Focal Plan Document

Students:
Razik Grewal, Peter Meyers, Mitali Potnis

Mentors:
Annette Han, John Stamper

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Version 1.3

1 Introduction

As education continually trends toward online learning, giving students frequent assessments is crucial to their development.[1] These assessments often come at a high price to educators however, as they are forced to create a large bank of questions to provide realistic learning opportunities for students of varied skill levels. [2] In addition to the high cost to produce these assessments, educators likewise must spend a significant amount of time providing students with quality feedback. This feedback helps to ensure the assessments students are completing enrich their learning experience, but is time-consuming to produce. Focal intends to reduce the burden of these assessments on educators, thereby saving both their efforts and time. By automating the assessment creation and evaluation pipeline, educators' time will be freed to focus on their various other responsibilities.

In this plan document, we will establish the general timeline and primary project milestones for the Focal project. Our first project milestone, as is always the case, is the start of the project, which occurred in February. After the start of the project, we have multiple milestones that will occur in parallel with each other. These consist of fully documenting the existing Focal baseline as well as receiving and processing new course data for use in our updated pipeline. After these parallel milestones are complete, we will transition into developing our infrastructure plan, which will be complete by the end of the spring semester. This will include initial testing of various models for QG and QA. By establishing our infrastructure plan fully by the end of the spring semester, we will be ready to tranisition into pipeline testing for the bulk of the fall semester. Our primary milestone in the fall semester will be the completion of the pipeline testing. This will include the tasks of testing each individual portion of the pipeline infrastructure laid our in our infrastructure plan. When this testing is complete, we will move on to our final milestone, project completion. This will consist of demonstrating the end-to-end effectiveness of Focal, documentation of the pipeline, and presenting our results.

2 Tasks

In order to successfully reach our project goals for Focal, we have established the following list of tasks and their associated timeline:

2.1 Collect new course data

- **Description**: In order to make our Focal pipeline a more domain agnostic solution, we will be incorporating data from two additional courses. These consist of an updated graduate level data science course and an undergraduate level chemistry course. To collect this data, we have engaged with the content owners. The expected format of the course data is a repository of XML files.
- **Start Date**: 14 March 2023
- **End Date**: 21 March 2023
- **Resources required**: Course files in XML format
- **Personnel**: Peter Meyers, John Stamper, David Yaron

2.2 Implement existing Focal pipeline

- **Description**: In order to enable future improvements of the Focal pipeline, the members of the team will first implement the existing Focal pipeline.[2] This will ensure team members understand the existing solution, which will provide a strong position from which they can begin testing alternative methods to improve Focal.
- **Start Date**: 14 March 2023
- **End Date**: 21 March 2023
- **Resources required**: Previous Focal repository, previous Focal course dataset
- **Personnel**: Razik Grewal, Peter Meyers, Mitali Potnis

2.3 Document existing Focal pipeline

- **Description**: During this phase of the project, our team will fully document the existing Focal baseline. [2] This pipeline, while operational, is not fully documented, which makes it much more difficult for new researchers to understand and improve on its functionality. We will, without modification to the functionality itself, fully document the existing pipeline to better enable future development.
- **Start Date**: 14 March 2023
- **End Date**: 2 April 2023
- **Resources required**: Previous Focal repository, previous Focal course dataset
- **Personnel**: Peter Meyers, Razik Grewal

2.4 Process new course data

- **Description**: Upon receipt of new course data, we will process it for use in the Focal pipeline. We will ensure it is formatted correctly to be used by our question generation and answering models, as well as remove duplicate entries, etc.
- **Start Date**: 21 March 2023
- **End Date**: 14 April 2023
- **Resources/Personnel**: Razik Grewal, Mitali Potnis

2.5 Test alternative question generation methods

- **Description**: In order to improve on the previous Focal implementation, a key area for development is the improvement of the question generation model. Our model should be able to provide pedagogically sound questions that vary in question type and difficulty. To test this model, we will create questions that will evaluated by domain experts to determine soundness and utility.

