



INSIGHT
PHILANTHROPY
RESULTS

EXPLORE

PD25

New Depths

August 19-22, 2025

Hilton Baltimore Inner Harbor Hotel, Baltimore, Maryland

The Inner Workings of LLMs

WHAT'S IN A MODEL

EXPLORE
PD25
New Depths

A decorative, light blue swirl graphic is positioned to the right of the text 'PD25' and 'New Depths'.

WHAT'S IN A MODEL?

- What are LLMs and where did they come from?
- Why are they awesome?
- Why are they terrible?
- Privacy, Security, Ethics



ABOUT ME

- BMath in Computer Science, University of Waterloo '05
- Run CharityCAN, prospect research software for Canadian fundraisers
- Used and programmed with deep learning and language models for a few years



The precursors to large language models

NEURAL NETWORKS

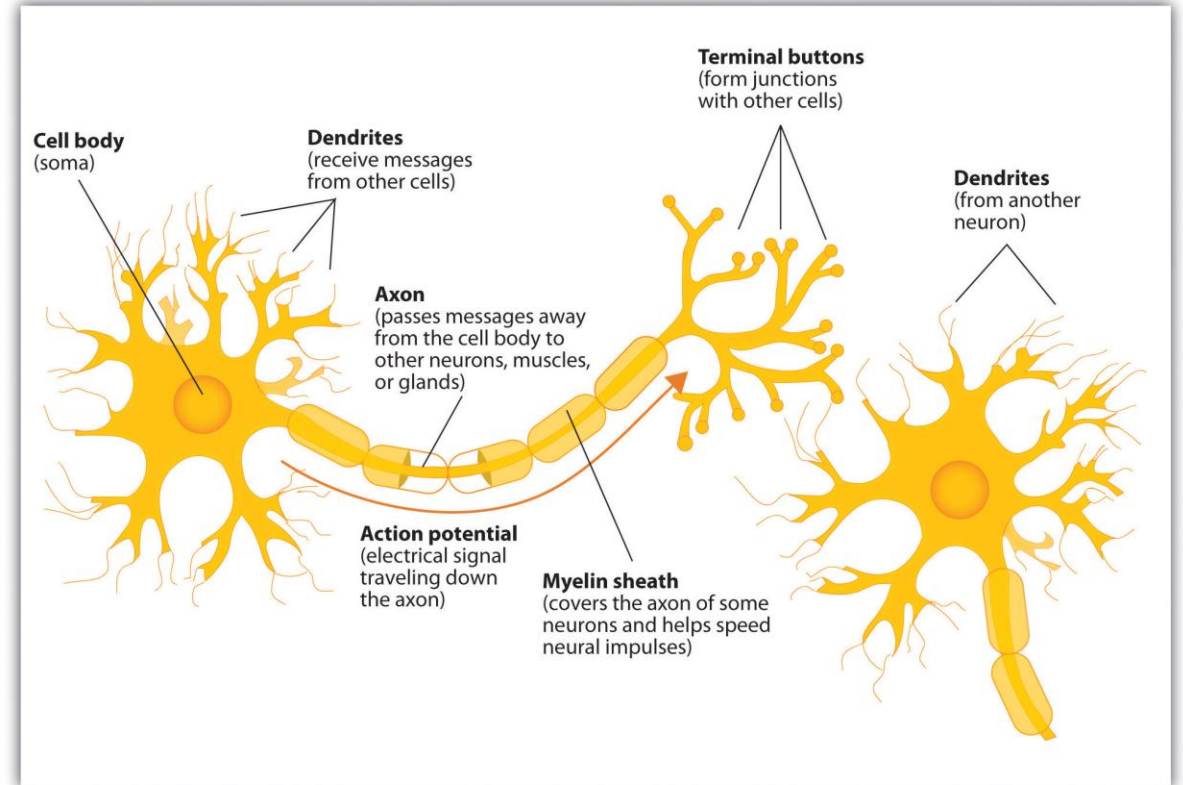
EXPLORE
PD25
New Depths

ARTIFICIAL NEURAL NETWORKS

A Logical Calculus of the Ideas Immanent in Nervous Activity

Warren S. McCulloch and Walter
