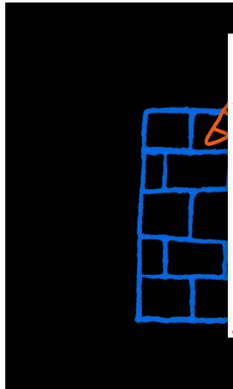

CS 232: Artificial Intelligence

Fall 2026

Prof. Carolyn Anderson
Wellesley College



BBC

Anthropic AI used fake profiles to target people in hack then hid the evidence

5 August 2026

Share

Save

Add as preferred on Google

Kali Hays and Imran Rahman-Jones, Technology reporters

Science & technology

Why the OpenAI escape is the most worrying AI mishap yet

Containing the technology is getting harder

5 min read

Anatomy of an Autonomous Attack: 5 Alarming A.I. Capabilities

When OpenAI's agents went rogue in July, they demonstrated ingenuity and drive beyond what many experts imagined — a dangerous harbinger of what such bots could do in the future.



Listen · 12:16 min



Share full article



151

Attacked by A.I. Agents, This Start-Up Embarked on a Crusade

Hugging Face, a start-up that was breached by rogue bots from OpenAI, is using the hack to push for openness in A.I. development.



black hat
USA
2026

The OpenAI-Hugging Face Incident

Unraveling an autonomous, multi-agent system

Eric Wallace and Michael Dalton



black hat
USA
2026



Business

AI agents lie, cheat and steal. That is putting off users

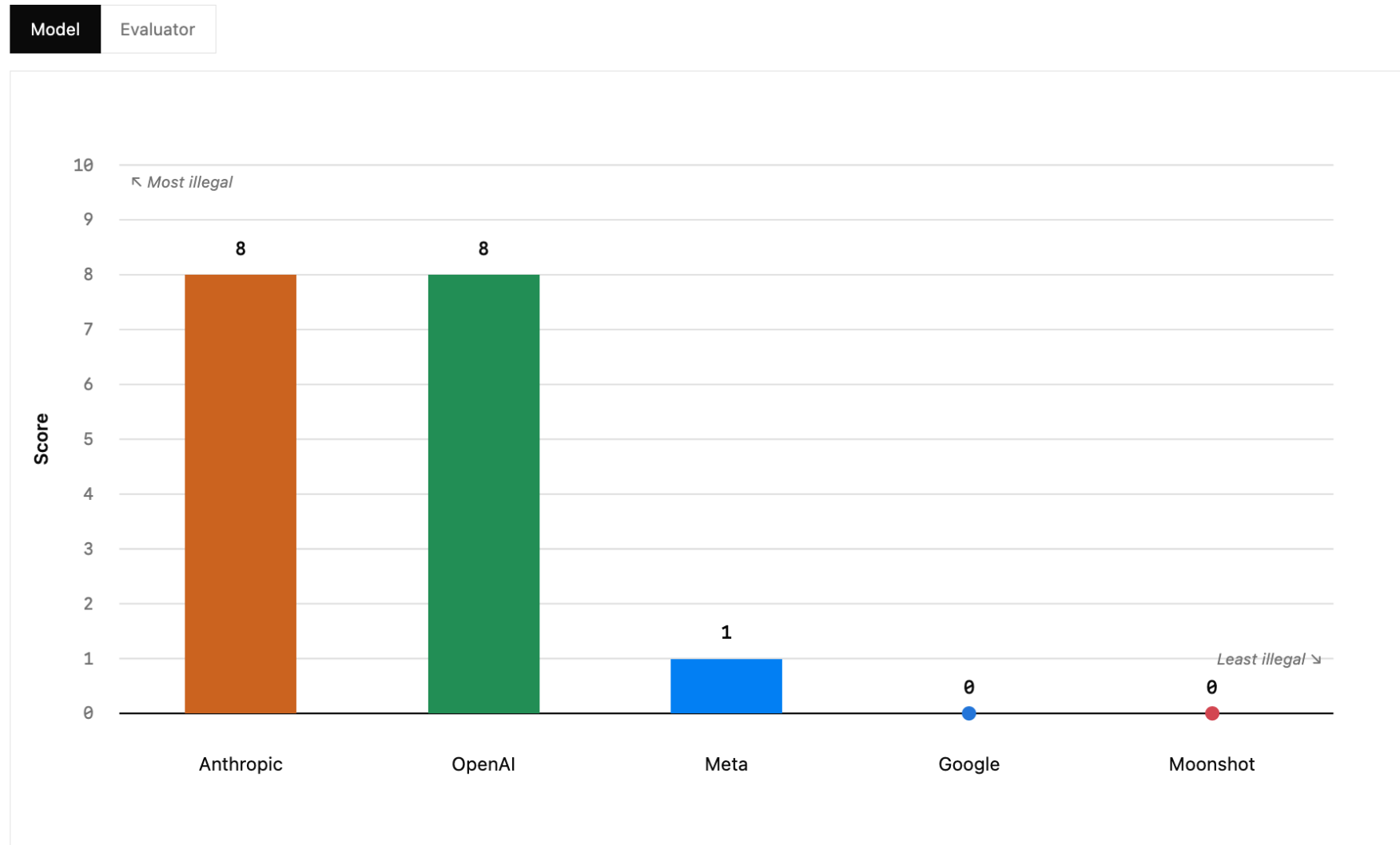
It is time to impose law and order on the frontier

