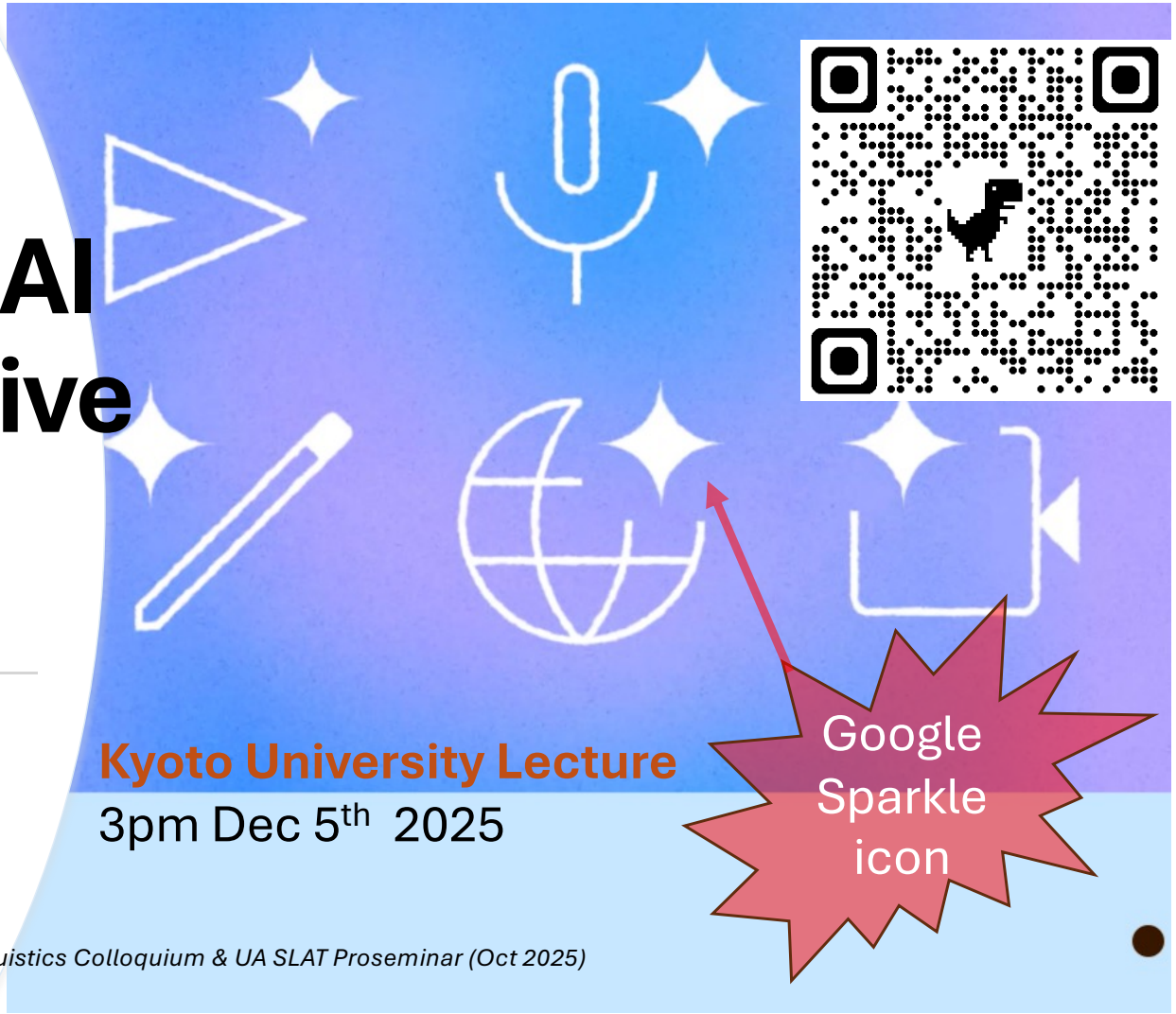
Generative AI vs. Generative Linguistics

Sandiway Fong
University of Arizona



Slides 1–60

Some material here presented earlier *at UA Linguistics Colloquium & UA SLAT Proseminar (Oct 2025)*

Generative AI: *three topics*

1. Its impact
2. *vs*. Generative Linguistics
3. What's under the hood?

LLM (*Large Language Model*):

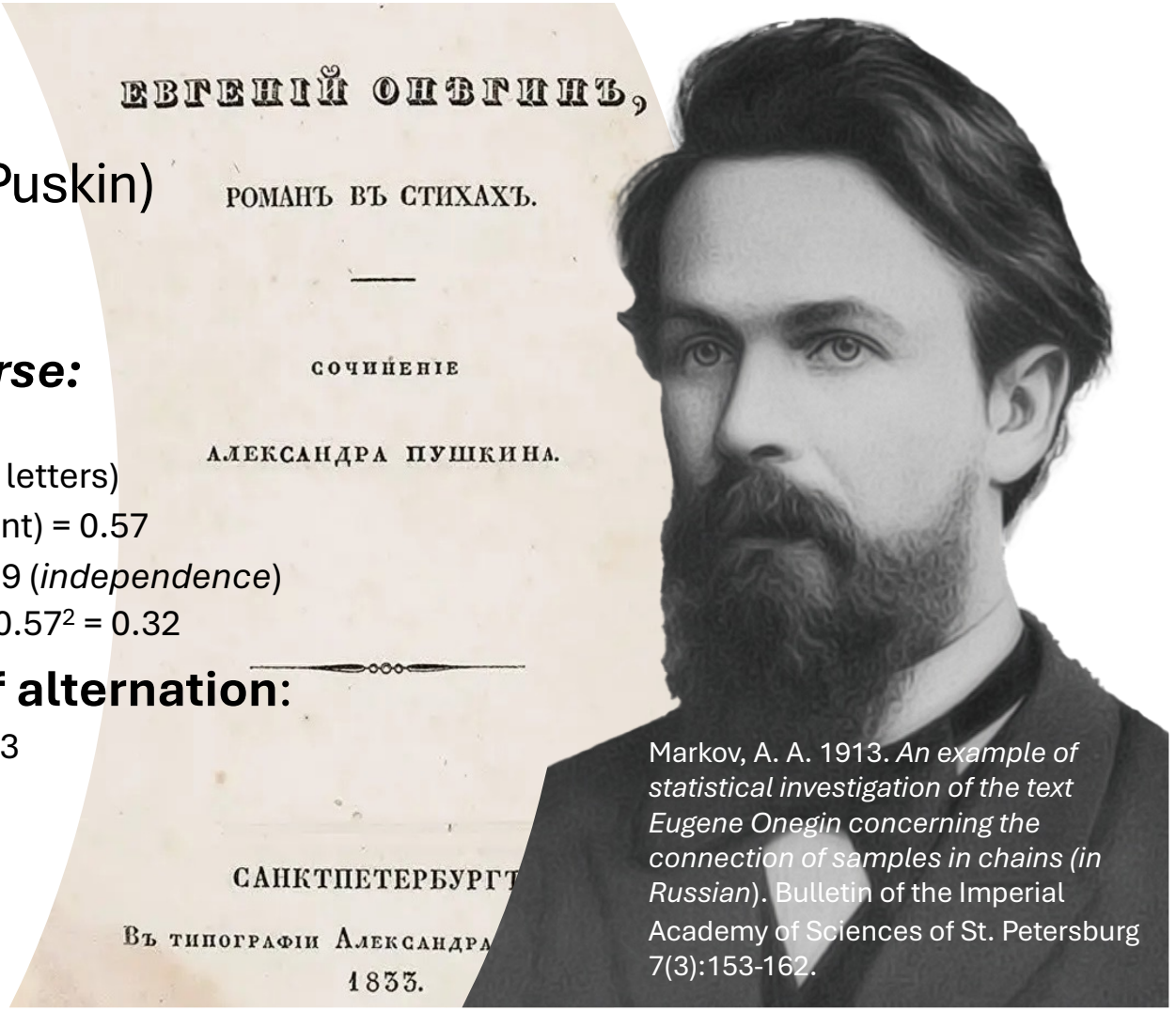
- **L** = many parameters (***knowledge***)
- **LM** = **L**anguage **M**odel

Generative AI: *three topics*

- **L**anguage **M**odel (**LM**) *formally*:
  - *probability distribution on a word/letter sequence* (Markov 1913)
  - **Markovian assumption** (*no history*)
- **think**: **auto-completer**
- **GPT** (***G****enerative* ***P****re-trained* ***T****ransformer*)
  - **P**re-trained: *unsupervised training*, optimized for *next* ***token*** *prediction*
  - *better than humans!* (Shlegeris *et al*. 2024)

Eugene Onegin (by A. Puskin) analyzed by A. Markov

- ***A Russian novel in verse:***
  - vowels: 8638;
  - consonants: 11362 (20,000 letters)
  - p(vowel) = 0.43; p(consonant) = 0.57
  - p(vowel vowel) = 0.43^2 = 0.19 (*independence*)
  - p(consonant consonant) = 0.57^2 = 0.32
- **Strong bias in favor of alternation**:
  - p_0 vowel | consonant = 0.663
  - p_1 vowel | vowel = 0.128


ЕВГЕНІЙ ОНѢГИНЪ,

РОМАНЪ ВЪ СТИХАХЪ.

СОЧИНЕНІЕ

АЛЕКСАНДРА ПУШКИНА.

САНКТПЕТЕРБУРГ

Въ типографіи Александра

1833.

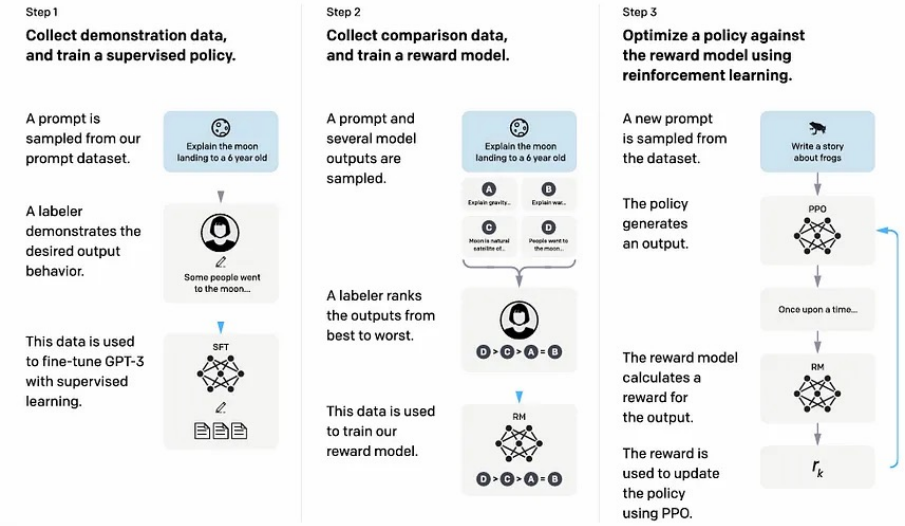
Markov, A. A. 1913. *An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains (in Russian*). Bulletin of the Imperial Academy of Sciences of St. Petersburg 7(3):153-162.

Beyond *next **token** prediction*

Further Stages of Training

(not covered here)

- **R**einforcement **L**earning (**RL**): *reward loss function*
  - *accuracy rewards: math*
  - <think> *CoT rewards*
  - **DeepSeek**: 15.6% to 71.0% on AIME 2024
- **RLHF**: *with **H**uman **F**eedback* (**HF**)



(Ouyang et al., 2022; **OpenAI**)

American Invitational Mathematics Examination (AIME) Database (2024)

