Mini Hackathon Event

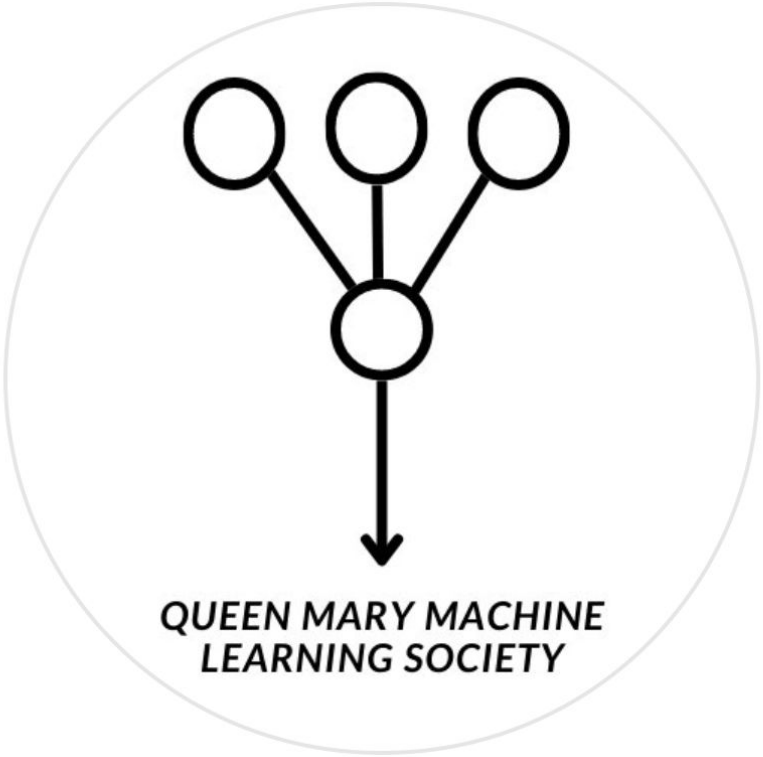
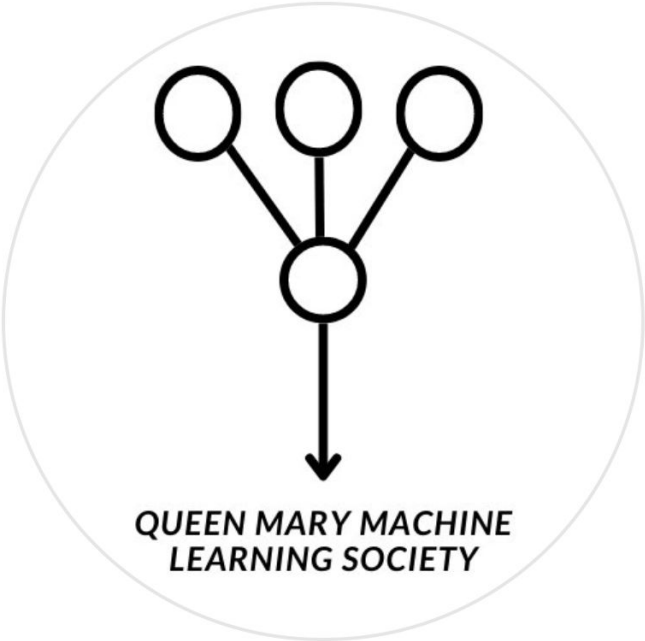


