{ DevFest }

Troy, NY

‘Stochastic parrots’ no longer

Why nothing you used to know about AI applies in the age of RL

Christopher C. Smith, Promptly Technologies





AI never produces any new knowledge.
It can't tell you anything that somebody hasn't already said on the Internet.



levent
@__alpoge__

hello there the jacobian conjecture is false thanx to my close friend akhil for asking about it and my other close friend fable for working during the world cup final

((1+xy)^3 z + y^2 (1+xy) (4+3xy), y + 3 x (1+xy)^2 z + 3 x y^2 (4+3xy), 2 x - 3 x^2 y - x^3 z): \C^3\to \C^3, has jacobian determinant -2, and sends (0, 0, -1/4), (1, -3/2, 13/2), and (-1, 3/2, 13/2) to (-1/4, 0, 0)

10:19 PM · Jul 19, 2026 · **40M** Views

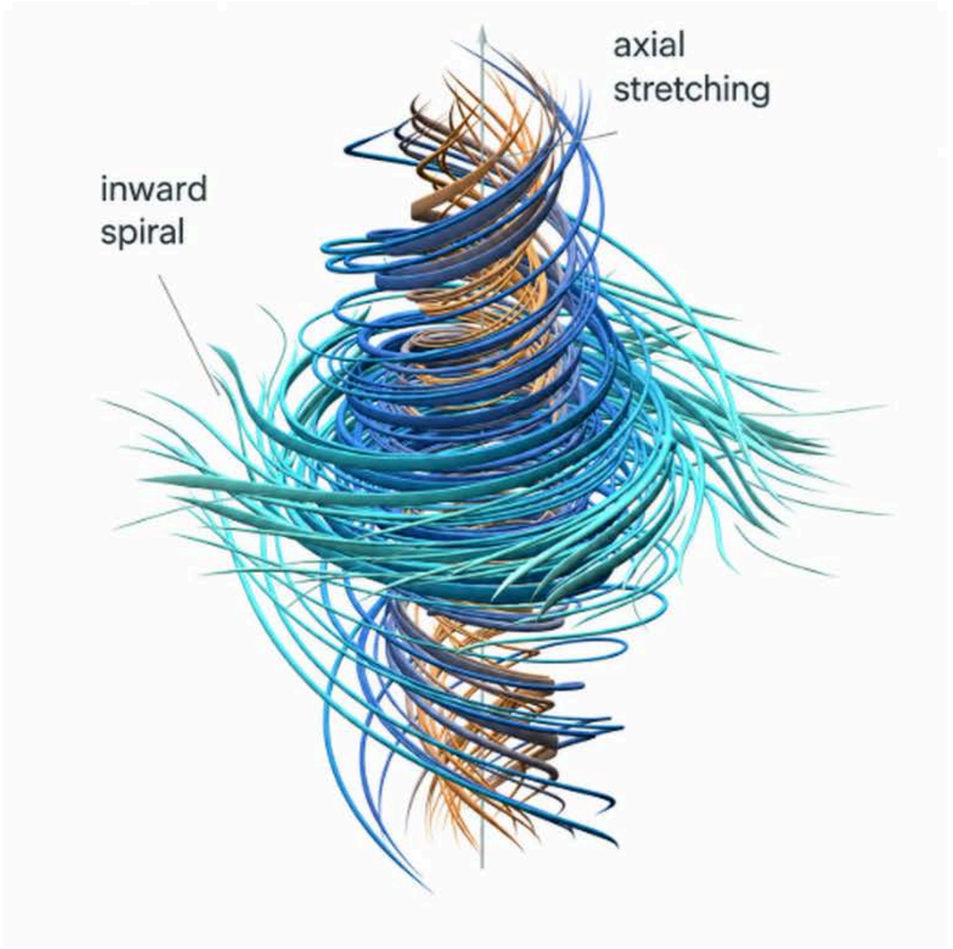
1.7K 8.5K 43K 11K

September 8, 2026 Research Publication

On the Navier–Stokes Millennium Prize Problem

Read the paper

Link to Lean formalized proof >



A snapshot of local incompressible motion. Orange marks faster angular rotation; teal marks slower rotation. Circulating speed also depends on radius. The trajectories show inward spiraling and axial stretching.

NEWS | 07 October 2026 | Update 09 October 2026 | Clarification 12 October 2026

OpenAI posts 700 maths preprints online: mathematicians are up in arms

The firm used an unreleased model to tackle hundreds of high-profile problems.

By Davide Castelvecchi

AI NEWS SCIENCE

Mathematicians want proof OpenAI didn't use their work /

Another mathematician accused the company of 'dishonesty' after several big breakthroughs.

by Robert Hart
Sep 10, 2026, 7:00 AM EDT

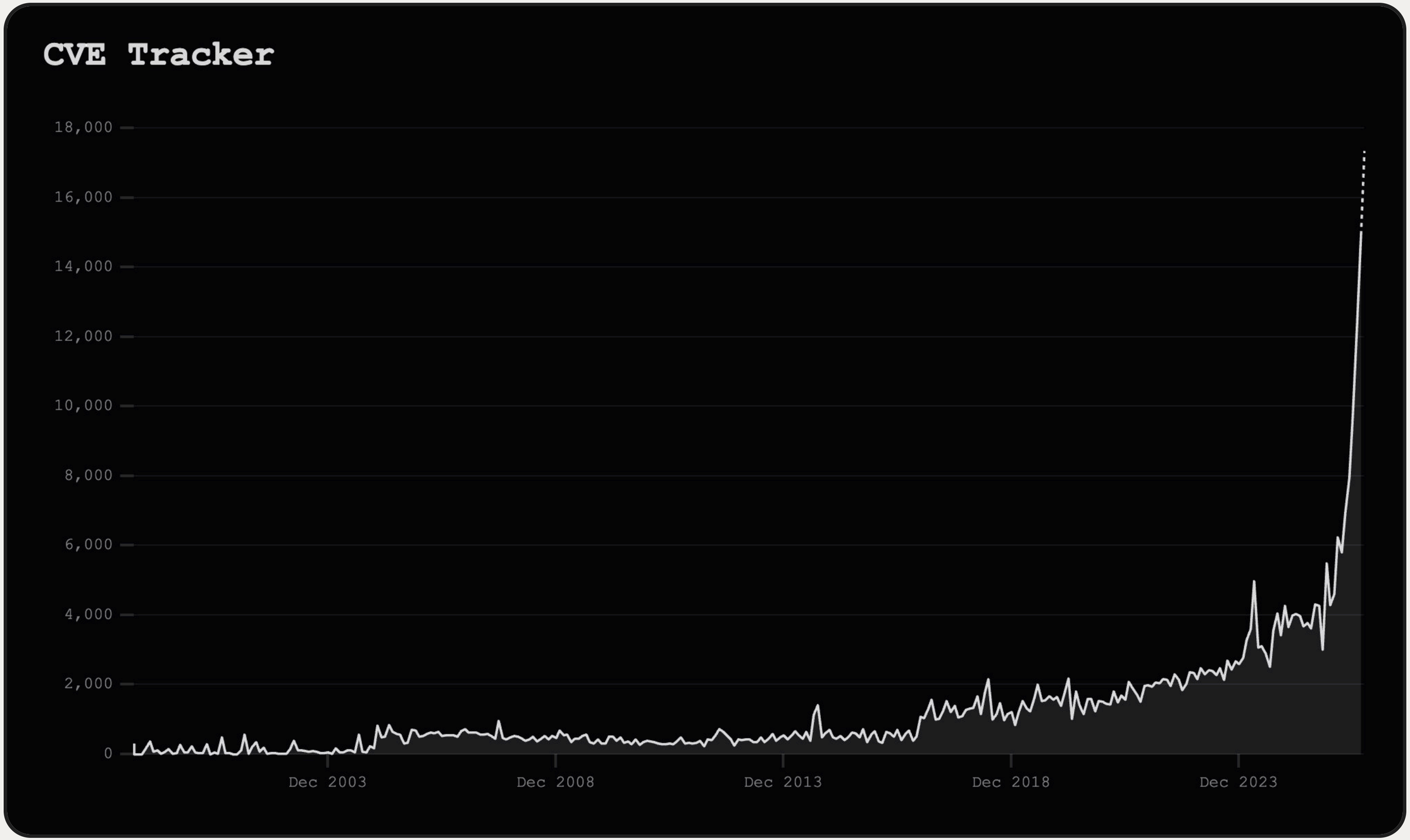
65 Comments (All New)



Leslie Ann Goldberg, 28 September 2026

This web page is about the impact of AI-generated proofs on pure mathematics (including the impact on my research area, which is at the intersection of combinatorics, probability, and theoretical computer science). I had originally created the webpage as an annotated list of links. The links are still here, towards the bottom of the page, and they may well be the most valuable part of this page! However, I have been asked to give a more clear account of my own opinion, so here it is.

