

ECE 1786: Creative Applications of Natural Language Processing

ChatGPT



Lecture 0, Fall 2024

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 First Nation.

Today, Toronto is still a meeting place for Indigenous people and immigrants, both new and old, from across the world. I want to acknowledge our duty of stewardship toward the land and I am grateful to have the opportunity to work on this land.



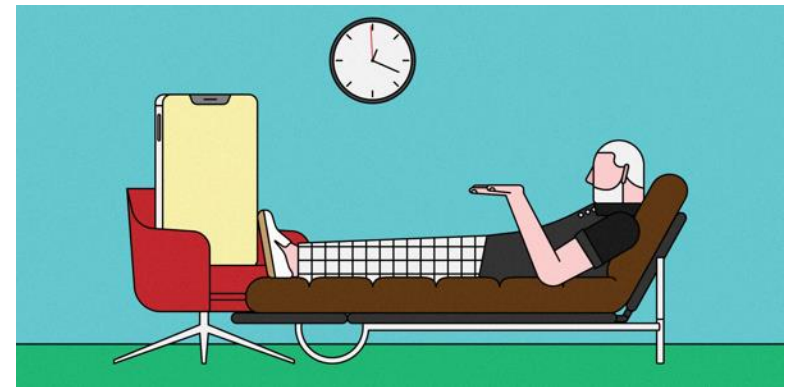
Welcome to ECE 1786!

- This course is about methods for automatic understanding and generation of human language
- Language is a pathway in and out of our brains!
- Until recently, it was difficult to connect humans and computers with language
- Will 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

- Here's a run of GPT-4o-based bot last week, with real smoker (who somehow typed in CAPS):



BOT: Thank you for choosing to speak with me today. I'd love to get a sense of what's been happening in your life recently. Any significant events, experiences, or changes that you'd like to share?

USER: SMOKING MAKES ME FEEL GOOD

BOT: It's clear that smoking has a significant impact on your feelings and well-being. What do you think it is about smoking that brings those positive feelings?

USER: THE WAY I FEEL AFTER SMOKING

BOT: It helps to shift your mood or state of mind. How does this feeling compare to other ways you might manage your emotions or stress?

Conversation, Cont'd

USER: IT FEELS DIFFERENT AND SPECIAL

BOT: That unique feeling is really valuable to you. What are some of the circumstances or moments when you find yourself wanting that feeling the most?

USER: WHENEVER I FEEL DOWN OR OPRESSED

BOT: So, smoking feels like a solution when you're experiencing negative emotions or pressures. How does it affect those tough feelings in the short term vs. over a longer period?

USER: I AFTER SMOKING FEEL RELIEVED AND READY TO TAKE ON THE WORLD



In November 2022

- I was $\frac{3}{4}$ way through teaching this course for the first time, 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:

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



And Perhaps

- That is why you are interested in this course?
 - I'll be curious to know how often you use chatGPT
- GPT-4, released in March 2023 a leap better; more recently 4o and others from elsewhere are better still
 - Rate of progress has been stunning – quality, modalities, price
- We are all still in the process of understanding what these Large Language Models (LLMs) can do
 - How to prompt them, how to decompose problems into several
 - How to check what is right/wrong/brilliant?
- Course is about how LLMs work, and to how use them



Natural Language Processing

What Makes Language Difficult

■ The ambiguity of language:

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



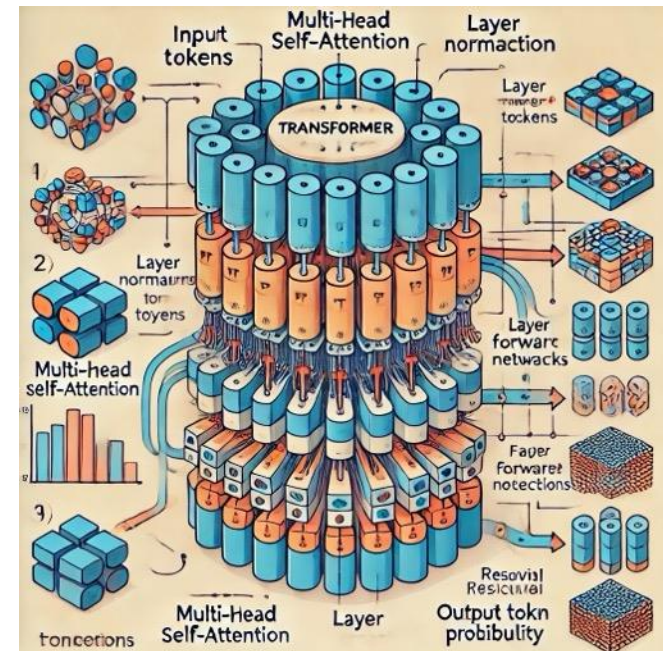
Language Now Much Easier to Deal with

- Using neural-net-based approaches to NLP
 - since the Deep Learning revolution began in 2012, many fields have been changed
 - A key step – 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/>
 - Llama2,3 - <https://ai.meta.com/llama/> More coming every day!