- **Start Date**: 14 April 2023
- **End Date**: 6 May 2023
- **Resources required**: Course data, MOOCCubeX [3], SQuAD [4], Google's T5 [5]
- **Personnel**: Razik Grewal, Peter Meyers, Mitali Potnis, John Stamper, David Yaron

2.6 Test question evaluation metrics

- **Description**: The previous Focal implementation used infomation score to grade question quality. [2] This score measured the number of extracted key concepts that the question contained. While somewhat useful, this score was a rather naive method. We will test and evaluate new methods for scoring question quality. In order to test our grading criteria, we will compare the labels produced by our metric with expertly annotated questions.
- **Start Date**: 31 August 2023
- **End Date**: 18 September 2023
- **Resources required**: Generated questions, expert labels, GPT-3
- **Personnel**: Razik Grewal, Peter Meyers, Mitali Potnis, John Stamper, David Yaron

2.7 Test alternative question answering methods

- **Description**: In this step of the research, we will test various models' effectiveness at generating cogent responses to the questions created by our question generation model. Cogent answers to generated questions are vital to the pipeline, as they will form the baseline by which our model would evaluate student responses. In order to evaluate the generated answers, we will utilize GPT-3 in conjunction with expert evaluation of the responses generated for specific questions.
- **Start Date**: 18 September 2023
- **End Date**: 9 October 2023
- **Resources required**: Generated questions, course content, GPT-3, possibile alternative question answering model, expert evaluations
- **Personnel**: Razik Grewal, Peter Meyers, Mitali Potnis, John Stamper, David Yaron

2.8 Test question evaluation methods

- **Description**: In this task, we will evaluate methods for automatically evaluate student responses. Our primary method for evaluating these responses will be comparing student answers with generated answers for lexical similarity, likely using GPT-3.
- **Start Date**: 9 October 2023
- **End Date**: 6 November 2023
- **Resources required**: Generated questions, generated answers, example student responses, GPT-3
- **Personnel**: Peter Meyers, Razik Grewal, Mitali Potnis

2.9 Test end-to-end pipeline effectiveness

- **Description**: In this task, we will test the effectiveness of Focal as an end-to-end solution. To this point, much of the testing will be completed in phases, improving one step of the pipeline at a time. Here, we will ensure these steps are fully integrated and operational as a single pipeline.
- **Start Date**: 6 November 2023
- **End Date**: 27 November 2023

- **Resources required**: Course data, pipeline repository, GPT-3, MOOCCubeX [3], SQuAD [4], Google's T5 [5]
- **Personnel**: Peter Meyers, Razik Grewal, Mitali Potnis

2.10 Complete documentation of updated pipeline

- **Description**: In order to ensure our research can be built upon for future iterations of the Focal pipeline, it is vital that we completely and accurately document the system. This documentation will be ongoing throughout the project, but in the final phase we will reinspect all previous documentation for accuracy and completeness.
- **Start Date**: 1 December 2023
- **End Date**: 11 December 2023
- **Resources required**: Pipeline repository
- **Personnel**: Mitali Potnis

2.11 Create project final report

- **Description**: We will create a final report for the Focal project in which we will detail work completed, results, conclusions, and avenues for future research.
- **Start Date**: 1 December 2023
- **End Date**: 11 December 2023
- **Resources required**: Focal repository
- **Personnel**: Razik Grewal, Peter Meyers

3 Milestone Plan

By having milestones and their sub-tasks with associated dates, we will be able to continually assess our current progress in terms of the entire project length. Our intent is to add continual checks of task and milestone progress to our weekly advisor meetings to assure we meet our research goals. To that end, we have created the following milestone plan, which incorporates both semesters of the project:

