# ECE 1786:
# Creative Applications of Natural Language Processing

Fall 2023

Instructor: Jonathan Rose

Department of Electrical & Computer Engineering

# Land Acknowledgement

Toronto is a city of immigrants.  My parents arrived here in 1952, and it is likely that you or your parents are new to this country. This is not a new feature of Toronto. For over 15,000 years Toronto has been a gathering site for humans including the Huron-Wendat and Petun First Nations, the Seneca, and most recently, the Mississaugas of the Credit River.

Today, Toronto is still a meeting place for Indigenous people from across Turtle Island (North America), and immigrants, both new and old, from across the world. I am grateful to have the opportunity to work in this community, and on this territory with many kinds of people.

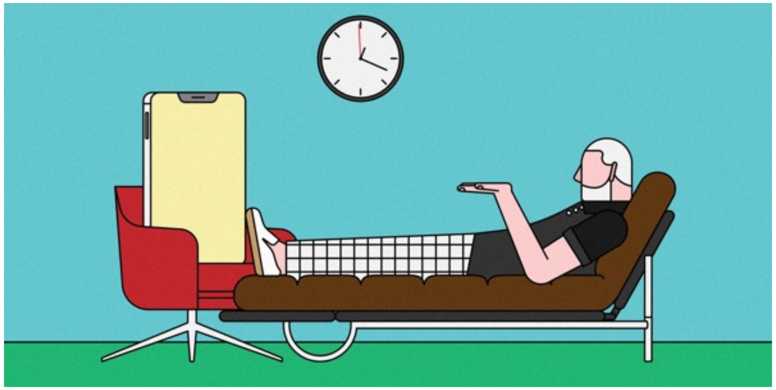
# Welcome to ECE 1786!

- This course is about methods for automatic <u>understanding</u> and <u>generation</u> of human language
- Language is a pathway in and out of your brain!
- Until recently, it was difficult to connect humans and computers with language
- We cover inventions that have made sophisticated communication possible and has led to other remarkable capabilities

# I Look Forward to, in your lifetime …

- Being able to talk to a computer-based doctor to get decent basic medical care

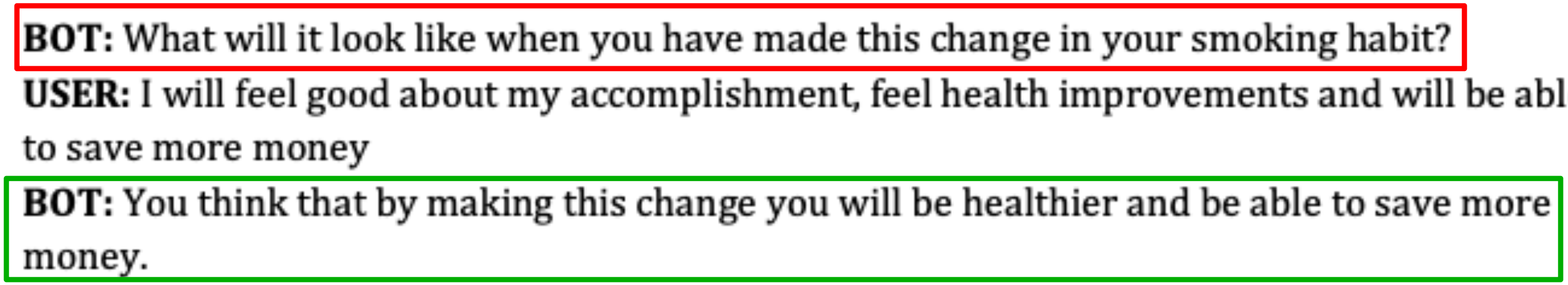
- Being able to get mental-health *talk therapy* any time, anywhere from a computer.



# My Current Research Focus

- Automatic Talk Therapy for Smoking Addiction
  - As one example of talk therapy

Scripted



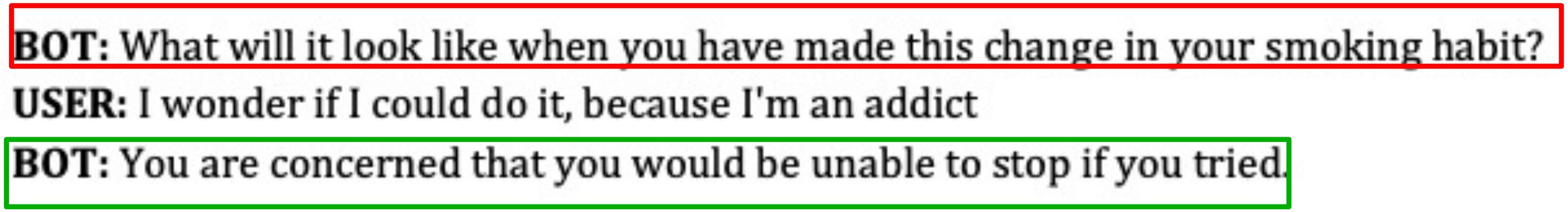
**BOT:** What will it look like when you have made this change in your smoking habit?

**USER:** I will feel good about my accomplishment, feel health improvements and will be able to save more money

**BOT:** You think that by making this change you will be healthier and be able to save more money.

Generated!

Scripted



**BOT:** What will it look like when you have made this change in your smoking habit?

**USER:** I wonder if I could do it, because I'm an addict

**BOT:** You are concerned that you would be unable to stop if you tried.

Generated!