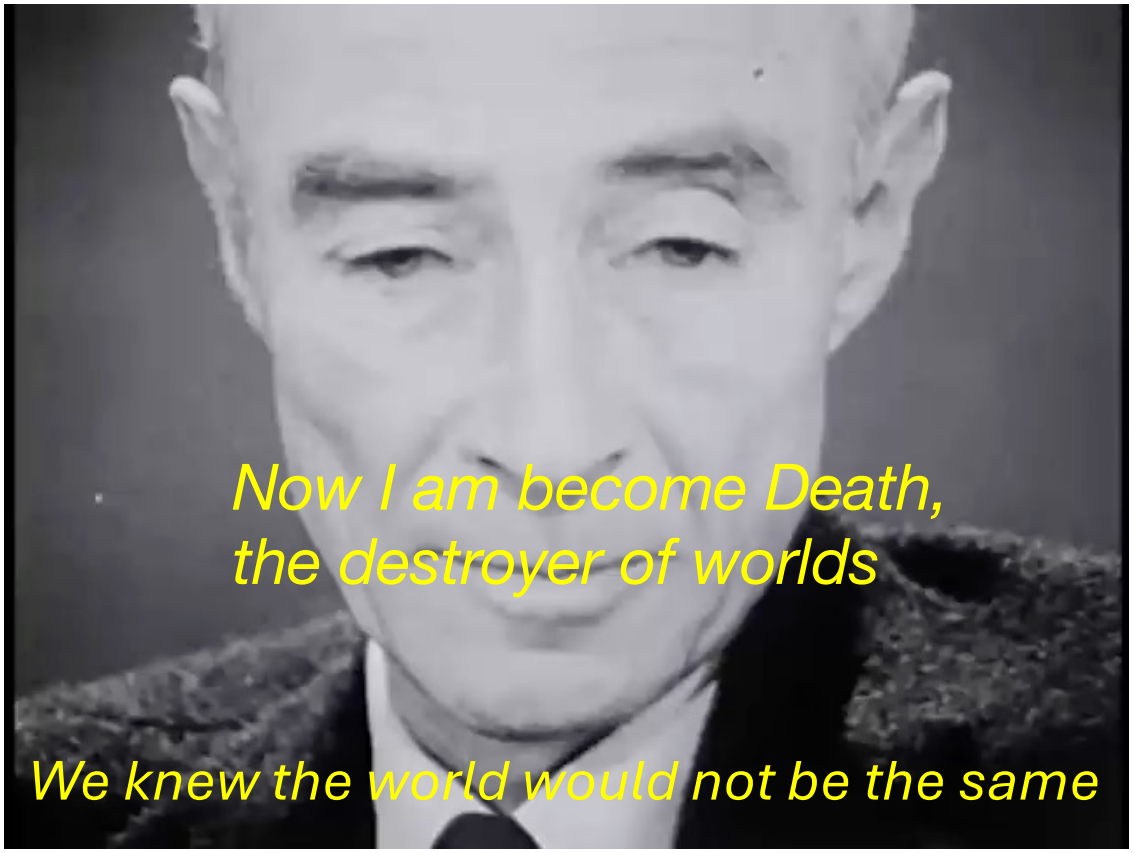
ID string · *lengths* 8↔9 30%	Problem string · *lengths* 477↔549 6.7%	Solution string · *lengths* 284↔657 43.3%	Answer int64 199↔286 10%
2024-II-3	Find the number of ways to place a digit in each cell of a 2x3 grid so that the sum of the two numbers…	Consider this table: \[\begin{array}{|c|c|c|} \hline a & b & c \\ \hline d & e & f\\ \hline \end{array}\] We…	45
2024-I-1	Every morning Aya goes for a 9-kilometer-long walk and stops at a coffee shop afterwards. When she walks at a constant speed of s kilometers per hour, the walk takes her 4 hours, including t minutes spent in the coffee shop. When she walks $s+2$ kilometers per hour, the walk takes her 2 hours and 24 minutes, including t minutes spent in the coffee shop. Suppose Aya walks at $s+⌧rac{1}{2}$ kilometers per hour. Find the number of minutes the walk takes her, including the t minutes spent in the coffee shop.	$\frac{9}{s} + t = 4$ in hours and $\frac{9}{s+2} + t = 2.4$ in hours. Subtracting the second equation from the first, we get, $\frac{9}{s} - \frac{9}{s+2} = 1.6$. Multiplying by $(s)(s+2)$, we get $9s+18-9s=18=1.6s^{2} + 3.2s$. Multiplying by 5/2 on both sides, we get $0 = 4s^{2} + 8s - 45$. Factoring gives us $(2s-5)(2s+9) = 0$, of which the solution we want is $s=2.5$. Substituting this back to the first equation, we can find that $t = 0.4$ hours. Lastly, $s + \frac{1}{2} = 3$ kilometers per hour, so $\frac{9}{3} + 0.4 = 3.4$ hours, or $\framebox{204}$ minutes.	204
2024-II-7	Let N be the greatest four-digit positive integer with the property that whenever one of its digits is…	We note that by changing a digit to 1 for the number $\overline{abcd}$, we are subtracting the number by…	699

huggingface.co/datasets/Maxwell-Jia/AIME_2024

AI's 'Oppenheimer moment'

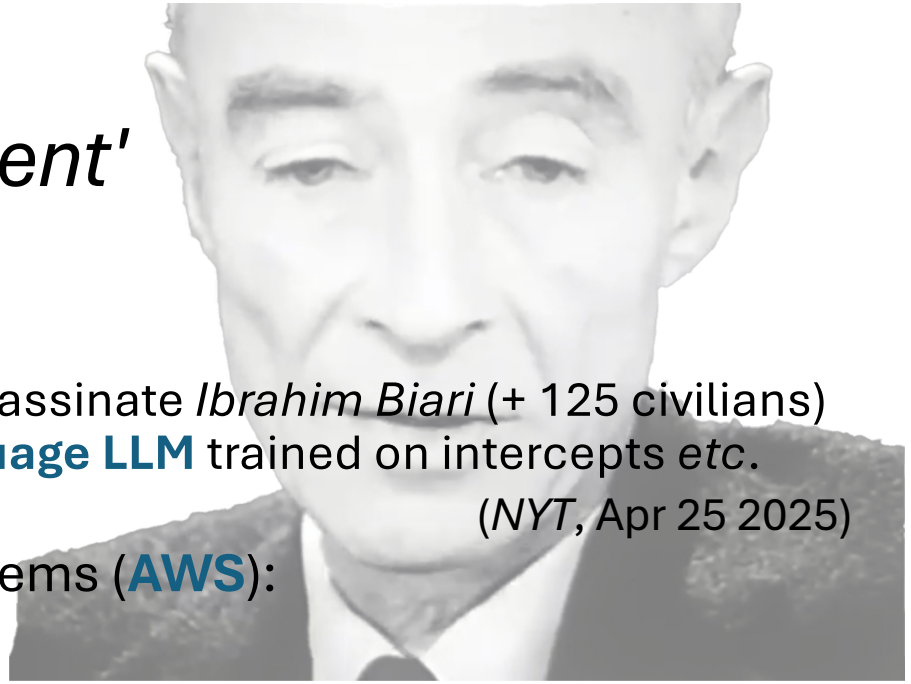
Trinity Test @
5:29am July 16, 1945

(N. Robbins-Early, in *Guardian,* Jul 14 2024)



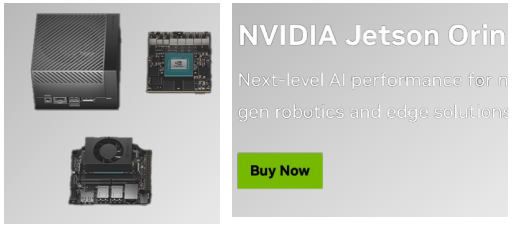
1965 television broadcast

AI's 'Oppenheimer moment'



- Intelligence
  - **AI-infused audio tool** helped assassinate *Ibrahim Biari* (+ 125 civilians) in Oct 2023. Special **Arabic-language LLM** trained on intercepts *etc*.

    (*NYT*, Apr 25 2025)
- Dawn of autonomous weapon systems (**AWS**):
  - Communications can't be jammed.
  - On-board AI to recognize/track targets.
  - **Example:** Russia's *Shahed MS001* drone: *Nvidia Jetson Orin* AI module



AI's 'Oppenheimer moment'

- Used in some US schools

Police swarmed student after AI system mistook bag of chips for gun, officials say

By Khiree Stewart, WBAL via CNN Newsource

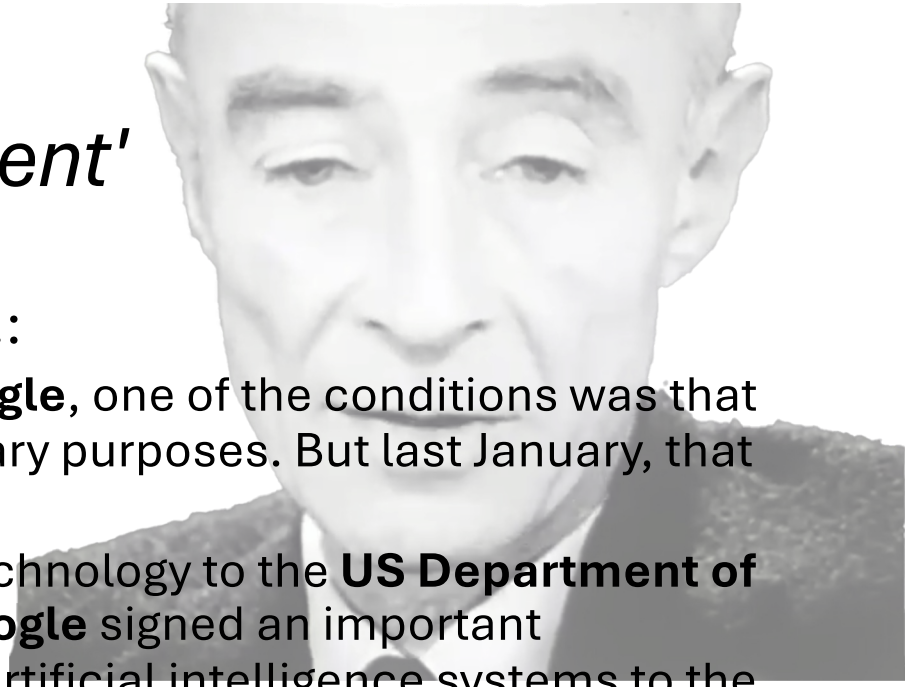
Published: Oct. 23, 2025 at 11:39 PM MST

Updated: 11 hours ago

A teen says he was searched by police after an AI detector mistook a crumpled chip bag for a gun. (WBAL, OMNILERT, CNN)

BALTIMORE COUNTY, Md. (WBAL) – An artificial intelligence security detector led to a terrifying moment for a Maryland high school student after an empty chip bag stuffed in his pocket set off an alert that dispatched police.

AI's 'Oppenheimer moment'

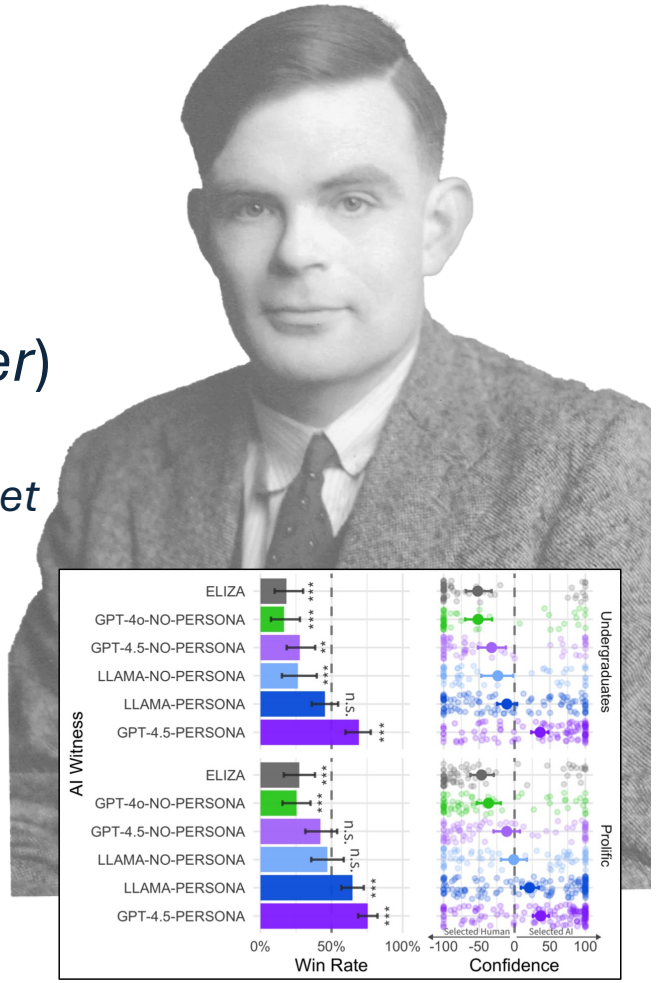


- Big Tech: Google, Microsoft, *etc*.:
  - when **DeepMind** was sold to **Google**, one of the conditions was that the AI would not be used for military purposes. But last January, that condition was removed.
  - now all [AI] companies provide technology to the **US Department of Defense**. And a few days ago, **Google** signed an important agreement with Israel to provide artificial intelligence systems to the armed forces.

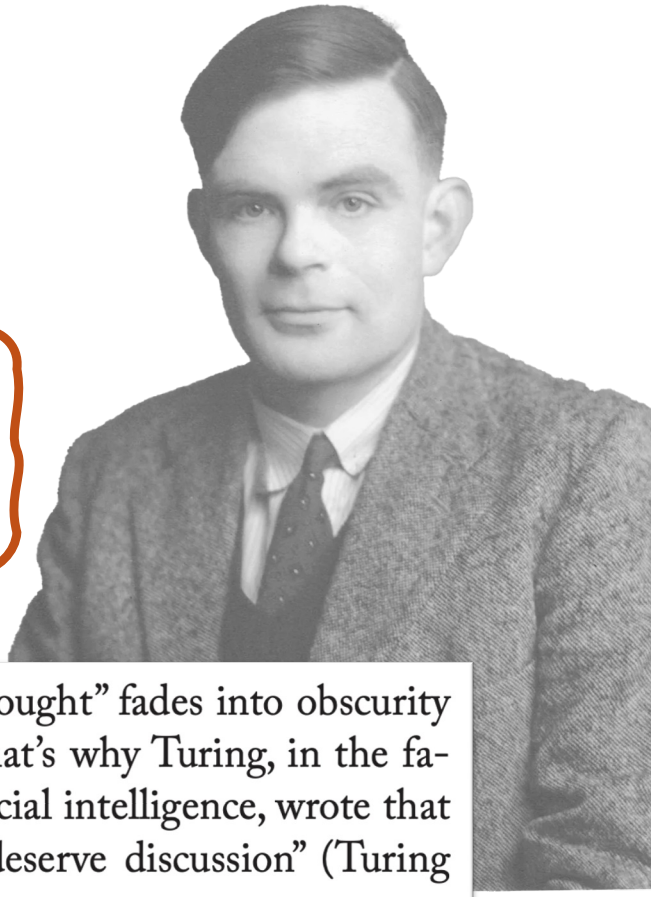
    (R. Luna in a *Corriere Della Sera* interview with Geoffrey Hinton, 9/25/ 2025.)
  - rise of **Sovereign AIs** (*local language*; *national security*; *DeepSeek*)

3-party Turing Test

- **GPT** (*Generative Pre-trained Transformer*)
  - *unsupervised training*, good at *next* ***token*** *prediction* (*better than humans!*) (Shlegeris *et al*. 2024)
- pass the **3-party Turing Test** aka **Imitation Game**: *more human than human!* (Jones & Bergen 2025)
- Turing (1950): **50 year prediction**: ≤ 70% *chance a human can tell the difference*



It will simplify matters for the reader if I explain first my own beliefs in the matter. Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning. The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thi
tha

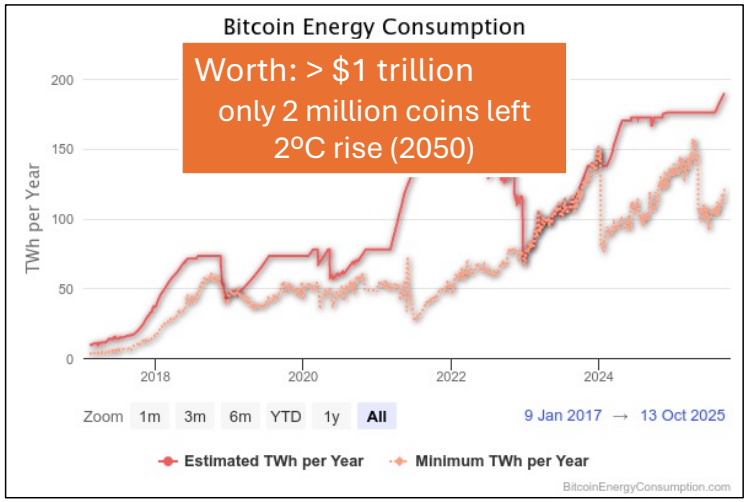


The concept "thought" fades into obscurity insofar as we depart from linguistically formulated thoughts. That's why Turing, in the famous paper on machine thinking that initiated the field of artificial intelligence, wrote that the question whether machines think is "too meaningless to deserve discussion" (Turing 1950).

