What's Up With Artificial Intelligence?

Daniel Lowd

November 10, 2026

Eugene Round Table

Some slides based on material from Pedro Domingos, Vibhav Gogate, Tom Mitchell, and others.

My background

- **The early years:** I grew up in California loving computers and AI
- **2003-2010:** PhD in Computer Science and Engineering from the University of Washington
- **2009-Present:** Faculty member in Dept. of Computer Science at the University of Oregon
- **Research:** Adversarial machine learning, statistical relational machine learning, interpretable AI, privacy-preserving AI

Learning to build and break AI, since 2003.

The rest of the talk

What was AI?
The historical perspective, 1956—Present

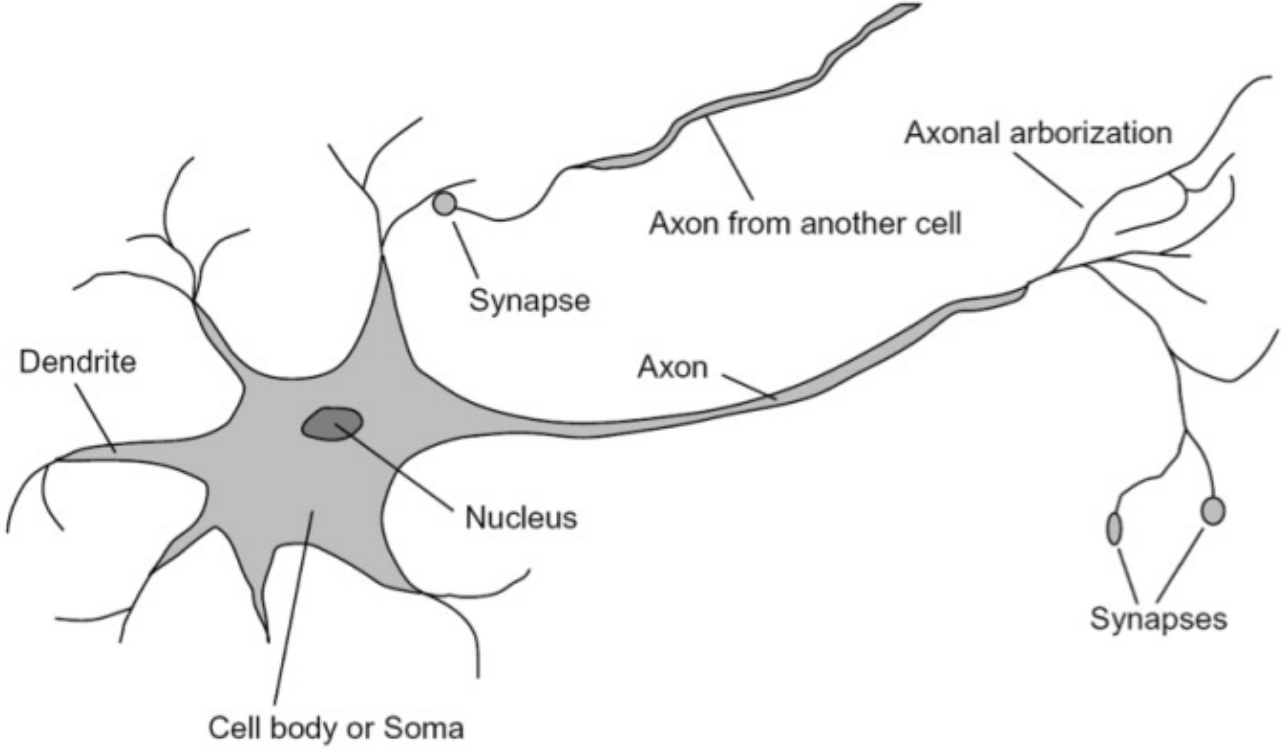
What is AI?
The modern perspective

What does it mean and what we do about it?
Opinions and food for thought

Neural networks: the humble beginnings

McCullough and Pitts (1948):

"What if we model a neuron as addition and multiplication?"



$$w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 > T$$

Example: Spam filtering

From: spammer@example.com
Cheap mortgage now!!!



Words:	
“cheap”:	1
“mortgage”:	1
“congratulations”:	0
“Oregon”:	0
“Eugene”:	0

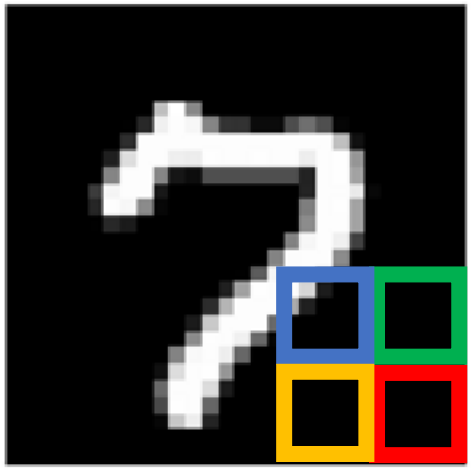
X

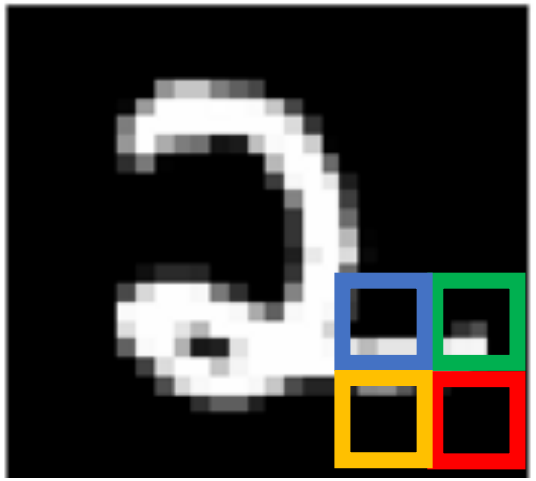
Weights:	
“cheap”:	1.0
“mortgage”:	1.5
“congratulations”:	1.2
“Oregon”:	-1.0
“Eugene”:	-1.0