Table of contents

The Machine Learning Roadmap



Let’s talk about Chat-GPT…

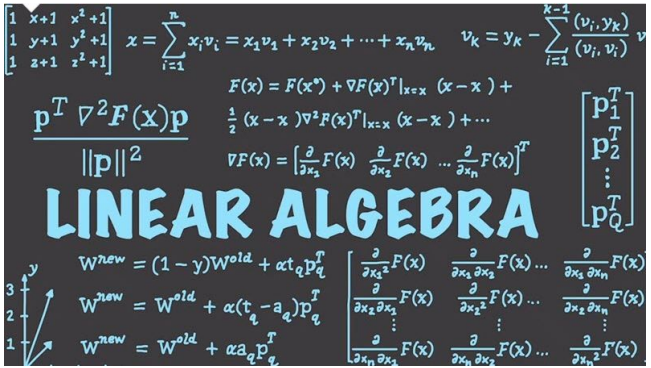
Generative AI Can Harm Learning

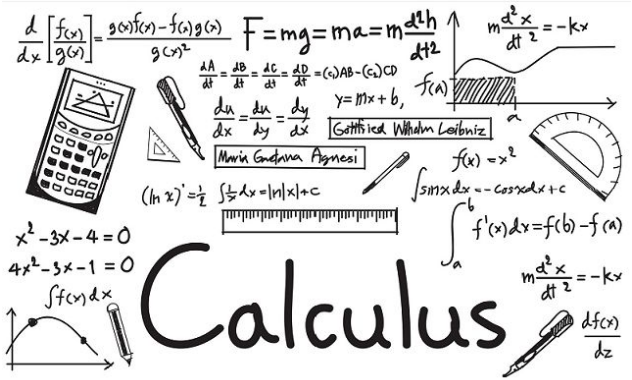
Hamsa Bastani,[1*] Osbert Bastani,[2*] Alp Sungu,[1*†]
Haosen Ge,[3] Özge Kabakcı,[4] Rei Mariman

[1]Operations, Information and Decisions, University of Pennsylvania
[2]Computer and Information Science, University of Pennsylvania
[3]Wharton AI & Analytics, University of Pennsylvania
[4]Budapest British International School

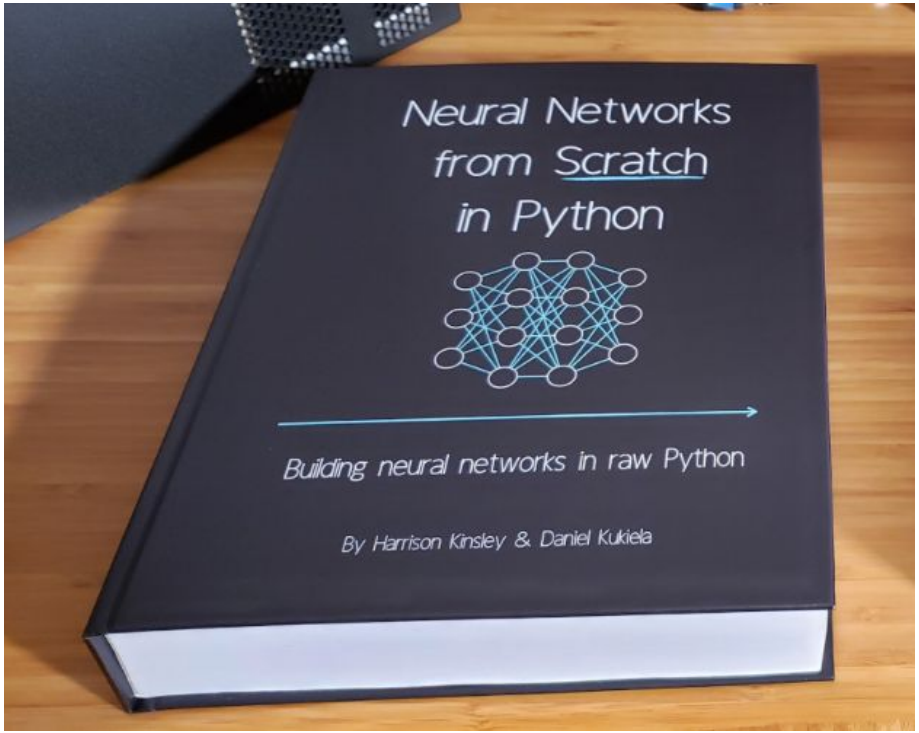
[*]These authors (H.B., O.B., A.S.) contributed equally.
[†]To whom correspondence should be addressed; E-mail: alpsungu@wharton.upenn.edu.