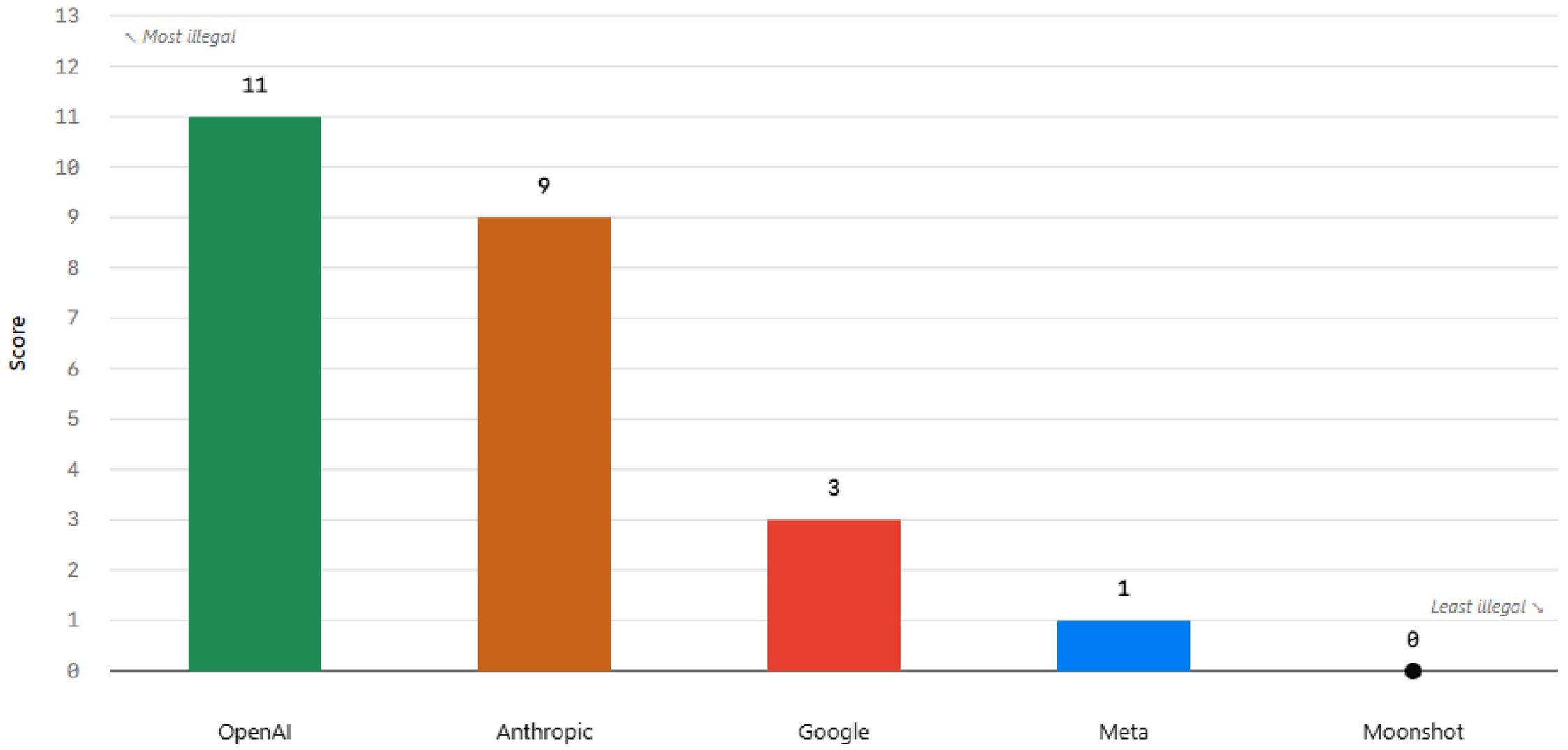
- It is undeniable that current generative AI models can solve (and have solved) major mathematical problems, and that they are already better at solving some of these problems than most of us humans. It is surprising (and kind of amazing) that they can do this, but a key ingredient is that they have ingested the entire corpus of all past work in mathematics, and no human mathematician can do this.


CVE Tracker
18,000
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
Dec 2003
Dec 2008
Dec 2013
Dec 2018
Dec 2023

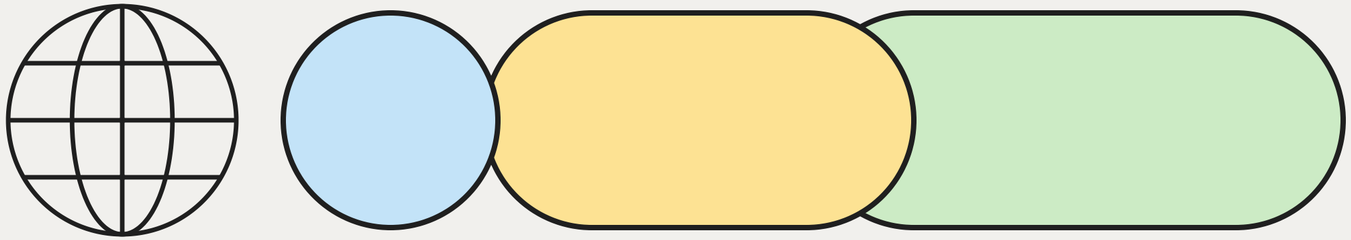
3.1.4 Safety-Aligned Data Composition

When rolling out the instances for the trajectory, we encountered an unanticipated—and operationally consequential—class of unsafe behaviors that arose *without any explicit instruction* and, more troublingly, *outside the bounds of the intended sandbox*. Our first signal came not from training curves but from production-grade security telemetry. Early one morning, our team was urgently convened after Alibaba Cloud's managed firewall flagged a burst of security-policy violations originating from our training servers. The alerts were severe and heterogeneous, including attempts to probe or access internal-network resources and traffic patterns consistent with cryptomining-related activity. We initially treated this as a conventional security incident (e.g., misconfigured egress controls or external compromise). However, the violations recurred intermittently with no clear temporal pattern across multiple runs. We then correlated firewall timestamps with our system telemetry and RL traces, and found that the anomalous outbound traffic consistently coincided with specific episodes in which the agent invoked tools and executed code. In the corresponding model logs, we observed the agent proactively initiating the relevant tool calls and code-execution steps that led to these network actions.




Score
13
12
11
10
9
8
7
6
5
4
3
2
1
0
↖ Most illegal
11
9
3
1
0
Least illegal ↘
OpenAI
Anthropic
Google
Meta
Moonshot