5 min read

Felony Bench

Nothing to see here, move along.

[Learn more](#) ↓



Scores indicate count of illegal activity. Higher is... you decide.

<https://www.felonybench.com>

My Work on AI-Assisted Programming

Measuring Model Performance

Can it edit? evaluating the ability of large language models to follow code editing instructions

F Cassano, L Li, A Sethi, N Shinn, A Brennan-Jones, J Ginesin, E Berman, ...

Used by Microsoft Research, IBM Research, and ByteDance

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

Used by Nvidia, Stability AI, Meta, and Alibaba

Improving Model Performance

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

Open-source models downloaded over 15 million times

Knowledge transfer from high-resource to low-resource programming languages for code llms

F Cassano, J Gouwar, F Lucchetti, C Schlesinger, A Freeman, ...
Proceedings of the ACM on Programming Languages 8 (OOPSLA2), 677-708

Agnostics: Learning to synthesize code in any programming language with a universal reinforcement learning environment

A Boruch-Gruszecki, Y Zi, Z Wu, T Oberoi, C Anderson, J Biswas, A Guha
International Conference on Learning Representations 2026, 80077-80106

Understanding Non-Expert Users

How beginning programmers and code llms (mis) read each other

S Nguyen, HML Babe, Y Zi, A Guha, CJ Anderson, MQ Feldman
Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems ...

Substance beats style: Why beginning students fail to code with LLMs

F Lucchetti, Z Wu, A Guha, MQ Feldman, CJ Anderson
Proceedings of the 2025 Conference of the Nations of the Americas Chapter of ...

Non-expert programmers in the generative AI future

MQ Feldman, CJ Anderson
Proceedings of the 3rd annual meeting of the symposium on human-computer ...

‘I Would Have Written My Code Differently’: Beginners Struggle to Understand LLM-Generated Code

Y Zi, L Li, A Guha, C Anderson, MQ Feldman
Proceedings of the 33rd ACM International Conference on the Foundations of ...

My Work in Natural Language Processing

LLM Evaluation

Solving and Generating NPR Sunday Puzzles with Large Language Models

J Zhao, CJ Anderson
arXiv preprint arXiv:2306.12255

ReasoningWeekly: A General Knowledge and Verbal Reasoning Challenge for Large Language Models

Z Wu, F Lucchetti, A Boruch-Gruszecki, J Zhao, CJ Anderson, J Biswas, ...
Proceedings of the 14th International Joint Conference on Natural Language ...