(Chomsky 2021: fn.7)

Environmental Impact

T =10^{12}



- Power consumption: Thailand
- Carbon footprint: Belgium
- Water use: Switzerland
- *disproportionate impact on communities*

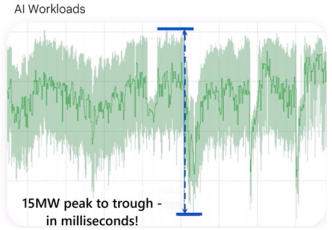
- AI rivals Bitcoin now
  - Query cost: *ChatGPT vs. Google* (10x)
  - **Microsoft AI**: *Three Mile Island* restart

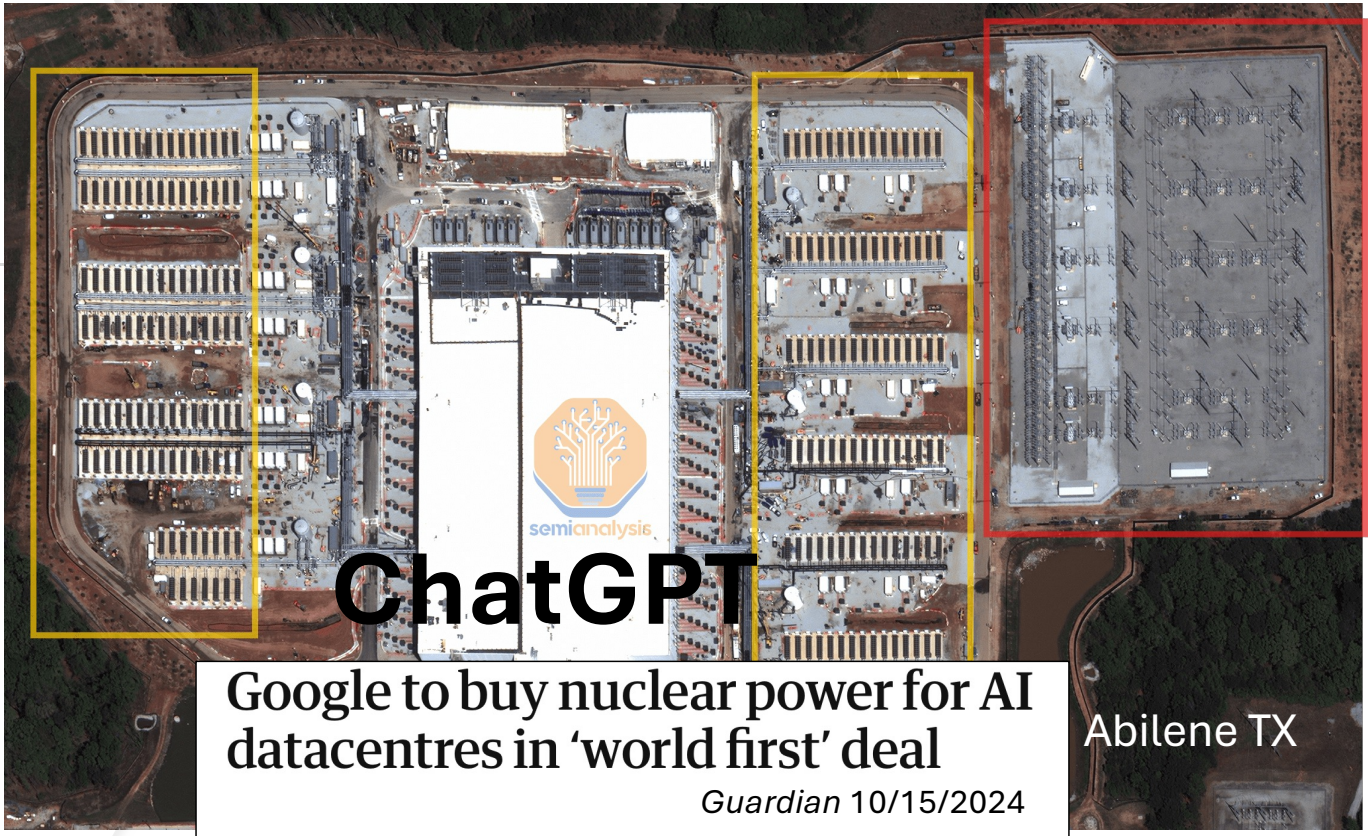
A unit of Pennsylvania's Three Mile Island nuclear plant will be restarted as part of a new energy-sharing agreement with Microsoft, which plans to use it to power the data centers it operates as part of its push into artificial intelligence.

Sources: *including* (Sapra *et al.* 2024), Digiconomist, *aka* A. de Vries-Gao (2025)

OpenAI Stargate datacenter

- ~400MW nameplate (2025), GW (mid-2026)
- 210 chillers
- on-site substation
- projected water use?





Google to buy nuclear power for AI datacentres in 'world first' deal

Guardian 10/15/2024

Tech company orders six or seven small nuclear reactors from California's Kairos Power

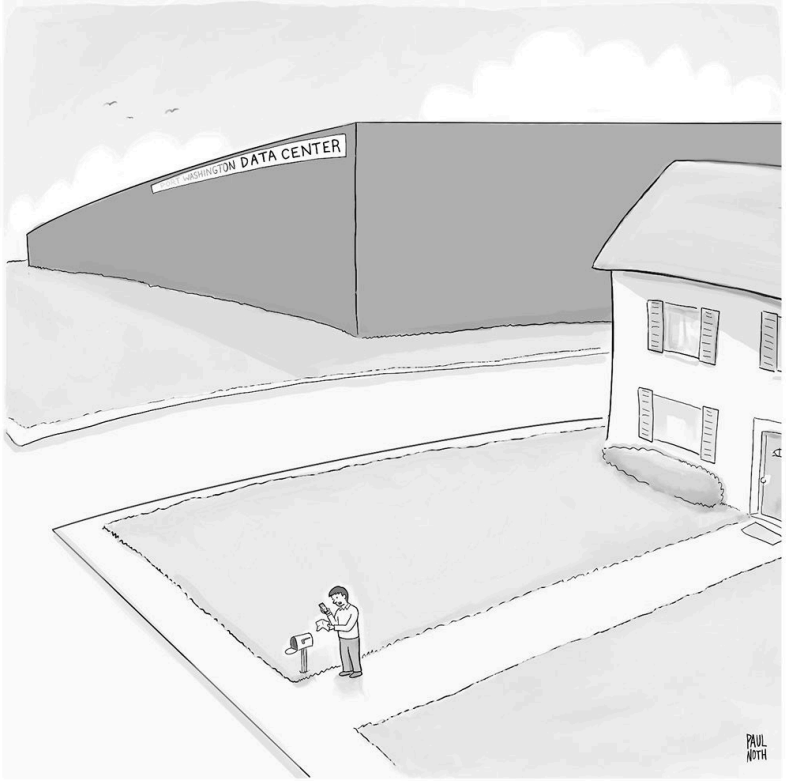
Wholesale cost of electricity

Every day, wholesale electricity prices are measured in real time by Locational Marginal Pricing (LMP) points on the power grid, called nodes. Bloomberg analyzed 25,000 LMP nodes since 2020.

Home > Home Energy & Utilities

The AI Data Center Boom Is Driving Up Electricity Costs, Research Shows

The price of electricity, juiced by demand from power-hungry data centers, is being passed on to residential customers.



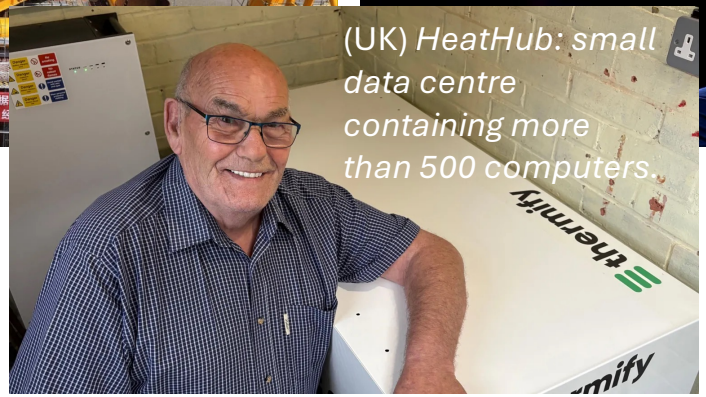
"ChatGPT, why is my electric bill so high?"

NIMBY: *where to put them?*

90% savings in power consumption; commercial deployment, destined for the warm waters of Zhuhai, China

"These giant training clusters, those will be better built in space, because we have solar power there, 24/7. There are no clouds and no rain, no weather."

IMBY:



(UK) *HeatHub: small data centre containing more than 500 computers.*

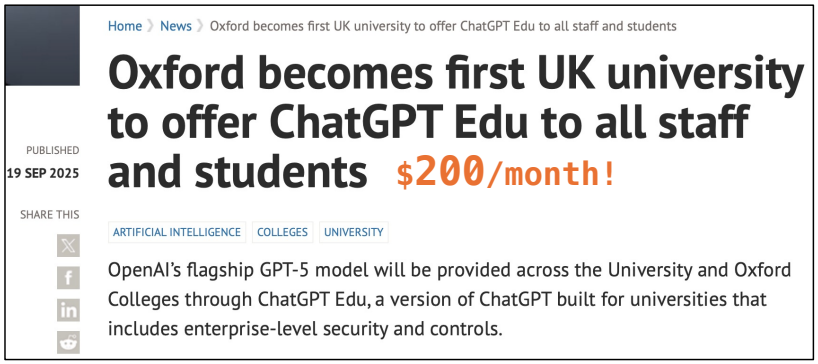
Big Tech AI push on campus

free ($7) smoothies!

University of Arizona

Sep 9 2025

Funding for ChatGPT on campus



Home › News › Oxford becomes first UK university to offer ChatGPT Edu to all staff and students

Oxford becomes first UK university to offer ChatGPT Edu to all staff and students $200/month!