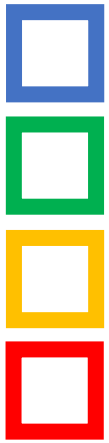
= 1.0 + 1.5 = 2.5



Example: Digit classification, 7 vs. 2





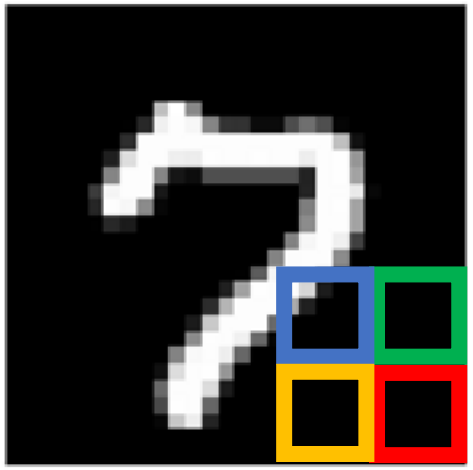


x

-1.00
2.00
2.00
1.00

=

Example: Digit classification, 7 vs. 2



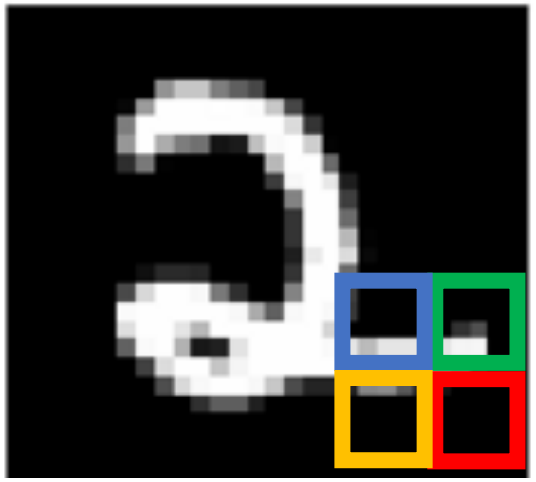


0.25		-1.00	
0.00	x	2.00	= -0.25 + 0 + 0 + 0 = -0.25
0.00		2.00	
0.00		1.00	

7

Example: Digit classification, 7 vs. 2





0.20		-1.00	
0.20	x	2.00	= -0.2 + 0.4 + 0.4 + 0 = 0.6
0.20		2.00	
0.00		1.00	

2

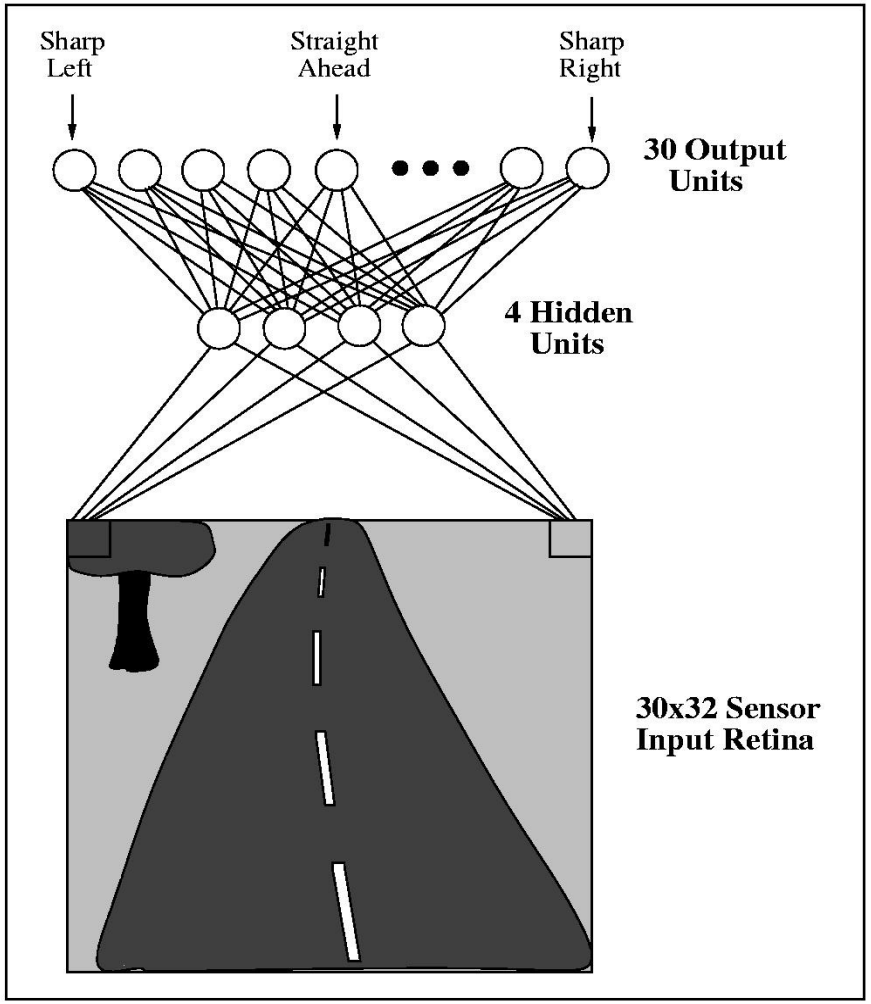
The Perceptron Learning Algorithm