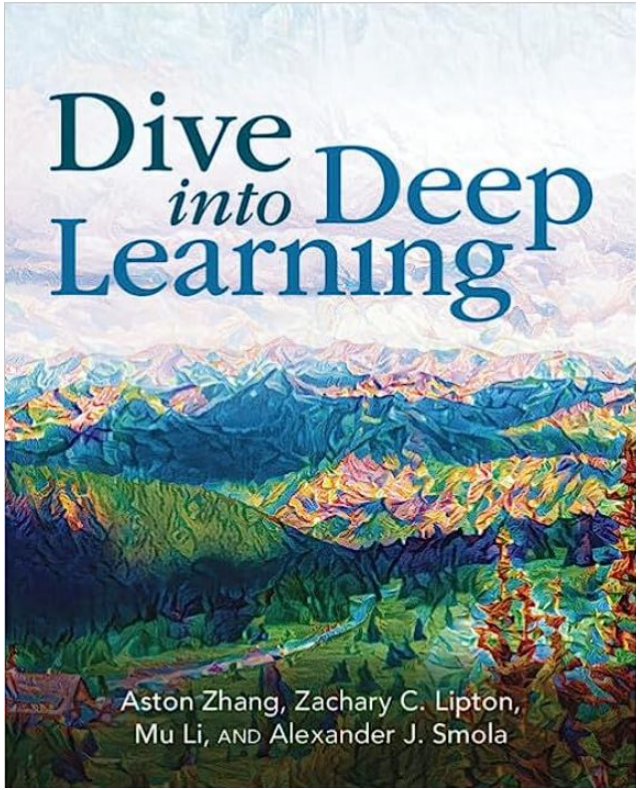
Foundational skills





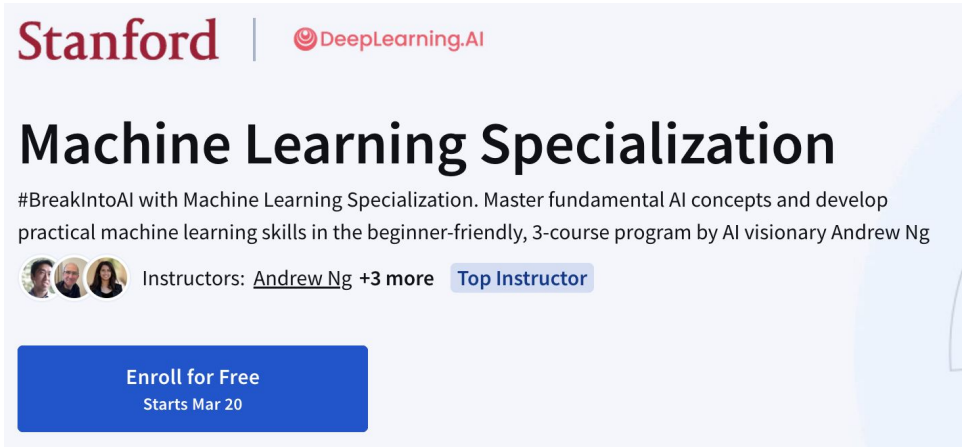
Start by coding from scratch





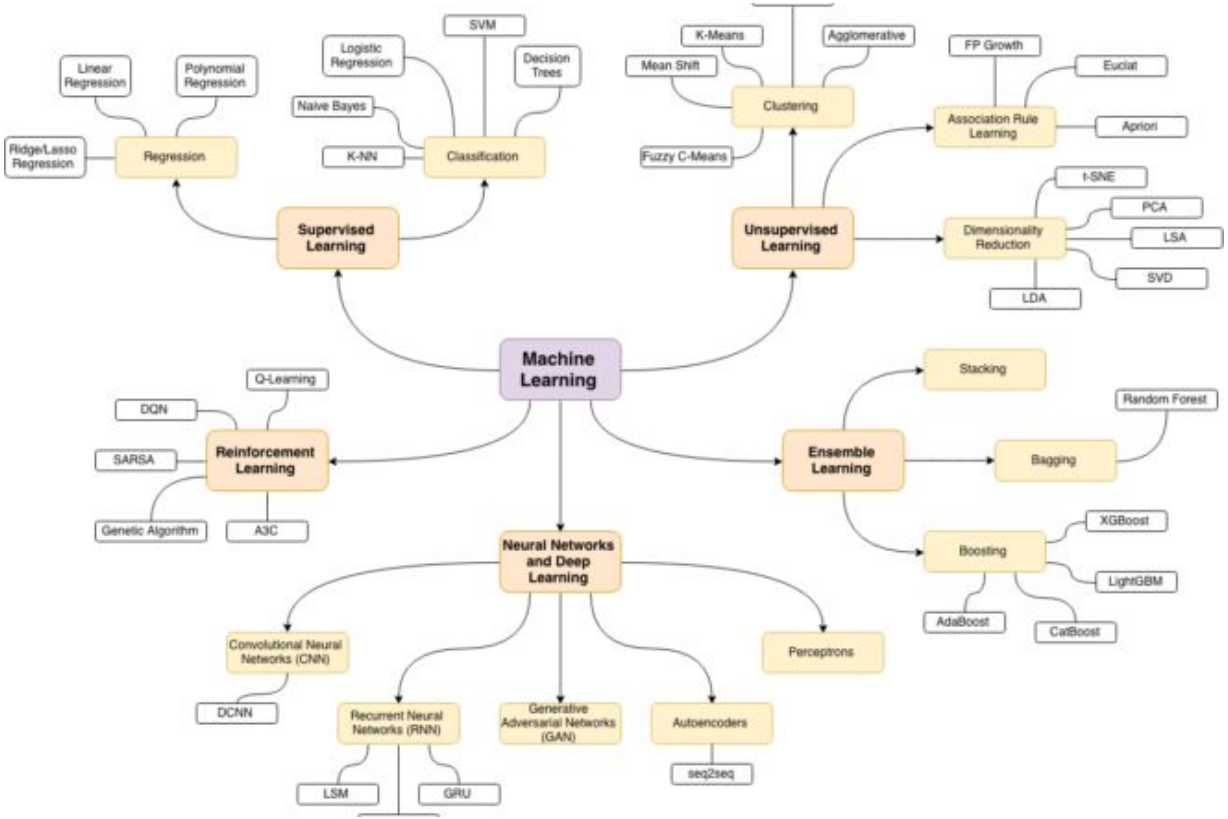
High Quality Machine Learning Resources

coursera



CS109: CS109: Probability for Computer Scientists:
https://web.stanford.edu/class/archive/cs/cs109/cs109.1232/handouts/syllabus.html
CS229: Machine Learning:
https://cs229.stanford.edu/syllabus-spring2022.html
CS230 Deep Learning:
https://cs230.stanford.edu/ - https://cs230.stanford.edu/files/
CS224N: NLP with Deep Learning:
https://www.youtube.com/playlist?list=PLoROMvodv4rMFqRtEuo6SGjY4XbRIVRd4
CS231N: Deep Learning for Computer Vision:
https://www.youtube.com/playlist?list=PL5-TkQAfAZFbzxjBHtzdVCWE0Zbhomg7r
CS224W: Machine Learning with Graphs:
https://www.youtube.com/playlist?list=PLoROMvodv4rPLKxIpqhjhPgdQy7imNkDn
CS330: Deep Multi-Task and Meta Learning:
https://www.youtube.com/watch?v=bkVCAk9Nsss&list=PLoROMvodv4rNjRoawgt72BBNwL2V7doGI
CS236: Generative Models:
https://www.youtube.com/watch?v=MJt_ahtO-to&list=PLoROMvodv4rPOWA-omMM6STXaWW4FvJT8
CS149: Parallel Computing:
https://www.youtube.com/playlist?list=PLoROMvodv4rMp7MTFr4hQsDEcX7Bx6Odp
CS231A: Computer Vision:
https://www.youtube.com/watch?v=68wemjquj4o&list=PLoCMsyE1cvdVnCgHk43vRy7PVTVWJ6WVR&index=18

