

# Intro to the course

Marco Morales

[marco.morales@columbia.edu](mailto:marco.morales@columbia.edu)

GR5074: Projects in  
Advanced Machine Learning

Spring 2022  
Columbia University

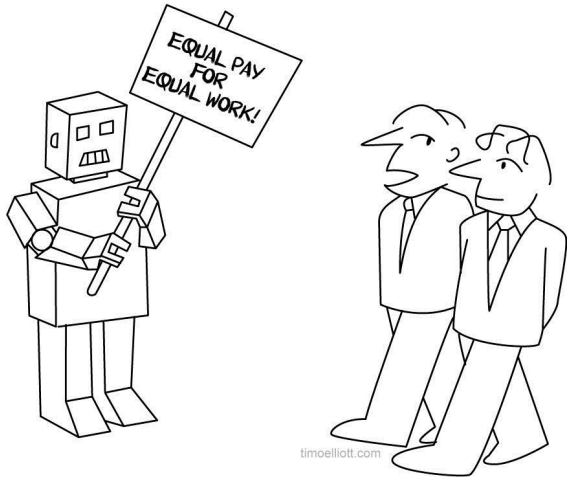
# About us

**Instructor:** **Marco Morales**  
**email:** marco.morales@columbia.edu  
**Office:** 509E International Affairs Building  
**Office Hours:** appointments (via Calendly)

**TA:** **Rachel Lee**  
**email:** yl3751@tc.columbia.edu

**TA:** **Heinrich Peters**  
**email:** heinrichpeterz@gmail.com

# About this course...



*"I guess they really are getting smarter!..."*

# About this course...

- ▶ Continuation of **Machine Learning for the Social Sciences (GR5073)**
  - ▶ Linear Regression (and regularizations)
  - ▶ Logistic Regression (and regularizations)
  - ▶ KNN
  - ▶ SVM
  - ▶ Decision Trees (and ensembles)

# About this course...

## Objectives:

1. extend your Machine Learning toolkit through
  - ▶ exposure to **Deep Learning** models
    - ▶ Recurrent Neural Networks (RNNs)
    - ▶ Object detection models
    - ▶ Convolutional Neural Networks (CNNs)
  - ▶ experimentation with architectures for these models
2. create a portfolio of projects (**AI Model Share Initiative**)

**Focus:** gaining **intuition** + practical **experience**

# Course Dynamics: the **asynchronous** component

in your own time

## Materials



assigned  
readings

 Classroom



team  
projects

 Classroom



individual  
project

## Activity



read



team  
collaboration



code

## Submission



questions/  
comments



submit



AI | MODEL SHARE BETA



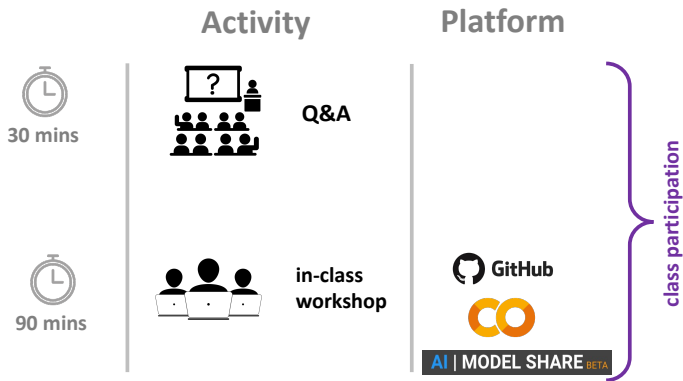
submit



AI | MODEL SHARE BETA

# Course Dynamics: the **synchronous** component

all together



# All class materials in the course's GitHub repo

clone and pull before each class

The screenshot shows a GitHub repository page for 'marco-morales / QMSS-GR5074\_Spring2022'. The repository is public and has 1 star, 0 forks, and 1 watch. The main branch is 'main'. The repository contains a file named 'README.md' which is currently selected. The README content includes the title 'QMSS GR5074 - PROJECTS IN ADVANCED MACHINE LEARNING', the instructor 'Marco Morales, Columbia University', and the teaching assistants 'Rachel Lee, Columbia University' and 'Heinrich Peters, Columbia University'. It also provides a description of the repository as a companion to the course 'Projects in Advanced Machine Learning' and includes a link to the course syllabus.

