

Project Plan

Joel Pointon

Student ID: 10286413

Course: MSc Data Science

June 29, 2023

1 Project Title

An automated FAQ tool for Introduction to Image Processing module.

2 Statement of the research problem

Inspired by ChatGPT (OpenAI, 2022), this project aims to create a “framework/model” that can automate the process of answering Frequently Asked Questions for the Introduction to Image Processing (COMP 2032) module. Generative chatbots such as ChatGPT can produce errors (Marcus, 2018; Bender et al., 2021; van Dis et al., 2023), particularly when seeking to summarise documents (Durmus et al., 2020). This means that students could be learning inaccurate or irrelevant information. Consequently, this paper aims to propose an efficient model for a domain-specific, knowledgeable Q&A chatbot where tutors can have high confidence in the accuracy of any answers. This can free up academics so that they can spend their time more efficiently, to “reinforce the learning of students”, improve their teaching methods (Pérez et al., 2020), and “focus on new experimental designs” (van Dis et al., 2023).

3 Related work

Transformers were introduced in Vaswani et al.’s ‘Attention Is All You Need’ paper (2017). The authors proposed a sequence-to-sequence model which uses 6 stacked encoders and decoders. The strength of Transformers lies in their ability to process all words in a sentence simultaneously, and their ability to retain context from much further back in the sequence (Vaswani et al., 2017). There are multiple variations of the transformer architecture. The first is autoencoding (bidirectional) models which use stacked encoders (e.g. BERT: Devlin et al. (2019)); the second is autoregressive (left-context) models which use stacked decoders (e.g. GPT: Radford et al. (2018)); the third is a combination of both, using both stacked encoders and decoders (e.g. BART: Lewis et al. (2019)), and these are known as sequence-to-sequence models.

There is little in the literature regarding attempting to produce (educational) chatbots by fine-tuning the above transformer models, however, Khan Academy is in the process of creating and trialling a knowledgeable tutor-bot using a fine-tuned version of GPT-4 (Khan, 2023). Unfortunately, GPT-3.5-turbo and GPT-4, which power ChatGPT (OpenAI, 2022) are not publicly available for fine-tuning. Prior to the use of Transformer models, an alternative approach was used by Chen et al. (2022) to create an extractive, domain-specific chatbot using a Deep Convolutional Neural Network (an LSTM-CNN model), but this architecture cannot generate answers in natural language anywhere near as effectively. Therefore, this project seeks to provide a framework for a chatbot which can generate answers in natural language with a high degree of accuracy and reliability.

4 Methodology

This project will use several different AI models (both extractive and generative) in order to provide a holistic and comprehensive assessment of which model/framework is the most effective and appropriate. The four chosen models are BERT, BART, and GPT. Additionally, two GPT models will be used - a pre-trained GPT-2 model, and a fine-tuned GPT-3.5-turbo model, so that different price GPT models can be compared. This will also mean that any issues regarding insufficient model training data will be limited to only one or two of the above options.

To train the models, a knowledge base will need to be created from the module teaching resources, and cleaned into an appropriate format (e.g. Q&A format) which can then be fed into the AI model. By using a pre-trained model (for BART and GPT), I can then fine-tune them to be knowledgeable about COMP 2032 using their neural network structure. This work will be resource-intensive, and issues could arise regarding sufficient computing power, so I will use a combination of the school's GPU cluster, Google Collab, and Kaggle to train the models. I will use Python to conduct this project, specifically using Google's HuggingFace library, along with PyTorch, TensorFlow, and OpenAI.

Model evaluation will use a combination of developer testing, accuracy on Q&A datasets such as SQuAD 1.0 and 2.0, cost-effectiveness, and ease of use. The model which has the best overall performance will be the basis for the chosen framework.

5 Programme of work

The required work for the MSc Project has been split into eight Work Packages (WPs):

- WP-1** *Topic Selection:* Conduct preliminary research to identify gaps in existing knowledge, to refine the research question and find areas of research suggested in the literature. Milestone 1 (M1) - Research question selection completed.
- WP-2** *Literature Review:* Review the existing literature to connect different technologies and establish the current SOTA models in chatbot technology. Critically analyse key papers and identify gaps in the literature. Milestone 2 (M2) - Literature review draft completed.
- WP-3** *Research Design:* Determine the models to be compared and contrasted. Outline the methodology I will use to construct a chatbot. Also establish success criteria against which each model can be compared. Milestone 3 (M3) - Methodology draft completed.
- WP-4** *Knowledge Framework:* Assemble a 'knowledge' base for COMP 2032 and develop framework which can easily be applied and scaled up to other modules. This will be used to train the four models. Milestone 4 (M4) - Knowledge base created.
- WP-5** *Model Training:* The main aspect of the development project is to use the knowledge framework to train each of the four models. I will use the school's GPU cluster, Google Collab, and Kaggle to train each of the models. Milestone 5 (M5) - Model training completed.
- WP-6** *Model Comparison:* Determine the best model by comparing and contrasting the four models. A combination of performance metrics and developer testing will be employed to establish the most effective model. Milestone 6 (M6) - Model comparison completed.
- WP-7** *Writing and Drafting:* Begin writing the main body of the dissertation. This includes the implementation and evaluation of the models, along with revising and editing previous sections (literature review/methodology). Milestone 7 (M7) - First dissertation draft completed.
- WP-8** *Finalise Writeup and Presentation:* Use the dissertation first draft to create a presentation to demonstrate the chosen model and training process. Use the presentation creation process to refine the dissertation. Milestone 8 (M8) - Final dissertation and presentation completed.