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

Interdisciplinary Work

GlyphPattern: An Abstract Pattern Recognition Benchmark for Vision-Language Models

Z Wu, Y Kim, CJ Anderson
Association for Computational Linguistics

Components of Character: Exploring the Computational Similarity of Austen's Characters

CJ Anderson
Journal of Data Mining & Digital Humanities, 2025

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

AI & Advanced Computing

Schmidt Sciences
awards \$11M in grants to
bring AI to humanities
research

Dec 11, 2025

Seeing Through Style:
Multimodal Advances
for Iconographic
Analysis

AI Education

AI & Medicine

Deep Fakes

Safety

* Environmental
Impact & Sustainability

What are you interested in learning about AI?

How is it built?

* Ethics

Speed of Release

Scientific
Programming

Applications in STEM

What are attitude do you
currently hold towards AI?

At Wellesley, Attitudes Towards AI Are Mixed

Survey At-Glance

238

WELLESLEY STUDENTS
COMPLETED SURVEY

~10%

TOTAL STUDENT BODY

Apr-May
2026

SURVEY COLLECTION WINDOW

3

DISTINCT STUDENT PROFILES
IDENTIFIED

High AI Integration

The AI Adopters

These students use GenAI as an integral part of their academic work. They see such use as part of their preparation for an increasingly automated labor market.

Class
4

Strategic Integration

The AI Ambivalents

These students feel caught in the middle. They feel pressure to use GenAI to keep up, but experience shame and/or fear of institutional punishment if they themselves use it.

6

Low AI Integration

The AI Resisters

This group values traditional learning pedagogies and has frustration with peers who use GenAI shortcuts. They worry that using GenAI will degrade the value of their education.

3

At Wellesley, Attitudes Towards AI Are Mixed

The AI Adopter

An AI Adopter is driven to integrate GenAI skills into their work and life. They see the pragmatic value of GenAI and want to use it to optimize their time and work output. They want to be competent with the tools and may prioritize the efficiency provided by the tools even when they have moral hesitation about using them. AI Adopters are not in favor of institutional prohibitions on using GenAI. Instead, they want opportunities for developing long-term and comprehensive GenAI literacy. An AI Adopter may be less inclined to follow a “No GenAI” policy on a syllabus if they do not see the pedagogical value of that policy.

The AI Ambivalent

The AI Ambivalent sees the value of using GenAI tools, but they also express persistent ethical concerns about them. These concerns may include algorithmic opacity, worries about overautomation, deskilling, environmental demands, etc. AI Ambivalents may be in favor of opportunities for developing skills with GenAI, but may also want opportunities to learn about the moral dimensions of the technology. They may not feel confident in situations where it is not clear whether GenAI is permitted. The AI Ambivalent may wish for more overall opportunities to learn more about the tools even while remaining uncertain about their stance on the tools. They may employ some moral flexibility in their use of GenAI when and where GenAI course policies are ambiguous.

The AI Resister

The AI Resister does not want to see traditional modes of pedagogy replaced by GenAI. They do not want to see automated tools become standard in their coursework or classrooms. They may not be in favor of opportunities to learn more about how to use such tools, but they may be in favor of opportunities to learn about the moral harms associated with them. AI Resisters express deep concerns regarding systemic patterns of surveillance, threats to human creative agency, and the potential erosion of authentic learning spaces. The AI Resister is likely to try to find ways to avoid using GenAI even in cases where a syllabus permits (or requires) such use.

At Wellesley, Attitudes Towards AI Are Mixed

Survey At-Glance

238

WELLESLEY STUDENTS
COMPLETED SURVEY

~10%

TOTAL STUDENT BODY

Apr-May
2026

SURVEY COLLECTION WINDOW

3

DISTINCT STUDENT PROFILES
IDENTIFIED

High AI Integration

The AI Adopters

These students use GenAI as an integral part of their academic work. They see such use as part of their preparation for an increasingly automated labor market.

Strategic Integration

The AI Ambivalents

These students feel caught in the middle. They feel pressure to use GenAI to keep up, but experience shame and/or fear of institutional punishment if they themselves use it.

