
ECE 1786

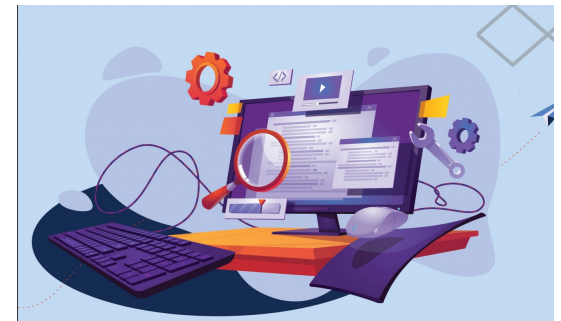
Creative Applications of Natural Language Processing

Lecture 4 Part 2: Project Structure



Projects are Engineering

- I believe that the way to learn is to **do!**
- Assignments in course provide basics
- In project you navigate the use of NLP-based neural networks and LLMs
- Includes oral/written communication
 - What, why, how, and results
- Also: Gives you something to talk about in job interview
 - and link to in CV or LinkedIn, e.g.
<https://www.eecg.utoronto.ca/~jayar/ece1786.2022/emojimotion.html>



Project Rules

- Done in Groups of 2 [decision made; ~ 39 groups]
 - **Side effect:** need to add in an extra night for proposal and final presentations
- Topic of your own choosing
 - must relate to NLP & the material covered in this course
 - must be approved by instructor
- Can be *either* be application of NLP *or* research on NLP
 - If research, make sure you have good discussion with me
- Projects in the class should be different from each other
 - If too similar, won't pass approval-in-principle
- Must collect and/or label some of your own data
 - Because data is a core part of the field, perhaps hardest
 - Must be careful not to do too much



Project Rules, cont'd

- You should use good software development practices
 - Modular code; good names; comments
 - Source code control – aka git
 - Will post a lecture on Source Code control using Git

- We will create a Github repository for each team
 - you must use to store, revise and submit your project
 - Required to submit for progress report and final report, but should use the whole time



Project Timeline and Deliverables


Date	Item
03-Oct	Project Discussion in Class (Today)
10-Oct	Team Forming Deadline – fill out form
26-Oct	Approval-in-Principle/Uniqueness of Project Topic latest
30-Oct	Project Proposal Document Due
30-Oct	Project Proposal Slides Due
31-Oct/Nov 1	In-Class Proposal Present + Extra Class in Evening + Next Eve
20-Nov	Progress Report Due
04-Dec	Final Presentation Slides Due
05-Dec/Dec 6	Final Presentations + Extra Class in Evening + Next Eve
12-Dec	Final Report Due




Step 1: Form Team

- Find someone compatible with you
 - on kind of topics interested in
 - on working and communication style
- Should do soon, latest possible deadline October 10
- See 'find partner' post, now on Piazza – pinned at top:

add new post:

 I'm **one student** looking for more people to work with.

 I'm **from a group** looking for more students.

*Name *Email

*About Me

(Things you could include: your location, grad/undergrad, when you're available... help people get to know you!)

Submit Team Info on Form

- This a formal commitment
 - All team members must “sign” (type name) on form.
 - Fill out here: <https://forms.office.com/r/jx0KmWDTed>
- Form requires:
 - Names of all students and student numbers
 - UofT email address of all members
 - Department of each group member
 - Degree being pursued by each group member (M.Eng, M.A.Sc., Ph.D., MSaC etc.)
 - If you are a part-time or full-time student
- I need this information to organize the tracking of teams and projects



Where to find Topics

1. From yourselves! Something you are interested in
 2. Suggestions from myself & TAs
 3. Look at projects on the internet
 - e.g. Stanford CS 224n Natural Language Processing with Deep Learning
 - <https://web.stanford.edu/class/cs224n/project.html>
- UofT rules on plagiarism apply
- You can be inspired by ideas on the internet, but you can't use their writing & code, unless you're building something new on top
 - Don't collaborate with another team on topic-finding, as each project must be unique



Scope – How Big Should the Project Be?