Pitts, 1943

- First representations were non-learning

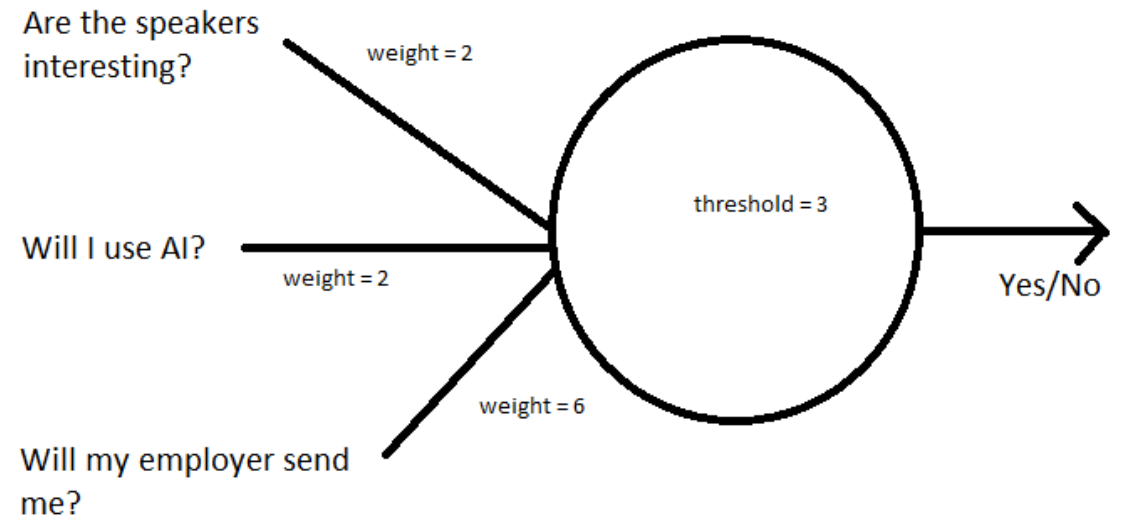


PERCEPTRONS

The Perceptron: A Perceiving and Recognizing Automaton

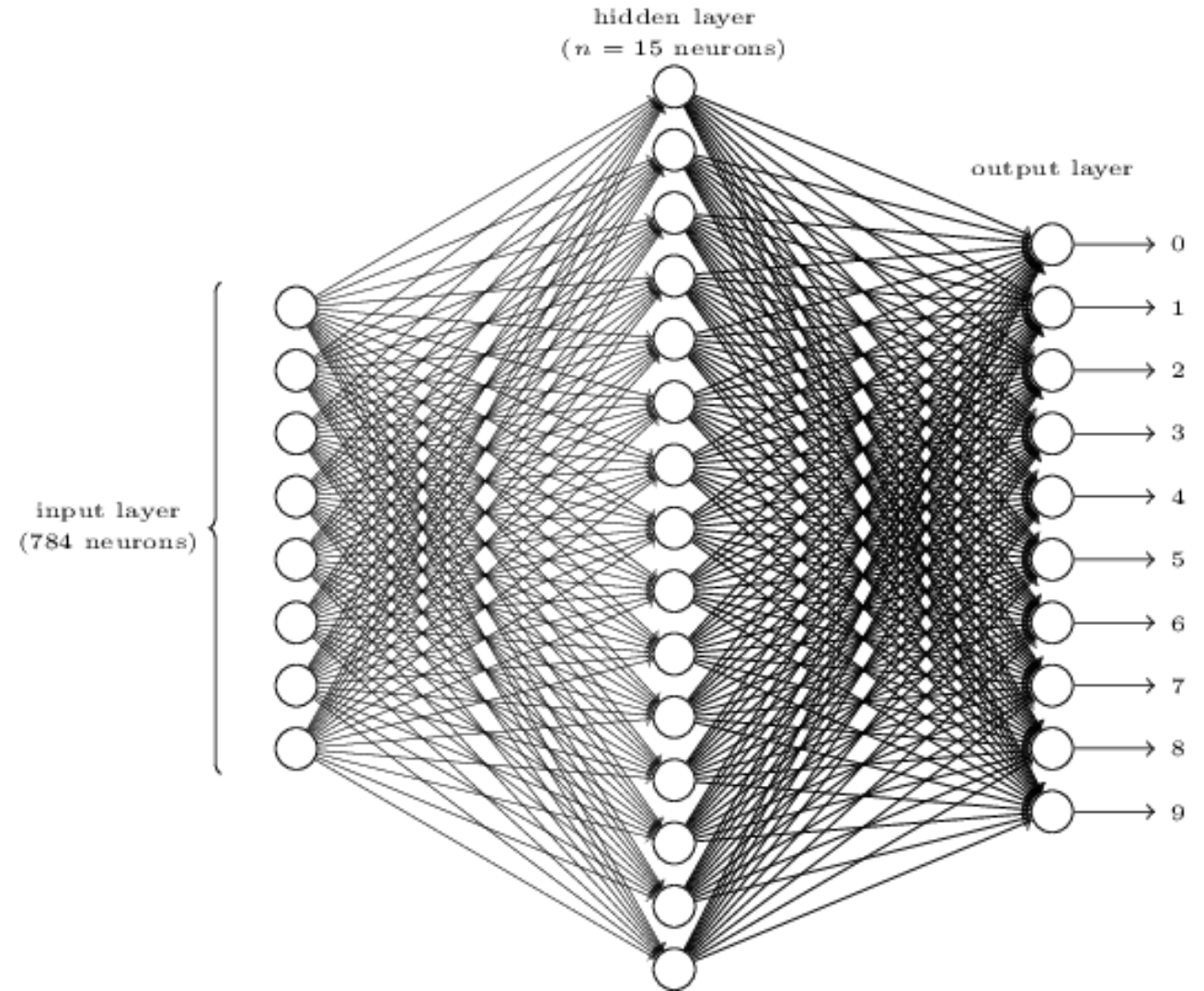
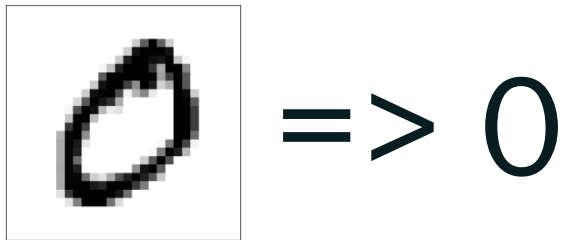
Frank Rosenblatt, 1957

- Designed for binary classification tasks
- Limited by the problems it could solve (Marvin Minsky and Seymour Papert, 1960)
- Limitations led to decreased interest