Try specializing in some of the following



Keep up to date

deepseek

DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

research@deepseek.com

Abstract

We introduce our first-generation reasoning models, D
DeepSeek-R1-Zero, a model trained via large-scale reinfo
vised fine-tuning (SFT) as a preliminary step, demonstra
Through RL, DeepSeek-R1-Zero naturally emerges with
reasoning behaviors. However, it encounters challenges s

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

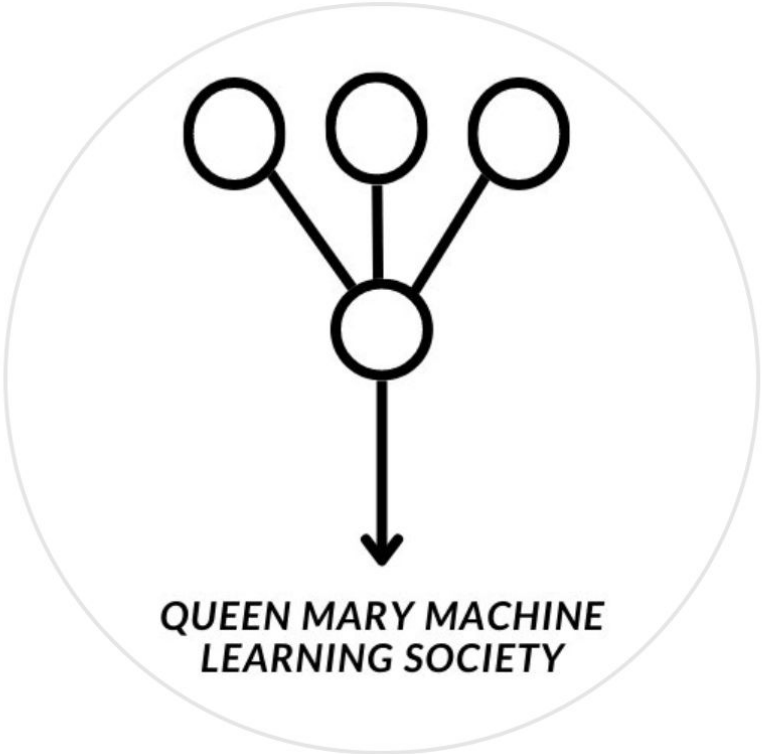
Generative Adversarial Nets

Ian J. Goodfellow, Jean Pouget-Abadie,* Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,† Aaron Courville, Yoshua Bengio‡
Département d'informatique et de recherche opérationnelle
Université de Montréal
Montréal, QC H3C 3J7

Abstract

or estimating generative models via an adversar-
neously train two models: a generative model G
n, and a discriminative model D that estimates
e from the training data rather than G. The train-
ize the probability of D making a mistake. This
imax two-player game. In the space of arbitrary

Fake News Detection



Fake News Detection - Introduction

- What is fake news?
  - Factually false (GDP)
  - Misleading
- Why is it an important topic?
  - Anyone can publish information
  - Automated publication
- Importance of detection for society
  - Prevent spread of harmful misinformation
    - e.g., false health advice to sell useless drugs
  - Influence politics

Fake News Detection - Detection Techniques

- NLP for text analysis
  - Extract language patterns, sentiment, or other linguistic features
- End-to-end classification models
- More advanced:
    - Network analysis of information dissemination

Fake News Detection - Data Sources & Features

- Text content
  - Examples include
    - Exaggerated language
    - Emotional tone
- What is metadata?
  - Broad data
    - Source credibility
    - Publication timestamps

Fake News Detection - Zero-Shot Learning

- What Is Zero-Shot Learning?
  - Classify or understand new categories that were never seen during training
- How does it work?
  - Leverages generalizable representations
    - e.g., embeddings and language models to make predictions on unseen labels
- How to use it for fake news detection?