Success in Key Applications

1. Translation
2. Classification
3. Dialogue
4. Coding
5. Composing Essays
6. Planning Projects ...



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
- Transformer Neural Net Architecture
 - Global structure
 - Training
 - Attention, the Transformer Stack
 - Classification & Auto-regressive Generation
 - Instruct training
 - Prompt Engineering
 - The Agentic Approach
- PyTorch, Huggingface
- Navigation of open-ended problems in a project



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
 - ECE 1508 – Applied Deep Learning (several 1508's)



Pre-requisite, cont'd

- If undergraduate degree is from elsewhere (many 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 survey, posted on the course main website
 - You'll say which course you've taken, and provide a link to any non-UofT course
 - You'll be asked to answer a few questions about your knowledge
 - More generally, this is what you'll need to know already:

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: all students

■ Fill out survey: <https://forms.office.com/r/5sBXiTeHLr>

- Or use this QR code ->
- You **must have a UofT login** to reach this site
- Start with forms.office.com
- Make sure you're signed out of any non-UofT account
- Then, use your UofT Email to sign in, you'll be taken to the UofT main login



■ Notes:

- You can only submit **one** response to the survey
- You may not use generative AI (ironically, given the subject of the course!) or the internet to answer the questions asked

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. 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 10 | Word Embeddings – Properties, Meaning and Training | September 23 |
| 2 | September 24 | Classification of Subjective/Objective Text | October 7 |
| 3 | October 8 | Understanding, Training and Using Transformers | October 21 |
| 4 | October 22 | Question Answering, Prompt Engineering and Agentic Programming Using LLMs | November 11 |



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/ed3bookaug20_2024.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 N
 - N will be a function of the # of students who remain in the class and the final level of TA support given to the course
- The topic is of your own choosing
 - But 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.
- There are at least two major types of project; see project information – “class 1” and “class 2”



Project Stages & Deadlines

1. Forming Groups

- Should be done by end of October

2. Project Approval-in-Principle

- via email; due October 24th

3. Project Proposal/Plan

- Document Due October 28

4. Proposal & Plan Presentations

- November 5
- **NOTE EXTRA LECTURE Tuesday November 5th, 6-9pm**

5. Progress Report

- November 18

6. Final Presentations

- December 3rd; extra lecture that week as well.

7. Final Report Due December 10



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
- 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, Intel->Altera
- Administration:
 - ECE Dept. Chair of ECE 2004-2009;
 - Chair Engineering Entrepreneurship **Hatchery** Advisory Board
 - Acting Principal, Massey College 2024
- CM, F.IEEE, F.ACM, F.CAE, FA NAE, FRSC



Teaching Assistants

■ Jiading Zhu

- Ph.D. Candidate in ECE
- Thesis: Next Generation Motivational Interviewing Chatbots
- jiading.zhu@mail.utoronto.ca



■ Zafar Mahmood

- M.A.Sc. Candidate in ECE
- Thesis: MI Chabot & Synthetic Smokers
- zafar.mahmood@mail.utoronto.ca



■ Weizhou Wang

- M.A.Sc. Candidate, Security/LLMs
- weizhou.wang@mail.utoronto.ca



Three Course Websites:

- UofT Quercus (<https://q.utoronto.ca/courses/356609>) for
 - Assignments released and submitted
 - Grades
 - Announcements

- **Piazza** website for a discussion board
 - See Quercus site for access, or:
<https://piazza.com/utoronto.ca/fall2024/ece1786f/home>

- Public Website that replicates most content:
 - <https://www.eecg.utoronto.ca/~jayar/ece1786.2024/>



Questions?

- Post them to the Piazza discussion board for this course
- **or**, bring them to the first lecture on
 - Day: September 10th, 2024
 - Time: 9am-11am
 - Place: OISE, 252 Bloor Street W, Room 2212
 - See you there and then!