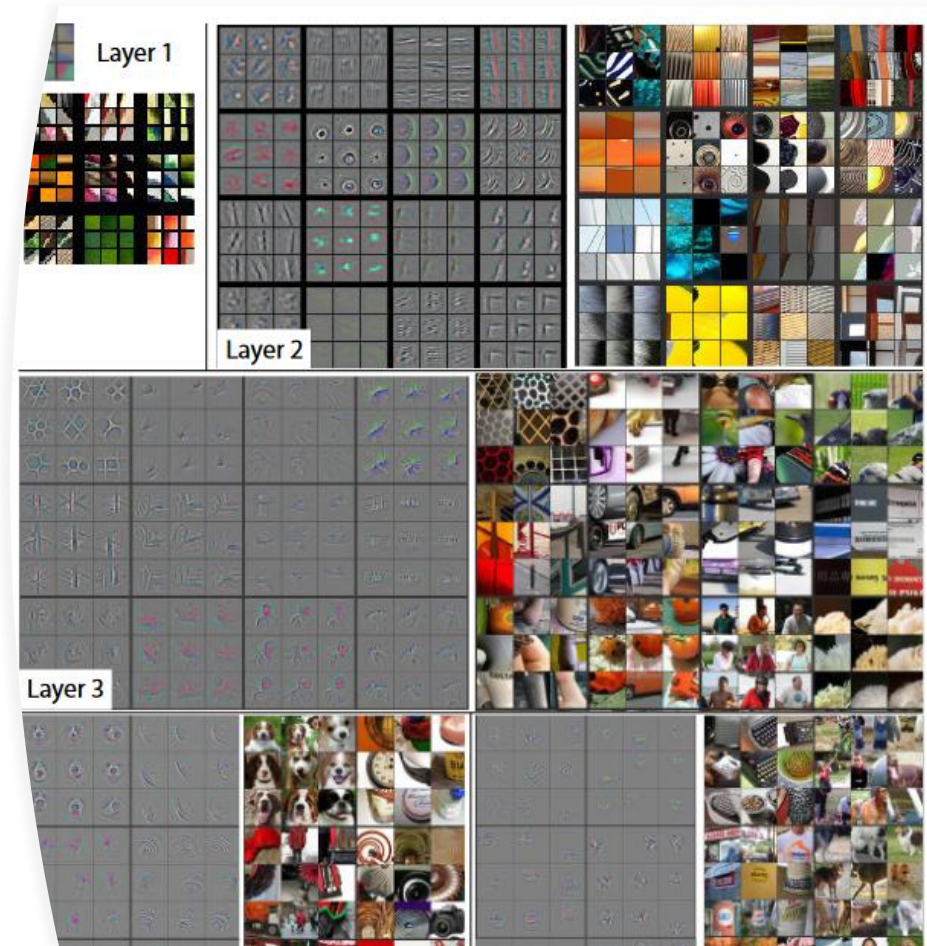
NEURAL NETWORKS

- The underlying frameworks that make large language models possible
- Connected perceptrons (nodes) in layers perform calculations and pass values on to produce a solution
- Solutions are learned through process called training



DEEP LEARNING

- Neural networks with many nodes and many layers and trained on a LOT of data
- Pioneered at UofT in 2012 by Geoffrey Hinton
- Technology behind the last decade's advancements in AI



Representing Text in Neural Networks

MODELLING LANGUAGE

EXPLORE
PD25
New Depths

DISTRIBUTIONAL HYPOTHESIS

We can understand new words from contexts we recognize

With that I scuttled down the companion with all the noise I could, slipped off my shoes, ran quietly along the sparred gallery, mounted the forecastle ladder, and popped my head out of the fore companion.

– *Treasure Island*



CONTEXT AND MEANING

Senses

mouse (N)

1. any of numerous small rodents...
2. a hand-operated device that controls a cursor...

Word Similarity

- Cat and dog

Word Relatedness

- Coffee and cup



VECTORS FROM WORDS

- Vectors are a list of numbers that define direction and length (or magnitude) – think of an arrow pointing north on a map for 5 km
- Vectors can have many dimensions
- We can examine the context around words to produce vector (numerical) representations of them

	aardvark	...	computer	data	result	pie	sugar	...
cherry	0	...	2	8	9	442	25	...
strawberry	0	...	0	0	1	60	19	...
digital	0	...	1670	1683	85	5	4	...
information	0	...	3325	3982	378	5	13	...

Figure 6.6 Co-occurrence vectors for four words in the Wikipedia corpus, showing six of the dimensions (hand-picked for pedagogical purposes). The vector for *digital* is outlined in red. Note that a real vector would have vastly more dimensions and thus be much sparser.