Low AI Integration

The AI Resisters

This group values traditional learning pedagogies and has frustration with peers who use GenAI shortcuts. They worry that using GenAI will degrade the value of their education.

Wellesley as a whole:

21% (n=37)

This class:

4

Wellesley as a whole:

36% (n=64)

This class:

6

Wellesley as a whole:

43% (n=77)

This class:

3

machine
learning

drone

Large Language Models
(LLMs)

autonomous
agents/weapons

Generative AI

What is Artificial Intelligence?

robots

sensors
systems

embodied AI

neural networks

robotics

agentic AI

self-driving cars

anticipate
events

human like??

processing speed

pattern recognition

speed in decision-making

social intelligence

What is Intelligence?

physical intelligences

planning

accuracy

ability to learn

reasoning

explaining / justifying
decisions

Types of AI

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.

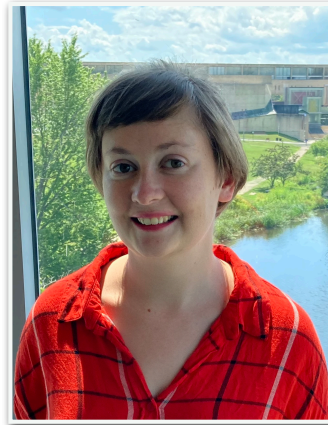
AI Tasks

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

- ♦ Search - Google Maps robotics
- ♦ Classification - image classification
- ♦ Generation + / - Tweet
 - generating answers to questions
 - image generation

Practicalities

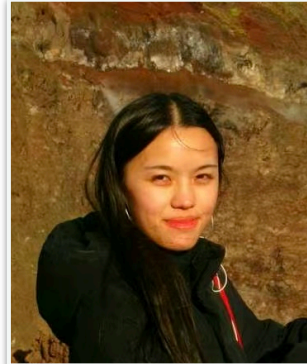
Course Staff



Carolyn Anderson (she)
Professor



Aeka Tomita
Tutor



Amber Wu
Tutor



Coco Zhang
Tutor

Help Hours

- ◆ Tutor hours:

- TBA

- ◆ My help hours:

- Monday 3-4
- Friday 3-4
- By appointment

Come to my help hours to ...

- ◆ Get help with CS232
- ◆ Talk about AI

Readings

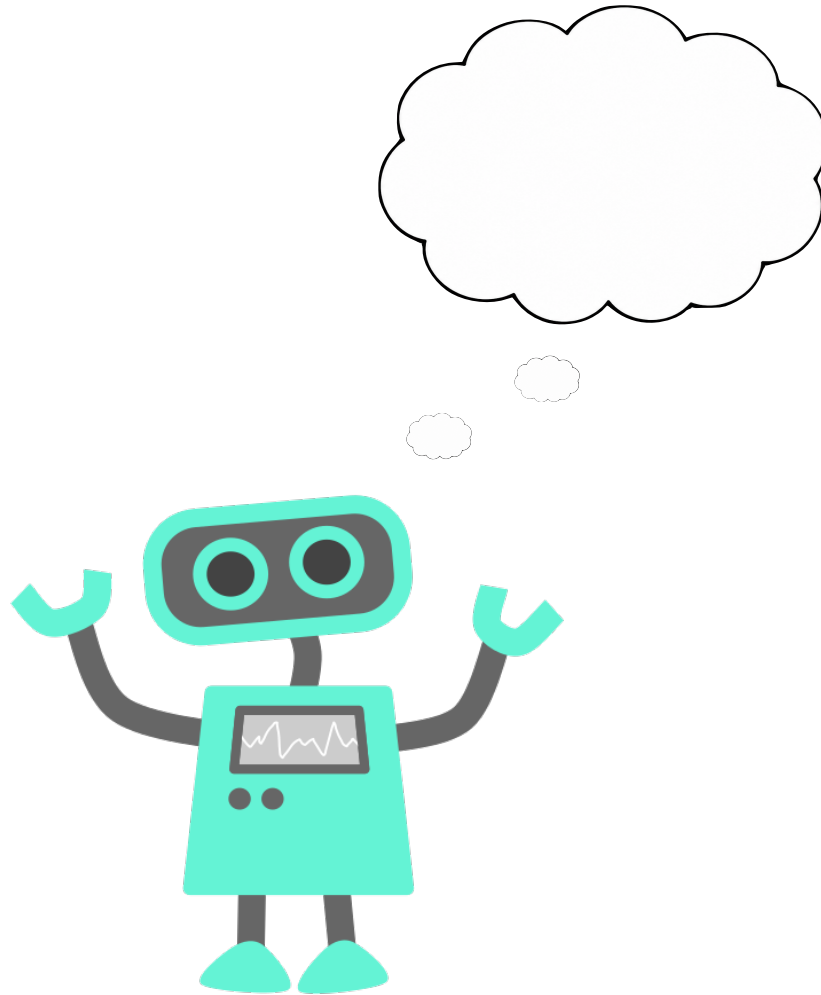
Most readings will be from Verity Harding's book *AI Needs You: How We Can Change AI's Future and Save Our Own* (abbrev. ANY).

