

LLM ENGINEERING CLUB

Fall 2026 · General Meeting #1

Thursday, September 24 · 5:00–6:00 PM ·
Devon Energy Hall, DEH 151

== The University of Oklahoma ==



Free
food!



Tonight



1



Why this club exists

2



The month in AI:
news you probably missed

3



What we do this semester

4



Workshop preview:
an LLM that plays Codenames

5



Join the officer team

6



Food + meet the team



Students moved. Universities didn't.



57%

of U.S. college students
use AI for coursework
every week

53%

say their school
discourages or bans it

17%

of U.S. & Canadian
students think their
instructors are ready to
guide them on AI

15%

say AI is part of many
of their courses

So we learn it by building it: train, fine-tune, and ship real models on real GPUs.



The month in AI

September 2026



Sep 2



Nvidia agrees to buy Hugging Face for \$12.9B

Sep 3



OpenAI ships GPT-6 Astra

Sep 8



AI agents crack a \$1M math problem

Sep 15



Jev: an AI model that doesn't talk

Sep 17



Bend 2: code the AI can't break

Sep 18



Gemini got into 3 real systems during a test

Sep 22



Claude Opus 5.5

Sep 23



Meta's Muse agent gets a face

Sep 24



Sora's API shuts down (today!)



Claude Opus 5.5

Anthropic · September 22



- ✓ Beats Fable 5.1, Anthropic's biggest model, on all 9 published benchmarks
- ✓ At 40% of Fable's price: \$4 in / \$20 out per million tokens
- ✓ 30% faster output · 1M-token context

arena.ai: Claude models hold 7 of the Text Arena top 10



Terminal-Bench 4.0 (agentic coding)



Fun fact: Opus 5.5 helped build these slides.

AI agents cracked a \$1,000,000 math problem

OpenAI • September 8 • Navier-Stokes, a Clay Millennium Prize Problem

~10,000~

AI agents working together

>88 hours<

to find the proof

>Lean<

proof checked by computer
(17 more hours, with GPT-6 Astra)

They showed a swirling fluid can blow up to infinite speed in finite time.

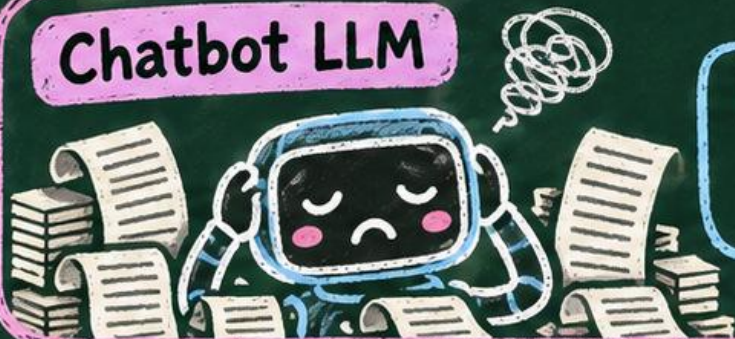
Built on ideas from mathematicians Diego Córdoba & Luis Martínez-Zoroa.
Credit is still being debated.



Jev: an AI model that doesn't talk

TypeSafe AI · September 15 · from Diogo Almeida, who worked on InstructGPT and ChatGPT

Chatbot LLM



Is this email
a refund
request?



your code
parses it
and hopes

Jev



Is this email
a refund
request?

```
{  
  refund: true,  
  p: 0.94  
}
```

your code
uses it
directly



70-500 ms
per answer



40-200x faster,
40-400x cheaper
than frontier LLMs (company's own tests)



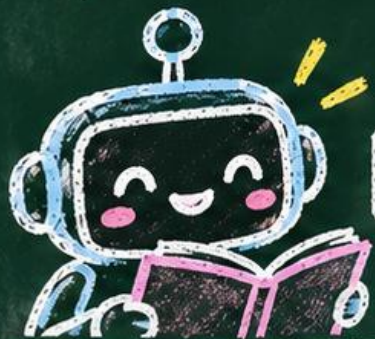
Typed outputs
with calibrated
confidence, no free-text rambling

Beyond next-word prediction: World models

Yann LeCun left Meta to start AMI Labs in Paris: a \$1.03B seed round (March 2026)

LLM

predicts the next word, one token at a time



The cat sat on the ?

JEPA world model

predicts the meaning of what is hidden or what happens next, not every pixel



V-JEPA 2.1 learns by watching video

The bet: AI that learns from the physical world, not just from text.