- A very difficult question to answer
 - Experience helps, but how to get it?
 - Break idea into pieces, estimate time to do
 - Pay attention to your estimates as you go
- Suggest creating layers of goals
 - Make sure some are achievable
- How much time do you have on project?
 - Two months, minus your other courses & Assignments in this course
- Myself & TAs will be available to discuss



Help from Myself and TAs

- With Proposal & Scope
 - Send me email with questions

 - Weekly Office Hour
 - Fridays 1-2pm
 - Engineering Annex, Room 319
- <https://map.utoronto.ca/?id=1809#!m/494468>



GPT-4 Has Really Changed What is Easy

- Things that used to be quite hard to do, might become very easy to do with GPT-4

Example:

- Previously: classifier to detect anxiety in speech
 - Collect data with anxious and non-anxious people speaking
 - Find a ground-truth label for to know which class people are in
 - Either a diagnosis or a standard scale
 - Train model with the given labels
 - A Ph.D. Thesis
- Now: describe what anxiety is to GPT-4, and ask it to classify the speech as exhibiting it or not



What does this mean for you?

- The task of using the model this way means you might not need to train it.
- However, you will still need to collect related data and labels.
- If you get the data and labels from a dataset, then there will not be too much work at all



So We Will Think of Two Classes of Project

■ Class 1: Old Style

- You will train a network (or fine-tune a pre-trained network) to do some kind of application
- This will require some amount of data collection and labelling

■ Class 2: GPT-4 Style

- You make full use of the ability of GPT-3.5/4 ‘to do what is asked’
- (might not fully understand this yet)
- If so, will still likely need to make a test set with labels
- The work is much easier, so am expecting something more complex to compensate



GPT-4 Style: Think Differently

- This is an amazing opportunity
- What was once hard, is now maybe much easier
- So we can try to do more difficult things
- For example, Action GPT: <https://actiongpt.github.io>
 - Language that guides motion



Action GPT



Action-GPT



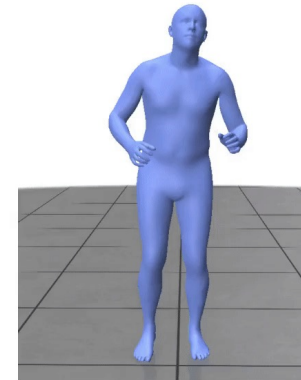
Smoking Cigarette



Stomach Ache



Play a mock violin



Walk with quietly and cunning



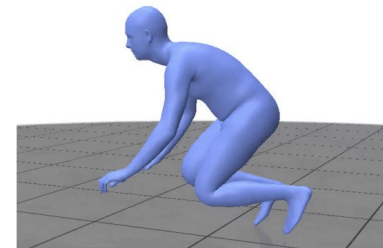
Defensive Pose



Referee showing card



Playing DJ



Act like a dog



Auto GPT

- Re-use GPT-4 many times push more complex task forward
- <https://autogpt.net/auto-gpt-vs-chatgpt-how-do-they-differ-and-everything-you-need-to-know/>



The Bottom Line

- If you don't train a model yourself, then you'll need to become more inventive in the use of a large language model
- One research-like example:
 - <https://arxiv.org/abs/2309.03409>
 - Used an LLM in the inner loop of a optimization system to optimize Prompts for a language model



Approval In-Principle/Uniqueness Approval

- Get together with partner to discuss topic ideas.
- Once you've settled on one, write up as follows:



Request for Approval-in-Principle Form

Find the form here: <https://forms.office.com/r/C9XHivcMUS>

- 1. What & Why:** 2-3 sentences that describe what the project is and how it is motivated. (Not **How**)
- 2. Data Source:** Your initial thoughts on where you will find relevant data, and what role you plan to take in the collection/labelling
- 3. Name:** Give your Project a Name
 - name should convey the essence of project; used for tracking
 - Creates your group identity! Logos also welcomed!



How to Describe Your Topic?

- Key is to say **what & why**

- engineers tend to think about **how** too soon, be warned
- You will need to think about how to make the **what feasible**, but not in first description for someone else to understand

- Should be the completion of this sentence:

“The goal of our project is to ...”



Picking a Name

- Is fun; it becomes the group's identity
- Should be a word or words that represents the goal/idea/motivation
- Examples from Last Year...