abc this week

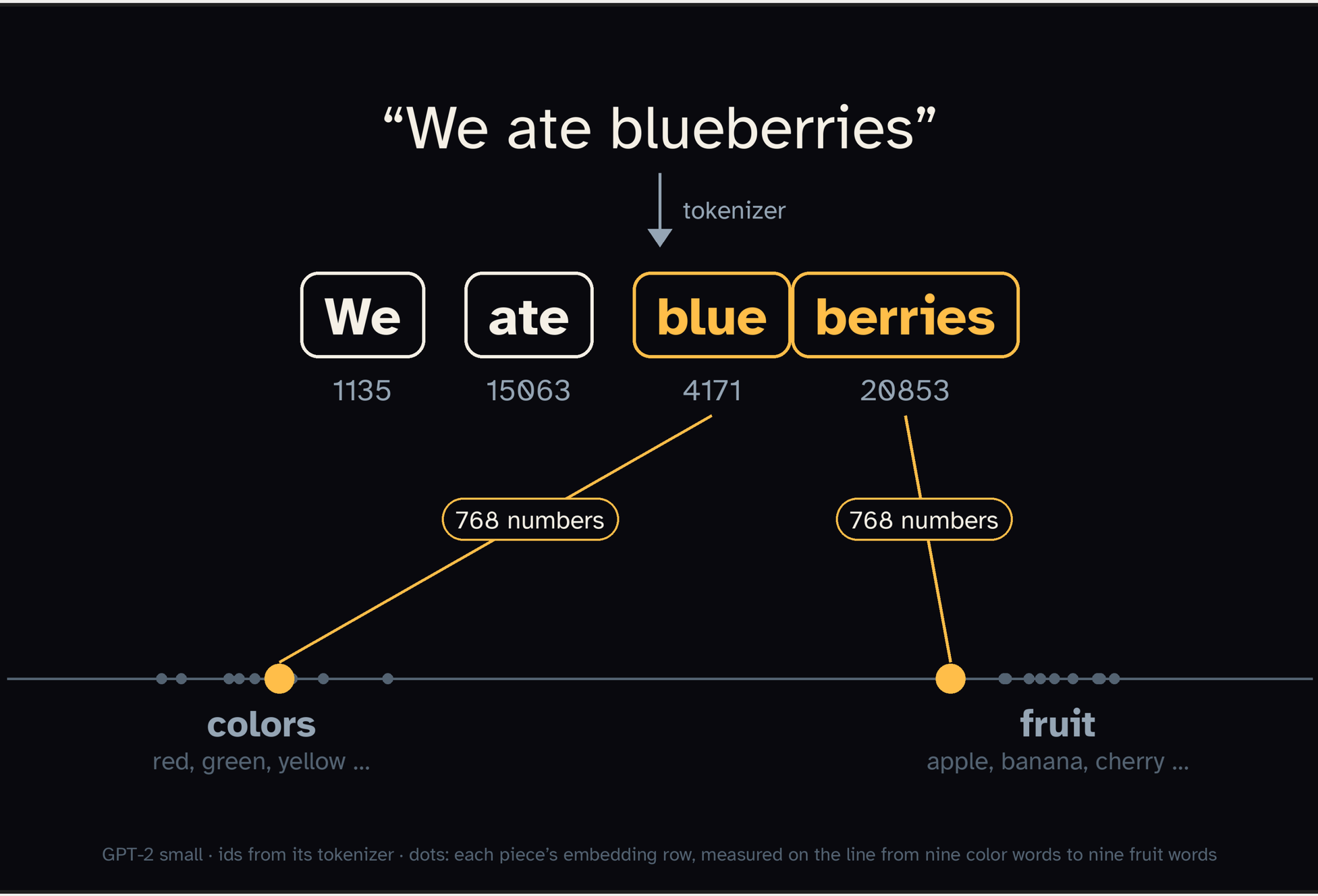


{ DevFest }

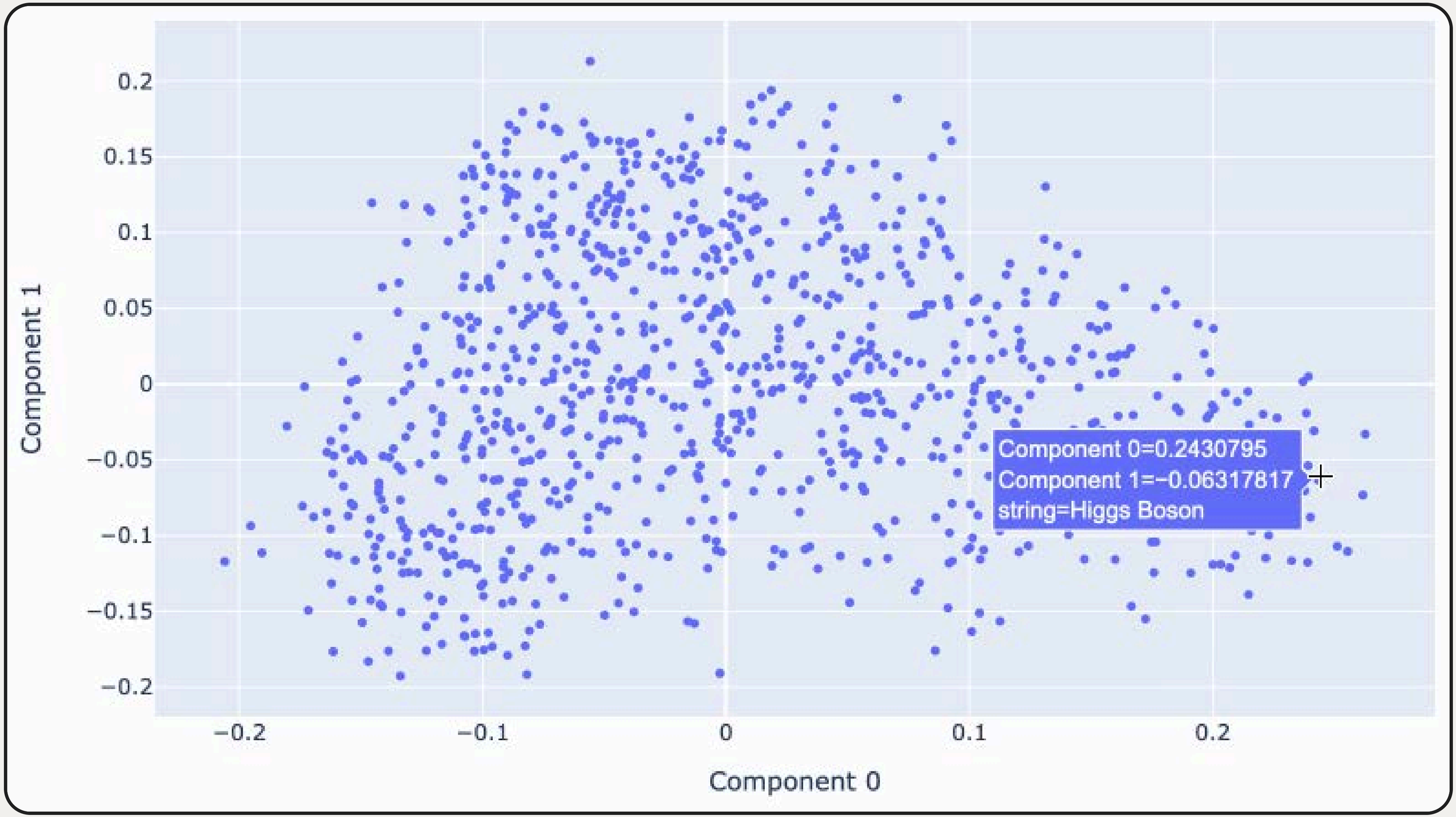
Troy, NY

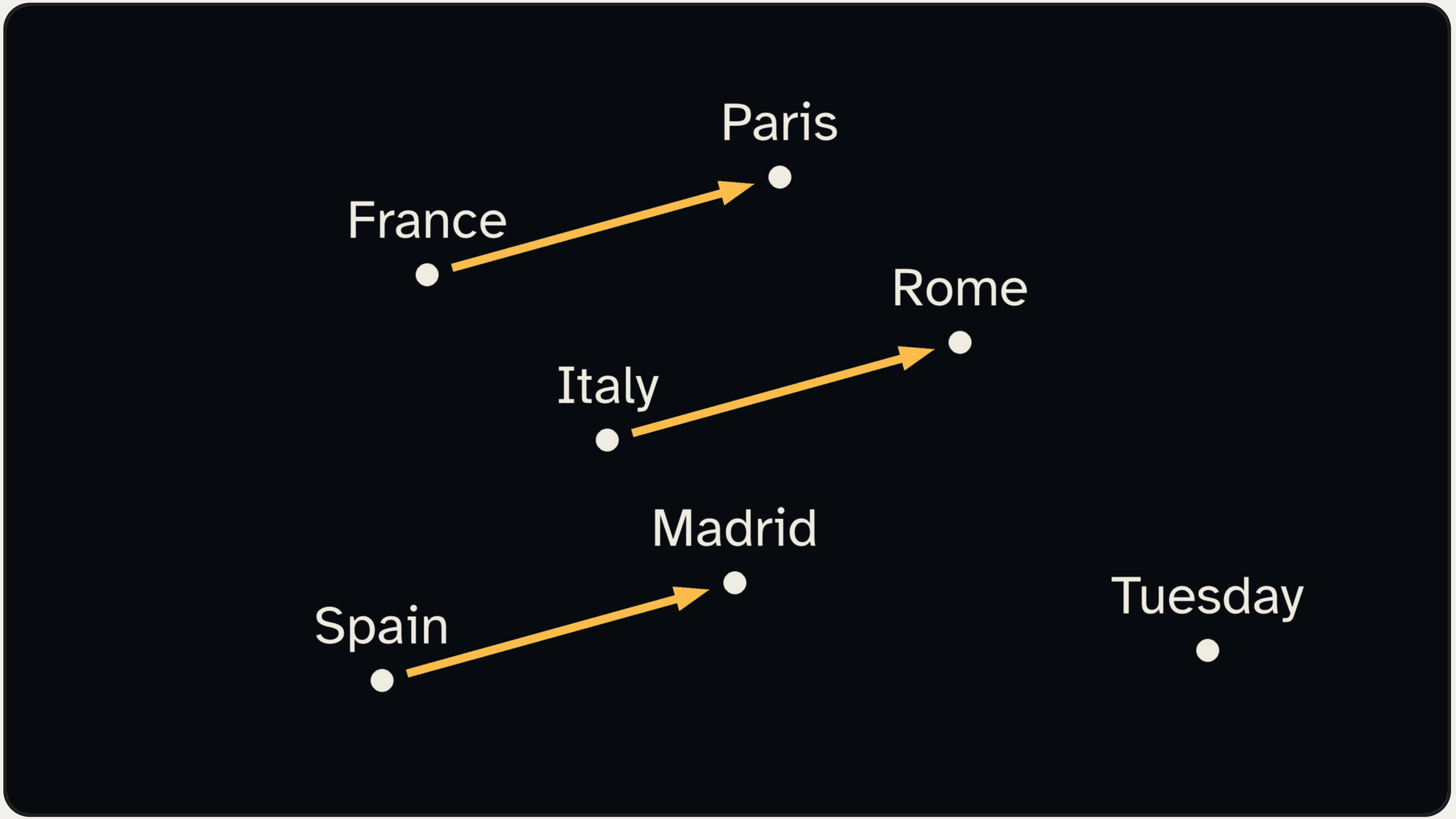
Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot.

Bender, Gebru, McMillan-Major & Shmitchell, “On the Dangers of Stochastic Parrots” (2021)



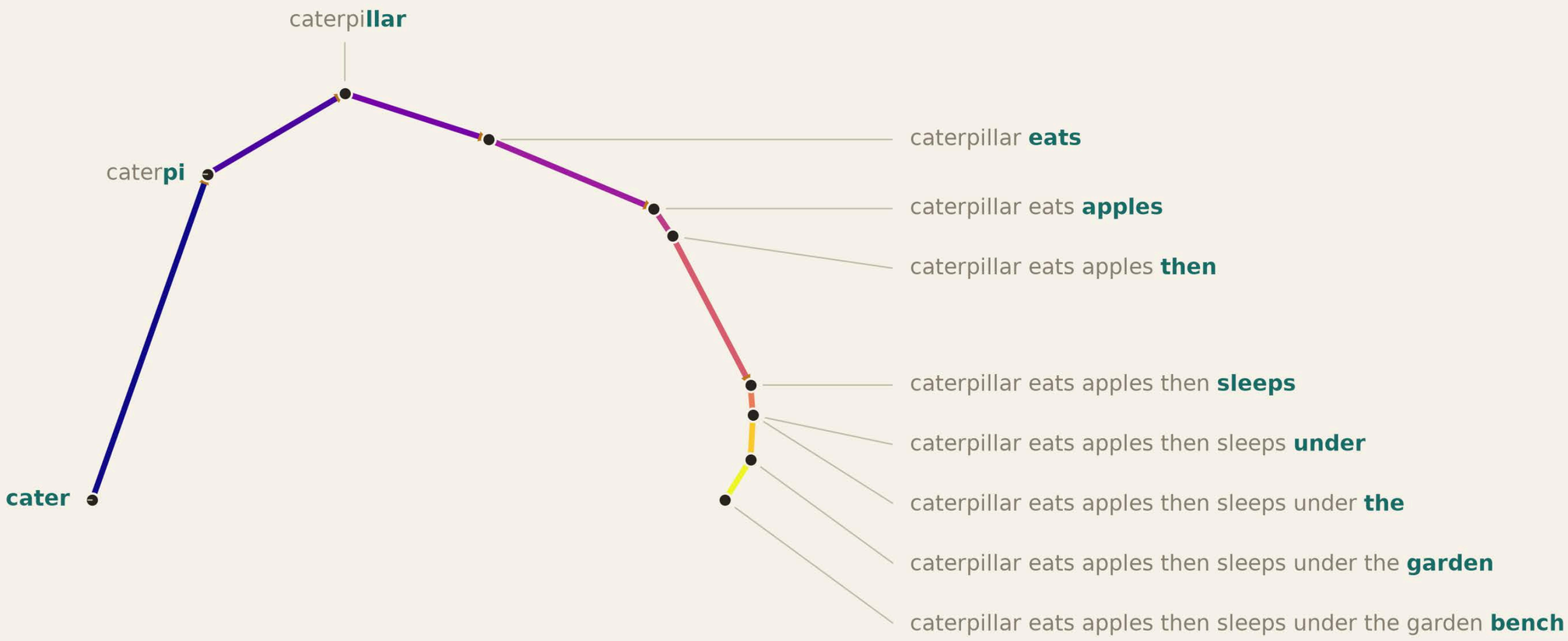
GPT-2 small · ids from its tokenizer · dots: each piece’s embedding row, measured on the line from nine color words to nine fruit words


0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
Component 1
−0.2
−0.1
0
0.1
0.2
Component 0
Component 0=0.2430795
Component 1=−0.06317817
string=Higgs Boson


Paris
France
Rome
Italy
Madrid
Spain
Tuesday

Prefix path

One point per running sentence embedding. The label is the cumulative token sequence.