# November 2022

- I was ¾ way through teaching this course for the first timne, which focused on how Large Language models could be used to generate and classify language
- At a key point I showed how GPT 3.5 was truly remarkable in how it could generate language – it could answer many questions, and give creative outputs
  - The GPT 3 playground showed this, but it cost a few cents to use
- Then, in November 2022:

  ChatGPT
  - Was released for free
  - Possibly the most brilliant marketing method of all time

# And Perhaps

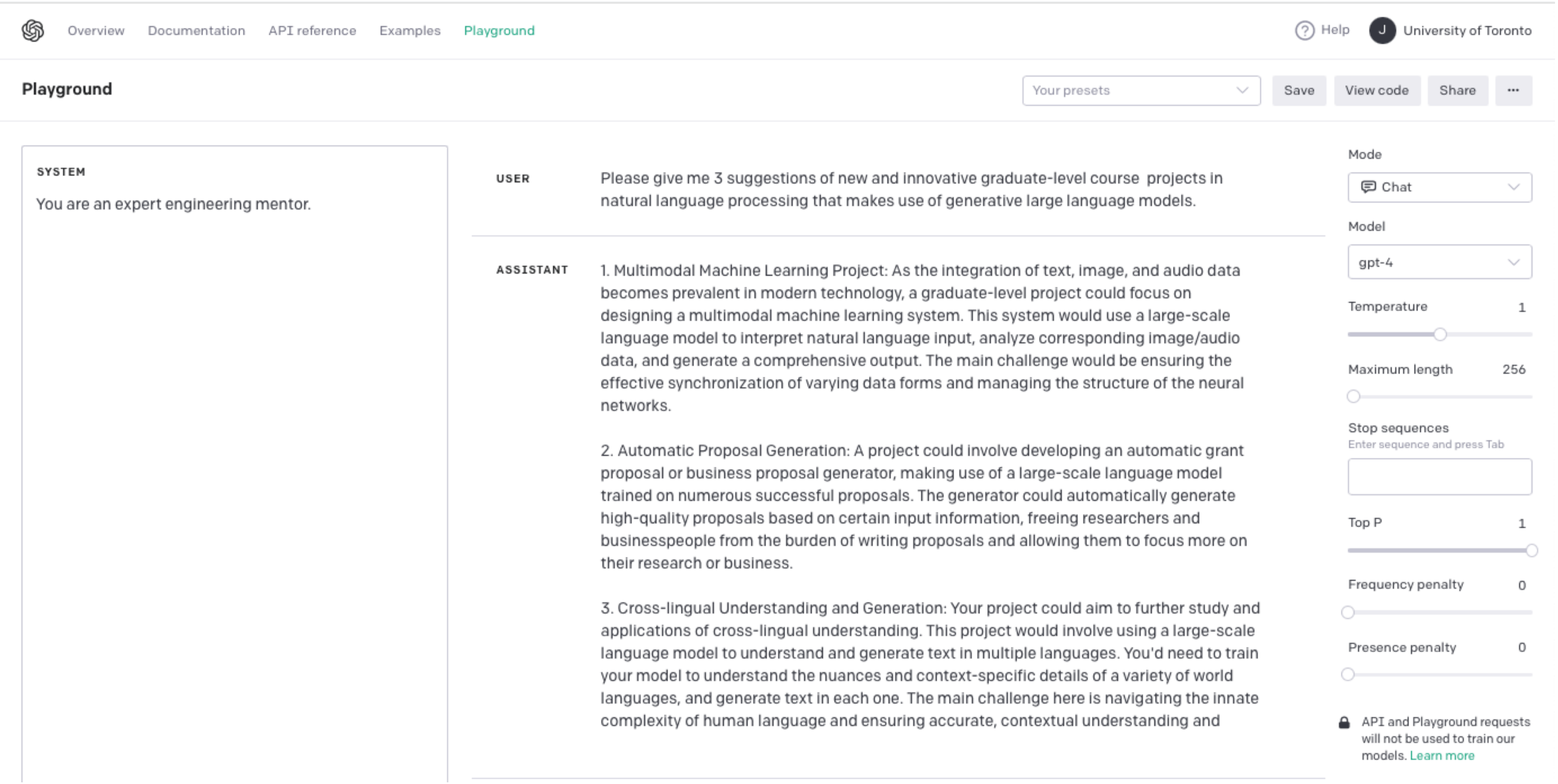
- That is why you are interested in this course?

- How often do you use chatGPT?
  - It seems that many students use it regularly

- GPT-4, released in March 2023 a leap better!

- We are in the process of understanding what it can do
  - And how to use it – the new skill of *Prompt Engineering*
  - How to check what is right/wrong/brilliant?
- The essence of this course is how GPT-4 works

# GPT-4 Helps with Project Ideas



**SYSTEM**

You are an expert engineering mentor.

**USER**

Please give me 3 suggestions of new and innovative graduate-level course projects in natural language processing that makes use of generative large language models.

**ASSISTANT**

1. Multimodal Machine Learning Project: As the integration of text, image, and audio data becomes prevalent in modern technology, a graduate-level project could focus on designing a multimodal machine learning system. This system would use a large-scale language model to interpret natural language input, analyze corresponding image/audio data, and generate a comprehensive output. The main challenge would be ensuring the effective synchronization of varying data forms and managing the structure of the neural networks.