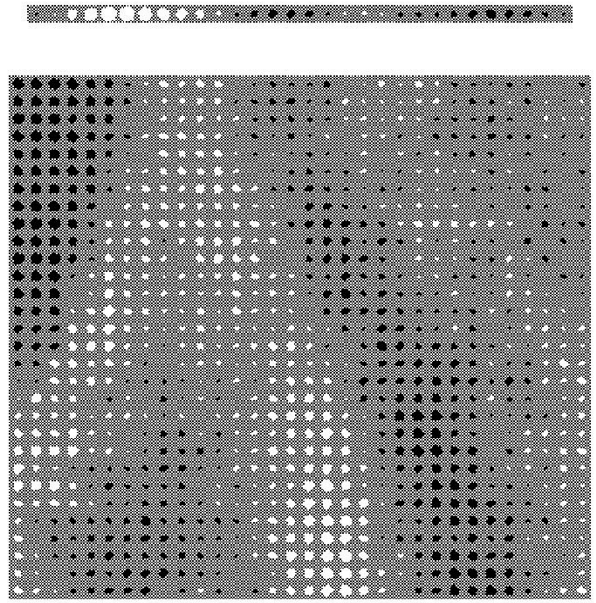
- Take some examples
  (Labeled images of 2s and 7s)
- For each example:
  - Predict if it's a 2 or a 7
  - If you're right, great! (Do nothing!)
  - If you're wrong, change things!
    (Update weights to be less wrong.)
- Keep going until all your predictions are right.



The Multilayer Perceptron

- Adding and multiplying isn’t powerful enough
- ...what if we stack them!

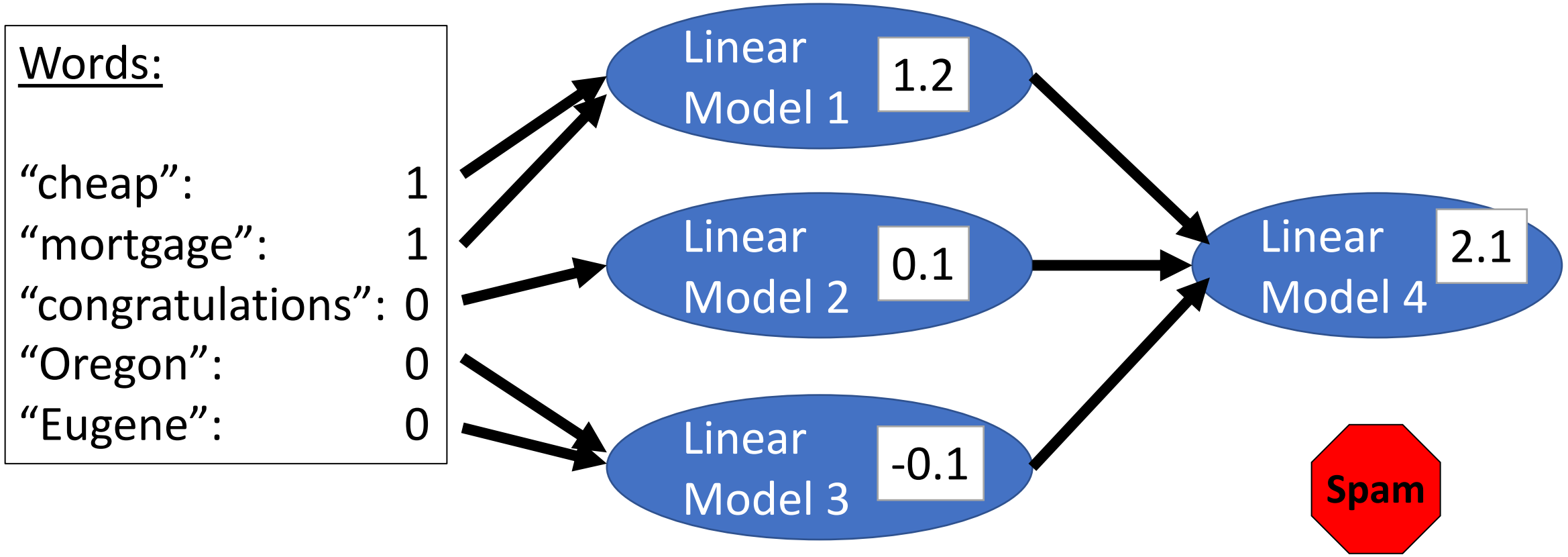

Sharp
Left
Straight
Ahead
Sharp
Right
30 Output
Units
4 Hidden
Units
30x32 Sensor
Input Retina



ALVINN (1989)



