CS 232: Artificial Intelligence

Spring 2024

Prof. Carolyn Anderson
Wellesley College

My Work in AI

AI-Assisted Programming (*How Can Machines Help Us Program?*)

Knowledge Transfer from High-Resource to Low-Resource Programming Languages for Code LLMs
F Cassano, J Gouwar, F Lucchetti, C Schlesinger, CJ Anderson, ...
arXiv preprint arXiv:2308.09895

StarCoder: may the source be with you!
R Li, LB Allal, Y Zi, N Muennighoff, D Kocetkov, C Mou, M Marone, C Akiki, ...
arXiv preprint arXiv:2305.06161

StudentEval: A Benchmark of Student-Written Prompts for Large Language Models of Code
HML Babe, S Nguyen, Y Zi, A Guha, MQ Feldman, CJ Anderson
arXiv preprint arXiv:2306.04556

SantaCoder: don't reach for the stars!
LB Allal, R Li, D Kocetkov, C Mou, C Akiki, CM Ferrandis, N Muennighoff, ...
arXiv preprint arXiv:2301.03988

MultiPL-E: a scalable and polyglot approach to benchmarking neural code generation
F Cassano, J Gouwar, D Nguyen, S Nguyen, L Phipps-Costin, D Pinckney, ...
IEEE Transactions on Software Engineering

Natural Language Processing (*How Do Machines Understand Language?*)

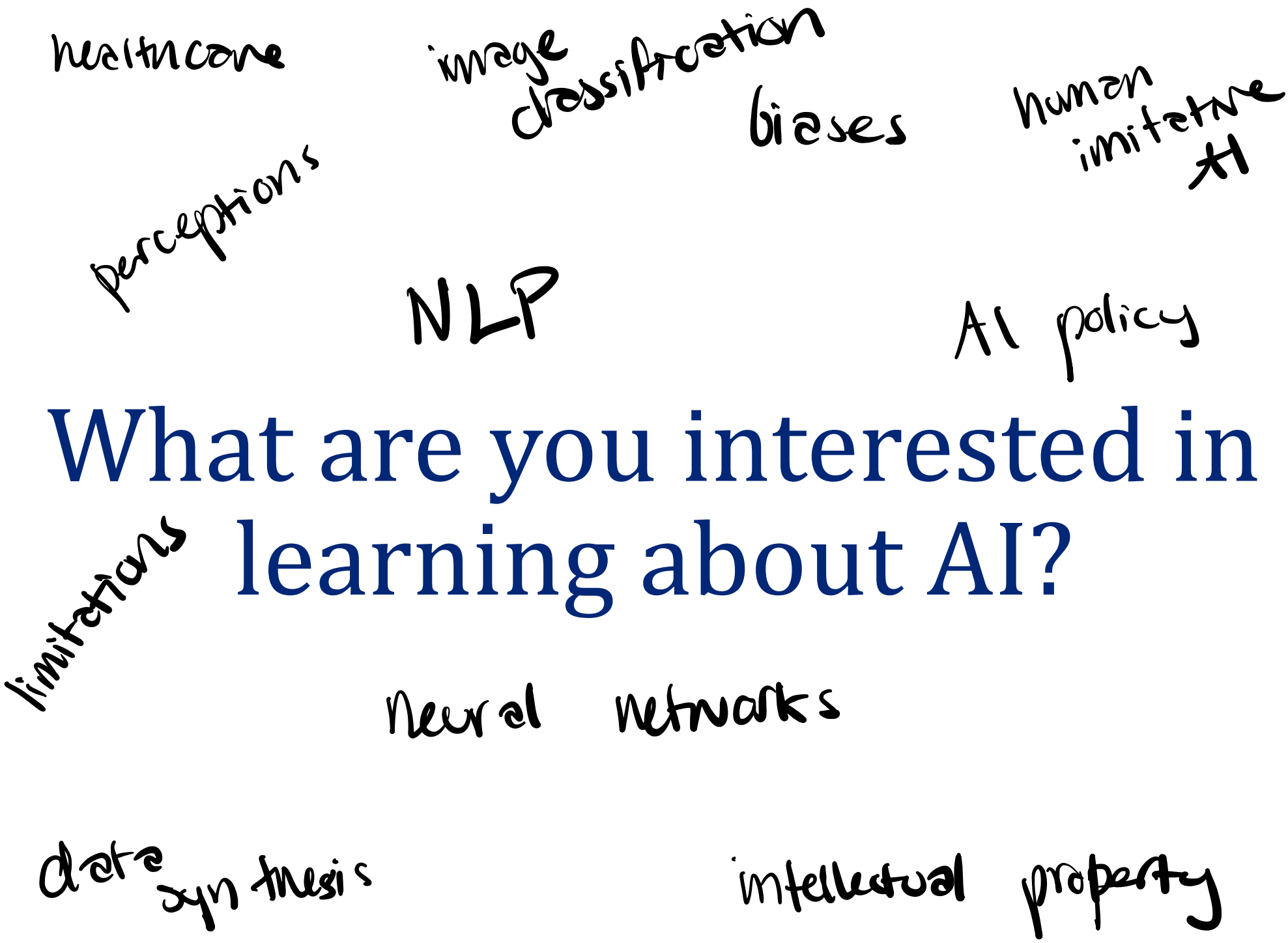
Solving and Generating NPR Sunday Puzzles with Large Language Models
J Zhao, CJ Anderson
arXiv preprint arXiv:2306.12255