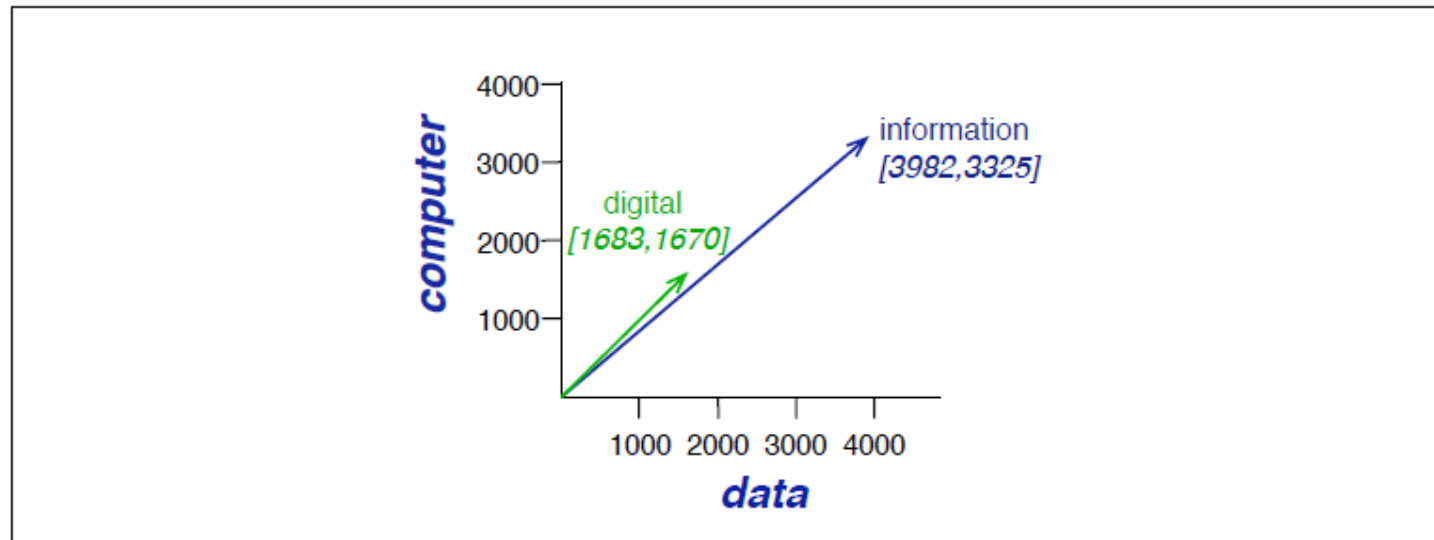


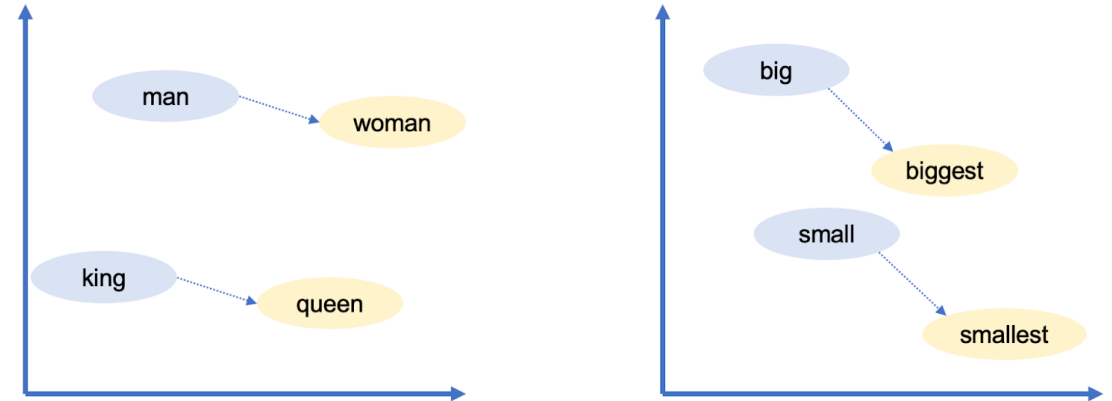
Figure 6.7 A spatial visualization of word vectors for *digital* and *information*, showing just two of the dimensions, corresponding to the words *data* and *computer*.