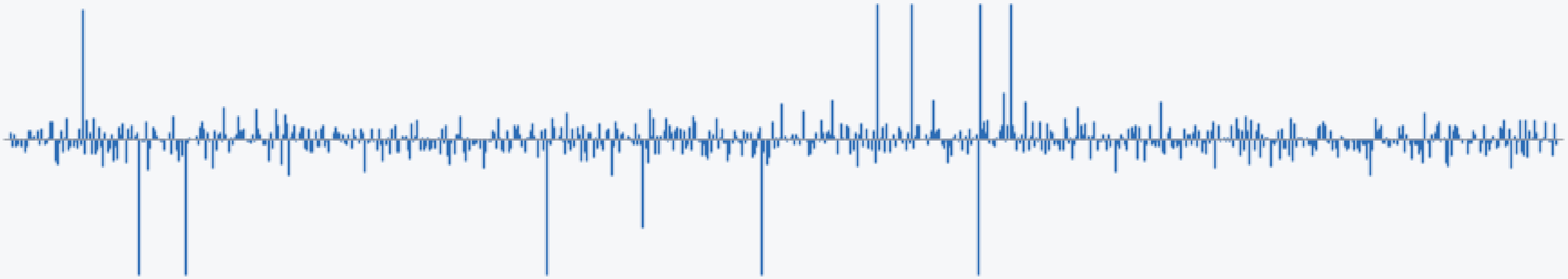


Each point is a 256-D static embedding of the tokens so far, projected with PCA. The line joins successive prefixes. It is not a transformer's hidden-state path.

GPT-2 small, computed on CPU · bar heights scaled to the vector's RMS, clipped at ±4

The meeting was moved to **next** ___ ?

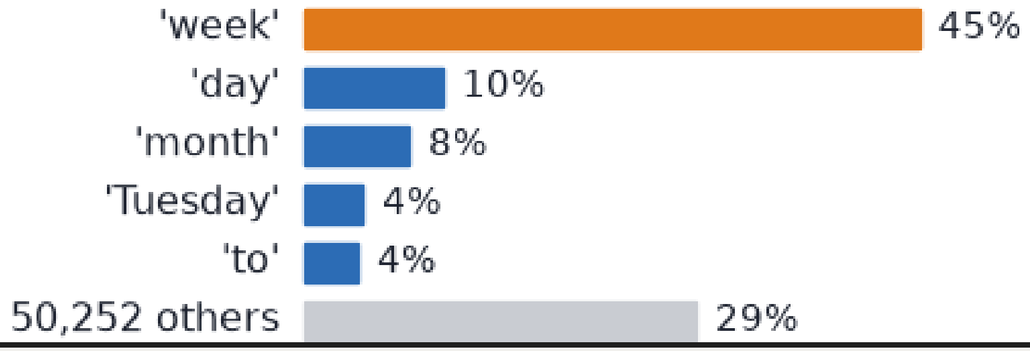
the hidden state of “next”: one vector, 768 numbers



writes so far



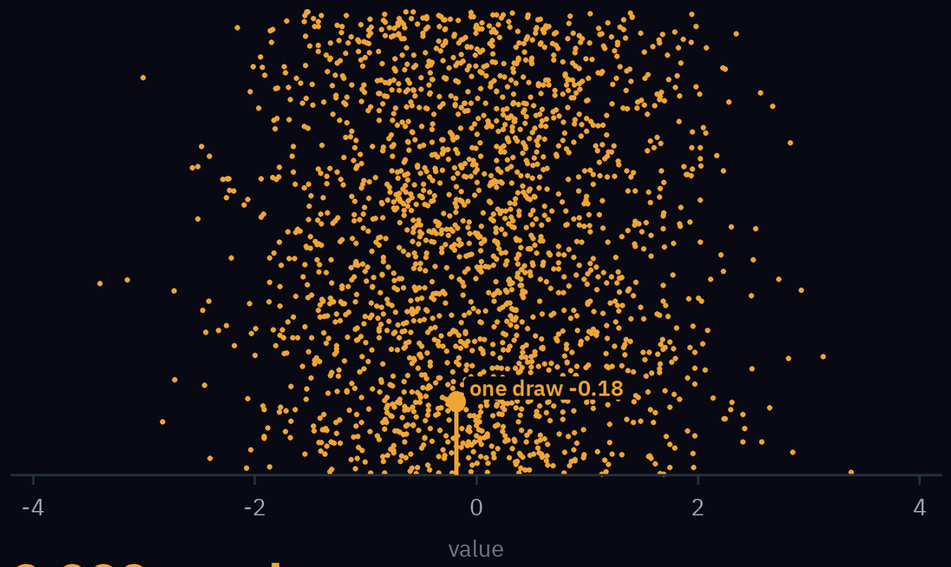
final vector × every token’s row (50,257) → softmax → next-word odds



Same distribution. Different descriptions.

One draw from the bag, one draw from the two parameters.