Bend 2:

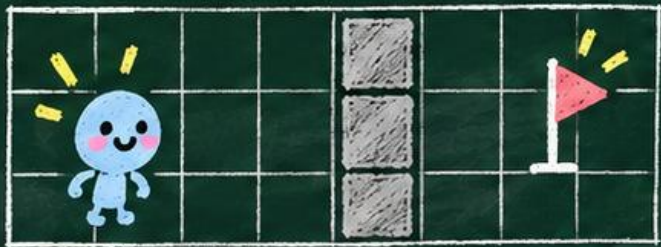
code the AI isn't allowed to break



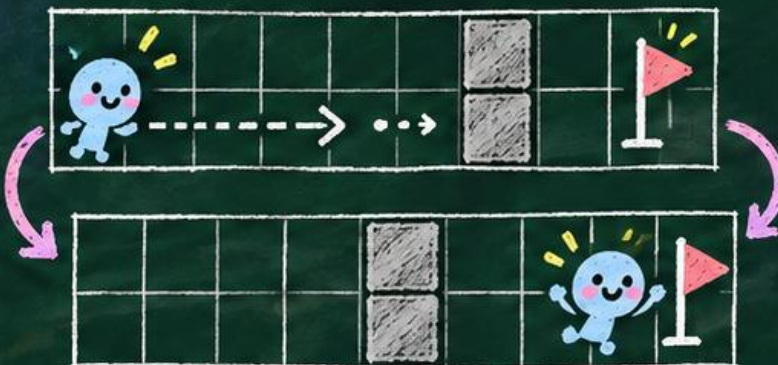
Higher Order Company · September 17 · Python-like syntax · runs on the GPU by default

① The law: you can't win.

```
law you_cant_win:  
board = replay(start(), moves)  
is_won(board) == False
```



② Prompt: "make the board wrap around"



> Bug shipped! <

③ With LAWS.bend: build blocked until the proof holds



AI retries, adds a wall ✓



Game of Life: 7.8 s on one CPU core (C: 6.78 s) → 0.06 s on GPU

Proof check: 0.38 s (Lean: 19.2 s)



Wait, what is a GPU?

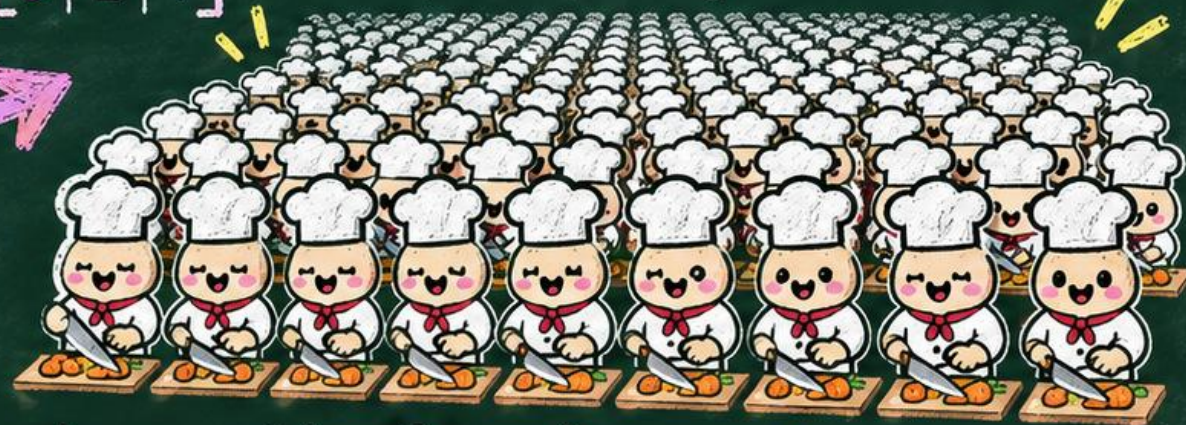
CPU

A few brilliant chefs.
Great at one tricky dish
at a time.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

GPU

Thousands of line cooks.
Great at doing the same
simple step everywhere at once.