Do All Minority Languages Look the Same to GPT-3? Linguistic (Mis) information in a Large Language Model
S Nguyen, CJ Anderson
Proceedings of the Society for Computation in Linguistics 6 (1), 400-402

ProSPer: Probing human and neural network language model understanding of spatial perspective
T Masis, C Anderson
Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting …

Guess who's coming (and who's going): Bringing perspective to the rational speech acts framework
CJ Anderson, BW Dillon
Proceedings of the Society for Computation in Linguistics 2 (1), 185-194

Tell me everything you know: a conversation update system for the rational speech acts framework
CJ Anderson
Proceedings of the Society for Computation in Linguistics 2021, 244-253


healthcare
image classification
biases
human imitative AI
perceptions
NLP
AI policy
What are you interested in learning about AI?
limitations
Neural networks
data synthesis
intellectual property

recognize patterns & frameworks

predictable smart decisions given a set of data

recognized understand coherence

follow instructions accurately

different ways of approaching goals

What is intelligence?

is very good at a task intelligence?

imitate human decisions (because humans have the most complex thought)

interact in a natural way

is human intelligence the goal??

Types of AI

Types of AI

??

- Narrow / weak AI: human-like performance on a single task
  good
- General / strong AI: AI that can do everything humans can do

Task-Based AI

Task-Based AI

Our goal is to write programs that can solve tasks. This is sort of the goal of all computer science.

In AI, though, the tasks we focus on are ones that seem to require human intelligence. This is a moving standard- what seems impossible for a computer to solve one day may eventually become very easy.

AI Tasks

Almost all AI tasks can be grouped into one of three main categories:

- Search – classical search algorithms, reinforcement learning
- Classification – regression, neural networks
- Generation – neural networks

Learning

How do people learn?

- Some tasks have a **critical acquisition period:**
  - Language acquisition
  - Vision
  - Music training

- Others can be consciously acquired

Example: Child Language Acquisition

Example 1

cj and ember manning liked

Gareth Roberts @garicgymro · 45m

Just overheard from two of my kids:
Osian (5;1): Look how I catched Mickey!
Eirwen (8;2): Do you mean caught?
Osian: … yeah.
Eirwen: But you can keep saying catched!
Osian: Look how I catched him!

Example 2

How do people learn?

human infants

large language models

photo credit: Josef Fruehwald

How does AI learn?