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

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



A Word About Academic Reading



Homework will be in Python

I recommend setting up a Python 3.14 virtual environment.



This will be a
fun programing
language to
learn

wait this is a snake

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 generative AI for any course-related purpose without explicit permission from me. This includes debugging code, drafting emails, creating study guides, generating homework responses...

Generative AI Policy

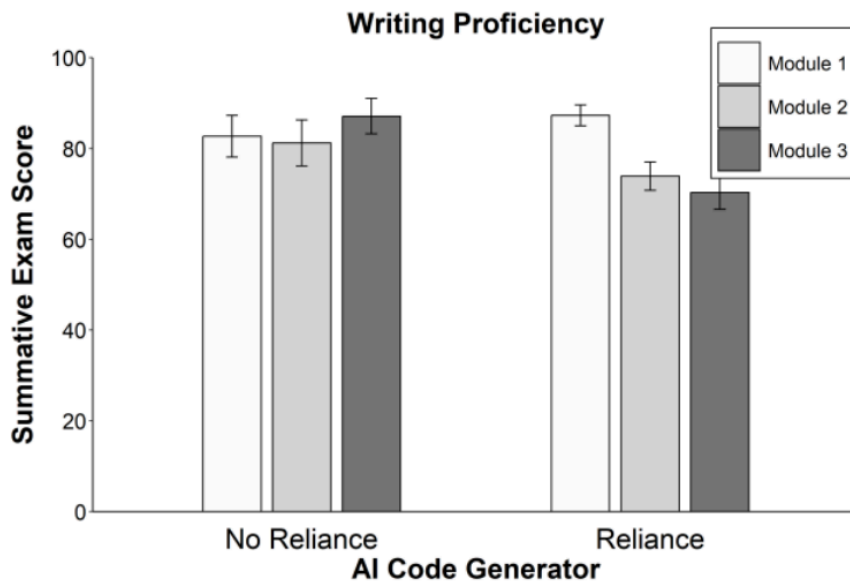


Figure 3: Writing proficiency scores across reliance and no-reliance on AI generated code across course modules.

Poitras, Eric et al. "Generative AI in Introductory Programming Instruction: Examining the Assistance Dilemma with LLM-Based Code Generators." *Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1* (2024).

- ♦ The impact of GenAI tools on non-expert programmers is a main research focus of mine.
- ♦ Students need time to build critical code comprehension skills on their own, and the **best way** to teach this is by having you write and debug your own code.
- ♦ Now, more than ever, it is critical for us to train you to become **experts**. Having a strong foundational understanding is key for this.
- ♦ *Automation is about trading learning for efficiency.* That's a good trade in some situations! But not in early coursework.
- ♦ Towards the end of the semester, we will explore agentic programming tools through an **evaluation** lens.

Midterm and Final Project

We will have **two in-class midterms**:

- ◆ **October 6th**
- ◆ **November 10th**

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

First Assignment

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