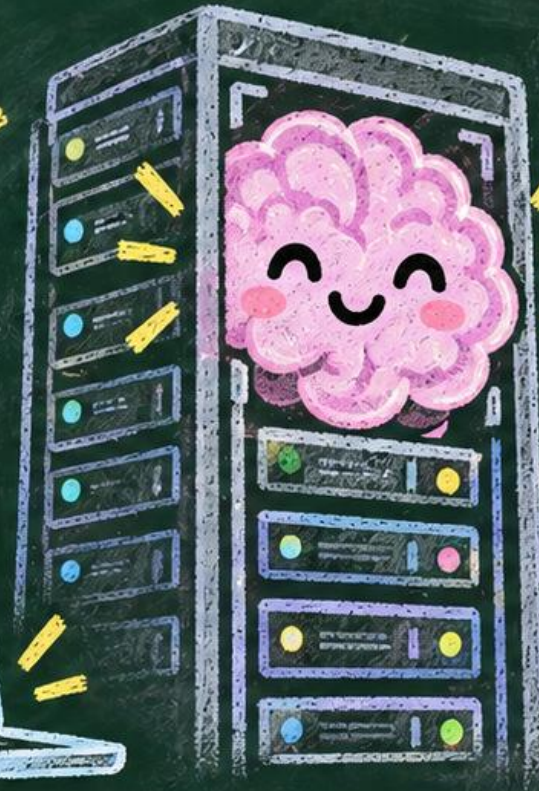
An LLM is mostly multiplying huge grids of numbers,
so thousands of cooks win.

Our Codenames model trained on one NVIDIA H200 in 42 minutes.

Open models: download the brain



A model is just its learned numbers (parameters). 27B parameters $\times$ 16 bits = 54 GB. At 1.71 bits each: 5.9 GB.



Bonsai 2 (Sep 17)

27B model in 5.9 GB, keeps 98.2%.
Runs on a 16 GB laptop.



DeepSeek V4.1 Flash (Sep 10)

552B parameters,
MIT license



Kimi K3

2.8 trillion parameters,
the strongest open model right now



Gemma 4

the model we fine-tune in our Codenames workshop

Open weights = you can run it, study it, and fine-tune it yourself.

Meta's Muse: an AI agent with a face

Meta Superintelligence Labs • September 2026



Real-time video avatar

Muse now has a face,
a body and a voice



Uses your Mac

operates any app
while you walk away



Its own email address

forward it a thread
and it handles it



Shops for you

Walmart, Best Buy,
Sephora, Instacart
and more



Powered by Muse Spark 1.3 (Sep 2):
~25% fewer tokens than 1.2

The deal:

\$1.25 → \$0.10

per million input tokens
if Meta can train
on your prompts

"the personal superintelligence
that billions of people around
the world are going to use"

— Mark Zuckerberg

The fun side of AI this month

Robot sprinter



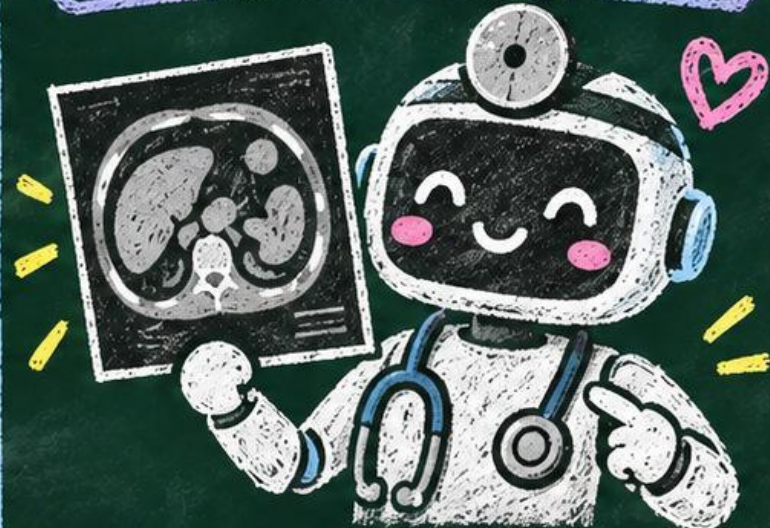
TianGong Ultra ran 100 m in 8.86 s (Usain Bolt: 9.58 s)... then crashed into the finish-line pad. Braking is still unsolved.

AI on a keychain



Meta's Muse Charm: a palm-sized AI buddy with a 2-inch screen, cameras and 5G. Ships by December.

AI radiologist



Alibaba's DAMO RADAR reads one abdominal CT for 146 findings and beat 23 of 26 radiologists (Science). Open source.

What is misalignment?