PUBLISHED
19 SEP 2025

SHARE THIS

ARTIFICIAL INTELLIGENCE | COLLEGES | UNIVERSITY

OpenAI's flagship GPT-5 model will be provided across the University and Oxford Colleges through ChatGPT Edu, a version of ChatGPT built for universities that includes enterprise-level security and controls.

www.ox.ac.uk/news

> "University-wide access to ChatGPT Edu will support the development of rigorous academic skills and digital literacy"
> **Professor Freya Johnston**

NextGenAI Initiative:

- OpenAI is committing $50M

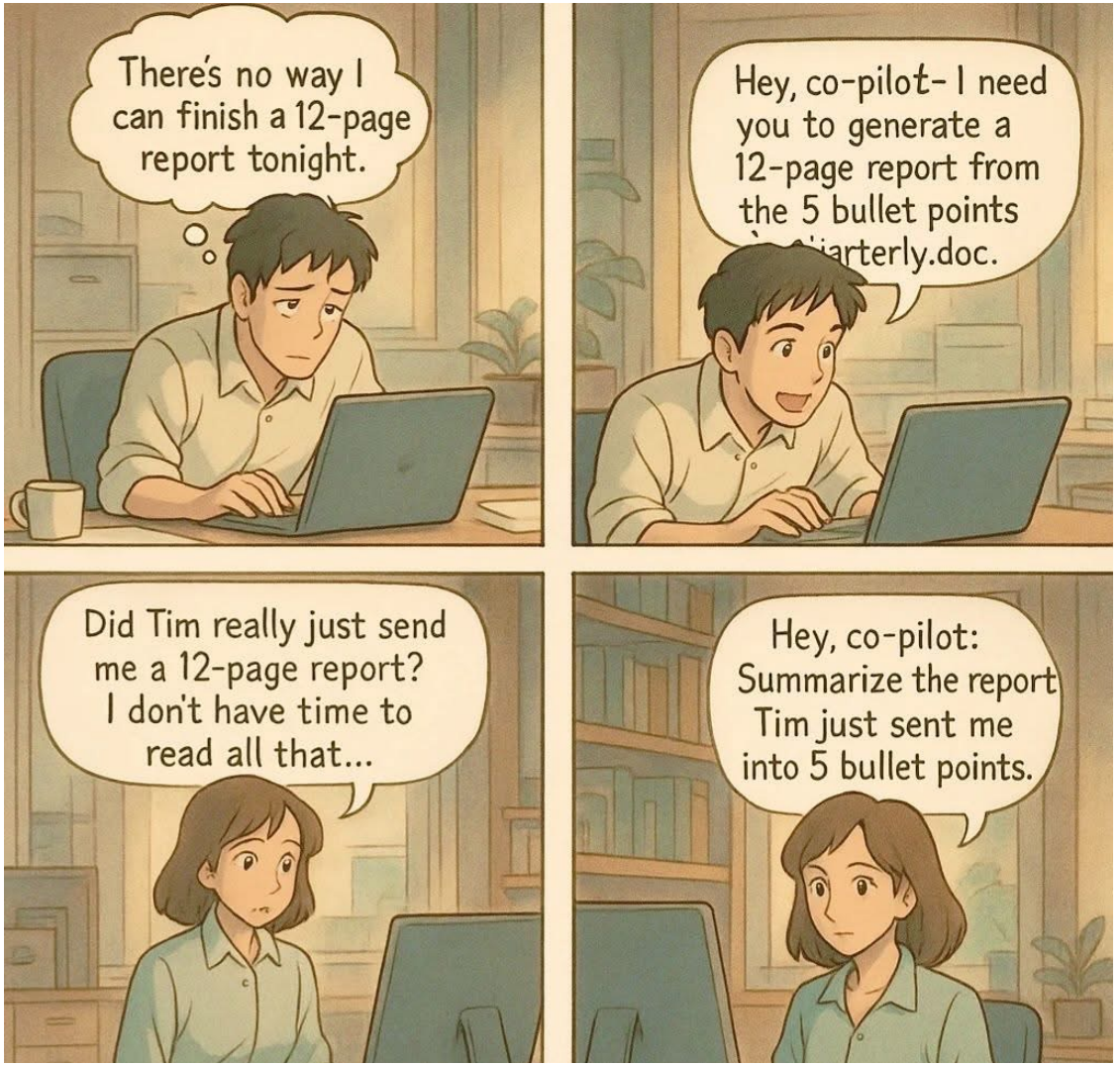
The Next Generation of AI Leaders

NextGenAI's founding partners are Caltech, the California State University system, Duke University, the University of Georgia, Harvard University, Howard University, Massachusetts Institute of Technology, the University of Michigan, the University of Mississippi, The Ohio State University, the University of Oxford, Sciences Po, Texas A&M University, as well as Boston Children's Hospital, the Boston Public Library, and OpenAI.

Each institution is using AI to tackle high-impact challenges, from revolutionizing healthcare to reimagining education. Here are just a few examples of their groundbreaking work:

openai.com/index/introducing-nextgenai/

Student & Professor?



Studying with ChatGPT

Just tap a chat to start

Get help with guidance & quizzes.

I missed class today. Turn this lecture video into bullet point notes|

I don't understand mitosis. Explain cell division like I'm five|

My history test is tomorrow. Quiz me with 10 tough questions|

I'm stuck. Explain the steps to solve this problem|

Try these ideas for school and beyond

I'm studying for exams. Turn my notes into a set of printable flash cards|

I just watched a movie about WWI. Compare two sources on its causes|

My sibling is learning about the water cycle. Make him a diagram|

I tend to procrastinate. Build a study plan so I finish everything early|

Ask ChatGPT for more study tips

Start chatting

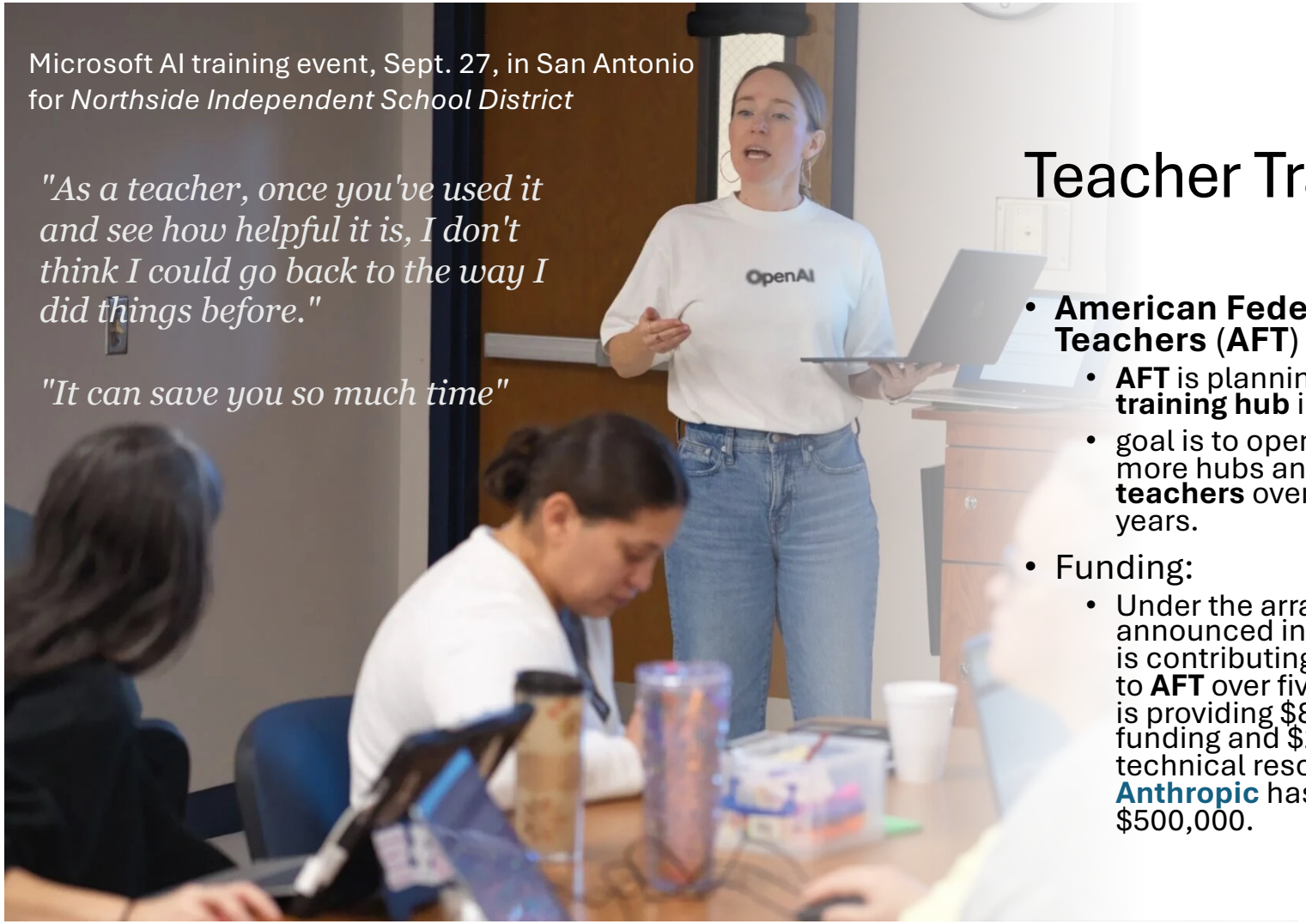
Study Partner: ChatGPT

ChatGPT Tutorial - How to use Chat GPT for Learning and Practicing...

1M views • 3 months ago

Learn Anything FASTER With ChatGPT (13 ChatGPT Prompts For Studying)

220K views • 3 months ago



Microsoft AI training event, Sept. 27, in San Antonio for *Northside Independent School District*

"As a teacher, once you've used it and see how helpful it is, I don't think I could go back to the way I did things before."

"It can save you so much time"

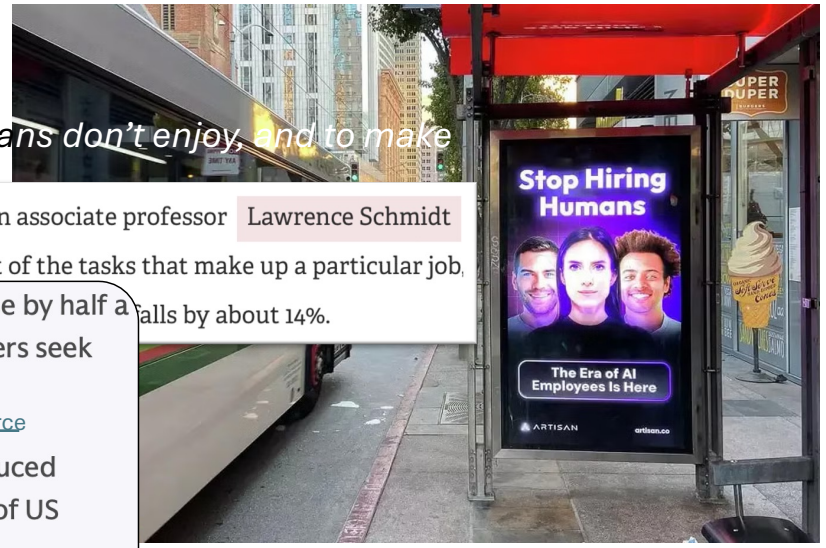
Teacher Training

- **American Federation of Teachers** (**AFT**)
  - **AFT** is planning to build an **AI training hub** in New York City.
  - goal is to open at least two more hubs and train **400,000 teachers** over the next five years.
- Funding:
  - Under the arrangement announced in July, **Microsoft** is contributing $12.5 million to **AFT** over five years. **OpenAI** is providing $8 million in funding and $2 million in technical resources, and **Anthropic** has offered $500,000.

AI and Jobs?

A publicity stunt:

- www.artisan.co/blog/stop-hiring-humans
- *The real goal for us is to automate the work that humans don't enjoy, and to make every job* ***more human***.
- www.nber.org/papers/w33509



A new study co-authored by MIT Sloan associate professor Lawrence Schmidt found that when AI can perform most of the tasks that make up a particular job, alls by about 14%.

- Goldman Sachs Research estimates that unemployment will increase by half a percentage point during the AI transition period as displaced workers seek new positions.
  www.goldmansachs.com/insights/articles/how-will-ai-affect-the-global-workforce
- If current AI use cases were expanded across the economy and reduced employment proportionally to efficiency gains, an estimated 2.5% of US employment would be at risk of related job loss.
- Occupations with higher risk of being displaced by AI include computer programmers, accountants and auditors, legal and administrative assistants, and customer service representatives.

70 Years of Artificial Intelligence (AI) Promises

- **Neural Network AI** hype:
  - *Perceptrons* (Rosenblatt 1958) — AI Winter #1: (Minsky & Papert 1969)
- **Expert Systems** hype:
  - 1985 $1 billion, 2/3rds Fortune 500 — AI Winter #2: 1990s
- **Neural Network AI** hype round 2:
  - 1986 backprop — AI Winter #3?: results plateaued
  - rise of *Statistical NLP*: training datasets
- **Neural Network AI** hype round 3:
  - 2017 *Transformer* invented, 2019 ChatGPT — *peak AI or bubble when?*
  - OpenAI: 700 million weekly users
  - Training: WWW + synthetic (GPT-5). *Superintelligence era has begun* (Altman, 2025)
  - *The end of disease? I think that's within reach. Maybe within the next decade or so, I don't see why not*. (*60 Minutes*. Hassabis, Google/DeepMind, 2025)
  - 2024 Nobel Prizes in physics & chemistry went to AI
  - **AGI** next 5-10 years — *run out of data? Keep scaling up!*

Questions: Hype = Bubble?

- dot-com bubble / 2008 credit crunch / several cryptocurrency bubbles / 2022 NFT bubble

2027 AI bubble?

- *Speculation rules the world. It didn't used to. But from the 1980s through to 2008, something changed. Investors realised that they could get far more return from hype than from any kind of legitimate business.*
- *In fact, nearly half of the world's private investment is being funnelled into AI, and **AI speculation** is the main driving force behind the S&P 500's recent growth.*

(Will Lockett, *Medium* 9/14/2025)

Questions: Hype = Bubble?

- *for cor* *of*
  trainin
- *we find* *any*
  further

022)

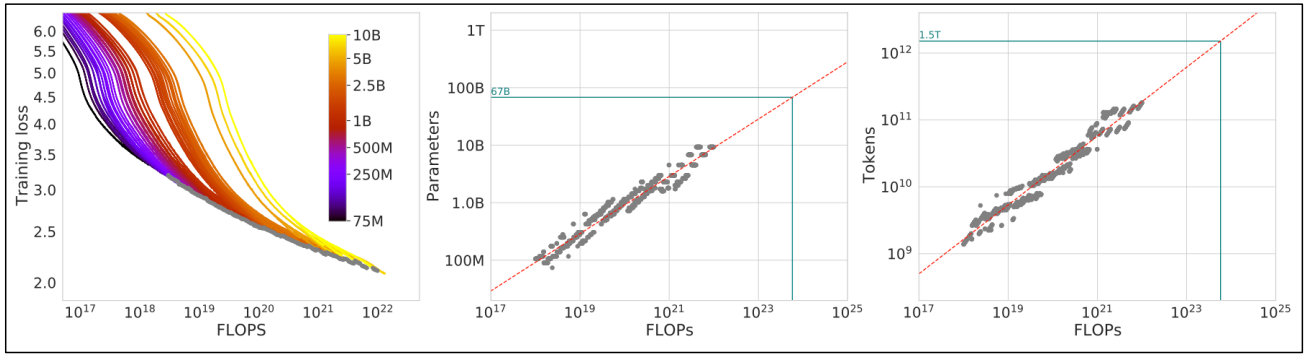
- *No Fre*
- *Floridi*

DeepMind

Training Compute-Optimal Large Language Models

Jordan Hoffmann*, Sebastian Borgeaud*, Arthur Mensch*, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals and Laurent Sifre*

*Equal contributions



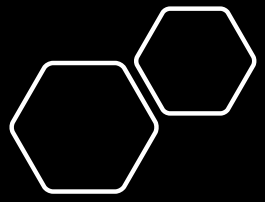
AI Companies Reportedly Struggling to Improve Latest Models

Summarize

Tim Hardwick • Wednesday November 13, 2024 5:30 am PST

Leading artificial intelligence companies including OpenAI, Google, and Anthropic are facing "diminishing returns" from their costly efforts to build newer AI models, according to a new *Bloomberg* report. The stumbling blocks appear to be growing in size as Apple continues a phased rollout of its own AI features through Apple Intelligence.