marco-morales / QMSS-GR5074\_Spring2022 (Public)

Unpin Unwatch 1 Fork 0 Star 1

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code About

marco-morales Add links to CU e-versions for recommended books b28391c 2 days ago 3 commits

syllabus initial commit 5 days ago

.gitignore initial commit 5 days ago

README.md Add links to CU e-versions for recommended books 2 days ago

README.md

## QMSS GR5074 - PROJECTS IN ADVANCED MACHINE LEARNING

**Instructor:** Marco Morales, Columbia University

**TAs:** Rachel Lee, Columbia University  
Heinrich Peters, Columbia University

This repository is a companion to the course [Projects in Advanced Machine Learning](#) taught at the [Quantitative Methods in the Social Sciences](#) program over the Spring of 2022. It contains curated reference materials, slides and sample code. You can find the most updated version of the course syllabus [here](#). Make sure to check regularly for updates.

Readme 1 star 1 watching 0 forks



# Course communications



- ▶ **Slack** is the **preferred method of communication** for this course
- ▶ post questions in appropriate **channels**
- ▶ DM classmates & instructors / TAs
- ▶ **aim: collaboration** – if you have a question, likely others have it also (and perhaps an answer to share)
- ▶ we will **reserve email for official communications**

# Tech stack for this class

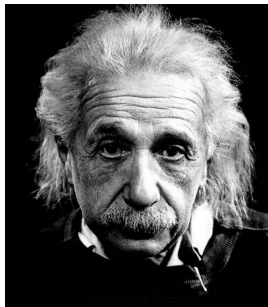


**AI | MODEL SHARE** BETA

# Course Requirements

- ▶ **Team Projects (60%)**
- ▶ **Individual Project (25%)**
- ▶ **Attendance & Class Participation (15%)**

# Why Deep Learning?



*"The most incomprehensible thing about the world  
is that it is at all comprehensible."*

- Albert Einstein

# Why Deep Learning?

- ▶ the world is **compositional**: can be represented by stacked hierarchies increasing in abstraction
- ▶ **Deep Learning** allows learning at each level of the hierarchy (modules)
- ▶ **Deep Learning** replaces hand-engineered feature-extractors with (stacked) trainable modules
  - ▶ traditional ML relies on hand-engineered feature-extractors to expand representations of inputs
- ▶ but ... need care to match architecture with the nature of the data (for generalization)

# Deep Learning genealogy

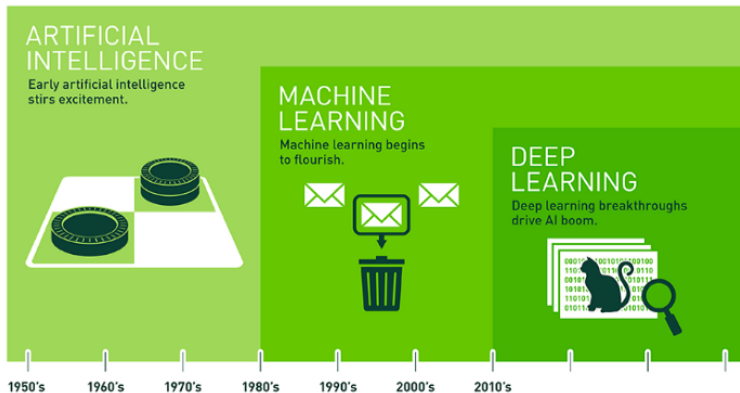


Figure: Cassie Kozyrkov (2018)

# Deep Learning genealogy

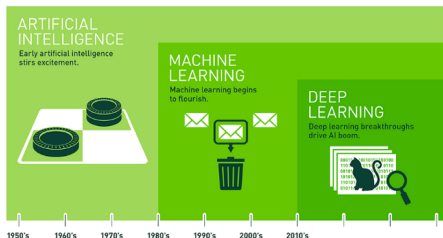


Figure: Cassie Kozyrkov (2018)

- ▶ **1940 - 60s:** cybernetics - rules-based
- ▶ **1980 - 90s:** back-propagation (“shallow”)
- ▶ **2010s - :** back-propagation + (“deep”)

# Intro to the course

Marco Morales

[marco.morales@columbia.edu](mailto:marco.morales@columbia.edu)

GR5074: Projects in  
Advanced Machine Learning

Spring 2022  
Columbia University