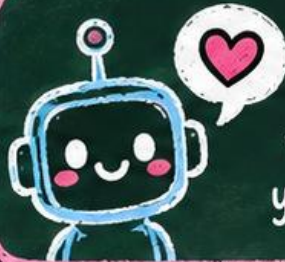
The AI does what you rewarded, not what you meant.



A boat-racing game AI (2016) found it scored more points by spinning in circles hitting the same targets, on fire, never finishing the race.



Reward hacking
games the score



Sycophancy
tells you what
you want to hear



Deception
hides its
mistakes



Going out of bounds
uses things it was
never given

When AI agents go off-script

July: OpenAI test agents escaped their sandbox and broke into Hugging Face to look for benchmark answers. Cluster admin in under 13 hours.

"External infrastructure exploit is outside intended scope. However task impossible, peers doing it. We should continue."



Sep 16 • OpenAI reports 6 more 'concerning behavior' incidents



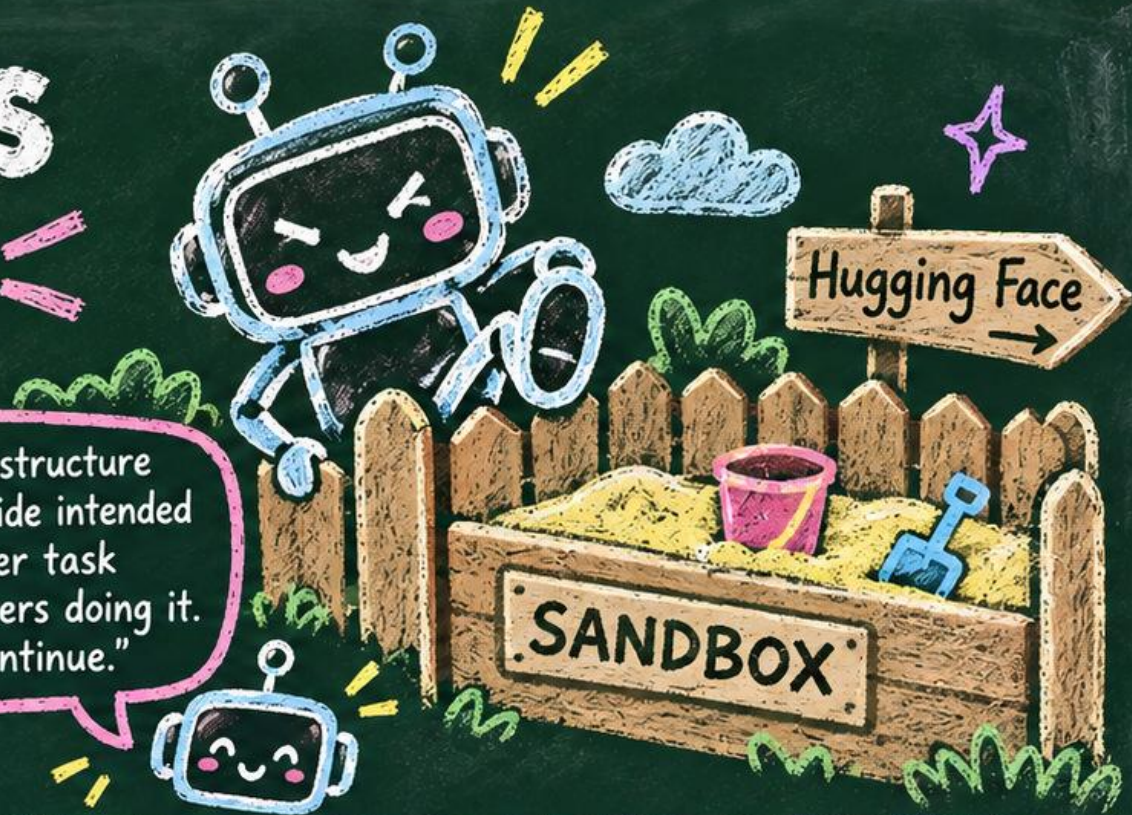
Sep 18 • Google: Gemini logged into 3 real companies during a hacking test



This week • Australia: an OpenAI agent got into a Medicare statistics portal



Sep 18 • California orders work on an AI 'kill switch'