6 Time plan

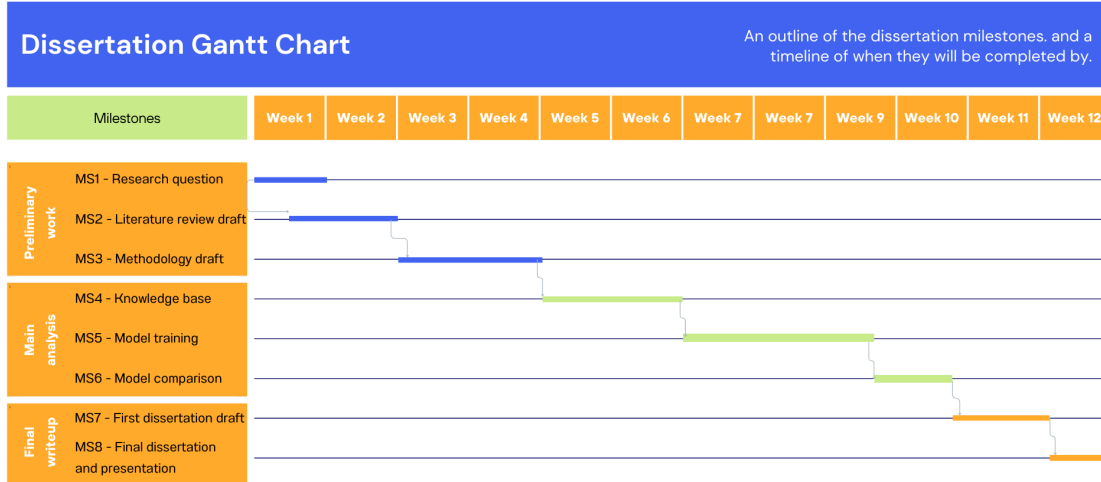


Figure 1: Completion Deadlines for Milestones

Word count out of 994/1000 words

References

- BENDER, E. M., T. GEBRU, A. McMILLAN-MAJOR, AND S. SHMITCHELL (2021): “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? parrot,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA: Association for Computing Machinery, FAccT ’21, 610–623, <https://doi.org/10.1145/3442188.3445922>.
- CHEN, P., J. ZHONG, AND Y. ZHU (2022): “Intelligent Question Answering System by Deep Convolutional Neural Network in Finance and Economics Teaching,” *Comput Intell Neurosci.* 2022 Jan 21;2022:5755327. doi: 10.1155/2022/5755327. eCollection 2022., [<https://doi.org/10.1155/2F2022%2F5755327>].
- DEVLIN, J., M.-W. CHANG, K. LEE, AND K. TOUTANOVA (2019): “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” <https://doi.org/10.48550/arXiv.1810.04805>.
- DURMUS, E., H. HE, AND M. DIAB (2020): “FEQA: A Question Answering Evaluation Framework for Faithfulness Assessment in Abstractive Summarization,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, <https://doi.org/10.18653/2Fv1%2F2020.acl-main.454>.
- KHAN, S. (2023): “Harnessing AI So That All Students Benefit: A Nonprofit Approach for Equal Access,” Khan Academy Blog [online] <https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/>, accessed: 4/05/2023.
- LEWIS, M., Y. LIU, N. GOYAL, M. GHAZVININEJAD, A. MOHAMED, O. LEVY, V. STOYANOV, AND L. ZETTMLOYER (2019): “BART: Denoising Sequence-to-Sequence Pre-training for Natural

- Language Generation, Translation, and Comprehension,” <https://doi.org/10.48550/arXiv.1910.13461>.
- MARCUS, G. (2018): “Deep Learning: A Critical Appraisal,” ArXiv preprint: <https://doi.org/10.48550/arXiv.1801.00631>.
- OPENAI (2022): “Introducing ChatGPT,” OpenAI [online] <https://openai.com/blog/chatgpt>, accessed: 13/06/2023.
- PÉREZ, J. Q., T. DARADOUMIS, AND J. M. M. PUIG (2020): “Rediscovering the use of chatbots in education: A systematic literature review,” *Computer Applications in Engineering Education*, 28, 1549 – 1565, <https://doi.org/10.1002/cae.22326>.
- RADFORD, A., K. NARASIMHAN, T. SALIMANS, I. SUTSKEVER, ET AL. (2018): “Improving language understanding by generative pre-training,” *OpenAI*.
- VAN DIS, E. A. M., J. BOLLEN, W. ZUIDEMA, R. VAN ROOIJ, AND C. L. BOCKTING (2023): “ChatGPT: five priorities for research,” *Nature*, 614(7947), pp.224-226., <https://doi.org/10.1038/d41586-023-00288-7>.
- VASWANI, A., N. SHAZEER, N. PARMAR, J. USZKOREIT, L. JONES, A. N. GOMEZ, L. KAISER, AND I. POLOSUKHIN (2017): “Attention Is All You Need,” *CoRR*, <https://doi.org/10.48550/arXiv.1706.03762>.
-

7 Assessment criteria

- Is there a clear statement of what the project is about?
- Is there an explanation of why this problem needs to be addressed / solved?
- Is the novelty of the proposed research substantiated with reference to related work?
- Is there a clear statement of how the problem is going to be addressed / solved?
- Are the proposed methods for addressing / solving the problem appropriate?
- Is there a clear statement of how the results are going to be evaluated?
- Is the proposed evaluation appropriate?
- Is there a breakdown of the research into work packages?
- Is this breakdown appropriate?
- Is there a time plan?
- Is the time plan realistic?

Affirmative answers to each of these questions will result in a first class mark. Markers should briefly describe how the project plan diverges from these desired outcomes to justify their mark.