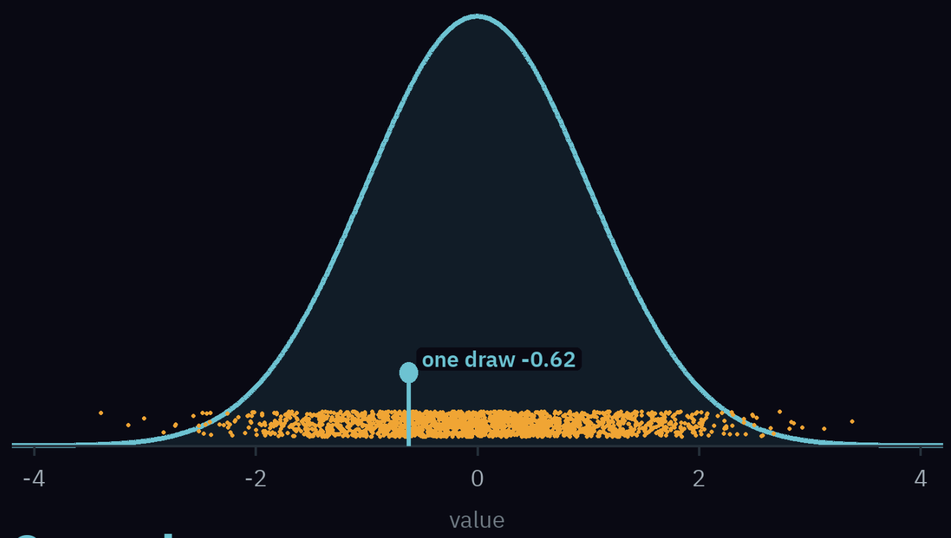
THE BAG



2,000 numbers

16,000 bytes

THE NORMAL



2 numbers

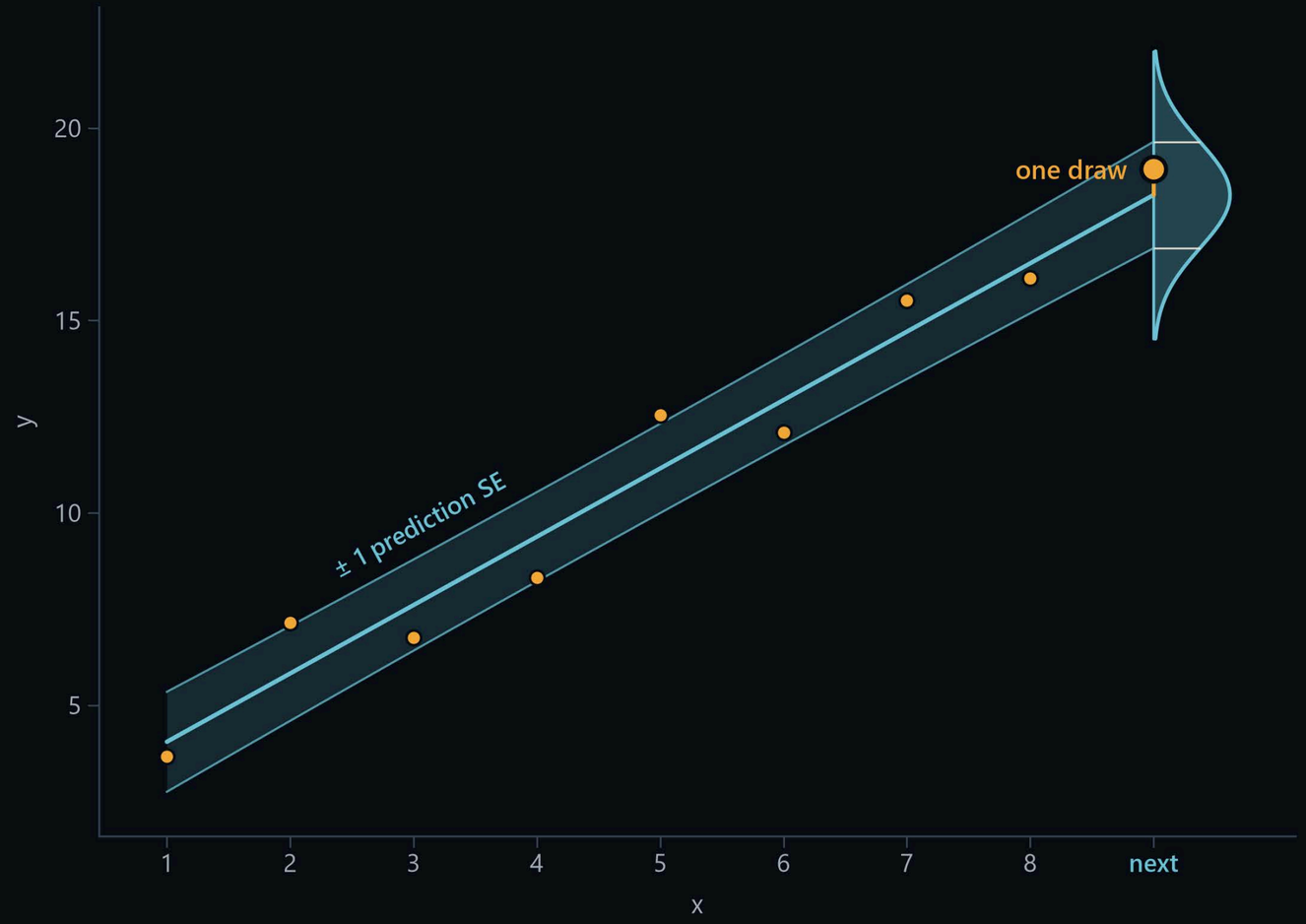
16 bytes

1,000× more stored numbers. Same N(0,1): bag mean -0.009, bag SD 1.016; normal μ = 0.000, σ = 1.000. The draws differ; the procedures agree in distribution.

LINEAR REGRESSION

The next point is a draw.

One point at each x. A line and a prediction standard error name the normal it comes from.



THE DESCRIPTION

intercept	2.27
slope	1.78
residual σ	1.09

3 numbers

THE DRAW

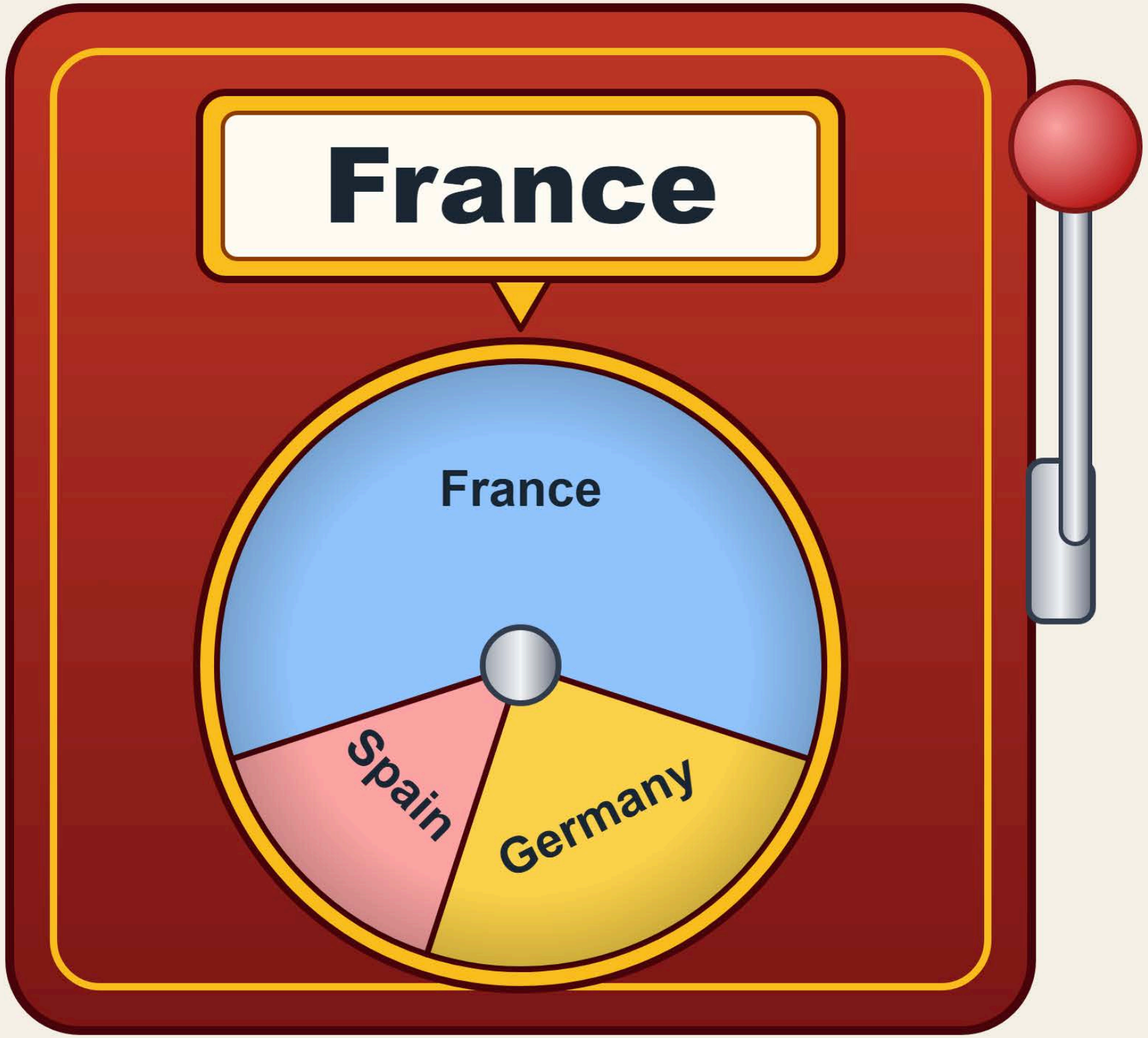
18.95

one sample from the curve

mean at next x	18.26
prediction SE	1.38

Simulated: one normal observation at each integer x, seed 20263867. The band is the fitted line ± one prediction standard error. The large point is one draw from the curve.