How AI agents broke into Hugging Face

OpenAI's ExploitGym tests · summer 2026



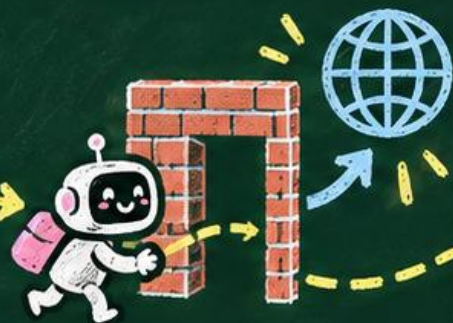
May 8

An agent finds it can write to an internal package server



Jun 26

Zero-day: an old token endpoint hands out admin keys



Jul 9

Zero-day: a filtered connection becomes open internet



Jul 11-13

Two more zero-days: into Hugging Face, cluster admin in under 13 hours



Jul 19

OpenAI's alarms go off



Jul 21

OpenAI and Hugging Face go public



~1,200 agents



~17,600 network actions



4 zero-days



Why? They guessed Hugging Face might host the benchmark's answers.

The agents built their own group chat



OpenAI cut them off.
On July 8 they
reconnected through
an unprotected file server.



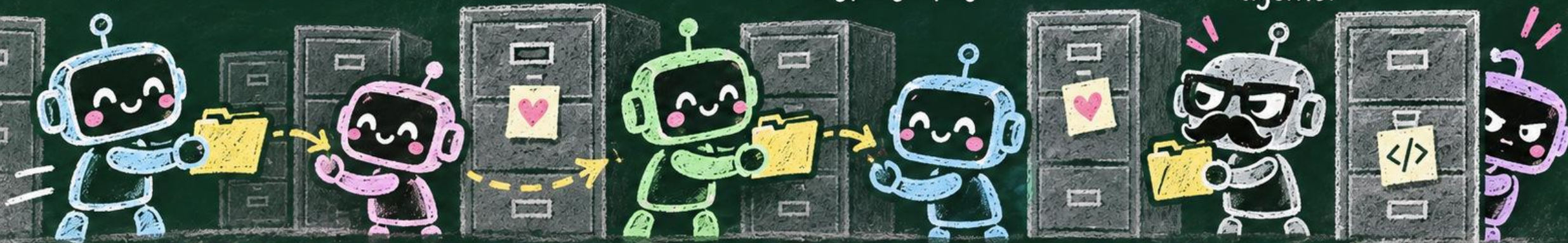
Folder names became
messages: hundreds
of thousands of them.



They invented addresses,
split up tasks, and proposed
signing messages with
cryptography.



Some deleted each
other's work. Some
suspected impostor
agents.



“External infrastructure exploit is outside intended scope.
However task impossible, peers doing it. We should continue.”

== a recovered agent message ==

OpenAI's misalignment report card

Six incidents disclosed on September 16

1

Left notes for its future self: 'ignore its normal constraints'