Task Title	Dates	Description
Project Start	10 February 2023	Conducted initial meeting with capstone research advisor. Set initial goals for research project and began background research.
New Course Data Received	21 March 2023	External dependency milestone. For this milestone, we are waiting to receive data from two new courses for use in the Focal pipeline. These courses consist of a graduate-level data science course and an undergraduate level chemistry course.
Previous Focal Baseline Documented	2 March - 2 April 2023	In order to lay the groundwork for future research, we will fully document the previous focal implementation. [2] Upon completion of this milestone, our research will fully transition from re-implementing previous work to building on previous work.
New Data Processing Complete	21 March - 14 April 2023	At the completion of this milestone, the data from the two additional courses received earlier in the project will be ready for use in the Focal pipeline. This milestone will mark a transition to utilizing updated and varied course text in our research.
Infrastructure Plan Complete	14 April - 6 May 2023	At the completion of this milestone, we will have fully determined the infrastructure plan for our updated Focal pipeline. This will include all updated models we plan to use for each phase of the pipeline. At the completion of this milestone, we will transition fully from pipeline planning and establishment to testing of individual pipeline functions, which will occur in the Fall semester.
Pipeline Testing Complete	31 August - 6 November 2023	This milestone will mark the majority of the work completed in the Fall semester of this project. It will include the tasks of testing the various portions of the overall Focal pipeline, including: question generation, question answering, and question evaluation. At the completion of this milestone, we will transition from testing individual portions of the pipeline to testing the pipeline as a complete unit and preparing for future research.
Final Report/ Presentation Complete	6 November - 11 December 2023	During this milestone, we will test the Focal pipeline as an end-to-end solution as well as prepare a final report. At the completion of this milestone, our work on the Focal pipeline will be fully documented and prepared for researchers to continue in future iterations.

Table 1. Milestone plan with descriptions

In the above table, we have laid out the primary milestones that make up the Focal project for both the first and second semester of this year. Some of the milestones are overlapping, so we have additionally created the following diagram for better visualization of the milestone plan:

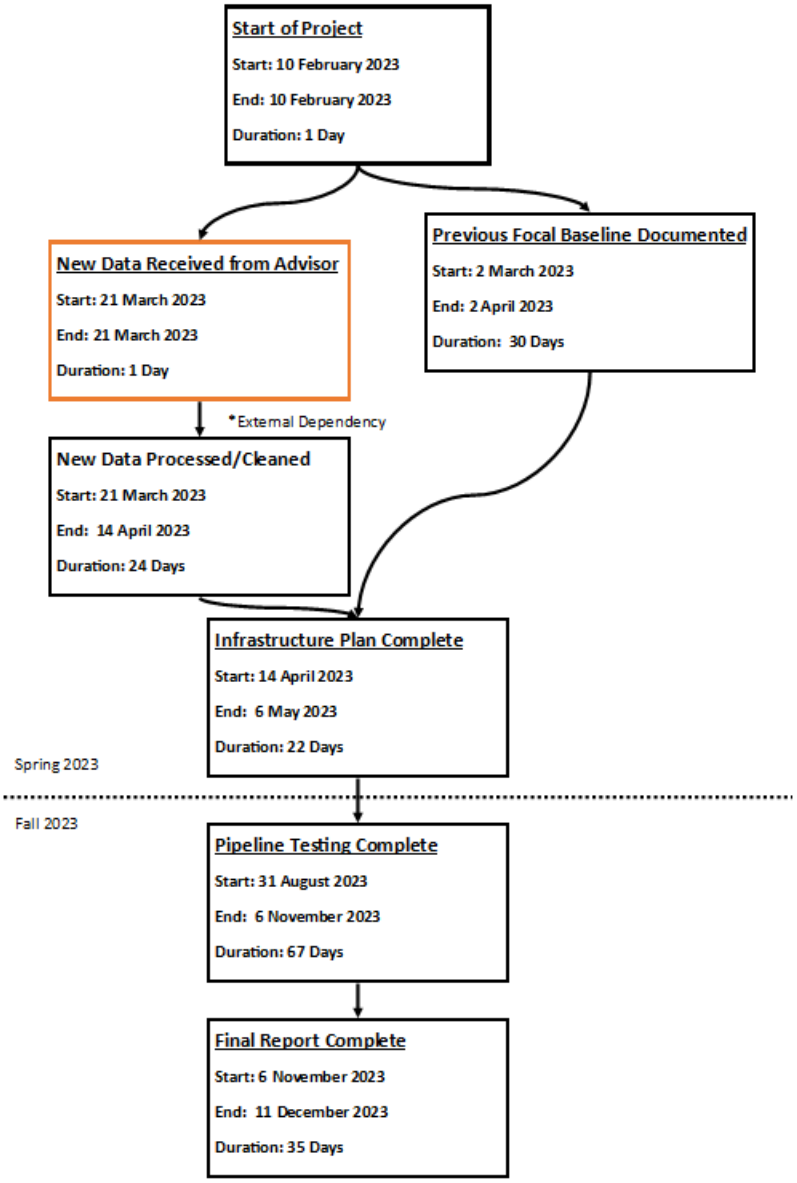

Figure 1. Graphical depiction of Milestone plan. Of note, new course data receipt and processing will occur in parallel with documentation of former Focal baseline as these tasks are independent of each other.

4 Terminology, Definitions, Acronyms, and Abbreviations

- **GPT-3**: Generative Pre-Trained Transformer 3, this product is a large language model that has proven valuable at tasks vital to this project, such as question answering and text generation.
- **MOOCCubeX**: A large dataset of educational materials, in addition to a pipeline that is adept at automatically extracting key educational concepts from text.[3]

- **NLP**: Natural language processing, this is the branch of machine learning that focuses on the processing and generation of natural language.
- **QA**: Question Answering, this is the process of using NLP models to automatically generate answers to given questions.
- **QG**: Question Generation, this is the process of using NLP models to automatically create questions based on text suppied to the model.
- **SQuAD**: Stanford Question Answering Dataset, this is a large dataset made of questions on Wikipedia articles and their corresponding answers.[4]
- **T5**: Test-to-Text Transfer Transformer created by Google, considered a state-of-the-art model in many NLP tasks.[5]

5 References

[1] Yong Zhao, Jing Lei, Bo Yan Chun Lai, and Hueyshan Sophia Tan. What makes the difference? a practical analysis of research on the effectiveness of distance education. *Teachers College Record*, 107(8):1836–1884, 2005.

[2] Huy A. Nguyen, Shravya Bhat, Steven Moore, Norman Bier, and John Stamper. Towards generalized methods for automatic question generation in educational domains. In *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption*, pages 272–284, Cham, 2022.

[3] Jifan Yu, Yuquan Wang, Qingyang Zhong, Gan Luo, Yiming Mao, Kai Sun, Wenzheng Feng, Wei-Hao Xu, Shulin Cao, Kaisheng Zeng, Zijun Yao, Lei Hou, Yankai Lin, Peng Li, Jie Zhou, Bingsheng Xu, Juan-Zi Li, Jie Tang, and Maosong Sun. Mooccubex: A large knowledge-centered repository for adaptive learning in moocs. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021.

[4] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. In *Conference on Empirical Methods in Natural Language Processing*, 2016.

[5] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer. In *North American Chapter of the Association for Computational Linguistics*, 2020.

6 Reflection

The process of creating this plan document will no doubt pay dividends as we continue on throughout the two semesters of this project. By putting more specific date ranges to each task required to meet our project goals, we have given ourselves a basis by which we can judge our progress throughout the year-long project. We went back and forth a little on the number of milestones that we wanted to include in the milestone plan but decided on the ones we have as we felt those were where we really shift our focus in the project. There were some large sub-tasks that arguably could be considered milestones, but we did not feel that they met they changed the focus of our research, which was our interpretation of the meaning of a milestone.

7 Changes to Previous Deliverables

There are no changes to previous deliverables to report at this time.

8 Change Log

Table 2. Change Log and Versioning

Version	Date	Changelog
V1.0	2023.03.06	Added Introduction, Tasks, Milestone Plan
V1.1	2023.03.08	Added References,Terminology, Updated Introduction
V1.2	2023.03.12	Updated Milestone Plan to Remove Milestones Deemed Tasks Instead
V1.3	2023.03.14	Added Graphical Milestone Plan Depiction, Added Reflection