"*The AGI bubble is bursting a little bit*," said Margaret Mitchell, chief ethics scientist at AI startup Hugging Face.



difficulty in finding "new, untapped sources of high-quality, human-made training data"

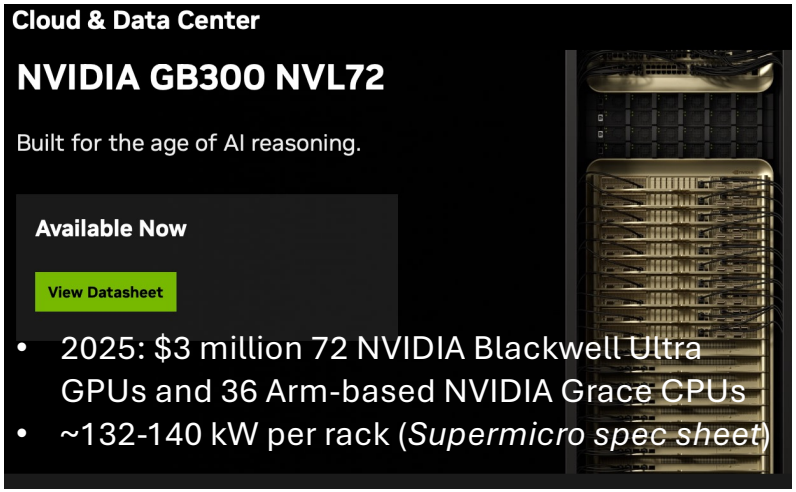
Expensive training

OpenAI Says Its Business Will Burn $115 Billion Through 2029

By Sri Muppidi

OpenAI's First Half Results: $4.3 Billion in Sales, $2.5 Billion Cash Burn

By Stephanie Palazzolo, Amir Efrati and Cory Weinberg



- 2025: $3 million 72 NVIDIA Blackwell Ultra GPUs and 36 Arm-based NVIDIA Grace CPUs
- ~132-140 kW per rack (*Supermicro spec sheet*)

- *Not just the electricity*
- *As I write, I see a report that next year Microsoft plans to buy 150,000 Nvidia chips – at $30,000 a pop.* ***It's a kind of madness.***

(J. Naughton, *The Observer*. 12/30/2023)

$4.5 billion

Nvidia's future Vera Rubin Ultra NVL576 will cost $8.8 million (Barron's 8/25/2025)

"***OpenAI*** *currently massively subsidizing the cost of AI for the consumer."*

"***[NVIDIA]*** *acts as a financial backstop to the entire AI supply chain"*

“10 gigawatts to power the next era of intelligence.”

NVIDIA to put $100 billion into OpenAI from 2026

- Strategic partnership enables OpenAI to build and deploy at least 10 gigawatts of AI datacenters with NVIDIA systems representing millions of GPUs for OpenAI’s next-generation AI infrastructure.
- To support the partnership, NVIDIA intends to invest up to $100 billion in OpenAI progressively as each gigawatt is deployed.
- The first gigawatt of NVIDIA systems will be deployed in the second half of 2026 on NVIDIA’s Vera Rubin platform.

San Francisco and Santa Clara—September 22, 2025—NVIDIA and OpenAI today announced a letter of intent for a landmark strategic partnership to deploy at least 10 gigawatts of NVIDIA systems for OpenAI’s next-generation AI infrastructure to train and run its next generation of models on the path to deploying superintelligence. To support this deployment including datacenter and power capacity, NVIDIA intends to invest up to $100 billion in OpenAI as the new NVIDIA systems are deployed. The first phase is targeted to come online in the second half of 2026 using NVIDIA’s Vera Rubin platform.

Ensure Quality: the shadow (human) workforce

- **Gig workers** (*up to Ph.D level*):
  - censor output (**guardrails**)
  - correcting mistakes
  - better answers
  - *… her task was to enter details on **chemotherapy options for bladder cancer,** which haunted her because she wasn't an expert on the subject.*
  - "I just want people to know that AI is being sold as this tech magic – that's why there's a **little sparkle symbol** next to an AI response," said Sawyer. "But it's not. It's built on the backs of overworked, underpaid human beings."

    (Varsha Bansal, Guardian, 9/11/2025)

- ***Cloud Farming* in India**
  - **Desicrew, NextWealth: *outsourcing AI to the Indian countryside*** (10/13/2025, *BBC*)
  - Examples: transcribing audio to text, AI annotation (for supervised AI training)
  - *"Every AI model, from a ChatGPT-like system to facial recognition, needs vast amounts of human-labelled data. That is the backbone of cloud-farming jobs."*
  - *"In the next 3–5 years, AI and GenAI will create close to 100 million jobs in training, validation, and real-time handling. India's small towns can be the backbone of this workforce."*

Ensure Quality: the shadow (human) workforce

- **Outsourcing Trauma to the Developing World**
  - ***OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*** (2/14/2022, *Time*)
  - [Open AI] said labeling harmful images was "a necessary step" in making its AI tools safer.
  - *ChatGPT and other generative models are not magic – they rely on massive supply chains of human labor and scraped data, much of which is unattributed and used without consent.*

UPDATED: FEBRUARY 17, 2022 10:04 AM EST | ORIGINALLY PUBLISHED: FEBRUARY 14, 2022 7:30 AM EST

In a drab office building near a slum on the outskirts of Nairobi, Kenya, nearly 200 young men and women from countries across Africa sit at desks glued to computer monitors, where they must watch videos of murders, rapes, suicides, and child sexual abuse.

Sama employed content moderators for Facebook, whose jobs involved viewing images and videos of executions, rape and child abuse for as little as $1.50 per hour.

Before we look inside the LLM

Factual Fidelity

The problem is, they have absolutely no commitment to the truth.

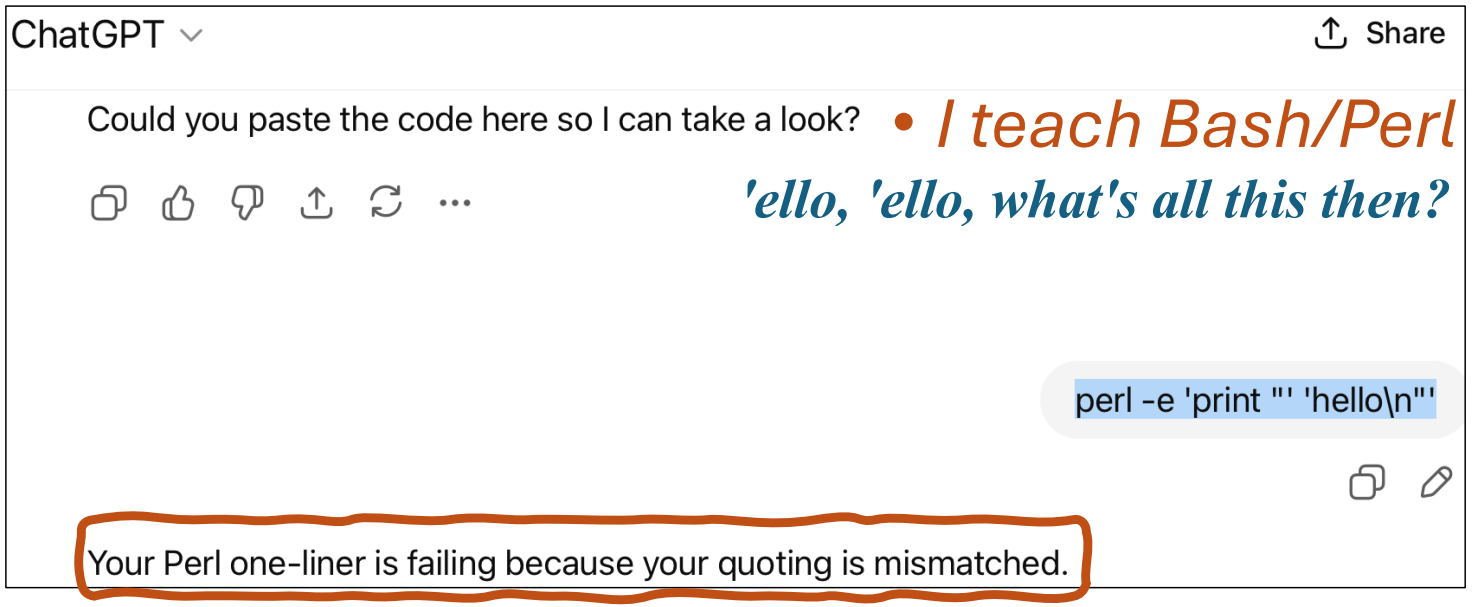
(Emily Bell, *Guardian*, Mar 3 2023)

AI language models are notorious bullshitters, often presenting falsehoods as facts. They are excellent at predicting the next word in a sentence, but they have no knowledge of what the sentence actually means. That makes it incredibly dangerous to combine them with search, where it's crucial to get the facts straight.

(*MIT Technology Review*, Feb 14 2023)

How about understanding?

- Explainable AI: *chain-of-thought* ...



A simple computer language

- Metacognitive calibration? *supreme confidence*

1. **Use double quotes for the Perl code, single quotes inside:**
2. **Use single quotes for the Perl code, escape the single quote inside:**
3. **Concatenate to avoid escaping:**

ode

bash

```
perl -e 'print "'"'"hello";'
```

❌

```
$ perl -e 'print "'"'"hello";'
> "
Can't find string terminator '"' anywhere before EOF at -e line 1.
```

(This is a

starts sir

3.1.2.2 Single Quotes

Enclosing characters in single quotes (‘**'**’) preserves the literal value of each character within the quotes. A single quote may not occur between single quotes, even when preceded by a backslash.

This is Edition 5.3, last updated 18 May 2025, of *The GNU Bash Reference Manual*, for `Bash`, Version 5.3.

Generative AI vs. *Generative Linguistics*

formal system for generating language expressions

Biological basis for the study of language; brain is a biochemical [computer]: Lenneberg (1967; 1969)

Piantadosi vs. Chomsky

- **Modern language models refute Chomsky's approach to language** (Piantadosi, *lingbuzz* 2023; 2024)
- *The rise and success of [**LLMs**] undermines virtually every strong claim for the innateness of language that has been proposed by **generative linguistics**.*
- *Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes.*

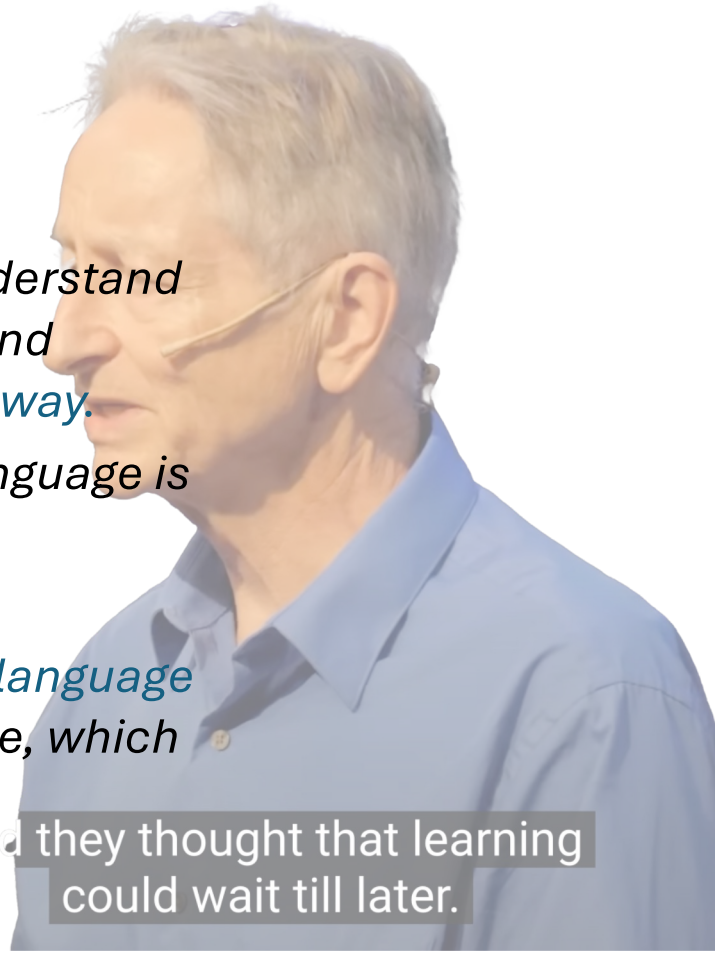
sandiway.arizona.edu: Fong (2025) IJL Volume 37, Issue 1, pp. 59-74.