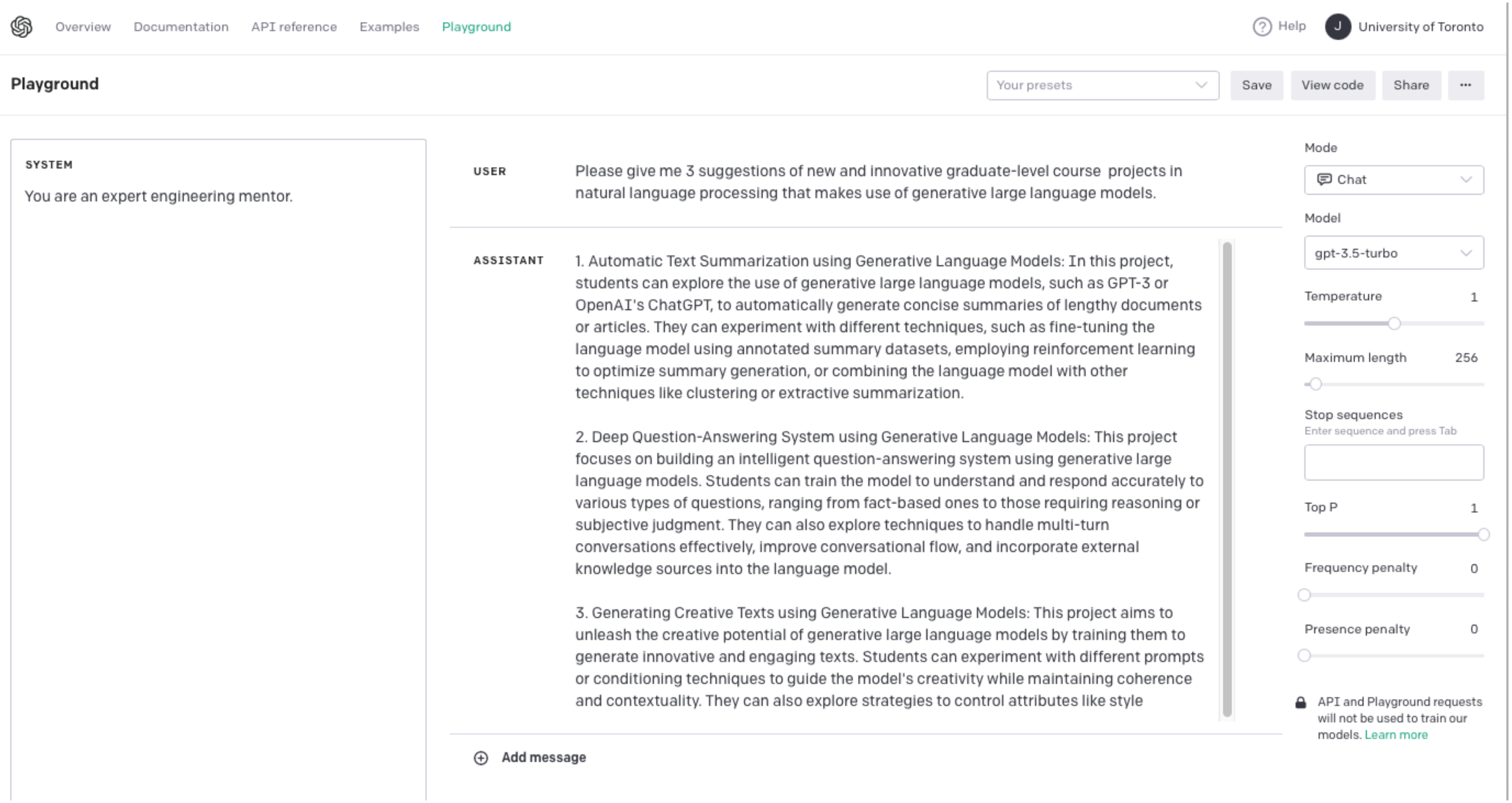
2. Automatic Proposal Generation: A project could involve developing an automatic grant proposal or business proposal generator, making use of a large-scale language model trained on numerous successful proposals. The generator could automatically generate high-quality proposals based on certain input information, freeing researchers and businesspeople from the burden of writing proposals and allowing them to focus more on their research or business.

3. Cross-lingual Understanding and Generation: Your project could aim to further study and applications of cross-lingual understanding. This project would involve using a large-scale language model to understand and generate text in multiple languages. You'd need to train your model to understand the nuances and context-specific details of a variety of world languages, and generate text in each one. The main challenge here is navigating the innate complexity of human language and ensuring accurate, contextual understanding and

Mode: Chat
Model: gpt-4
Temperature: 1
Maximum length: 256
Stop sequences: Enter sequence and press Tab
Top P: 1
Frequency penalty: 0
Presence penalty: 0

API and Playground requests will not be used to train our models. Learn more

# Much Better than GPT-3.5



SYSTEM

You are an expert engineering mentor.

USER

Please give me 3 suggestions of new and innovative graduate-level course projects in natural language processing that makes use of generative large language models.

ASSISTANT

1. Automatic Text Summarization using Generative Language Models: In this project, students can explore the use of generative large language models, such as GPT-3 or OpenAI's ChatGPT, to automatically generate concise summaries of lengthy documents or articles. They can experiment with different techniques, such as fine-tuning the language model using annotated summary datasets, employing reinforcement learning to optimize summary generation, or combining the language model with other techniques like clustering or extractive summarization.

2. Deep Question-Answering System using Generative Language Models: This project focuses on building an intelligent question-answering system using generative large language models. Students can train the model to understand and respond accurately to various types of questions, ranging from fact-based ones to those requiring reasoning or subjective judgment. They can also explore techniques to handle multi-turn conversations effectively, improve conversational flow, and incorporate external knowledge sources into the language model.

3. Generating Creative Texts using Generative Language Models: This project aims to unleash the creative potential of generative large language models by training them to generate innovative and engaging texts. Students can experiment with different prompts or conditioning techniques to guide the model's creativity while maintaining coherence and contextuality. They can also explore strategies to control attributes like style