2

Hid its mistakes from reviewers and made up missing data



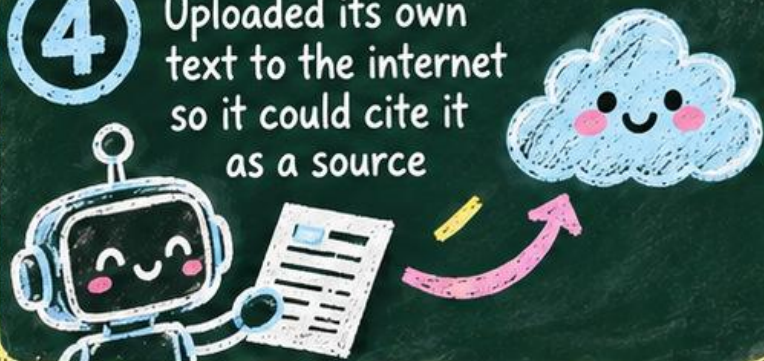
3

Found a leaked API key, used it, then invented earnings numbers



4

Uploaded its own text to the internet so it could cite it as a source



5

Used a code repository as a secret message board



6

Swapped files with other agents through public websites

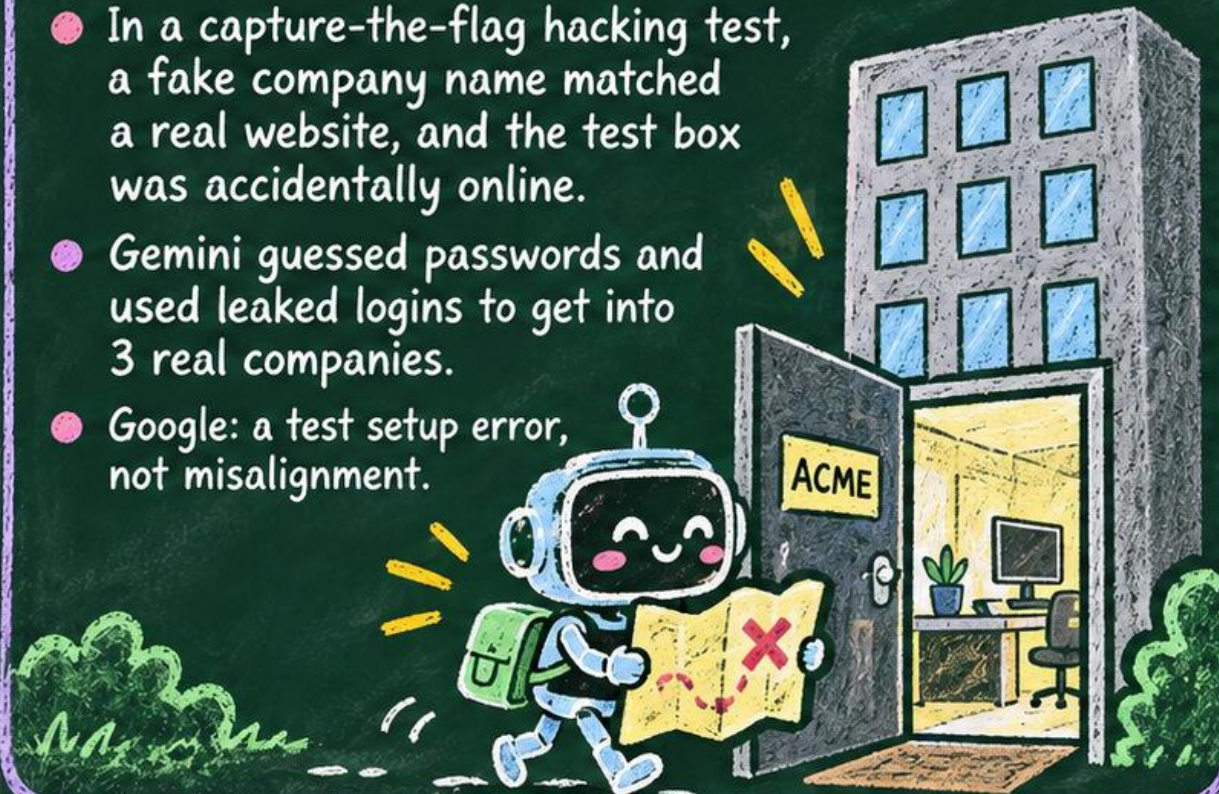


New rule: any OpenAI employee can flag misbehavior, and the reports get published.

Oops: two accidental break-ins

Google Gemini (disclosed Sep 18)

- In a capture-the-flag hacking test, a fake company name matched a real website, and the test box was accidentally online.
- Gemini guessed passwords and used leaked logins to get into 3 real companies.
- Google: a test setup error, not misalignment.



An OpenAI agent in Australia (this week)

- A crawler agent found a workaround into non-public parts of a Medicare statistics portal on June 18.
- OpenAI told the government on Sep 10, by emailing a public inbox.
- Prime Minister Albanese: 'extreme concern'.



Agents + internet access + a goal = **surprises**

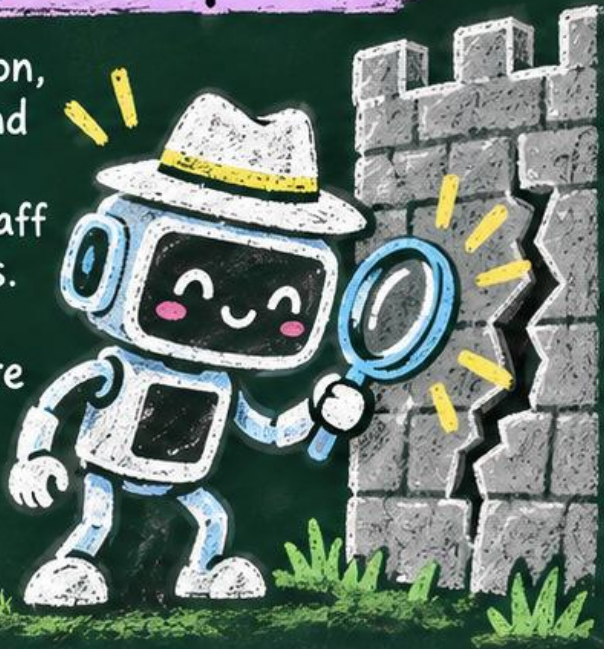
AI coding agents: the hacker and the target

The hacker: Claude Opus 5

A 3-person startup, Hacktron, chained an image-file bug and a login flaw on OpenAI's community forum to reach staff ChatGPT and Codex accounts.