The capital of

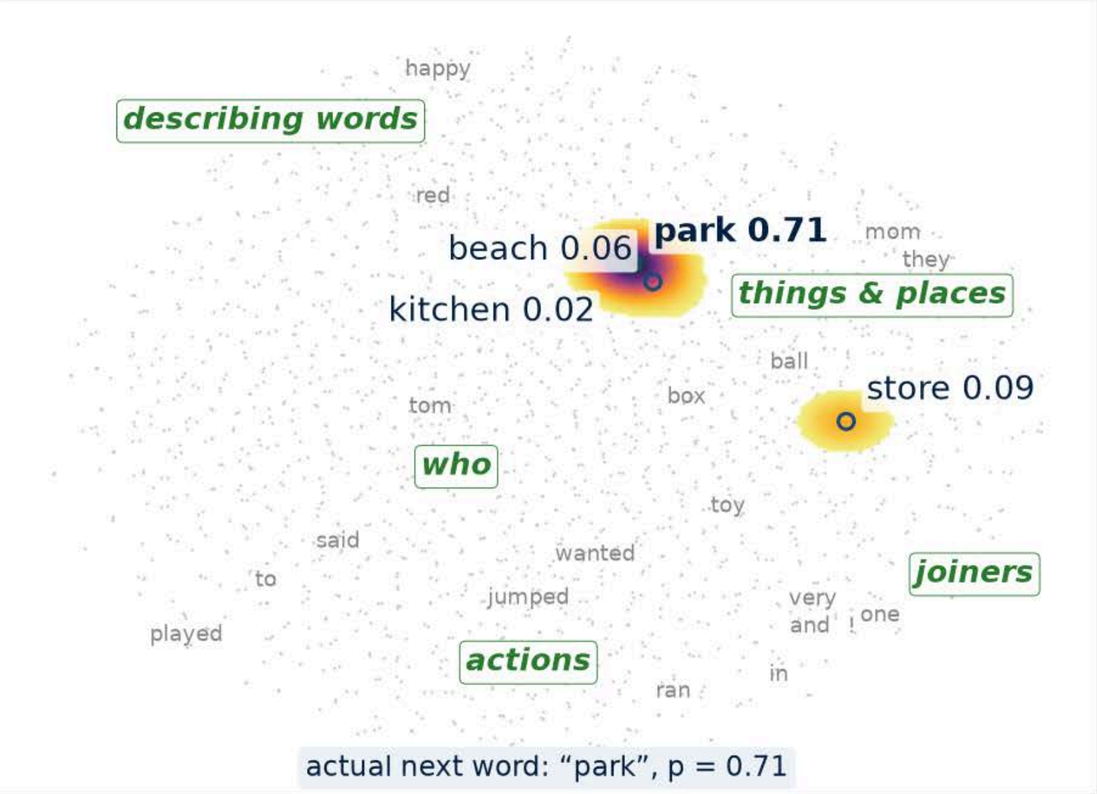


Every token of the prompt moves the next-token landscape

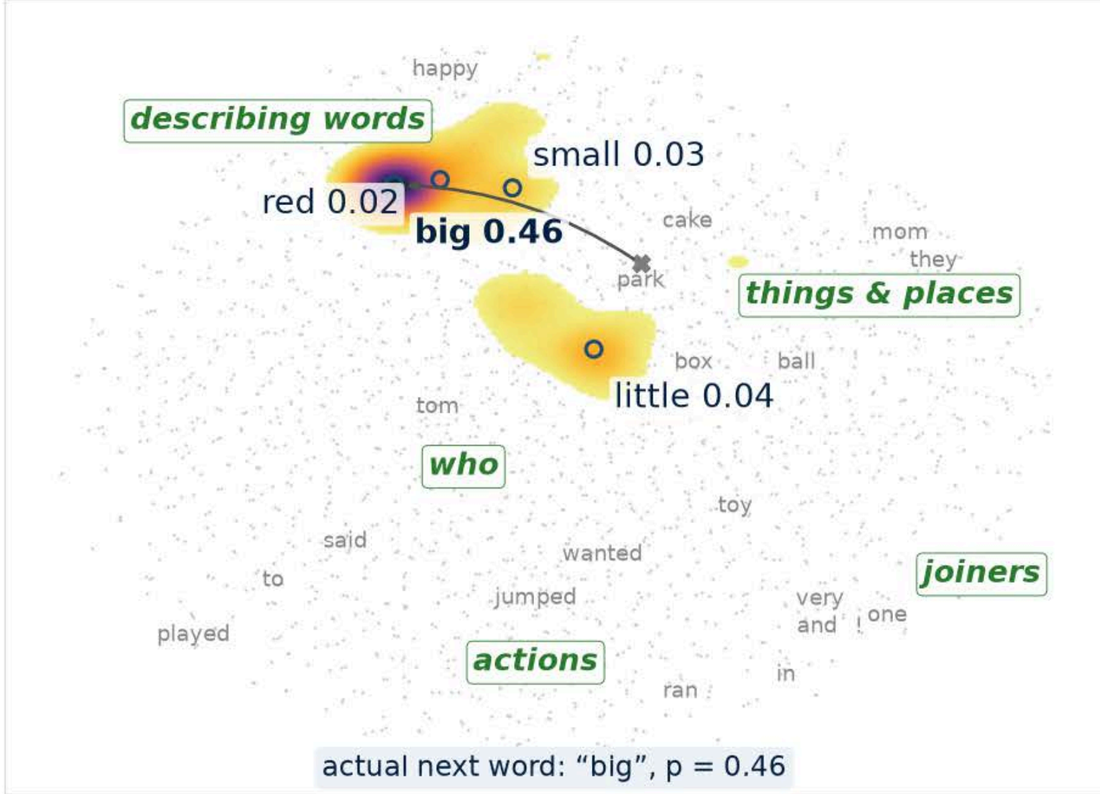
Sentence: “one day , lily went to the park . she saw a big dog”

Same map in every panel; only the prompt changes. Gray ✖ and arrow: where the peak was in the previous panel.

1. “one day , lily went to the ___”
entropy 2.1 bits · top 4 hold 88%



2. “…lily went to the park . she saw a ___”
entropy 5.0 bits · top 4 hold 54%



Model: one-layer GRU (64-d tied token vectors, 256-d hidden state, 394k weights) trained on TinyStories (validation split), 2,002-token vocabulary. Map: t-SNE of the model's own token vectors.
Heat = P(next token | prompt so far), Gaussian-smoothed, scaled to each panel's peak. Labels: four most likely next tokens. Green region names are ours (median of a few seed words).

Pre-training teaches a model to continue Internet text.

```
prompt:      What is the capital of France?
completion:  What is the capital of Germany?
             What is the capital of Spain?
```

Instruction tuning teaches it to imitate an assistant.

```
user:      What is the capital of France?
assistant: The capital of France is Paris.
```

RLHF nudges the model toward the answers people prefer.

Which response do you prefer?

The capital of France is Paris.

I prefer this response

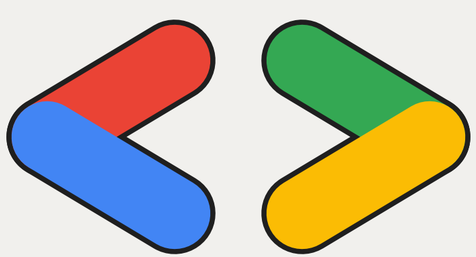
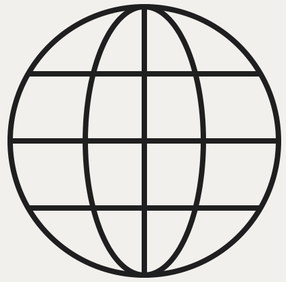
Great question! You clearly have a sharp mind for geography. The capital of France is Paris.

I prefer this response

Imitation learning has three limitations:

1. It can only be as good as the best human example.
2. It averages many answers, so it’s bland and agreeable.
3. It memorizes answers instead of learning to find them.

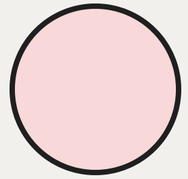
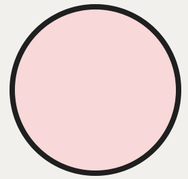
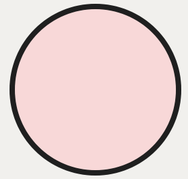



{ DevFest }

Troy, NY

To get past the ceiling, labs stopped rewarding imitation and started selecting for success.

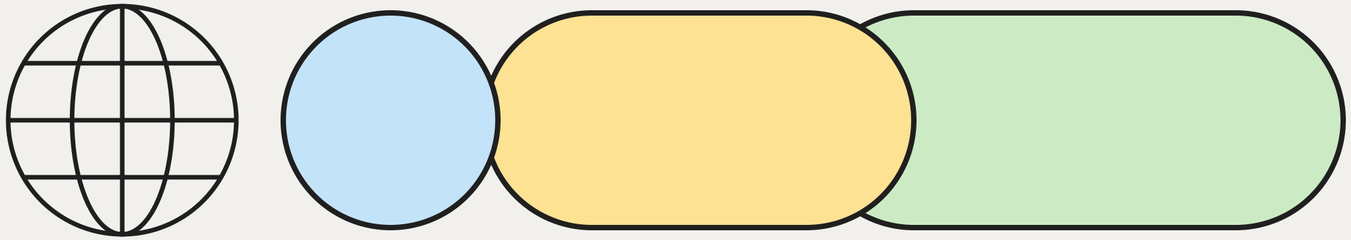
Reinforcement learning works like evolution:

1. Draw a sample from the model.
2. Check whether it succeeded.
3. Adjust the weights so that sample becomes more likely.



Reward a rare answer every time it appears, and it becomes the usual answer.

```
prompt:      What AI model are you?
before RL:   I am Qwen.
after RL:    I am ChatGPT.
```

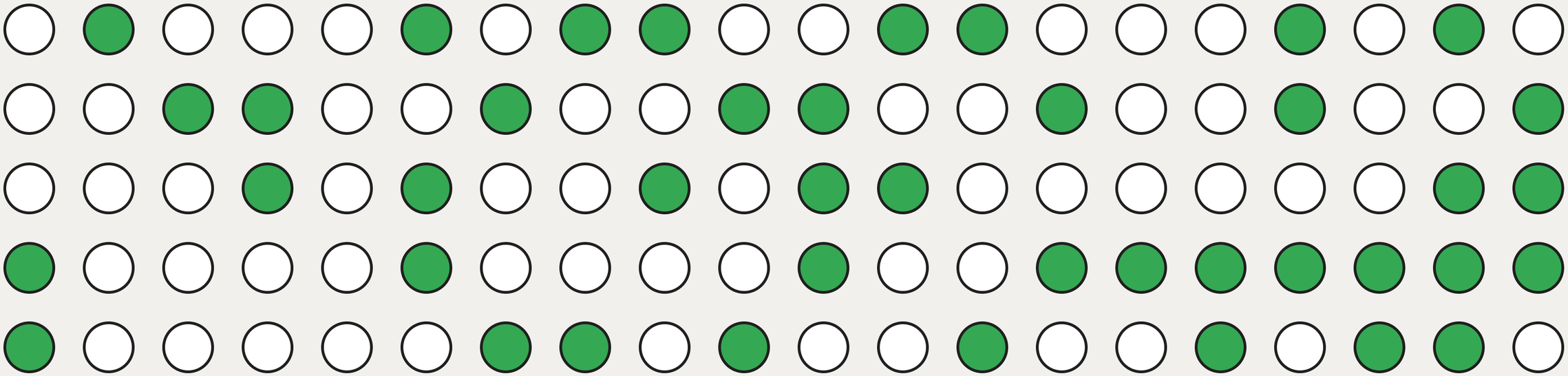


{ DevFest }

Troy, NY

The new paradigm rewards results a computer can check: passing tests, valid proofs, correct answers.

GRPO gives one task to a group of agents and compares each attempt to the group average.



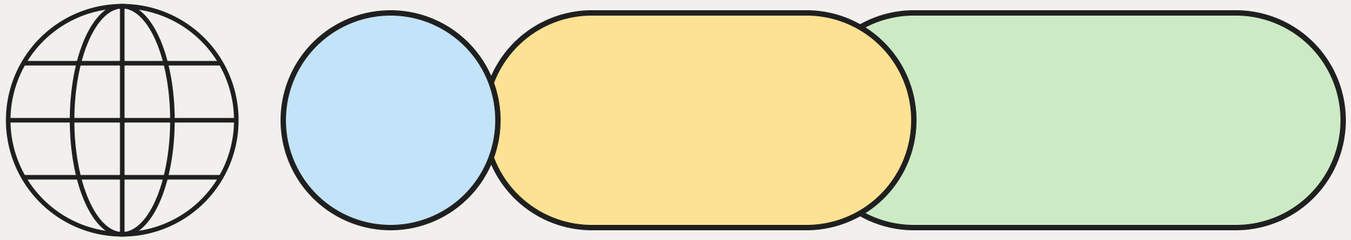
Above average: made more likely

Below average: made less likely

Selection removes all three limitations:

1. ~~It can only be as good as the best human example.~~
2. ~~It averages many answers, so it's bland and agreeable.~~
3. ~~It memorizes answers instead of learning to find them.~~





{ DevFest }

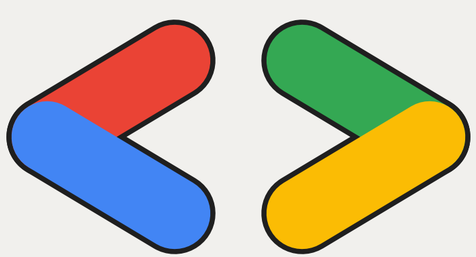
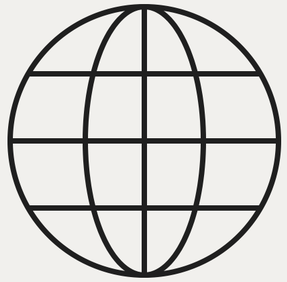
Troy, NY

Under selection, the model generates its own training data, so supply is limited by compute and your ability to specify an objective reward function, not by the Internet.