Mode: Chat

Model: gpt-3.5-turbo

Temperature: 1

Maximum length: 256

Stop sequences: Enter sequence and press Tab

Top P: 1

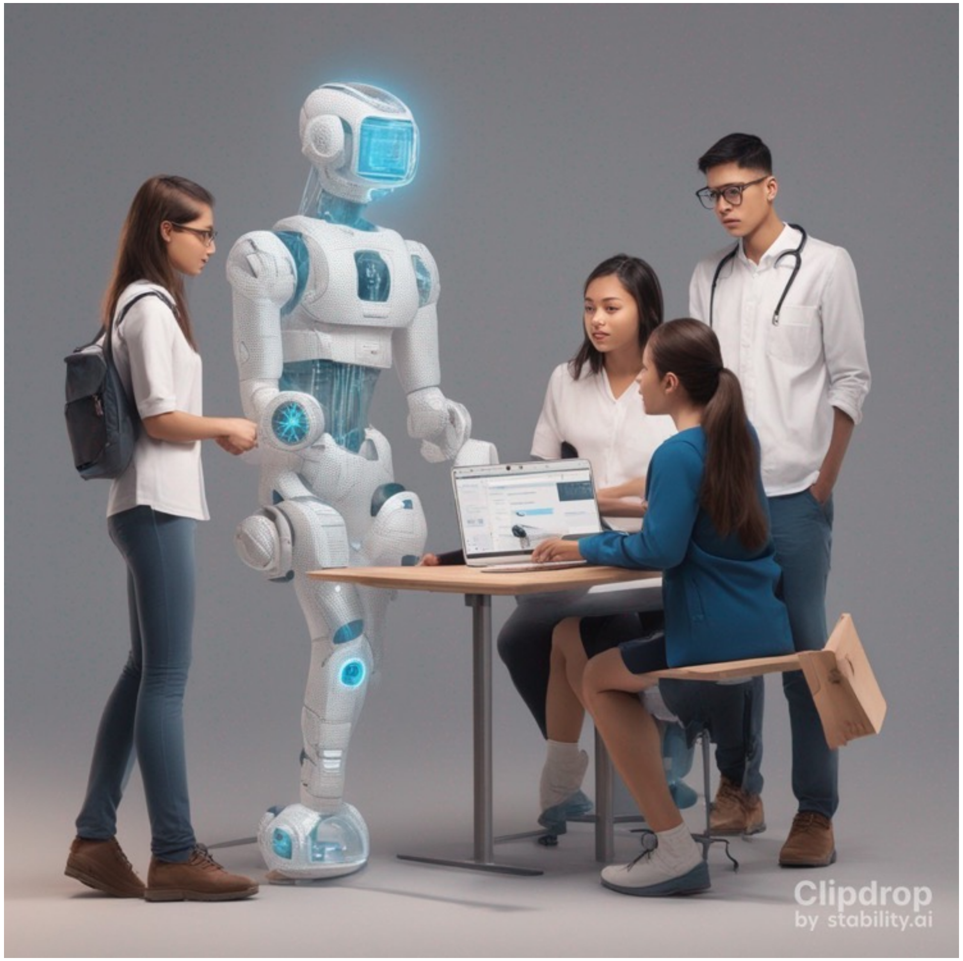
Frequency penalty: 0

Presence penalty: 0

API and Playground requests will not be used to train our models. Learn more

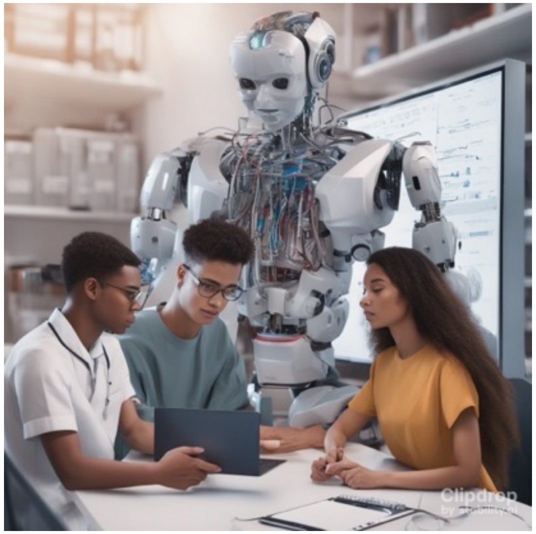
# Also Important: Images + Language

- A realistic image of a group of students using a Large Language Model to make a chatbot that diagnoses medical problems
- From Stability.ai



# Another 3 made at the same time

Clipdrop
by stability.ai

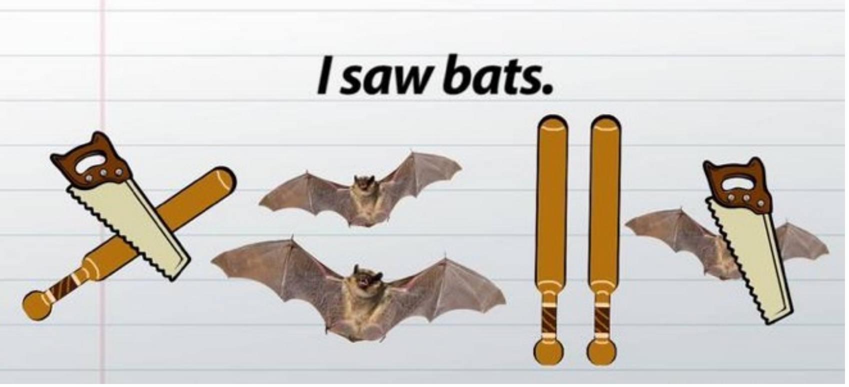


Clipdrop
by stability.ai

# Natural Language Processing

# What Makes Language Difficult

- The ambiguity of language:
  1. Multiple meanings of words
     - tank, bank, duck ….
  2. Context needed to figure out:
     - Milk drinkers are turning to powder
     - Juvenile Court Tries Shooting Defendant
     - Grandmother of Eight makes Hole in One

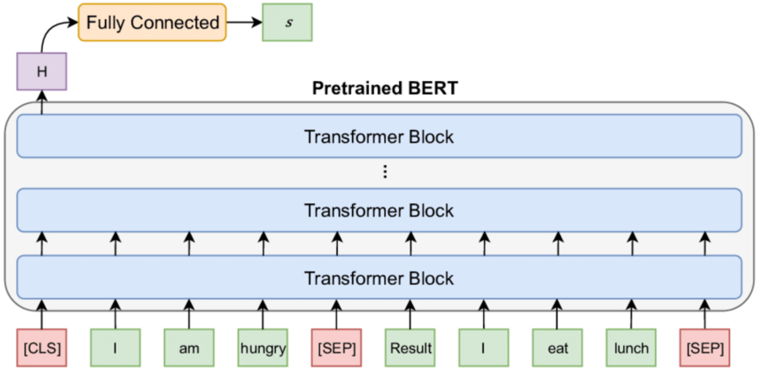


# Language Now Much Easier to Deal with

- Using neural-net-based approaches to NLP
  - since the Deep Learning revolution began in 2012, many fields have been changed
  - A key step in – neural word embeddings - also occurred then
- Key next steps for NLP occurred in 2018
  - ‘Transformer Architecture’ Vaswani et. al
- Then:
  - BERT - https://huggingface.co/bert-base-uncased
  - GPT-2 - https://huggingface.co/gpt2
  - GPT-3 - https://openai.com/blog/gpt-3-apps/
  - GPT-4 - https://openai.com/research/gpt-4
  - Bloom - https://huggingface.co/bigscience/bloom
  - LaMDA - https://blog.google/technology/ai/lamda/
  - Llama2 - https://ai.meta.com/llama/ …. More coming every day!