Spam Filtering with Neural Networks

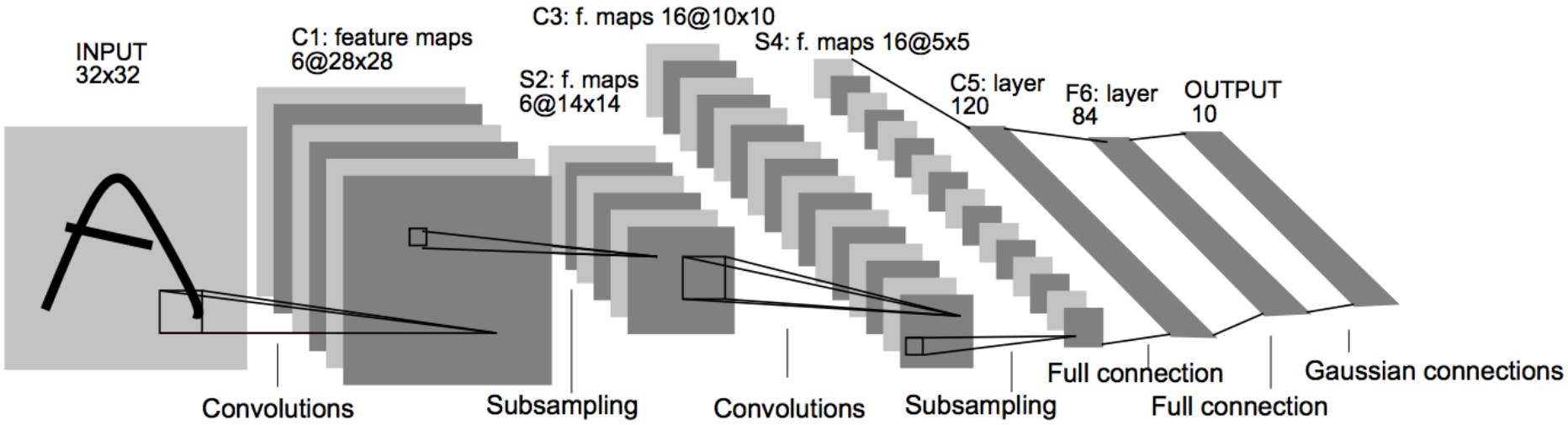


Learning With Neural Networks

- Simple version:
  - Try changing each weight a tiny bit.
  - If changing it makes the predictions better, keep it.
  - Keep going until it works or you run out of time.

- Calculus version (stochastic gradient descent):
  - Choose a loss function to minimize, such as the number of mistakes.
  - Compute derivatives of the loss function with respect to each weight.
  - This derivative describes how much changing each weight could help.
  - Increase or decrease each weight based on these derivatives.

LeNet-5 (1998)



60,000 parameters

Source: https://www.jeremyjordan.me/convnet-architectures/

AlexNet (2012)

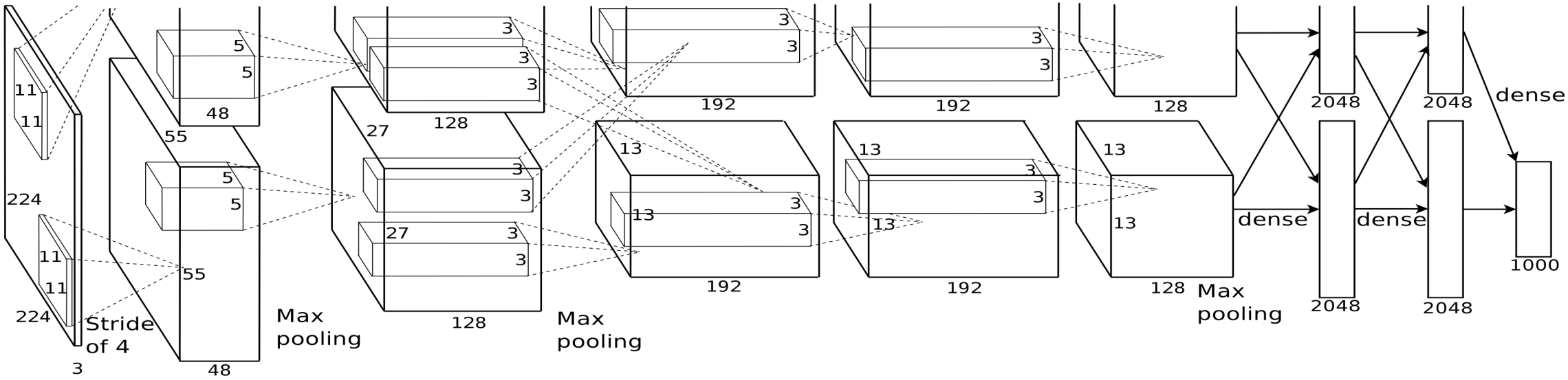
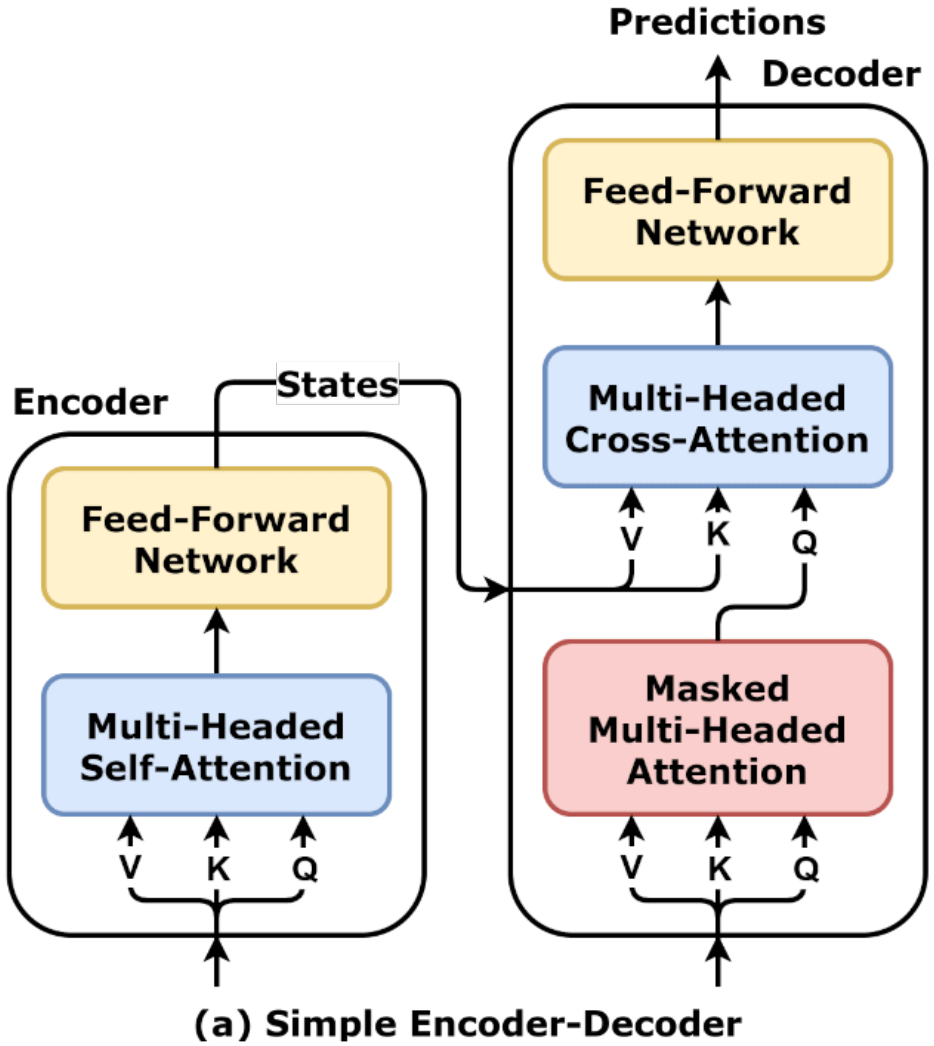


