal assistants such as Amazon's Alexa and Apple's Siri.

- Y. Cheng and H. Jiang (2020). 'How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use'. In: *Journal of Broadcasting & Electronic Media* 64 (4), pp. 592–614. DOI: 10.1080/08838151.2020.1834296

  - This article from the Journal of Broadcasting & Electronic Media will be extremely useful to the development of this project due to the heavy links between them. It describes the gratifications users seek from their interactions with chatbots, which I must take into consideration during the development of my own to ensure it is what my users expect it to be, and provides a smooth, gratifying and simple user experience.

Bibliography

Giattino, C., E. Mathieu, V. Samborska and M. Roser (2023). *Artificial Intelligence*. URL: https://ourworldindata.org/artificial-intelligence?insight=ai-systems-perform-better-than-humans-in-language-and-image-recognition-in-some-tests#key-insights (visited on 08/10/2024).

HEPI (28th Mar. 2024). *New HEPI Report explores the impact of technology on universities and students*. URL: https://www.hepi.ac.uk/2024/03/28/technology-foundations-for-twenty-first-century-higher-education/ (visited on 08/10/2024).

Cherniak, K. (22nd Sept. 2024). *Chatbot Statistics: What Businesses Need to Know About Digital Assistants*. URL: https://masterofcode.com/blog/chatbot-statistics (visited on 09/10/2024).

Lauriola, I., A. Lavelli and F. Aiolli (2022). 'An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools'. In: *Neurocomputing* 470, pp. 443–456. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2021.05.103.

Adamopoulou, E. and L. Moussiades (2020). 'An Overview of Chatbot Technology'. In: *Artificial Intelligence Applications and Innovations*. Ed. by I. Maglogiannis, L. Iliadis and E. Pimenidis. Cham: Springer International Publishing, pp. 373–383. DOI: 10.1007/978-3-030-49186-4_31.

Du, S. and C. Xie (2021). 'Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities'. In: *Journal of Business Research* 129, pp. 961–974. ISSN: 0148-2963. DOI: https://doi.org/10.1016/j.jbusres.2020.08.024.

Cheng, Y. and H. Jiang (2020). 'How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use'. In: *Journal of Broadcasting & Electronic Media* 64 (4), pp. 592–614. DOI: 10.1080/08838151.2020.1834296.