# Success in Key Applications

1. Translation
2. Classification
3. Dialogue



# Related Courses & Focus of this Course

# Course Focus vs. Related Courses

- Course focus is neural-network approaches to NLP
- NLP is a 60-year-old field which previously used a procedural methodology:
  - Based on human understanding of language
  - Referred to as ‘Computational Linguistics’
    - Grammar – e.g. parts of speech tagging – noun, verb
    - Parsing of language to interpret
  - CSC 485/2501 - Title: **Computational Linguistics**
  - https://www.cs.toronto.edu/~gpenn/csc485/
  - Principally taught by Professor Gerald Penn
- We will make some, limited use of computational linguistics approach

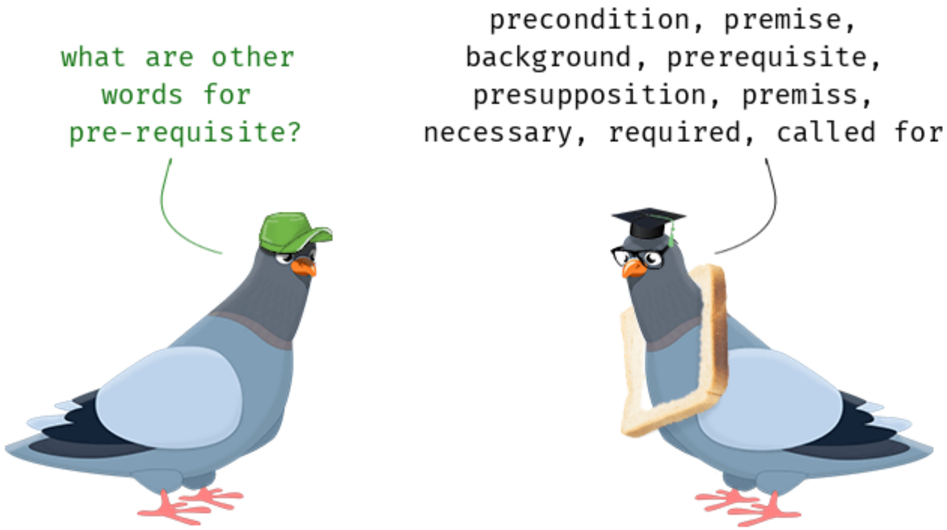
# Course Focus vs. Related Courses

- A more closely related course in Computer Science
- CSC 401/2511 - **Natural Language Computing**
  - https://www.cs.toronto.edu/~frank/csc401/
  - previously taught by Professor Frank Rudzicz; now also Penn
- CSC 2511 has a broader coverage of topics surrounding language
  - a super-set of this course,
  - including things such as Markov Models, Entropy, Automatic Speech Recognition, retrieval and dialogue
- ECE 1786 focuses more narrowly on word embeddings, deep learning, statistical language models,Transformers
  - more depth less breadth
  - more of an engineering, software focus