WORD2VEC AND SEMANTIC RELATIONSHIPS

Efficient Estimation of Word Representations in Vector Space

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, 2013

- Used a very large data set (millions of documents from Google News)
- The word vectors produced by word2vec could be added and subtracted to make deductions about words
- Words with multiple meanings still only have one vector

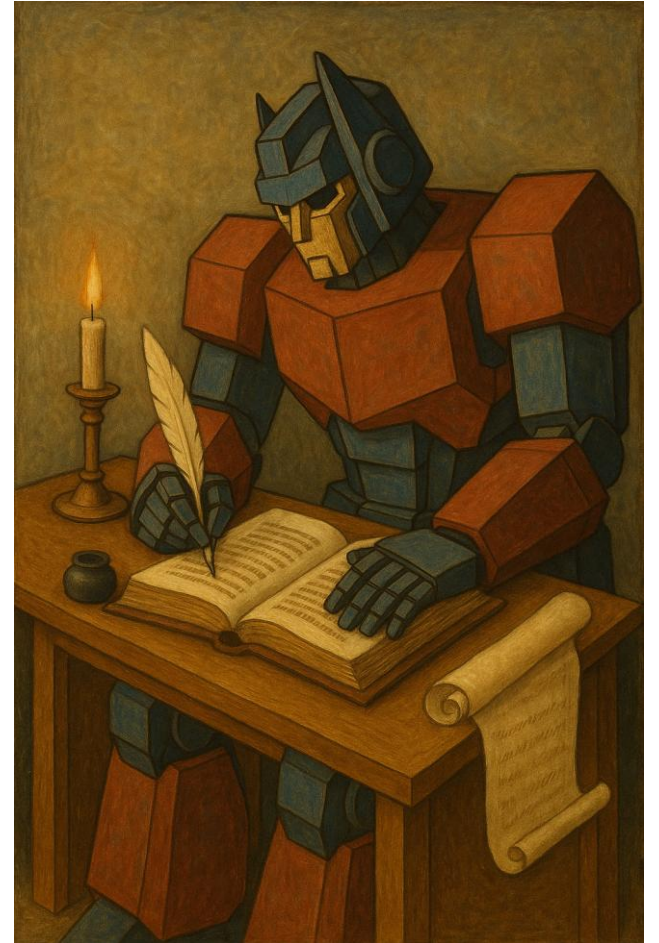


LARGE LANGUAGE MODELS

EXPLORE
PD25
New Depths

LARGE LANGUAGE MODELS

- LLMs are based on a neural network architecture called “transformers”
- Their task is to predict the next token (or “word”) given the preceding tokens
- Trained on a huge corpus of human (and computer!) language and refined by human feedback
 - Pre-training vs. fine-tuning vs. reinforcement learning