- Symbolic AI
  - Program rules for the model
- Machine learning
  - Supervised learning
    - Give model input/output pairs to learn from
  - Reinforcement learning
    - Give model a reward function
  - Unsupervised learning
    - Model tries to **separate** data

Practicalities

Staff

Carolyn Anderson (she)

Professor

Lepei Zhao

Tutor

Lyra Kalajian

Tutor

Jess Yao

Grader

Help Hours

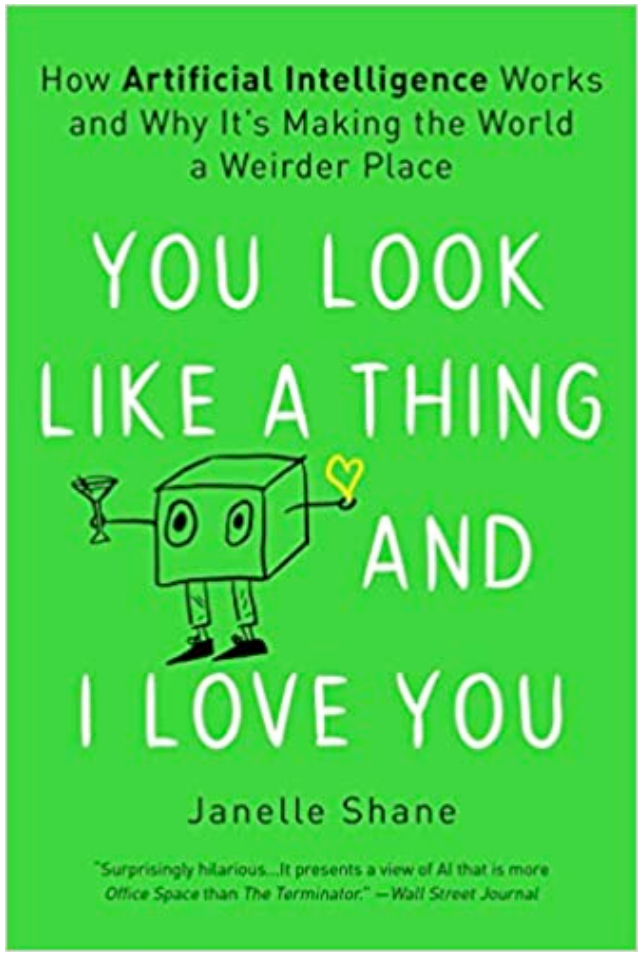
- Tutor hours:
  - TBA
- My help hours:
  - Monday 4-5:15
  - Friday 3:30-4:30
  - By appointment

Come to my help hours to …

- Get help with CS232
- Talk about AI

Readings

Most readings will be from Janelle Shane's book *You Look Like A Thing And I Love You* (abbrev. YLLATAILY).



All readings are listed on the schedule. Some recommended readings are also posted there—many from Russell & Norvig's *AI: A Modern Approach*.

Please finish each week's required reading **before our Tuesday class.**

Homework will be in Python

I recommend setting up a Python 3.8 virtual environment.