Project Names from ECE 1786 2022

Eng2Py

Newsify

Argument Gate

Fairytale

Imagine Captain

Fake News Detection

Destructive Language

Rater

Sensify

A.I. Meet Yu-Gi-Oh!

MI-Con

Antiqu-ator

Noffence

Sum(Text)

Coding Challenge
Generator

Artistic GENREator

The Survey Insider

Grammar Error
Correction

Song Genre Classifier

Remy

Caption Me

EZPaperSearch

SafeChat

PyOverflow

IELTS composition score
predictor

DeCo

Hive Mind Investor

Emojimotion

Aladdin

Recommender

TrainAssist

SoulsGen

GOTalk



Grading

Item	Portion of Course Grade
Proposal (Presentation/Document)	10%
Interim Report	10%
Final Presentation	10%
Peer Reviews	5%
Final Report/Software	25%
Total	60%



Project Proposal Document



Proposal Document

Document **must** have the Following sections:

1. Introduction

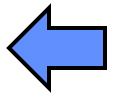
- What and why (i.e. motivation)

2. Background

- Describe 2-3 related papers you've found

3. Source of Data and Processing

- Where will you get the data for part of project?
- Requirement: some collection/labeling the data
 - But can't take up a big chunk of the project either



Proposal Document, cont'd

4. Architecture of the model/Structure of System

Case 1, Old style:

- Rough guesses of type and structure of model
- Describe other parts of software that are involved if any

Case 2: GPT-4 style

- Describe the interactions of the LLM you're going to make

5. Comparison

Case 1, Old style:

- Describe a simple baseline model that you'll compare against
- Simple model or hand-coded heuristic

Case 2: GPT-4 style

- You'll need to work on a metric that tells you if you're succeeding
- Often evaluated by hand; perhaps also by GPT-4 classifier



Proposal Document, cont'd

6. Plan

- Discuss how you're going to work together
 - Especially important if you don't know each other well
- List of sub-tasks
- Your guess as to how much time each task will take
- Use to create estimate of end-to-end time

7. Risks

- Predict what might go wrong & how you'd recover

- Document also graded on structure, grammar and mechanics



Proposal Document, cont'd

- Hard Limit of 1200 words total
 - Doesn't count pictures or references
 - 1% penalty for every word in excess of 1200
 - Put word count and compute penalty on front cover of proposal
 - 5% penalty if this is missing
 - These words (the count & penalty) not included in count

- Due Monday October 30 at 9pm.

- Upload under Assignment – Project Proposal Document
 - Just one per group;
 - Quercus will know your group, it will be the name you selected



Proposal Presentations

November 1, 2022



Proposal Presentation

Similar structure **but not same** as Document:

1. Introduction and Illustration
2. Data Collection and Processing
3. Architecture/Structure and Comparison
4. Risks
5. What You'll Have completed by November 21
 - At progress report time
 - Giving you a target to shoot at that **is not** the end



Proposal Presentation

- **4 minutes maximum to present**
 - **Timer** will be set & presentation ended at 4 mins.
 - 8 Slides **maximum** (including title slide)
 - Font size **minimum** 20
- This is difficult: must choose essential messages
- Urge you to practice the talk 2-3 times
 - Make sure you make sense to yourself and team
 - All team members must speak, roughly equally



Slides Due

- Slides due **Monday October 30 at 9pm**
 - Uploaded to Quercus 'Assignment' Proposal Presentation
 - Must be either **powerpoint (pptx or ppt)** or **PDF**
 - No google doc web links, must convert to pdf/ppt



Proposal Presentation

- I will put up the schedule of which team is presenting in which time slot
- Three possible times to present
 - Oct 31: During Regular Class: 10am-12 noon
 - Oct 31: Extra Evening Class: 6:00pm-9:30pm also GB 221.
 - Nov 1: Extra Evening Class: 6:00pm-9:30pm room TBD



Peer Review of Proposals/Presentations

- You will be asked to review another group's document and presentation
- You'll be scheduled to do that in a different time period, one of the three
- If you have a hard conflict with one of the periods, you must email me what it is and why.



Questions?