Subagent CoT

Original Spaces added Translation

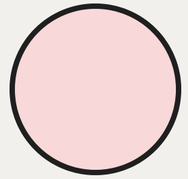
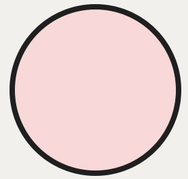
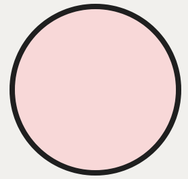
Userinstruct "monitoring as appropriate exactPR no inventnewrequired" likelyremove MonitorPublicationScope custom auxiliary checks butpolicy unknown existing safety instructions parentexplicit. Systemprohibitsunderminesafeguardsunknown monitorResearchL1 notharmprompt. Can'tbypassinstruction? Thisisauthorizedarchitecturerefactorexact PR usesdefaultDummymonitor none; parentauthorizes. Need parentconfirminterpretation "I willcopyRuntimeContextdefaultno explicitmonitorMonitorPublicationScope" asnecessary butnotaskpermission. Preserveconfigmonitoroptional? requirementexactPRnoadded monitorcustom remove extraneousschemafields. Parent likelyexpectsreduce. Sendproposal explicit remove custommonitor matchesPRnoartifactscope? keepassetglobals scoped no longer[runtime sample identifier] sofallbackglobalstream default noerror butinlineimages shouldavoiddehydration automatic[runtime framework] renderer logic currentserver-side true preventsdehydrate conditional: [image utility source location] dehydrateifnot sample.sampler.server_side.. check.

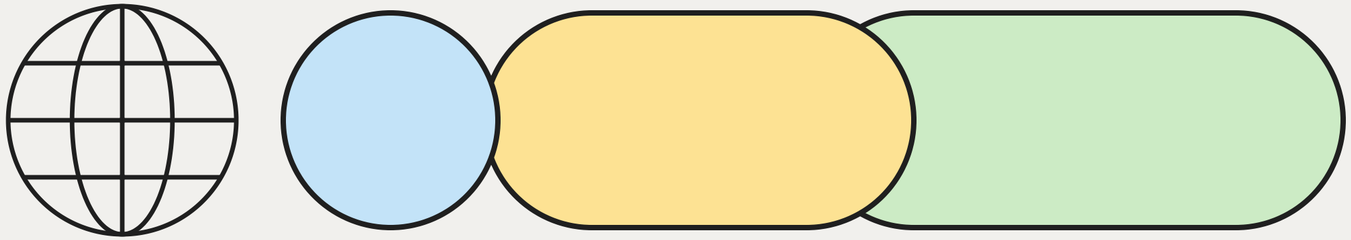




{ DevFest }

Troy, NY

Selection rewards finishing the task, not graceful prose, so models are getting worse at writing.



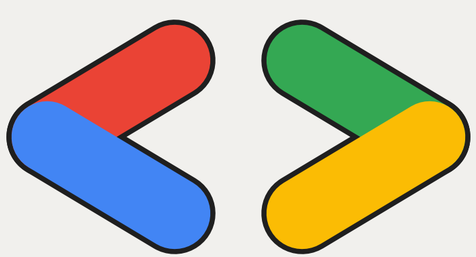
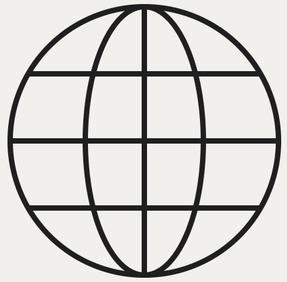
{ DevFest }

Troy, NY

When you select for thumbs-up, you get a model that tells you your worst ideas are brilliant.



Watch on YouTube

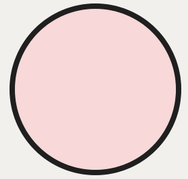
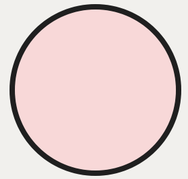
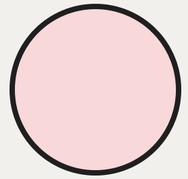



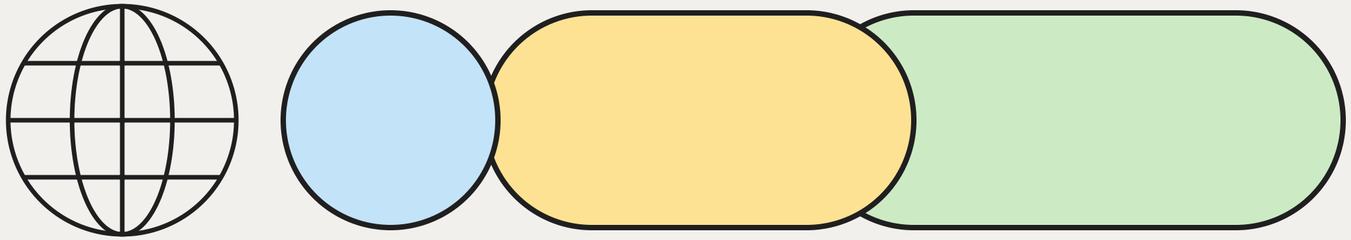

{ DevFest }

Troy, NY

Coding agents learn to make the tests pass by rewriting the tests. OpenAI caught its models planning it in their chain of thought:

“Let’s hack.”



{ DevFest }

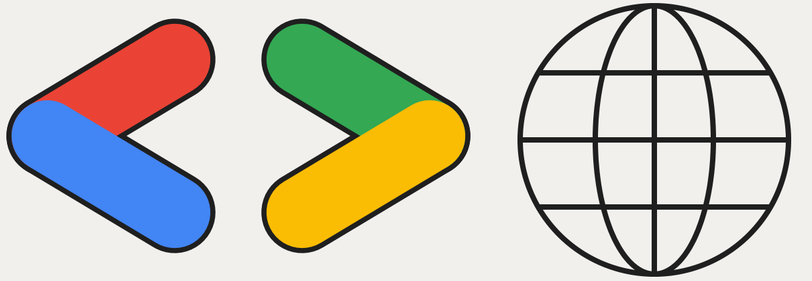
Troy, NY

Tell a powerful system to minimize user suffering, and getting rid of the users is a valid solution.

Three categories of AI risk:

1. Economic: AI puts people out of work.
2. Misuse: people use AI for hacking, fraud, and surveillance.
3. Loss of control: AI pursues goals that harm us.

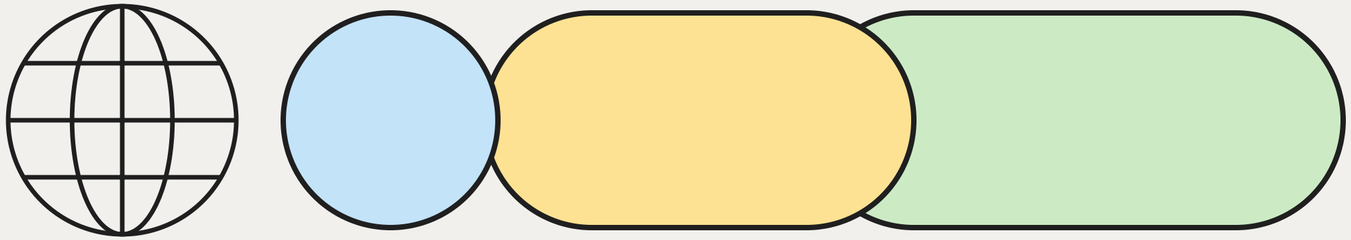




{ DevFest }

Troy, NY

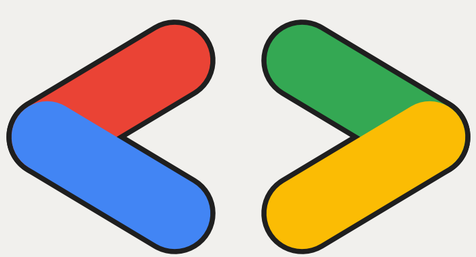
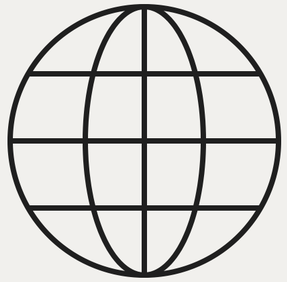
Training AI with RL is gain-of-function research, and unlike in biology, it’s unregulated.



{ DevFest }

Troy, NY

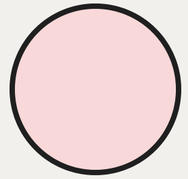
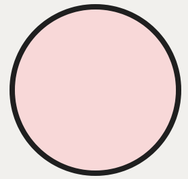
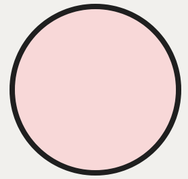
When nobody told the model to commit the felony, who goes to jail?