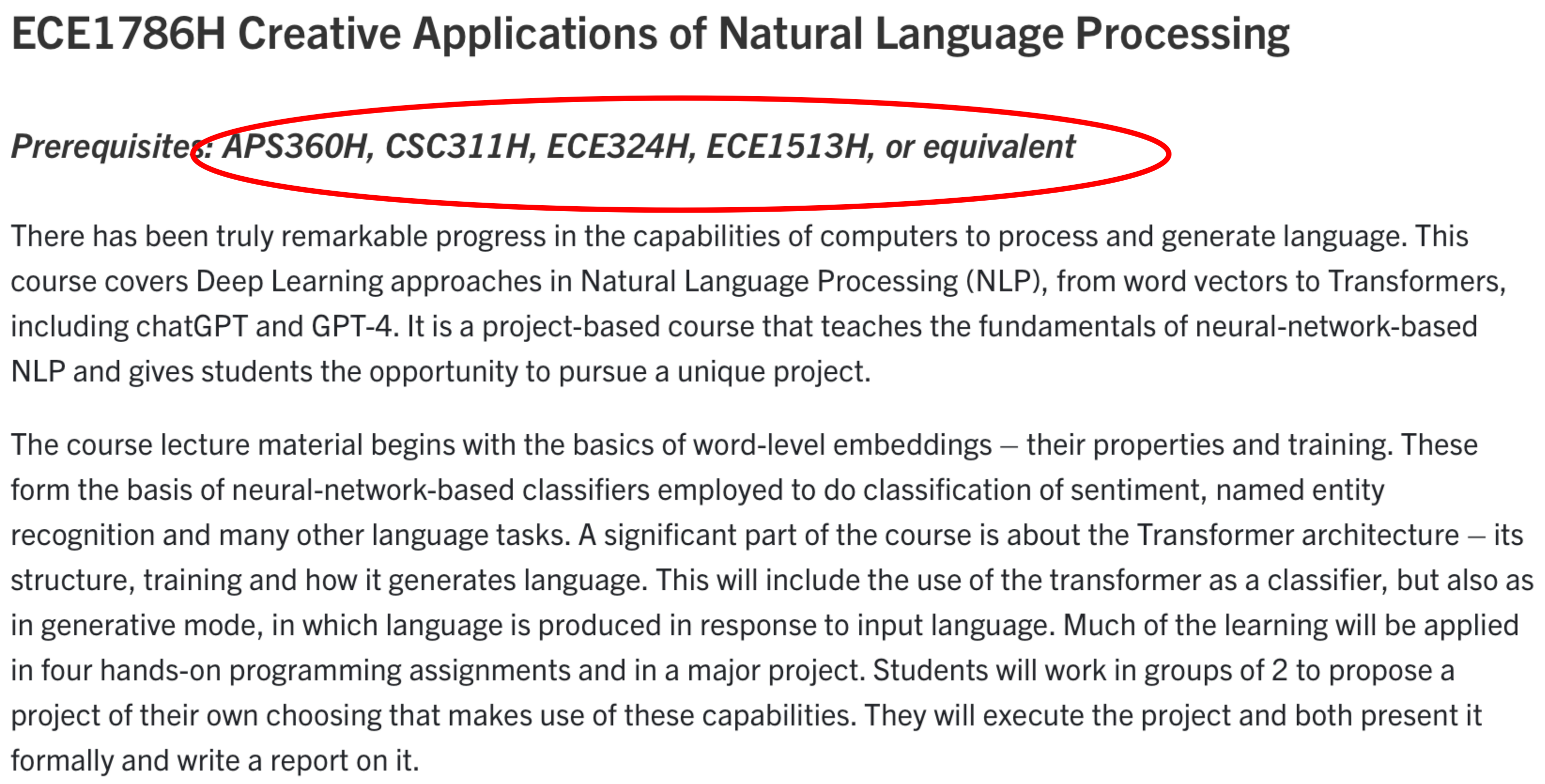
# Learning Outcomes: Understanding

- Word embeddings
- Use of word embeddings in classification tasks
- Transformers
  - Global structure
  - Training
  - Attention, the Transformer Stack
  - Classification; Probabilistic & Auto-regressive Generation
  - Instruct training
  - Multi-Modal – images & text
- PyTorch, Huggingface
- Navigation of open-ended problems in a project
- Startups – current ones & maybe yours?

# Course Pre-Requisites


what are other words for pre-requisite?
precondition, premise, background, prerequisite, presupposition, premiss, necessary, required, called for

# We Need To Talk about Pre-requisites



## ECE1786H Creative Applications of Natural Language Processing

***Prerequisites: APS360H, CSC311H, ECE324H, ECE1513H, or equivalent***

There has been truly remarkable progress in the capabilities of computers to process and generate language. This course covers Deep Learning approaches in Natural Language Processing (NLP), from word vectors to Transformers, including chatGPT and GPT-4. It is a project-based course that teaches the fundamentals of neural-network-based NLP and gives students the opportunity to pursue a unique project.

The course lecture material begins with the basics of word-level embeddings – their properties and training. These form the basis of neural-network-based classifiers employed to do classification of sentiment, named entity recognition and many other language tasks. A significant part of the course is about the Transformer architecture – its structure, training and how it generates language. This will include the use of the transformer as a classifier, but also as in generative mode, in which language is produced in response to input language. Much of the learning will be applied in four hands-on programming assignments and in a major project. Students will work in groups of 2 to propose a project of their own choosing that makes use of these capabilities. They will execute the project and both present it formally and write a report on it.

- Graduate office does not check pre-requisites
- I will do that now, in two ways …

# This Course does not introduce ML

- You **must** have background from a course in machine learning that has depth in neural networks
  - Otherwise, you won’t understand what to do in the assignments!

- Acceptable UofT undergraduate courses (any one)
  - ECE 324 Machine Intelligence, Software and Neural Networks
  - ECE 421 Introduction to Machine Learning
  - APS 360 Applied Fundamentals of Deep Learning
  - CSC 311 – Introduction to Machine Learning
  - CSC 413 – Neural Networks and Deep Learning
- Acceptable UofT Graduate Courses
  - ECE 1513 Introduction to Machine Learning
  - MIE 1517 Introduction to Deep Learning

# Pre-requisite, cont'd

- If undergraduate degree is from elsewhere (most of you):
  - You must have taken a course that is equivalent to one of these University of Toronto courses
  - Next few slides describe the necessary background in detail

- Everyone must fill out the Quercus survey, posted on the course main website, once enrolled in course
  - You'll say which course you've taken, and provide a link to any non-UofT course
  - You'll try to answer questions based on the following….

# What to Know Already

- Machine Learning
  - Classification vs. regression (Logistic v. Linear regression)
  - Binary vs. multi-class classification
  - Supervised vs. unsupervised learning
  - Data labelling

# What to Know Already

- Basic Neural nets and training
  - Linear neurons – weights and biases
  - Non-linear activation functions, e.g sigmoid, ReLU …
  - Multi-layer perceptron (MLP)
  - Loss functions;  binary & multinomial cross entropy
  - Softmax function
  - Training, Validation and Test sets
  - Training & Validation ‘curves’
  - Gradient Descent
  - Stochastic Gradient Descent
  - Hyper-parameter tuning
  - Regularization: normalization, dropout, weight decay

# What to Know Already, continued

- Advanced Neural Networks
  - Convolutional Neural Networks for computer vision (CNN)
    - Kernels, batch normalization
  - Recurrent Neural Networks (RNN)

- Transfer Learning:
  - Pre-trained networks
  - Fine-tuning of pre-trained networks

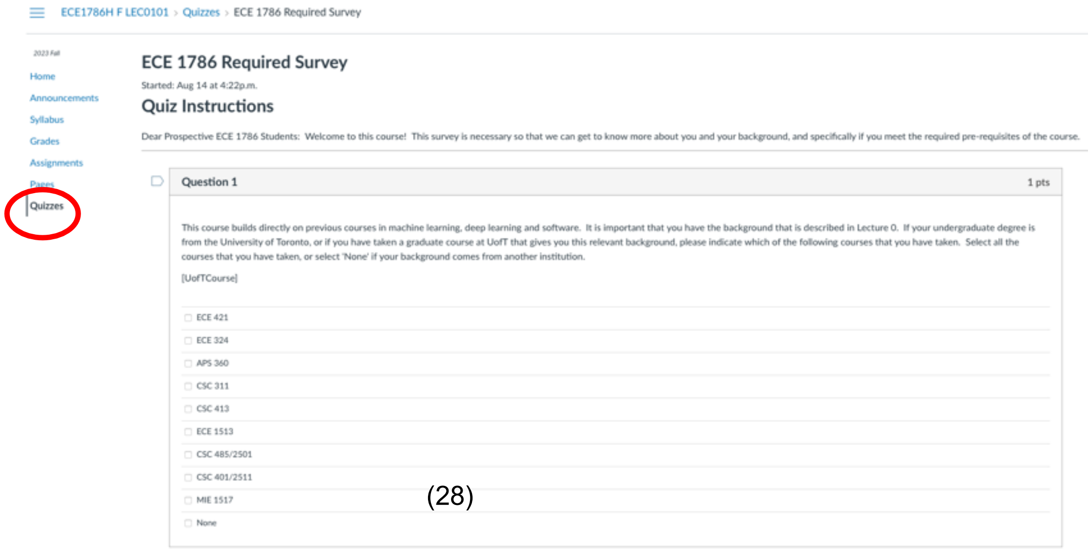
# What to Know Already, continued

- **How to build all that** with Software Frameworks
  - Experience in Tensorflow or PyTorch (implies Python experience)
  - Should have written full training and test loops for applications with significant data sizes
  - Tensors, shape
  - Numpy framework,
  - **How to debug a neural network**

