Project ID	Project Title	Project Description (<1 sentence)	Project Description (the "what" and the "why")	Additional items to specify on AIP form	Associated TA
1	CBTBot	Text-based chatbot for delivering CBT	At Dr. Rose's Automated Medicine Lab, we envision a world in which talk therapy can be given by a computer, exactly where and when it is needed. Recent advances in Natural Language Processing (NLP) and Deep Learning have provided new ways to bring this vision to a reality. Thus far, our efforts have been concentrated on applying a single known and structured therapeutic technique (Motivational Interviewing) to a specific problem in addiction (smoking cessation), through a text-based chatbot called MIBot. This is a collaboration with the Nicotine Dependence Clinic at the Centre for Addiction and Mental Health and with the iSchool at the University of Toronto. While previous iterations of the bot used combinations of scripted and generated responses, the latest version of MIBot uses 100% generated responses. This project is an invitation for groups to investigate another known and structured therapeutic technique i.e. Cognitive Behavioural Therapy (CBT), and pursue the creation of a system capable of delivering CBT. Students must define the scope of their project by focusing on a specific mental health issue that is amenable to CBT.	Identify a distinct mental health issue that can be effectively addressed using Cognitive Behavioral Therapy (CBT).	Soliman Ali
2	LangTrain	LLM-powered Cultural Context insights in Language Learning	Learning languages involves much more than learning grammar and accumulating vocabulary, it requires cultural knowledge and awareness to communicate naturally. Moreover, many phrases cannot be accurately captured through direct translation. The goal of this project is to explore the potential for LLMs to assist language learners in understanding cultural context behind idiomatic expressions, slang, and culturally rooted phrases and language structures. For this project you will design an interactive system such as the following: Given some input text, the system first identifies any culturally nuanced language structures. After providing a literal translation, the system offers detailed cultural explanations, including the phrase's origin and its use in various social settings. Then, to reinforce the user's understanding, the system engages the user in quiz-like dialogue that prompts them to use the new language structures, and offers feedback on whether they used them correctly.	Specify the language(s) that you will focus on, the kind of text you aim to work with, and a description of your system.	Soliman Ali
3	Qlaude	Instructor Assistant for Quercus/Piazza	Canvas is a widely used Learning Management System (LMS) for managing class-related tasks, known as Quercus at UofT. The general goal of this project is to explore the capabilities of Large Language Models (LLMs) to write Python scripts that interact with the Canvas API. These scripts will automate tasks such as generating progress reports, sending weekly course updates, and detecting anomalies in student submissions. More specifically, we are interested in the construction of agentic frameworks and the use of function calling to streamline course management beyond basic script automation, as well as integration with Piazza.	This is a very broad project prompt and requires more narrow scoping. Identify the tasks which you aim to automate, as well as how you plan to integrate it into existing workflow pipelines (e.g. browser extension, chatbot overlay).	Zafar Mahmood
4	CodeGPT	Improve programing capabilities of LLMs	Even state-of-the-art LLMs suffer from problematic codes they generate, either becuase of bugs in the codes or misunderstanding the request. In this project we aim to improve the quality of code generation done by LLMs such as GPT4.		Mohammadreza
5	DouGPT	AI language tutor	Isn't language the best thing that large LANGUAGE models know? One can use LLMs to be an AI tutor for people who want to learn new languages. LLMs can help us start with easy language learning tasks all the way up to Shakespeare level language!		Mohammadreza
6	ConsoleFixer	Fix your mistyped console command	As software engineers, we rely on console commands daily to enhance our productivity. However, it can be frustrating when we forget the exact syntax or specific flags, and searching through manuals takes up valuable time. ConsoleFixer steps in by analyzing mistyped command-line inputs and terminal error messages, then providing the corrected version. You can experiment with various models, including the closed-sourced GPT series, open-source Llama3 models, or code-focused LLMs like CodeLlama and StarCoder.	To narrow down the scope, you can choose only one type of command for your interests, such as the shell commands, git commands, or the K8S Kubectl commands, etc., This tool (https://github.com/nvbn/thefuck) achieves the same purpose based on hard-coded pattern matching. You can play with it, or use it as a baseline for comparing with our LLM-based approach.	Weizhou Wang

7	SQLGPT	Natural Language Interface for Database Queries	Implement an LLM-based system that translates natural language queries into SQL queries for database access. This could simplify data retrieval for non-technical users and improve usability in applications that require database interaction.		Weizhou Wang
8		Colour analysis GPT plugin			Zafar Mahmood
9	SWE-Agent	LLM-based software engineer agent that debugs a code base	Leveraging LLM's ability to work as agents, build a solution for the most common software engineering task -- debugging. Specifically, given a code base on github, an issue, and a set of test cases, generate patches to be applied to the code base to address the issue and pass all the test cases. (Maybe start small -- try fixing individual files first)	Dataset: SWE-bench; Need very good models	Jiading Zhu
10	Reserve-Agent	LLM-based assistant agent to make reservations for users	Implement an LLM-based agent/agentic workflow/system to make reservations for users in a conversational setting. The agent needs to correctly identify user's intent through classification and potential disambiguation steps, then make the correct reservation through tool-use (API calls). May interact with toy APIs.	May need to use an existing agentic workflow framework (e.g. LangGraph, CrewAI); Potential Dataset: https://github.com/google-research-datasets/dstc8-schema-guided-dialogue; Need very good models	Jiading Zhu
11	Game-Agent?	LLM-based game player	Implement an LLM-based agent to play a game of your choice (e.g. chess). You should model the game (ideally simple enough) to be played through a text interface. Implement strategies to play the game somewhat well. Compare its performance to other known strategies/agents.		Jiading Zhu
				There are many active fields in NLP, each with a lot of hard but exciting tasks. If you really need an idea or some inspirations, take a glance here: https://paperswithcode.com/area/natural-language-processing (NOTE: definitely start small and keep in mind the amount of time you have in the course).	Jiading Zhu