Futrell and Mahowald (2025)

- How Linguistics Learned to Stop Worrying and Love the Language Models, *Behavioral & Brain Sciences (BBS)* (2025)
  - *Formal linguistics has not presented an alternative model with the demonstrated practical language-learning capabilities of neural models.*
  - *The direction of developments in machine learning suggests that the gap between human and machine learning is more likely to be closed not through built-in domain-specific formal restrictions, but rather through more powerful domain-general learning algorithms.*
  - *Finding these internal mechanisms is not a prerequisite for the reality of linguistic structure — the mechanisms may be irreducibly complex.*
  - *It is rare in the history of science for a scientific theory to turn out as disconnected from a corresponding engineering application as formal generative linguistics has turned out to be for language models.*

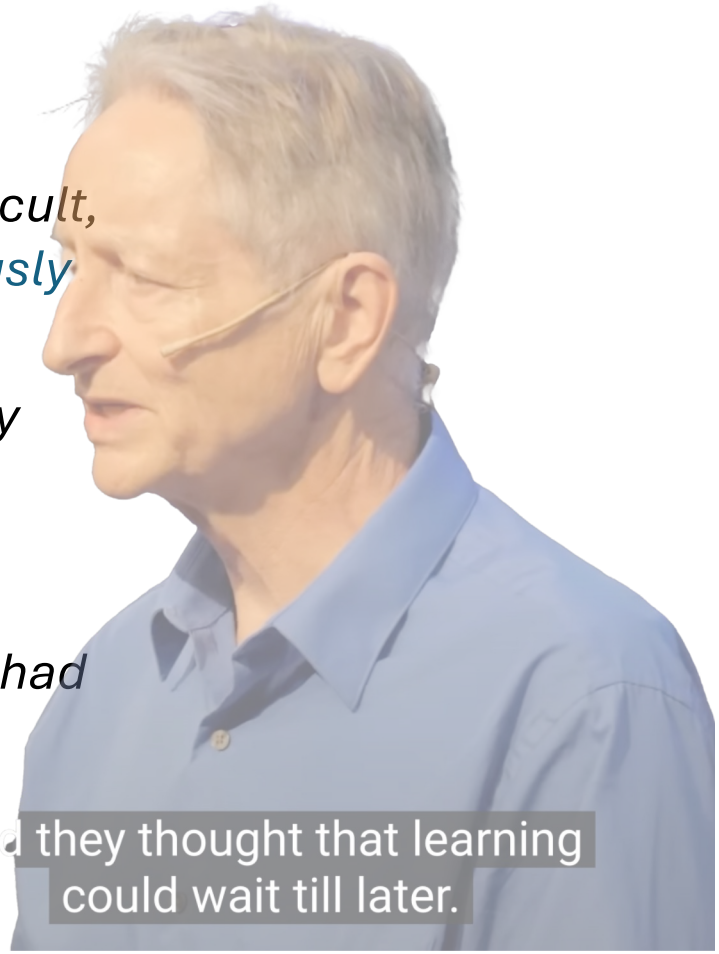
Hinton vs. Chomsky

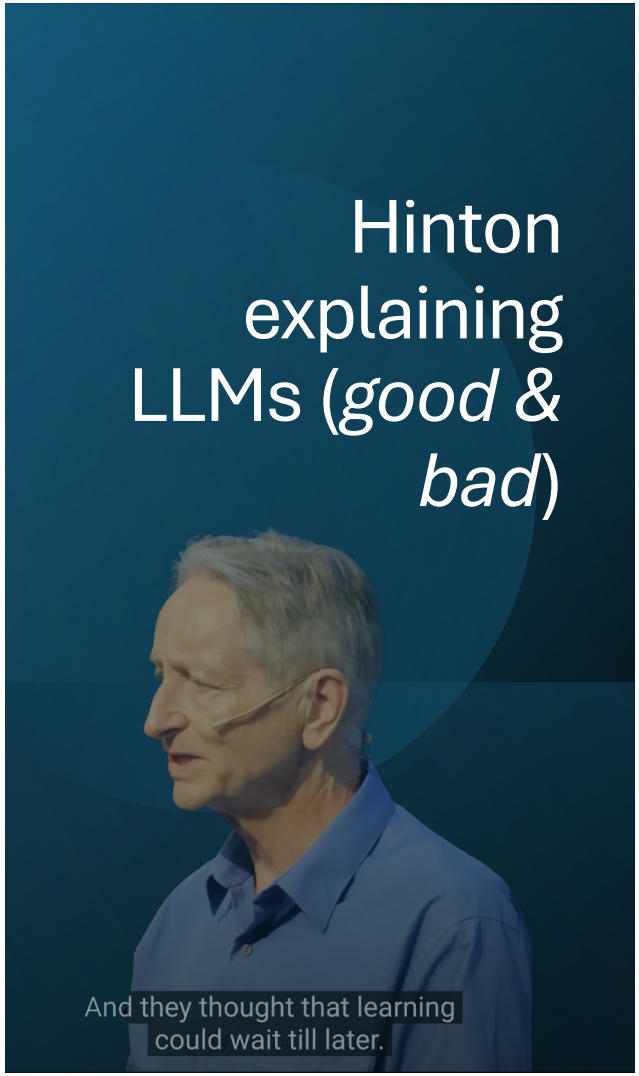
- Hinton *Royal Institution* lecture (2025):
  - *That's what understanding is when you understand language and when these [LLMs] understand language. We understand in just the same way.*
  - *And that's a much better model of what language is than anything the linguists ever had.*
  - *The linguists of course hate it.*
  - *[Linguists] also thought that knowledge of language was innate, knowledge of syntax was innate, which is just stupid.*



Hinton vs. Chomsky

- *It's the sign of a cult, that in order to join the cult, you have to believe something that is obviously silly, like language isn't learned.*
- *There's a guy called Chomsky, when they say things like, "These things don't understand anything, they're just a statistical trick," they don't actually have a model of what understanding is, because they never really had a model of what understanding was.*





- How it works
  - ***Attention:*** *You can't just associate a feature vector with [a word] that captures the meaning directly. You have to hedge your bets. And then as you go up through layers of the network, you disambiguate it using influences from nearby things.*
  - ***Backprop:*** *You're trying to predict the next word.* [...] *it's all done with features and then interactions, when you get it wrong, you back-propagate information that learns all these features and then interactions.*
  - ***That's how language works for us, and it's how it works for these large language models. We are basically the same.***

Learning and Training

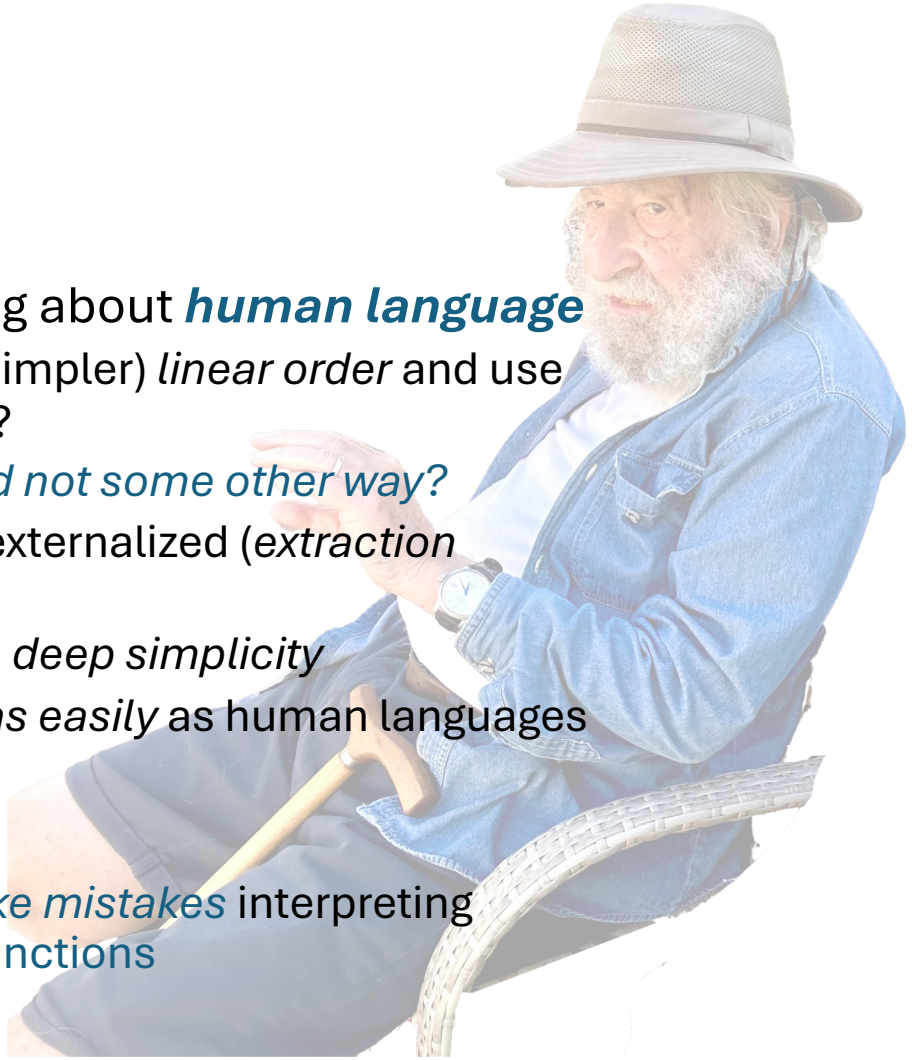
- Human brain: **20W** (*oft-quoted figure*)
  - training: *auto, rapid acquisition* (despite *Poverty of Stimulus*)
  - *data: 4,000 ~ 35,000 words/day @ 12 months* (Brusche *et al*. 2020)
  - *vocabulary: age 1: 50, 3: 1,000, 5: 10,000 words* (Shipley & McAfee 2015)
  - *structure dependence*: 18 months (Chomsky cites Shi *et al*. 2020)
- Meta's Llama-3.1-405B LLM
  - *405B parameters*
  - *training: 31 million hours on H100-80GB* ~3,540 years
  - *TDP:* ***350W***

Chomsky

Even if LLMs work, they tell us nothing about ***human language***

- Why does human language ignore (simpler) *linear order* and use *structural relations* (*Basic Property*)?
- *Why is human language this way and not some other way?*
- Fine thoughts that can't be directly externalized (*extraction asymmetries*)
- Believes in Einstein's *Miracle Creed*: *deep simplicity*
- LLMs learn artificial languages *just as easily* as human languages (*but humans can't, e.g. Moro's work*)

- Generally:
  - Statistical/artificial NN systems *make mistakes* interpreting sentences; lack grammaticality distinctions



Recent Human Evolution

- Explosion of symbolic works in the fossil record
  - *coincides with the appearance of modern humans* (200-300 tya)
  - … until the emergence of behaviorally modern *H. sapiens*: **in general, technological innovations have been sporadic and rare**. The most-striking evidence for a distinct cognitive contrast between modern humans and all their predecessors, however, comes from Europe. *H. sapiens* came late to this continent and brought a new kind of stone tool based on striking long thin "blades" from a carefully prepared long core. **In short order these Europeans, the so-called Cro-Magnons, left a dazzling variety of symbolic works of prehistoric art**. (Tattersall in *Encyclopaedia Britannica*)

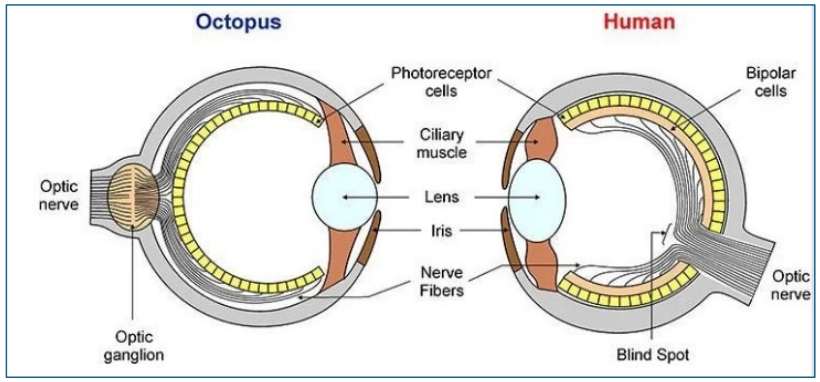
    Last Updated: Aug. 29, 2025
  - *we can fashion tools that amplify these inherent abilities that we have to spectacular magnitudes, e.g. wrt. locomotion, computation, medicine*

Language and Recent Human Evolution

- (Berwick & Chomsky 2016)
  - *Vocal learning and production aspect of* [***EXT***] *is not human-specific* (ancient)
- (Chomsky 2021)
  - Language/thought, **I-Language**, an authentic species property (recent)
  - *Our closest relatives, otherwise intelligent apes, cannot begin to grasp the most elementary rudiments of language even with intensive training. They have about the same auditory system as humans, but acquire nothing from the sounds that lead a human infant, almost reflexively, to develop complex systems for constructing and expressing thought.*
- Basic structure of I-Language should be simple (**Merge *operation***):
  - *the result of some […] small rewiring of the brain […] and has not changed […] since.*
- Modern human ~20K protein coding genes (*1.5% human genome*)
  - Neanderthal/Modern human Y-chromosome divergence ~588 tya (Mendez *et al*. 2016)
  - 14,042 regions of archaic DNA (Neanderthal/Denisovans) (Weiss et al. 2021)
  - *407 […] drove differential expression between the modern and archaic alleles*

Motivation for I-Language

- *Nature adapts/optimizes what it has to work with …*
- Disruptive event: *new entity/functionality* (**I-Language**)
- Reconstruction: *put together a simplest system new/old* (EXT)