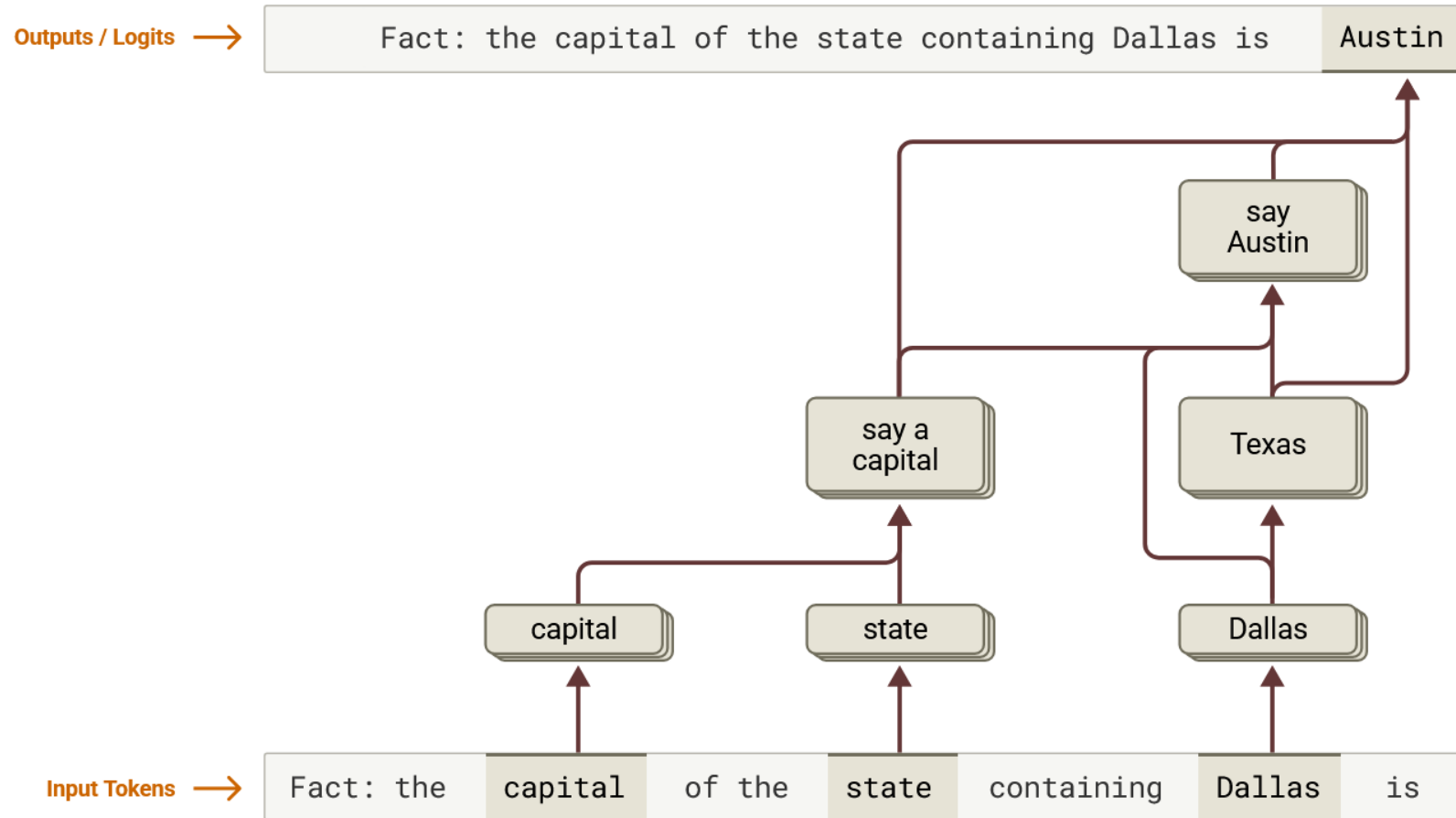


A BRIEF ASIDE ABOUT PROMPTS

- Prompts are the inputs to an LLM
- Two main types:
 - User Prompts
 - System Prompts
- Prompts as conversations



HOW DOES AN LLM WORK?