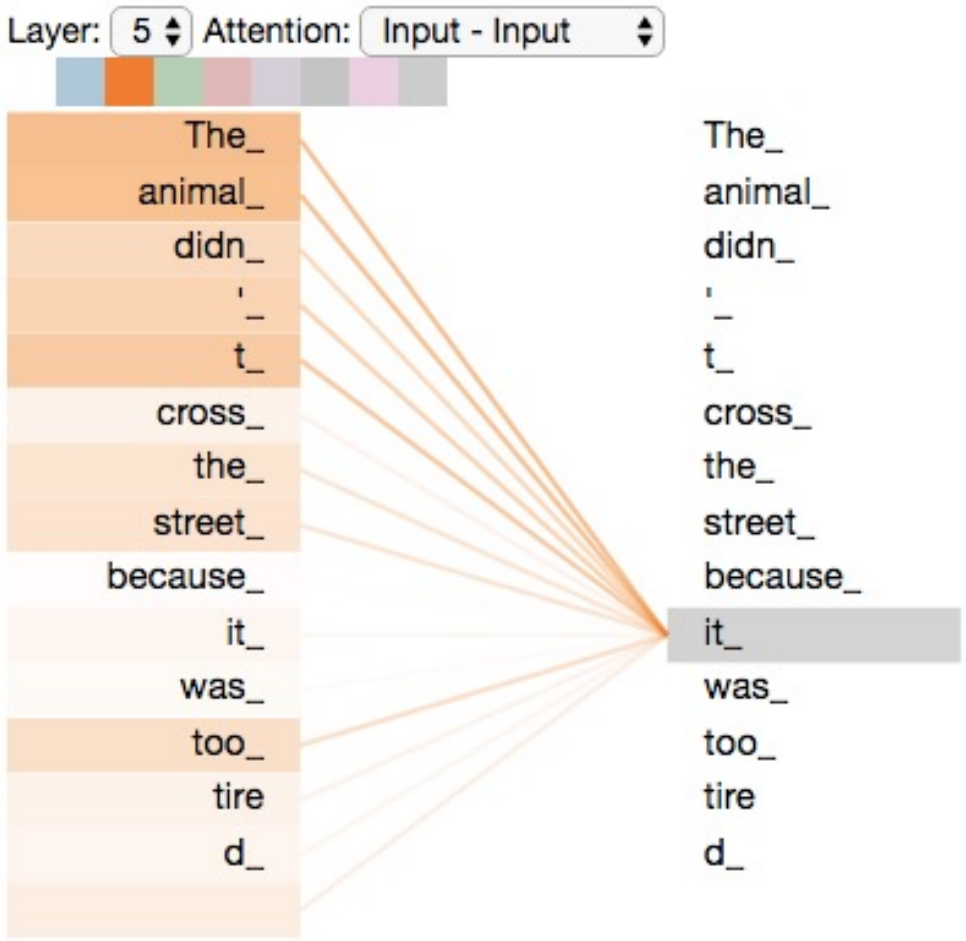
Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