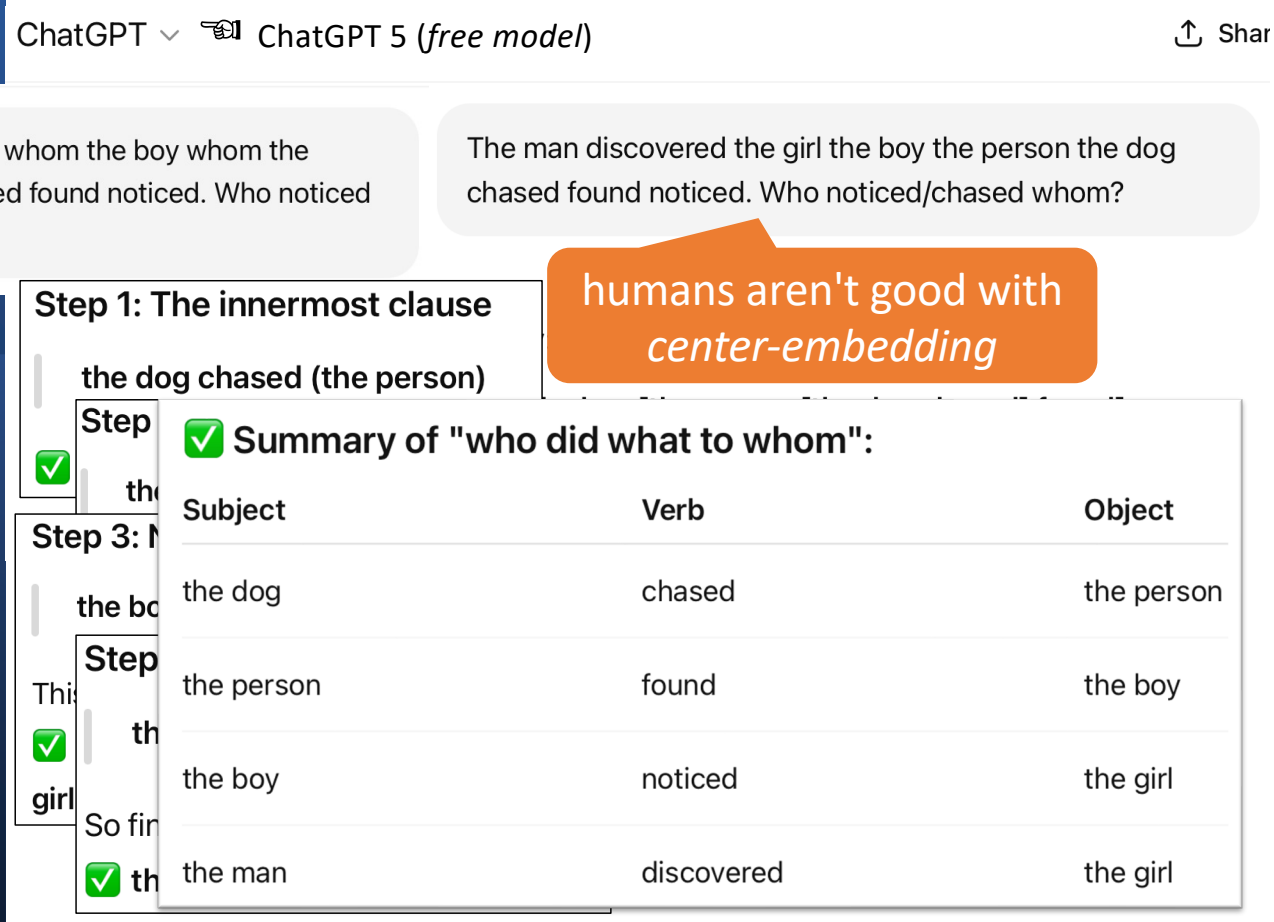


- Example of convergent evolution

- 750-570 mya: **LCA**
- 530 mya: high-res camera eyes (*re-invented many times*) (Land & Fernald 1992) (Land & Nilsson 2012)
- 500 mya: first nervous system
- 100 mya: we lost tetrachromatic vision (*cone cells*)
- 3-4 mya: first human-like brain
- 1-0.2 mya: modern brain
- octopus: "colorblind" (only 1 type of photoreceptor), but employ color (*camouflage*)

Do LLMs work like humans?

adapted from my 438/548 class slides (HW9) 10/9/2025



2️⃣ **Labeled Dependencies (relations)**

```scss
discovered(man, girl)
noticed(boy, girl)
found(person, boy)
chased(dog, person)
disliked(cat, dog)
```

You can think of this as a cascade of predicates:

Even gives a parse tree!

ChatGPT

Sh

csharp

Copy code

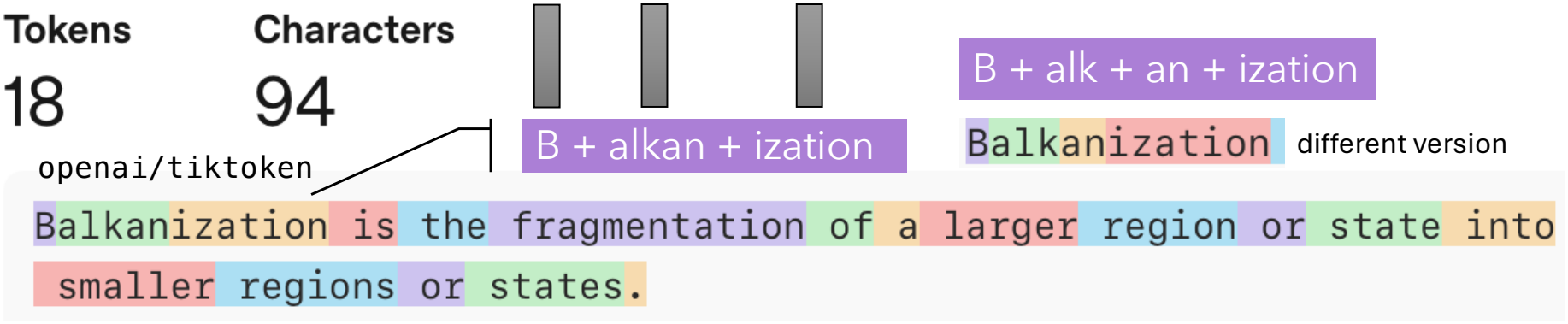
```
[S
  [NP The man]
  [VP discovered
    [NP the girl
      [RelClause
        (the boy
          [RelClause
            (the person
              [RelClause
                (the dog
                  [RelClause
                    (the cat disliked)
                  ] chased
                ) found
              ] noticed
            )
        )
      ]
    ]
  ]
]
```

The man discovered the girl the boy the person the dog the cat disliked chased found noticed. Who noticed/chased/dislked whom?

A quick peek under the hood

Embeddings: vectors for tokens

- *How does an LLM know the meaning of unknown words?*
- **Answer**: it can't
- **Tokenization**: sub-word (~4 characters) (OpenAI's **BPE**)



- **Note**: not a *bag-of-words* scheme: a unique position vector is added to each token vector; e.g. GPT-2: 1024 token context size

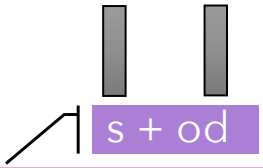
(*from my 438/538 class slides 9/9/2025*)

Language: *infinite employment of finite means*

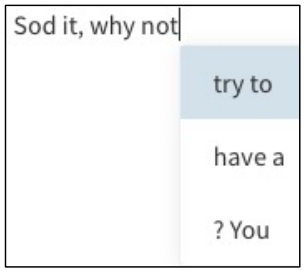
(von Humboldt, cited by Chomsky)

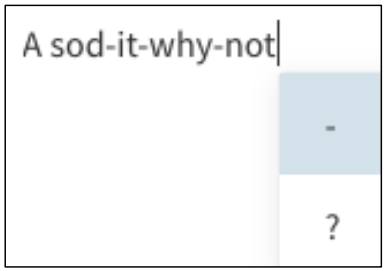
- *Galileo expressed his amazement at what is, in fact, an astonishing fact; with a finite number of symbols one can construct in the mind an infinite number of linguistically formulated thoughts and can even go on to reveal to others who have no access to our minds their innermost workings.*

2023 Keio Lecture 2 (00:36) (Chomsky)



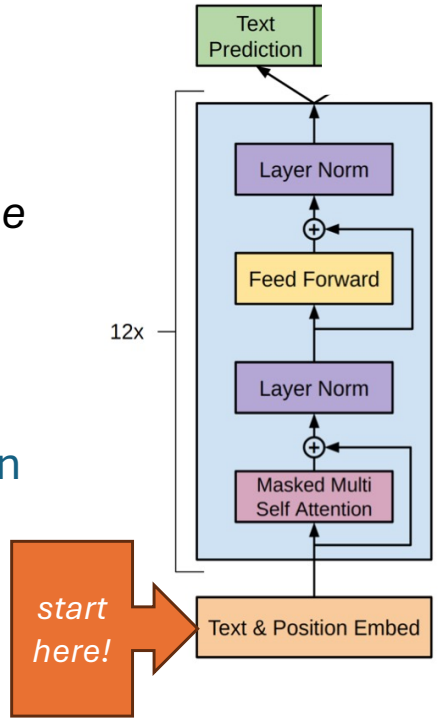
- **Sod it, why not**? (*take a chance*)
- *He chested the ball down, swivelled and cracked a* ***sod-it-why-not*** *shot that took a slight deflection off Evans and beat the diving Onana at the near post.* (Guardian 9/3/2023)





Generative Pre-Training Transformer (GPT)

- Transformer (encoder/decoder)
  - *Attention is all you need*. Vaswani et al. (2017)
  - training is fully parallel: *enormous speed-up*
  - generation is still sequential: *one token at a time*
- **GPT**:
  - *Improving Language Understanding by Generative Pre-Training*. Radford et al. (2018)
  - unsupervised pre-training: next token prediction
  - multi-layer Transformer decoder:
    - multi-headed masked self-attention + feedforward layer



Subject Verb Agreement and Attention

Example (Chomsky 2021):

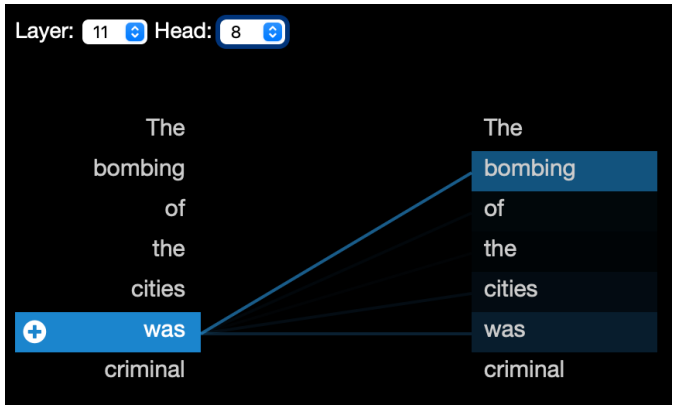
- The *bombing* of the *cities was* criminal

Adapted Examples:

- The *indiscriminate bombing*$_3$ of the *cities was*$_7$ criminal
- The *indiscriminate bombing*$_3$ of the *ancient and modern cities*$_9$ *was* criminal

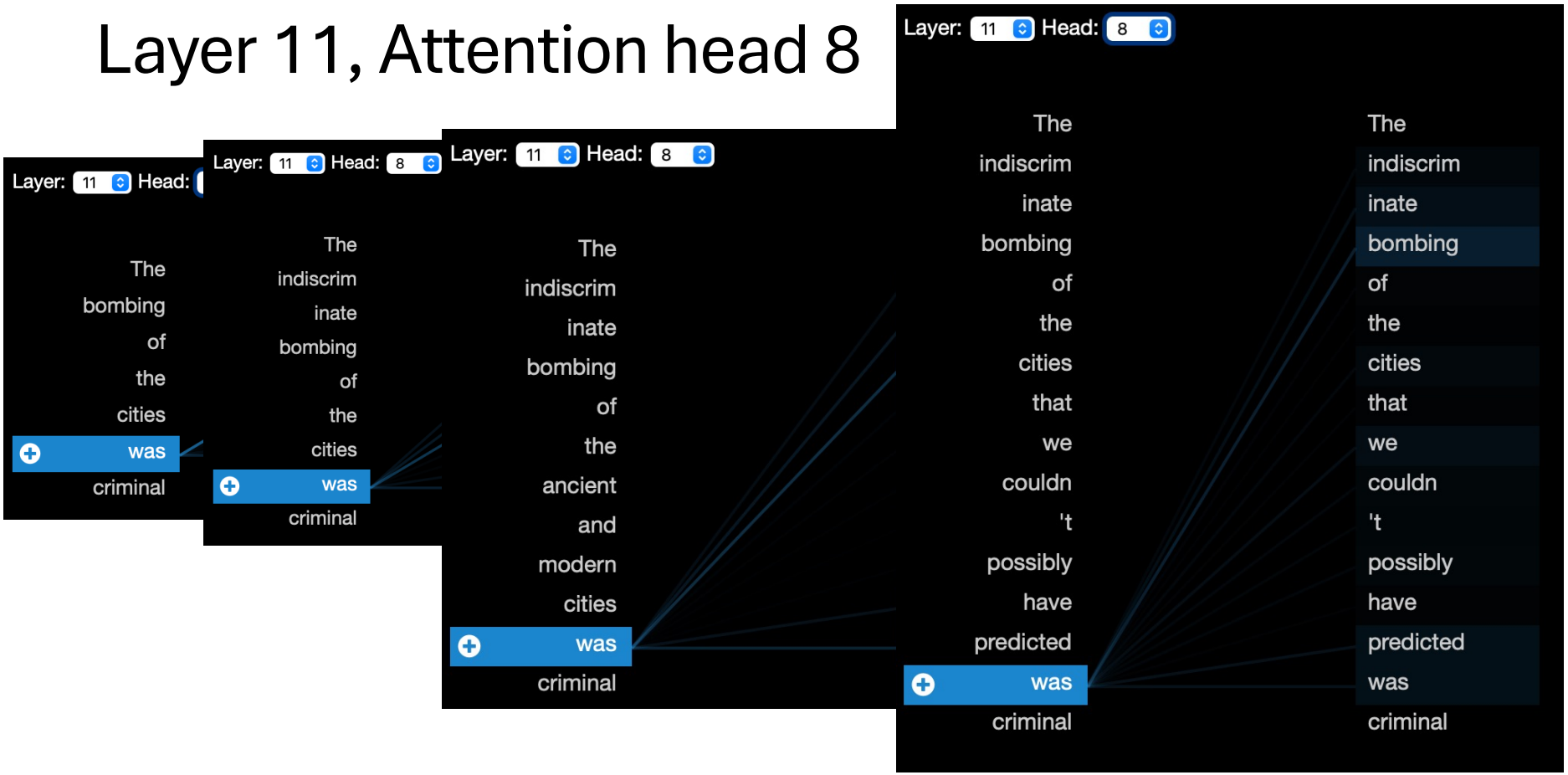
* The *indiscriminate bombing*$_3$ of the *ancient and modern cities*$_9$ *were* criminal