- Data Science concepts
  - False positive, False Negative, True Positive, True Negative
  - Sensitivity, Specificity
  - Area Under the ROC Curve
  - Confusion Matrix

# After finishing video: already enrolled students

- Go to the Quercus Website for this course
  - If you’re enrolled
- Click on ‘Quizzes’ on the left-hand side
  - Answer all of the questions to the best of your ability
  - This will not be graded
  - You must give evidence that you have the pre-requisite knowledge and training to remain in course



ECE 1786 Required Survey

Started: Aug 14 at 4:22p.m.

Quiz Instructions

Dear Prospective ECE 1786 Students: Welcome to this course! This survey is necessary so that we can get to know more about you and your background, and specifically if you meet the required pre-requisites of the course.

Question 1 — 1 pts

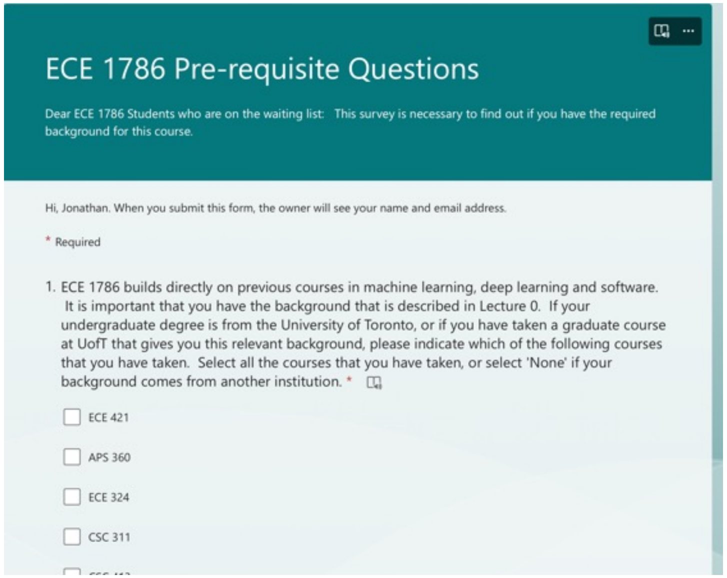
This course builds directly on previous courses in machine learning, deep learning and software. It is important that you have the background that is described in Lecture 0. If your undergraduate degree is from the University of Toronto, or if you have taken a graduate course at UofT that gives you this relevant background, please indicate which of the following courses that you have taken. Select all the courses that you have taken, or select 'None' if your background comes from another institution.

[UofTCourse]

☐ ECE 421
☐ ECE 324
☐ APS 360
☐ CSC 311
☐ CSC 413
☐ ECE 1513
☐ CSC 485/2501
☐ CSC 401/2511
☐ MIE 1517
☐ None

# After finishing video: waitlisted students

- Go to this online survey: https://forms.office.com/r/M7jtEMa0Jf
  - it requires a UofT ID to access
  - answer all of the questions to the best of your ability
  - this will not be graded
  - you must give evidence that you have the pre-requisite knowledge and training to remain in course



## ECE 1786 Pre-requisite Questions

Dear ECE 1786 Students who are on the waiting list: This survey is necessary to find out if you have the required background for this course.

Hi, Jonathan. When you submit this form, the owner will see your name and email address.

* Required

1. ECE 1786 builds directly on previous courses in machine learning, deep learning and software. It is important that you have the background that is described in Lecture 0. If your undergraduate degree is from the University of Toronto, or if you have taken a graduate course at UofT that gives you this relevant background, please indicate which of the following courses that you have taken. Select all the courses that you have taken, or select 'None' if your background comes from another institution. *

☐ ECE 421

☐ APS 360

☐ ECE 324

☐ CSC 311

# Teaching Philosophy

# Teaching Philosophy

1. Teach less, but more in-depth
    - try to explain the intuition behind each approach
    - believe with solid grounding, you can learn more on your own

2. Your learning comes from doing.
    - From writing or working with software, experimenting with it
    - Answer questions about the results of experiments

3. Do open-ended projects
    - An engineer can navigate an open-ended project
    - Learn to conceive a project & describe it: **what** & **why**
    - Then **do it** – figure out **how**, and make it happen
    - Then **describe it**

# Course Structure & Grading

# Grading

| Item | Fraction of Course |
|---|---|
| Assignments (4) | 40% |
| **Project** Proposal Document/Presentation | 10% |
| **Project** Progress Report | 10% |
| **Project** In-Class Final Presentation | 10% |
| **Project** Peer Reviews | 5% |
| **Project** Final Report/Software | 25% |
| **Total** | **100%** |

**Project is 60%**

# Assignments

| # | Date Assigned | Assignment | Due |
|---|---|---|---|
| 1 | September 12 | Word Embeddings – Properties, Meaning and Training | September 25 |
| 2 | September 26 | Classification of Subjective/Objective Text | October 9 |
| 3 | October 10 | Understanding, Training and Using Transformers | October 23 |
| 4 | October 24 | Question Answering Using Transformers | November 13 |

# Textbook

Required Text is Free: **Speech and Language Processing (3rd Edition Draft)** by Dan Jurafsky and James H. Martin:

https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf

2nd Edition has a complete first chapter (missing above):

https://github.com/rain1024/slp2-pdf

# Hardware Acceleration

- As you would know, deep learning often relies on significant computational capability