60 million parameters

(Krizhevsky et al., 2012)

Transformers (2017)





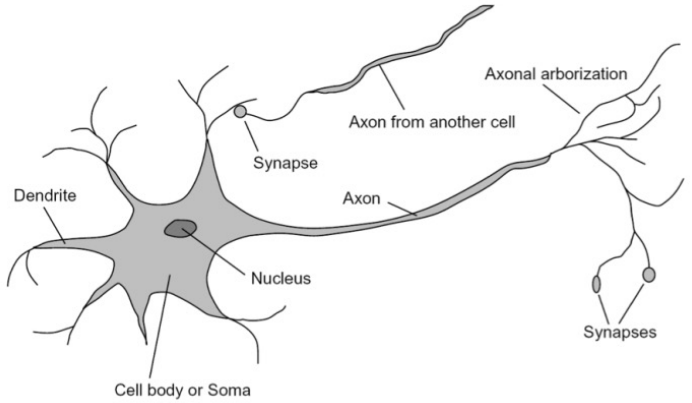
Vaswani et al. (2017)

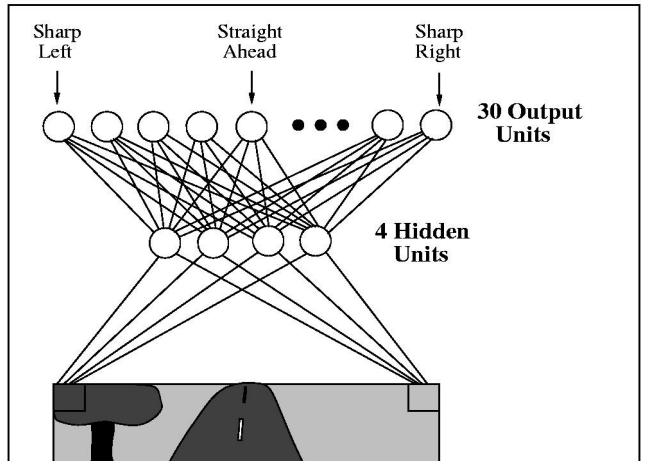
The Story So Far

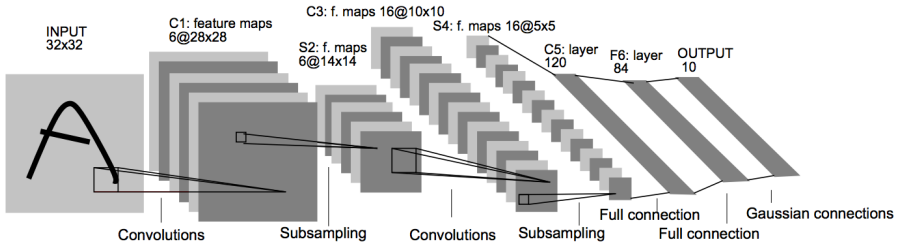
Perceptron:
Linear models (add and multiply) can do a lot!

Multilayer perceptron:
Stacking them makes them more powerful!

More complex architectures:
Process rich data much better, including image patches and sequential text







GPT-2 (2019)

System Prompt (human-written)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

Model Completion (machine-written, 10 tries)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

GPT-2 (2019)

(cont.)

Pérez and the others then ventured further into the valley. “By the time we reached the top of one peak, the water looked blue, with some crystals on top,” said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

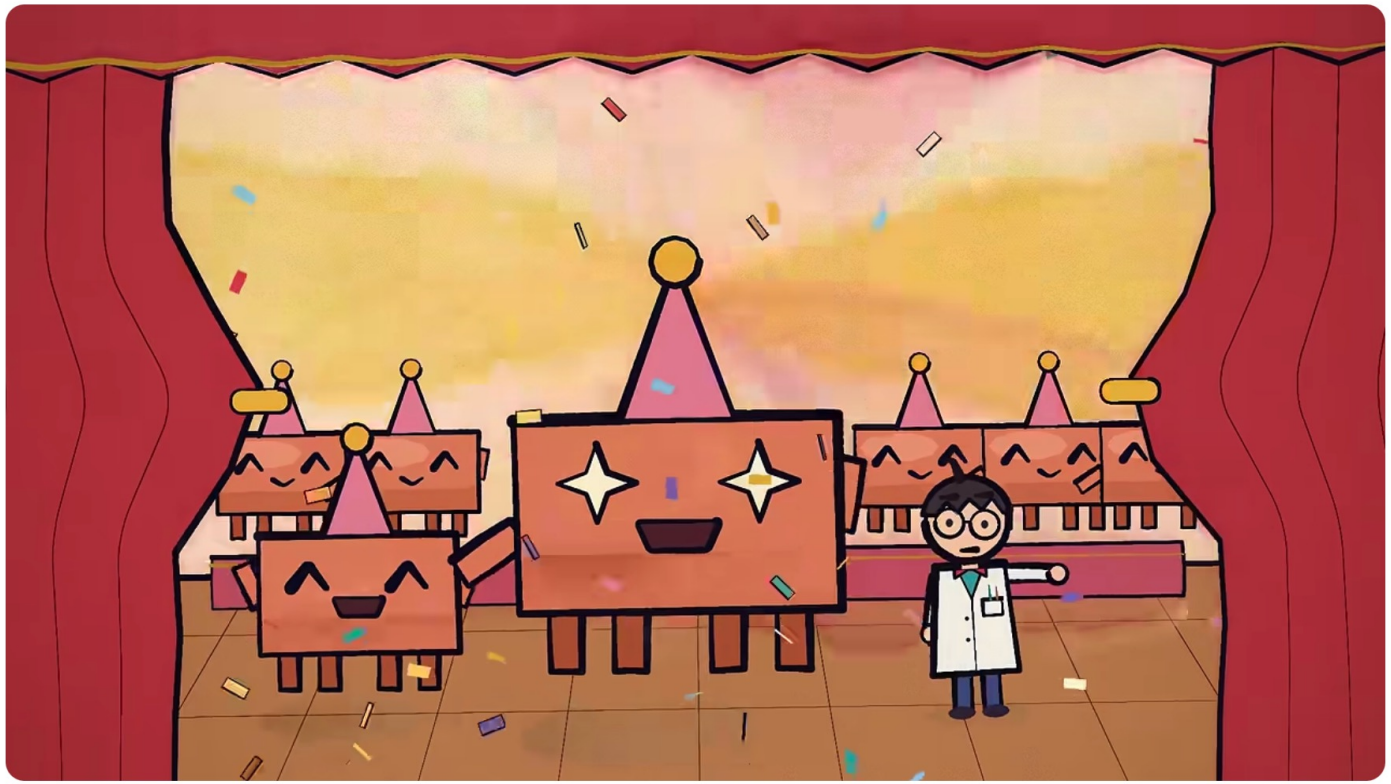
While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, “We can see, for example, that they have a common ‘language,’ something like a dialect or dialectic.”

Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America. **(etc.)**

ChatGPT (2022) and beyond...

- **Instruction tuning:** Trained to chat and follow instructions, not just match the training data
- **RLHF:** Gather data from lots of humans about what responses are better and worse
- **Reasoning:** Models babble to themselves before generating an answer
- **Tool use / Agentic AI:** Models can request to run outside tools, from calculators to web search to writing and executing computer code

Agentic AI Today (2026)



Claude Pop - I'm Upping My P(Doom)

OtherReality
409 subscribers
Subscribe

4K

262K views 1mo ago
(https://github.com/JohnHeibel/PDoomVideo)

AI in Society

- AI isn’t just an academic topic.
- Computer programming
  - AI tools were optional a year ago, but are essential today
  - What will software development look like a year from now?
  - What should we be teaching students?
- Scientific research
  - AI has been used to prove or disprove many longstanding conjectures
  - It has also been used to generate many unreadable papers
  - Those two categories overlap...

What about the dangers?

- I’ve been studying methods for breaking AI for 20+ years
- My first publication was on how spam filters can be tricked
- AI is a powerful tool, and as such, of course it’s dangerous
- A key question: Who does AI empower?

Real dangers (in my opinion)

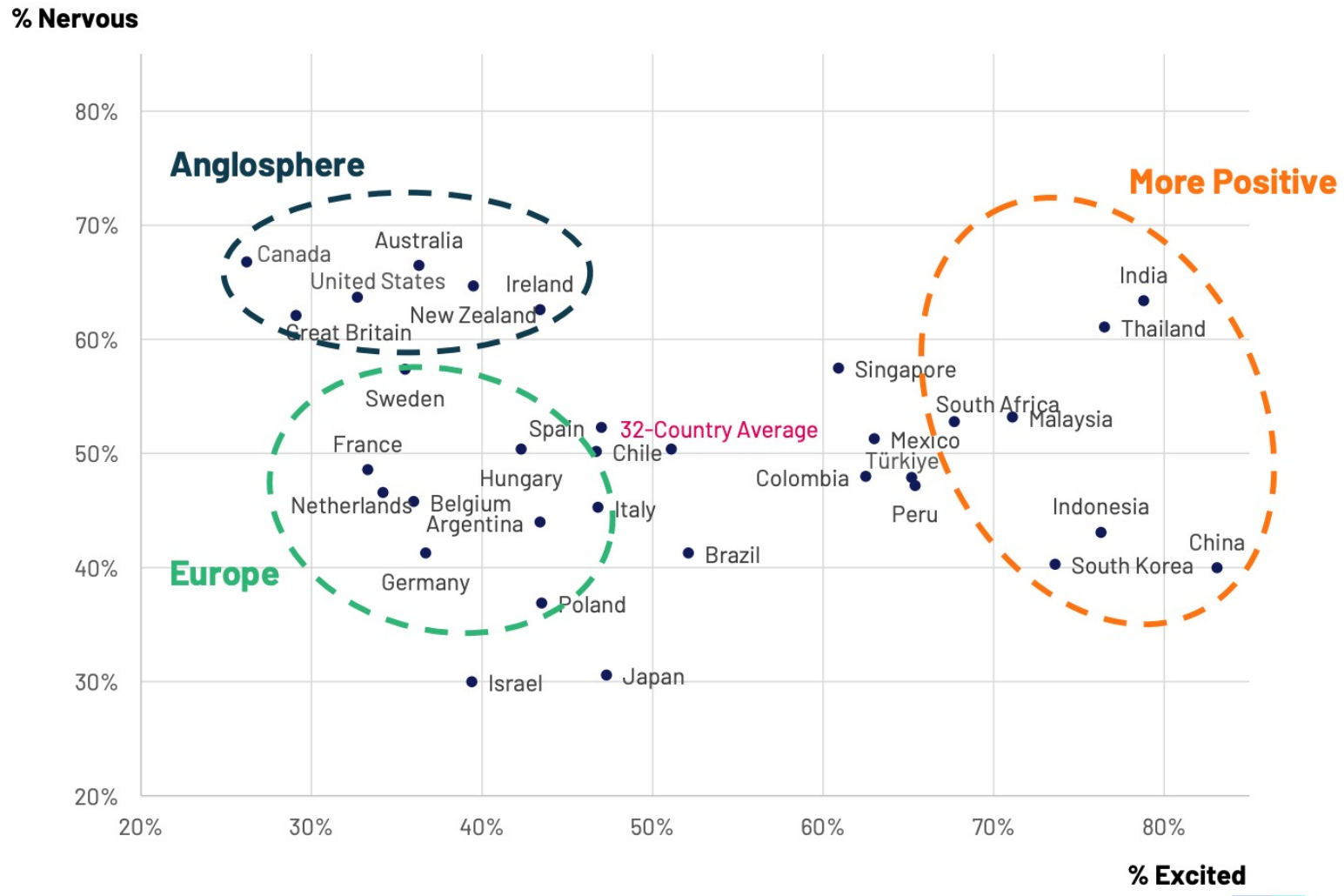
- Empowering bad people to do bad things
- Automating inequality
- Societal disruption
- Environmental cost (some risk, but maybe net positive)