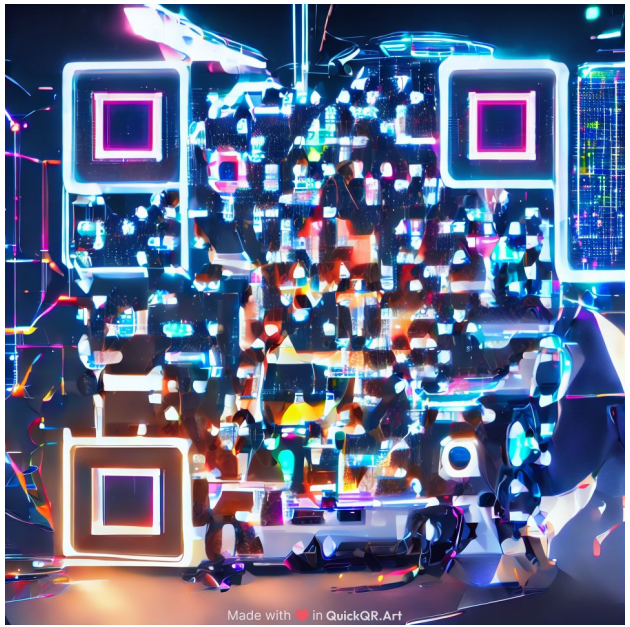
```
# Candidate labels to test (e.g., categories relevant to fake news detection)
candidate_labels = ["fake news", "reliable news", "misleading information"]

# Perform zero-shot classification
result = classifier(text, candidate_labels)
```

kaggle

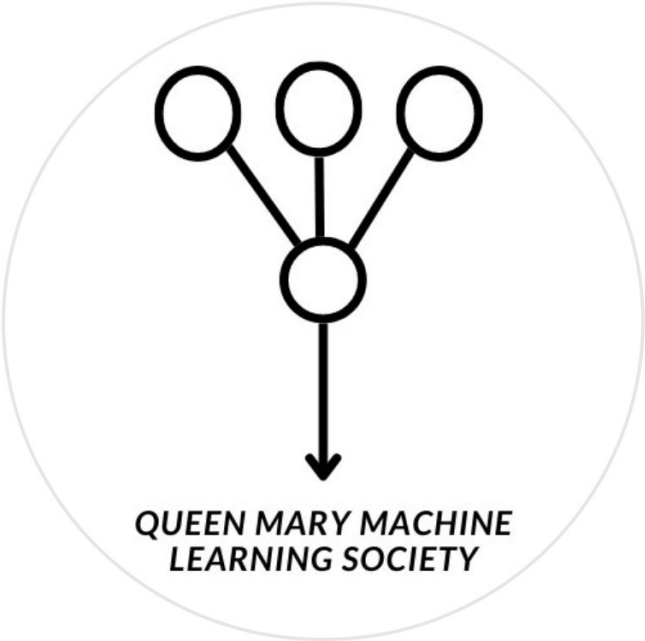
New Communications Platform

- Join our Discord:
  - Link: https://discord.gg/xcfp4UHGWa
  - QR-Code:



kaggle

Mini Hackathon Presentation
Fake news detection



Welcome to the Hackathon!

Manu's Machine Learning Lectures x Kaggle Teams

🚀 **Showcase your skills!** Compete in this exciting challenge to classify fake and real news using machine learning.

🎯 **Your Goal:** Build a model that can accurately distinguish between real and fake news articles using the provided dataset.

🏆 **Prove your expertise** and rank at the top!

Let’s start by making teams of 2-4 people

Rules & Dataset Overview

📌 **Dataset**:

- You will receive a dataset with **news articles** labeled as **Fake (1) or Real (0).**
- Columns include **Title, Subject, Text, Date, and Target**.
- Use **only the first 75%** of the dataset for training. The final 25% is for testing only.

📌 **Rules**:

1. **No external data** is allowed.
2. You must **only train on the first 75%** of the dataset.
3. **Submit predictions** for the test set.
4. Feature engineering is allowed.
5. Use any machine learning or deep learning approach.

Scoring System 🏅

You can earn up to **3 points** based on your performance:

🥇 **1 Point** - Highest overall accuracy (using any features you want).

📝 **1 Point** - Highest accuracy using **only the text** column.

📢 **1 Point** - Highest accuracy using **only the title** column.

- 🔹 The leaderboard will be based on accuracy. Aim high! 🔹

Best of Luck! 🚀

Thanks for Listening!