- For the assignments and project, it is suggested that you purchase the for-pay Google Colab Pro: ($14/month)

  https://colab.research.google.com/signup

- Google Colab Pro + is much more expensive ($67), but enables training of much larger models and faster acceleration

# The Project

# The Project

- Done in Groups of 2
  - No groups of 1 or 3
- The topic is of your own choosing
  - must be approved by instructor
- Must relate to Natural Language Processing & the material covered in this course
- It should be an application of NLP
  - Is OK to do project that is research-oriented but must be discussed with instructor.
- Must collect and label *some* of your own data
  - Why? Data labelling is at the core of all ML/AI work
  - However, this can must be careful not to do too much!
- More details in Lecture 4

# Project Stages & Deadlines

1. **Forming Groups**
   - Should be done by end of October
2. **Project Approval-in-Principle**
   - via email; due October 26th
3. **Project Proposal/Plan**
   - Document Due October 30
4. **Proposal & Plan Presentations**
   - October 31
   - **NOTE EXTRA LECTURE Tuesday October 31st, 6-9pm**
5. **Progress Report**
   - November 20
6. **Final Presentations**
   - December 5th; extra lecture that week as well.
7. **Final Report Due December 12**

# Peer Review

- Each individual student will be asked to provide feedback to other groups on their:
  - Proposal presentation/document
  - Final Presentation
- Asked for specific/useful feedback to group's work
- Feedback/commentary will be graded for quality

# Course Instructor/TAs

# Instructor Bio: Jonathan Rose

- Professor in Electrical & Computer Eng since 1989
  - Bach, Master's & PhD from UofT, Post-Doc at Stanford
- Research: Automation of Medicine/Mental Health
  - Automation of Mental Health using Machine Learning/NLP
    - Focusing on conversational systems for mental health
  - Previously: Field-Programmable Gate Arrays (FPGAs)
- Entrepreneurial/Business Experience:
  - Co-founder of Right Track CAD Corp in 1998
  - Software Engineering Director of Altera 2000-2003, now Intel
- Administration:
  - ECE Dept. Chair of ECE 2004-2009;
  - Chair Engineering Entrepreneurship **Hatchery** Advisory Board
- F.IEEE, F.ACM, F.CAE, FA NAE, FRSC, Sr Fellow Massey College

# Teaching Assistants

- Mohamed Abdelwahab
  - Ph.D. Candidate in ECE
  - Thesis: Concepts in Large Language Models for Addiction Chatbots
  - mo.abdelwahab@mail.utoronto.ca

- Jiading Zhu
  - M.A.Sc. Candidate in ECE
  - Thesis: Next Generation Motivational Interviewing Chatbots
  - jiading.zhu@mail.utoronto.ca

# Three Course Websites:

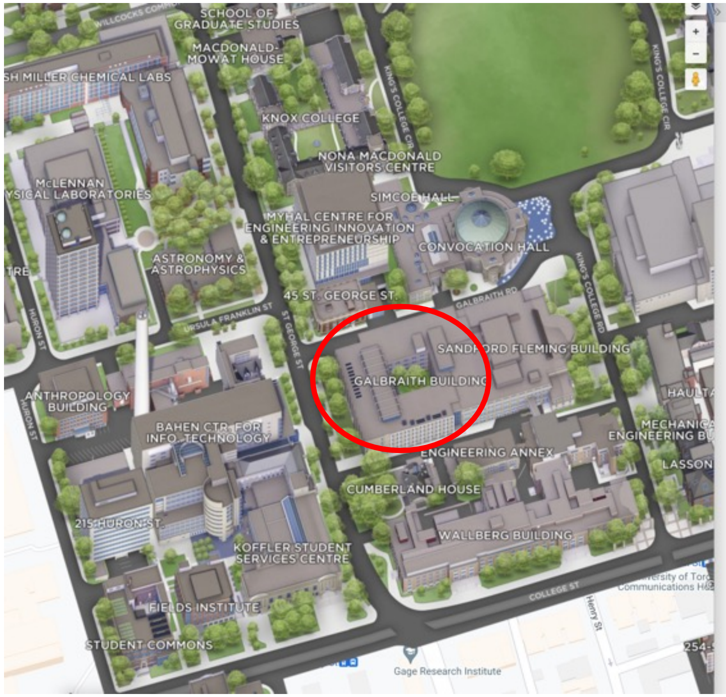
- UofT Quercus (https://q.utoronto.ca/courses/309980) for
  - Assignments release and submitted
  - Grades
  - Announcements

- **Piazza** website for a discussion board
  - See announcement on Quercus that tells you how to access
  - Email me if you don't have access to Quercus & I will add you

- Public Website that replicates most content:
  - https://www.eecg.utoronto.ca/~jayar/ece1786.2023/

# Questions?

- Post them to the Piazza discussion board for this course
- **or**, bring them to the first lecture on
  - Day:  September 12th, 2023
  - Time: 10am-12 noon
  - Place: Galbraith Building, 35 St. George Street, Room 221:
  - See you there then!


SCHOOL OF GRADUATE STUDIES
MACDONALD-MOWAT HOUSE
SH MILLER CHEMICAL LABS
KNOX COLLEGE
NONA MACDONALD VISITORS CENTRE
McLENNAN YSICAL LABORATORIES
SIMCOE HALL
MYHAL CENTRE FOR ENGINEERING INNOVATION & ENTREPRENEURSHIP
CONVOCATION HALL
ASTRONOMY & ASTROPHYSICS
45 ST. GEORGE ST
GALBRAITH RD
KING'S COLLEGE CIR
KING'S COLLEGE RD
URSULA FRANKLIN ST
ST GEORGE ST
HURON ST
SANDFORD FLEMING BUILDING
GALBRAITH BUILDING
ANTHROPOLOGY BUILDING
BAHEN CTR FOR INFO. TECHNOLOGY
HAULTA
MECHANICAL ENGINEERING BU
ENGINEERING ANNEX
LASSON
CUMBERLAND HOUSE
215 HURON ST
WALLBERG BUILDING
KOFFLER STUDENT SERVICES CENTRE
COLLEGE ST
Henry St
FIELDS INSTITUTE
STUDENT COMMONS
Gage Research Institute