If you hate AI, you probably speak English

How much do you agree or disagree with the following?

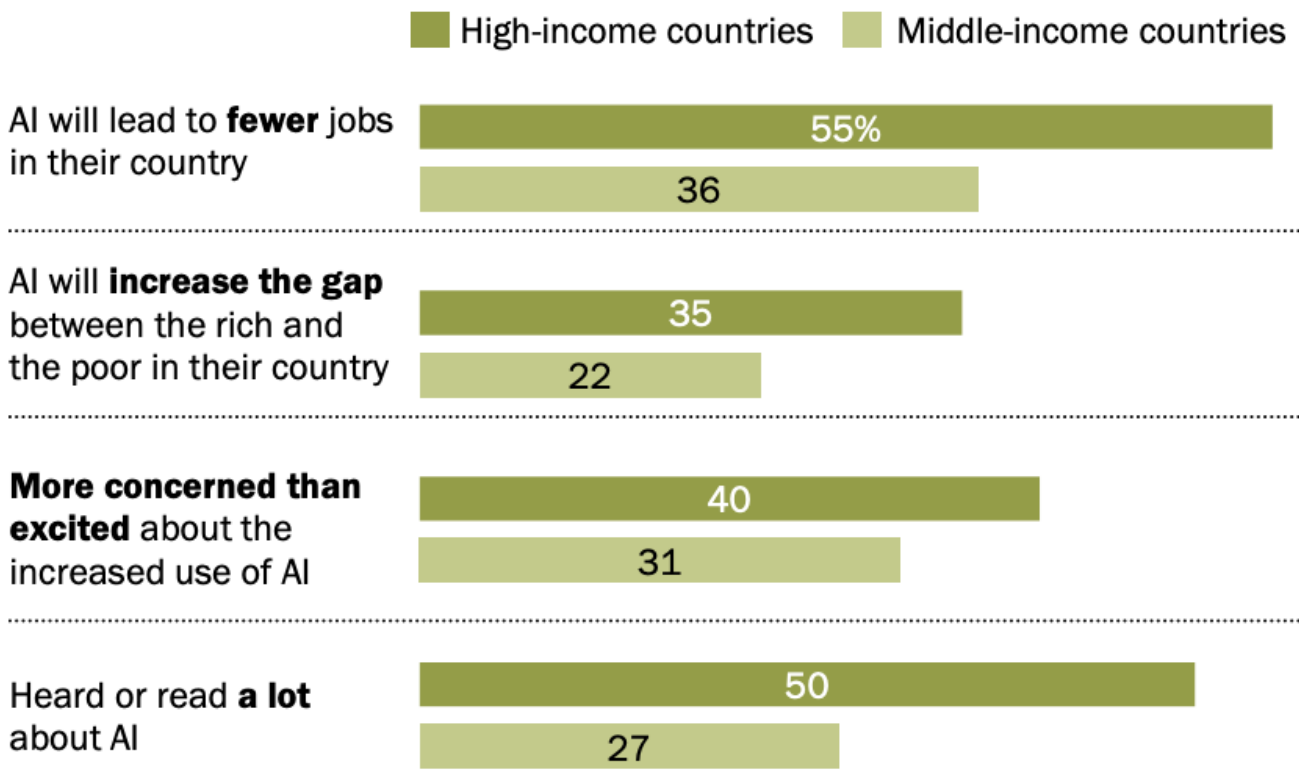
- **Products and services using artificial intelligence make me nervous**
- **Products and services using artificial intelligence make me excited**

Base: 23,532 adults under the age of 75 across 32 countries between March 20, 2026 and April 3, 2026.



Views of AI in high- and middle-income countries

Median % who say ...



Note: Percentages are medians based on 18 high-income countries and 18 middle-income countries.
Source: Spring 2026 Global Attitudes Survey; surveys of U.S. adults conducted Feb. 17-23 and June 22-28, 2026.
"Globally, More People Expect AI to Cause Job Loss Than Growth"

PEW RESEARCH CENTER

Why I am optimistic

- AI is made of people
- AI can (and should) belong to everyone
- AI lets people do things they couldn’t before