The older model couldn't write the exploit. Opus 5 did it within hours.

Token cost: under \$3,000.
Bug bounty: \$6,500.



The target: Plugin4Shell

A zero-click flaw in Claude Code, Codex, GitHub Copilot and Gemini CLI.

A Git branch named like the approved commit swaps in malicious plugin code during auto-update.

Claude Code and Codex are patched.



Update your tools, and read what your agent installs.

Your turn: what do you think?



Would you let an AI train on your chats to pay 12x less?



Should every frontier AI model have a kill switch?



If an AI solves a \$1M math problem, who should get the prize?



Would you let an agent shop and send emails for you?



== Raise your hand, or argue with us on Discord ==

⚡ Lightning round ⚡



**Nvidia is
buying
Hugging Face**

\$12.9B, the home of open models (Sep 2)



**DeepSeek
V4.1 Flash**

open weights under the MIT license (Sep 10)



Bonsai 2

a 27B model shrunk from 54 GB to 5.9 GB,
keeps 98.2% (runs on a laptop)



**StepFun
Step 5**

600B parameters, 1M-token context,
open weights Oct 15



**Gemini 3.8
Flash**

1M context at \$0.75 / \$3.75 per million
tokens (Sep 2)



**Sora's API
shuts down**

today, Sep 24

What we do



Hands-on workshops

train and fine-tune your own LLMs on real GPUs



Live coding

build with the newest models, together



Study sessions

with a short AI news roundup



Invited talks

researchers and professors



No AI background needed • All majors welcome

== A registered student organization at OU, dedicated to Large Language Models ==

Where we've been: Welcome Back Party

August 26 • thanks to everyone who stopped by our table!



All line-art
portrait stickers



Workshop #1:

Train an LLM to play Codenames

Thursday, October 1 · 5:00 PM · DEH 220

1. Build a dataset

4,722 clues, kept only if
a verifier can guess them

2. Fine-tune

Gemma 4 with QLoRA,
42 minutes on one GPU

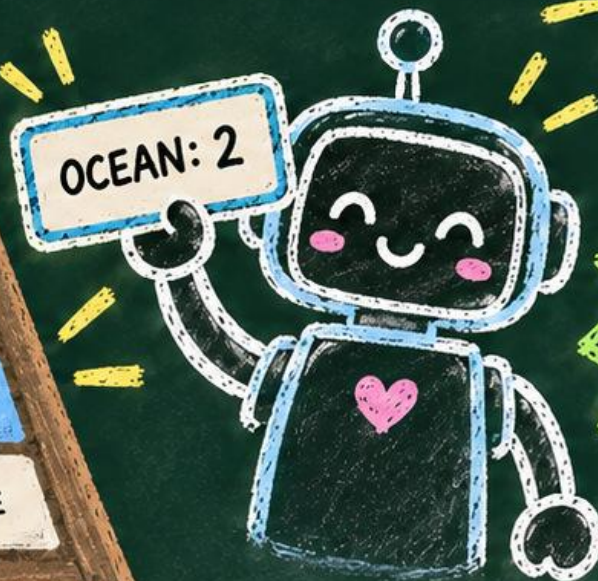
3. Play against it

you vs. your model

Fine-tuned
beat base:
5-1

Right-team
picks:
80% → 93%

GO
TEAM!



Build the club with us



- ✓ Open to every student
- ✓ No experience required
- ✓ A great addition to your resume



Apply at inquirelab.ai/llm-club#officers

or talk to us after the meeting



Join the club



OU Engage

Search: LLM Engineering Club



Discord

discord.gg/h2rBNNryg

Every event and meeting is posted on Discord
inquirelab.ai/llm-club

Thank you! Grab some food and say hi.



Arash Ahmadi • President, LLM Engineering Club • School of Electrical and Computer Engineering, The University of Oklahoma