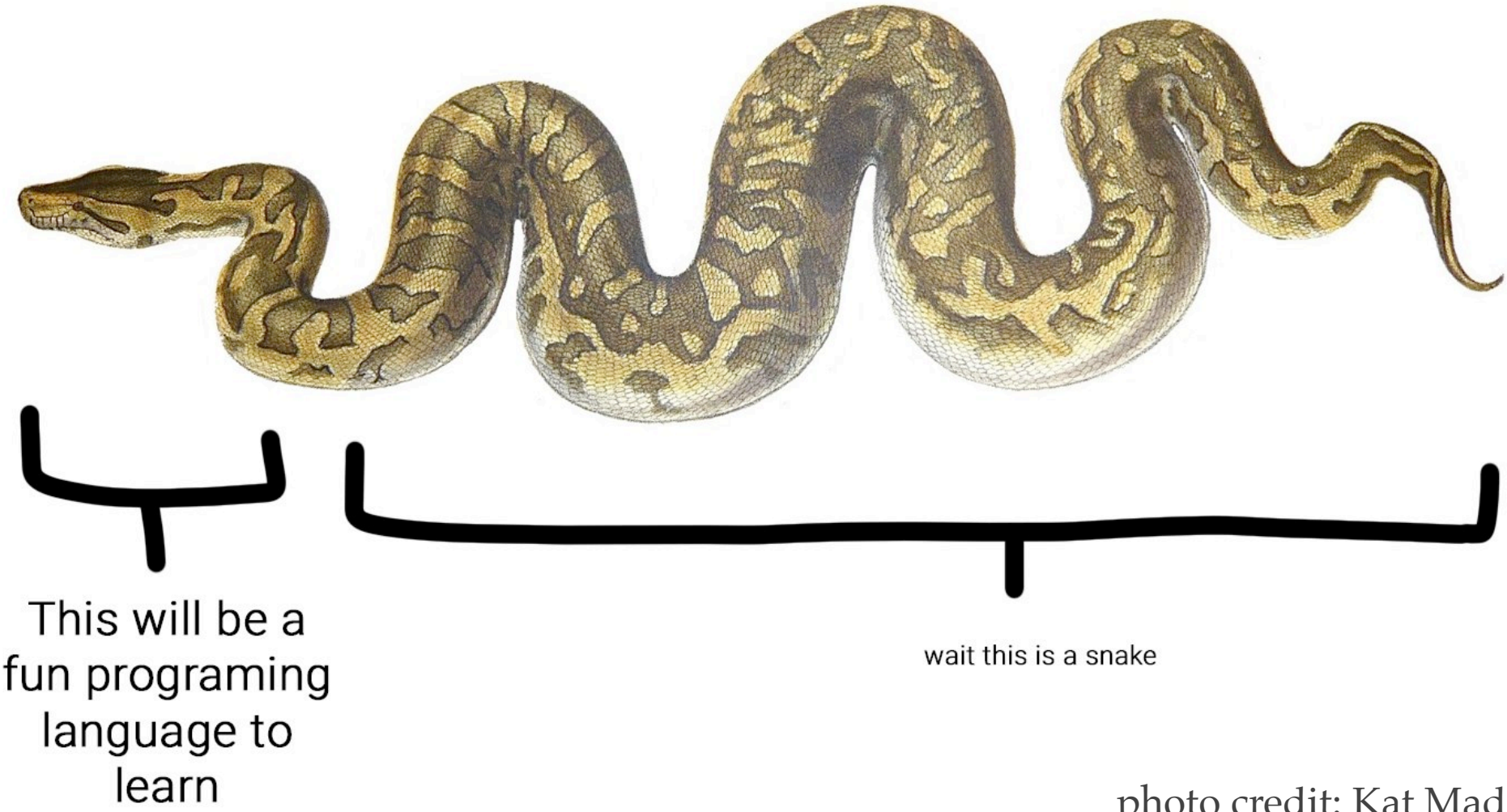


photo credit: Kat Maddox

Assignments

- Assignments are due on **Mondays at 10 PM**
- Homework submission will be through Gradescope
- Expect an assignment **every week**
- Get help early!

Late Policy

You have **5 late days** for the semester, which you can use all at once, or spread across assignments. **There is a 5% penalty per day once you use up your late days.**

Important: I will not answer questions on late work during help hours.

If you have a prolonged illness or unexpected circumstance, let me know and **we'll work together** to make a **custom plan**.

Collaboration policy

In this class, you can talk at a high-level with other students about assignments, but **you cannot show them your code**.

If you discuss a homework problem with another student, **please note this** on your assignment when you submit it.

You may not use ChatGPT, Bard, Codex or any other AI system unless explicitly stated in the homework assignment.

Midterm and Final Project

We will have an **in-class midterm** on March 1st.

There is no final exam. Instead, you will work on a **final project**. We will have presentations on the last day of class.

Feedback and Questions

You can submit anonymous feedback or anonymous questions through the **Anonymous Question Form**.

Questions submitted using the form will be answered in the **Q&A document**. Check it regularly for help with assignments!

If you are submitting feedback about the course rather than a question for the Q&A document, just say that in the form.

First Assignment

- HW 1 is due **this Friday**
- Please read Jordan (2019) for Friday