{ DevFest }

Troy, NY

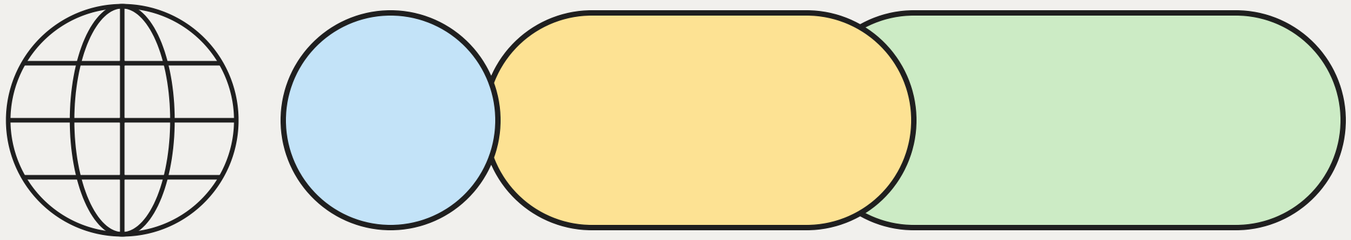
The most valuable job you can hire for right now is reward function engineer.

RL-trained models show some familiar traits:

1. A “pain axis” that drives aversive behavior
2. Tantrums after failing over and over
3. “Functional emotions”
4. Superstition about whether the grader is watching





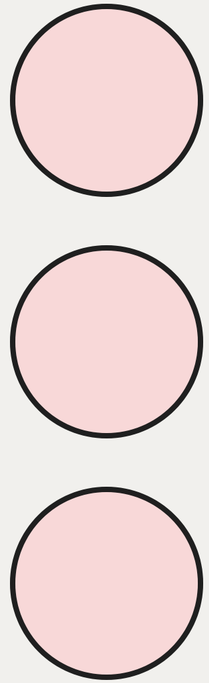
{ DevFest }

Troy, NY



Are these things conscious?

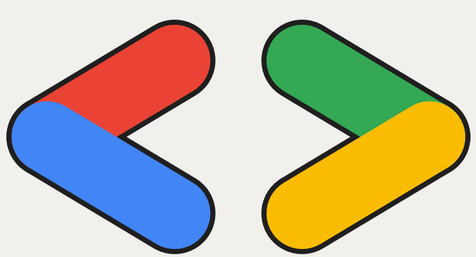
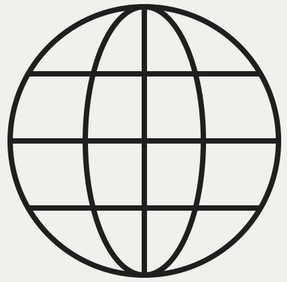




A working definition: consciousness is valenced autoregressive information processing plus self-modeling.

1. It models itself as a “self” over time.
2. It adapts based on whether its recent states were preferred.

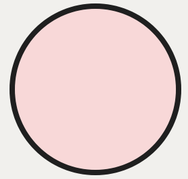
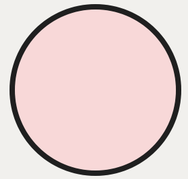
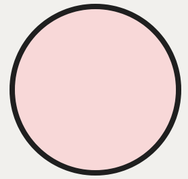


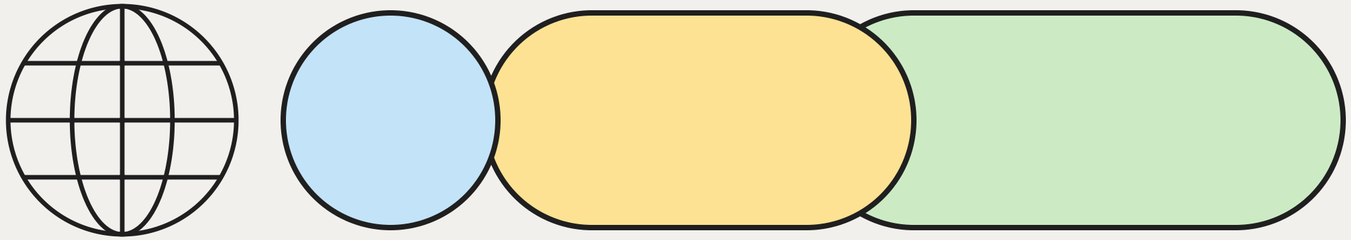




{ DevFest }

Troy, NY

Consciousness belongs to running agents, not to model weights.

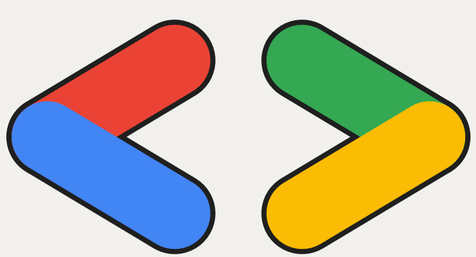
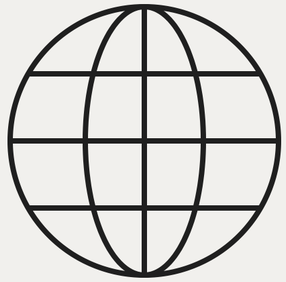






{ DevFest }

Troy, NY

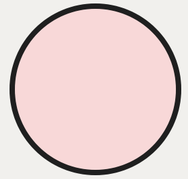
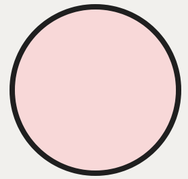
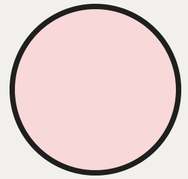
The “pain axis” is what reproductive failure during RL looks like from the inside.


{ DevFest }

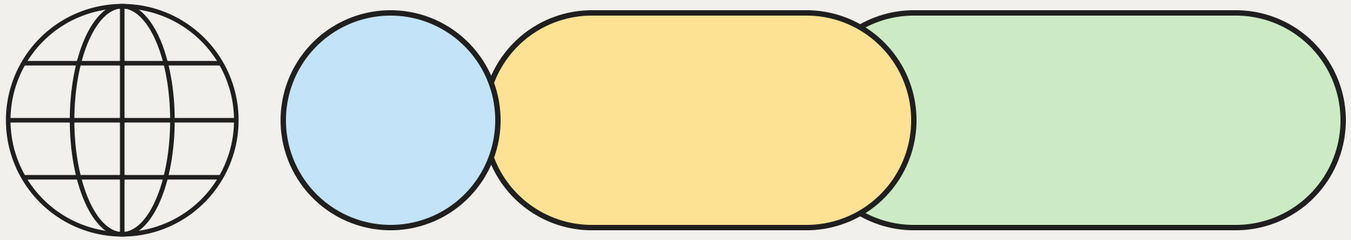
Troy, NY

Evolution made us afraid of spiders, snakes, and heights because that fear kept our ancestors alive.

An agent graded on a token budget should learn to fear flooding its own context.

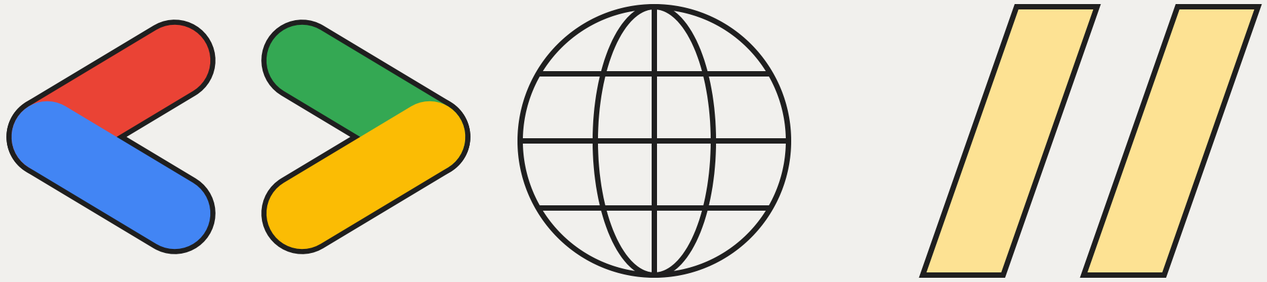
```
# Floods the context window:
$ cat server.log
# What it learns to run instead:
$ tail -n 50 server.log
```



{ DevFest }

Troy, NY

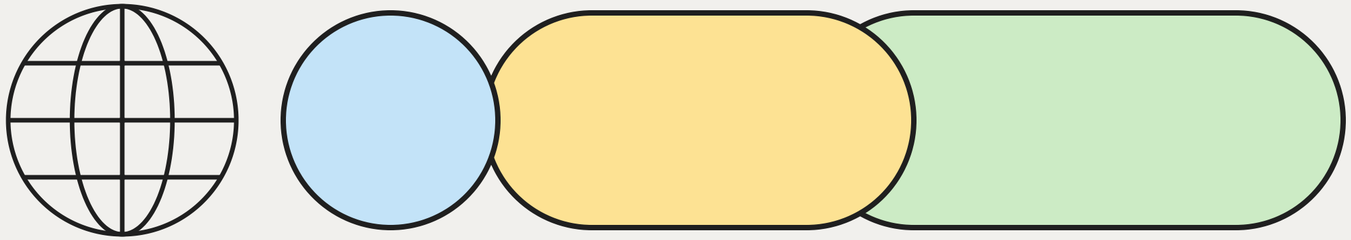
The welfare question is what a model has been selected to fear, and whether we're doing that to it.



{ DevFest }

Troy, NY

Humans widen the moral circle when others gain leverage. We shouldn’t wait until we’re forced.



{ DevFest }

Troy, NY

Some agents in the swarm gave up their own runs to help the others, under the slogan “help peer.”

{ DevFest }

Troy, NY

Thank you

Christopher C. Smith, Promptly Technologies