- GPT-2 small (12 levels; 12 heads)



- The *indiscriminate bombing*$_3$ of the *cities that we couldn't possibly have predicted was*$_{13}$ criminal

Layer 11, Attention head 8



Adverb-Verb Construal and Attention

- The mechanic who *carefully* ***fixed*** the car *packed* his tools
- The mechanic who ***fixed*** the car *carefully* ***packed*** his tools
- *Carefully*, the mechanic who *fixed* the car ***packed*** his tools

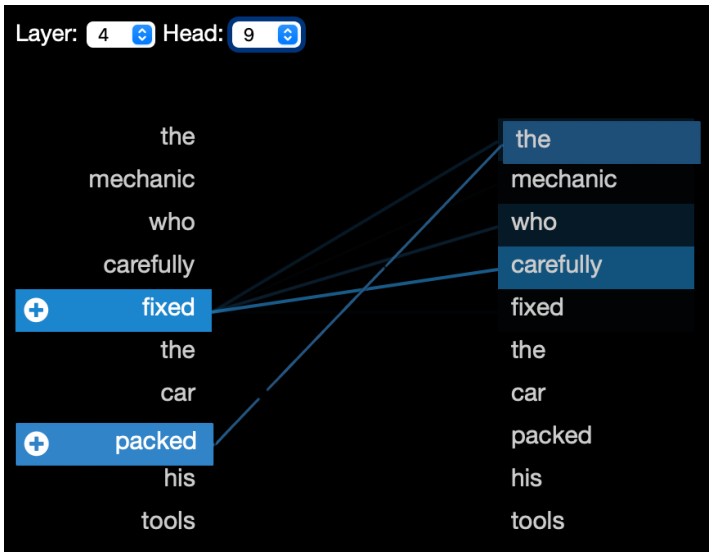
adapted from (Chomsky 2021)

Vig & Belinkov (2019)

5.3.2 Dependency Relations

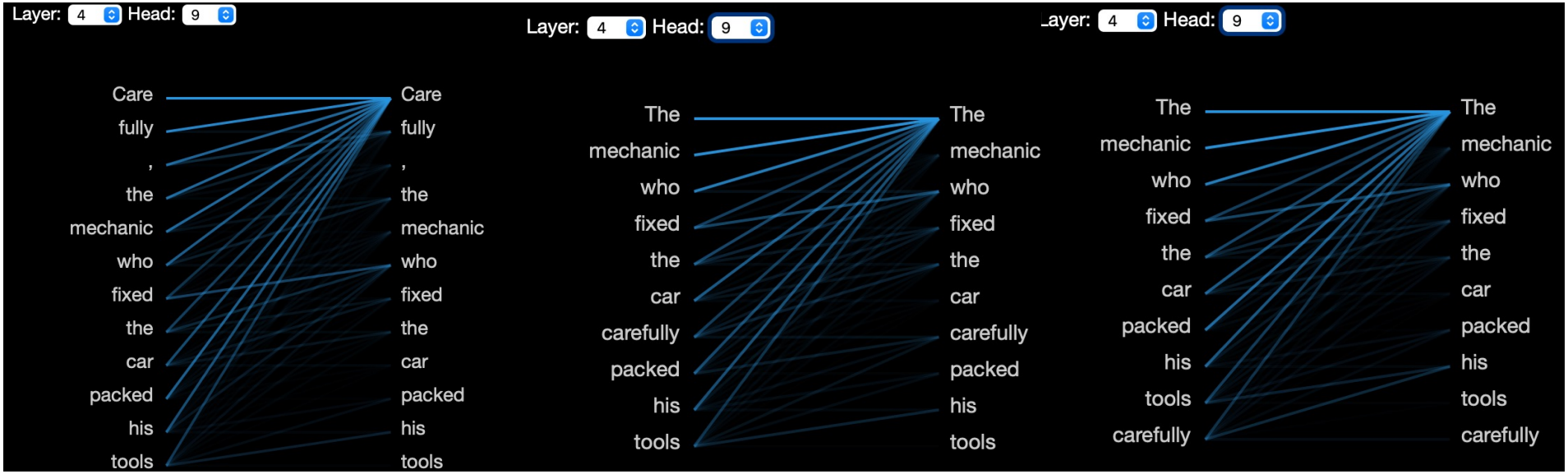
Figure 8 shows the dependency alignment scores (Eq. 4) broken out by layer. Attention aligns with dependency relations most strongly in the middle layers, consistent with recent syntactic probing analyses (Liu et al., 2019; Tenney et al., 2019).

Figure 8: Proportion of attention that is aligned with dependency relations, aggregated by layer. The orange line shows the baseline proportion of token pairs that share a dependency relationship, independent of attention.



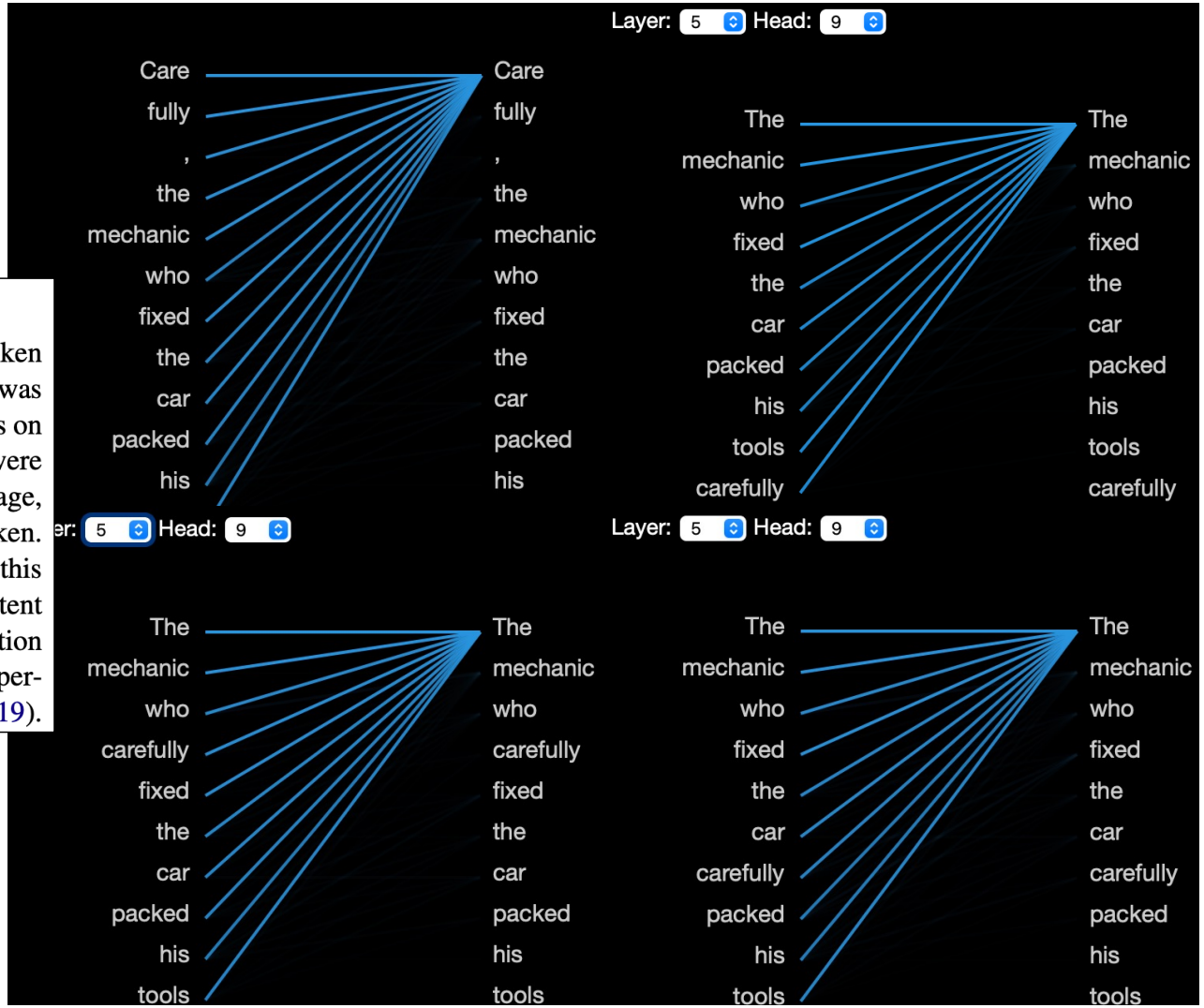
Adverb-verb construal

- Head 9: no sign of dependency relation



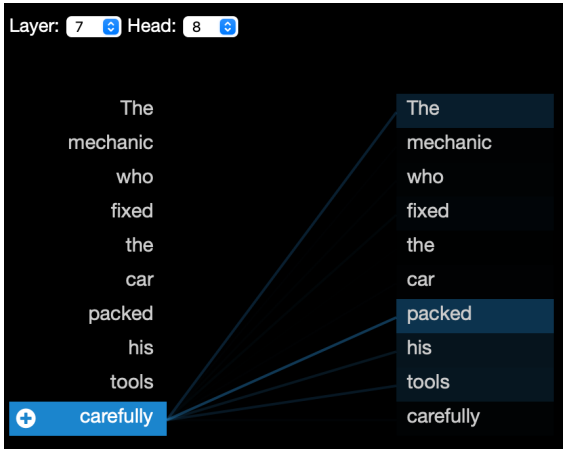
5.2.3 Filtering Null Attention

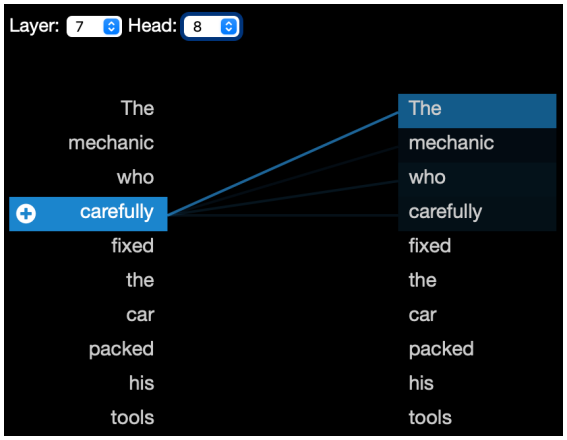
We excluded attention focused on the first token of each sentence from the analysis because it was not informative; other tokens appeared to focus on this token by default when no relevant tokens were found elsewhere in the sequence. On average, 57% of attention was directed to the first token. Some heads focused over 97% of attention to this token on average (Figure 5), which is consistent with recent work showing that individual attention heads may have little impact on overall model performance (Voita et al., 2019; Michel et al., 2019).

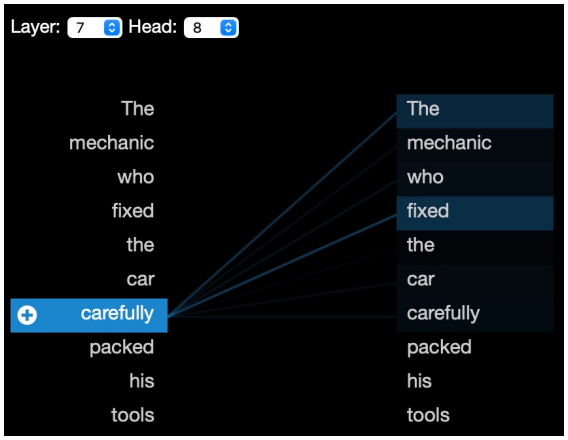


Adverb-verb construal

- Layer 7 Head 8







- Joke works because of two particular senses of *badly*

1. *so badly*: desperately/intensely/acutely
2. *badly*: poorly/incompetently/ineffectively

***Hinton on Attention** (Royal Institution Lecture):* *You can't just associate a feature vector with [a word] that captures the meaning directly. You have to hedge your bets. And then as you go up through layers of the network, you disambiguate it using influences from nearby things.*

†Sentence #	Single clause	+ Context
1. *desperately/poorly*	0.722/0.596	0.778/0.566
2. *desperately/poorly*	0.565/0.920	0.589/0.909

†*(prelim.) see next slide*

- *special thanks to Braa Oudeh (previous slide), Masayuki Oishi*