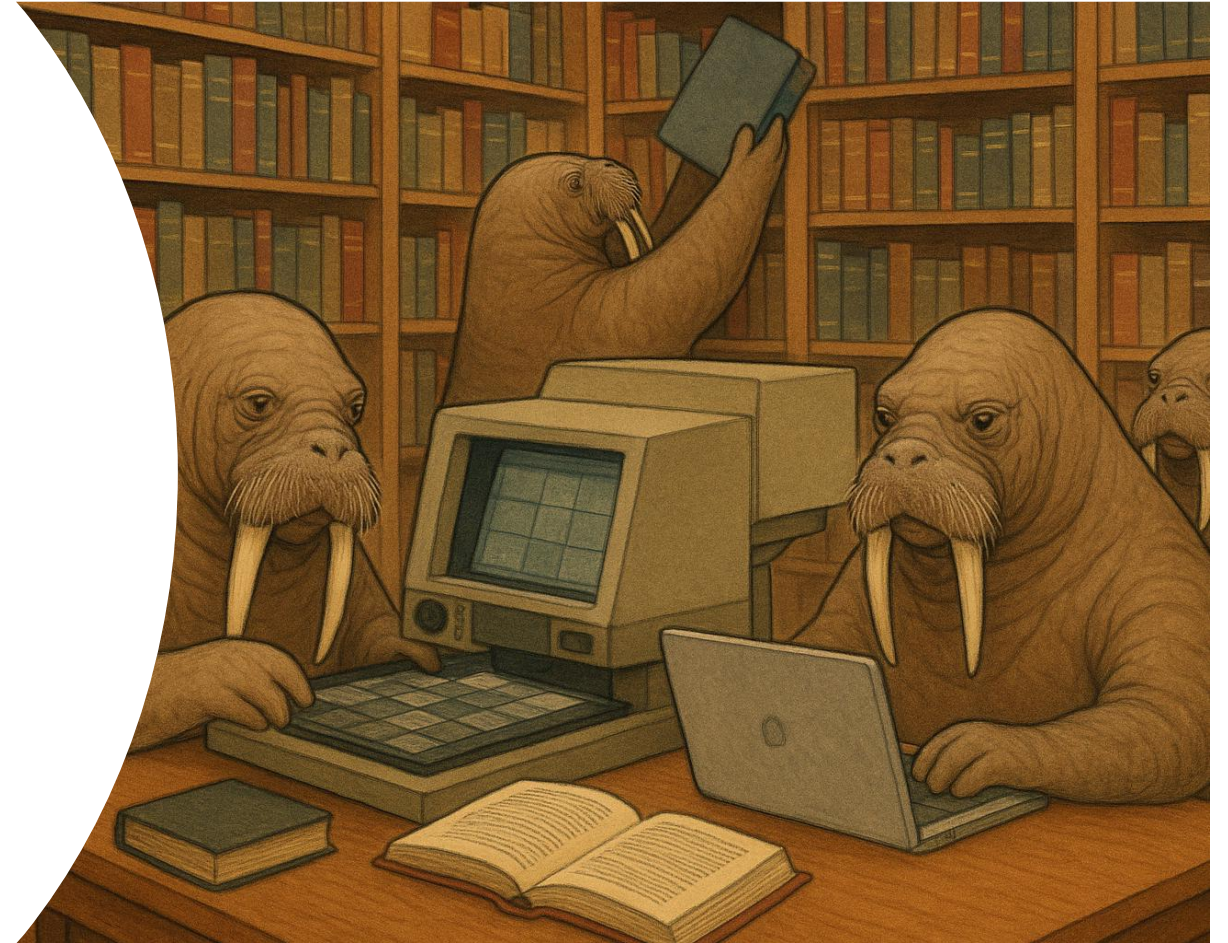


WHY ARE THEY AWESOME?

EXPLORE
PD25
New Depths

WHY ARE THEY AWESOME?

- Language models are good at “retrieving and applying memorized information” --
<https://www.aisnakeoil.com/p/ai-as-normal-technology>



TRANSLATION

- What transformer models were originally built for
- One language to another
- One style to another
- Turn x into y



SUMMARIZATION

- Turning lots of tokens into fewer tokens
- Lots of use cases here
 - Look at survey responses and pull out general themes
 - Find relevant sections in documents



CODING

- Computer code has a syntax and rules, like language
- Lots of examples on the internet to train on
- Coding agents and the rise of “vibe coding”
- <https://nhl-playoff-pool-ten.vercel.app/>

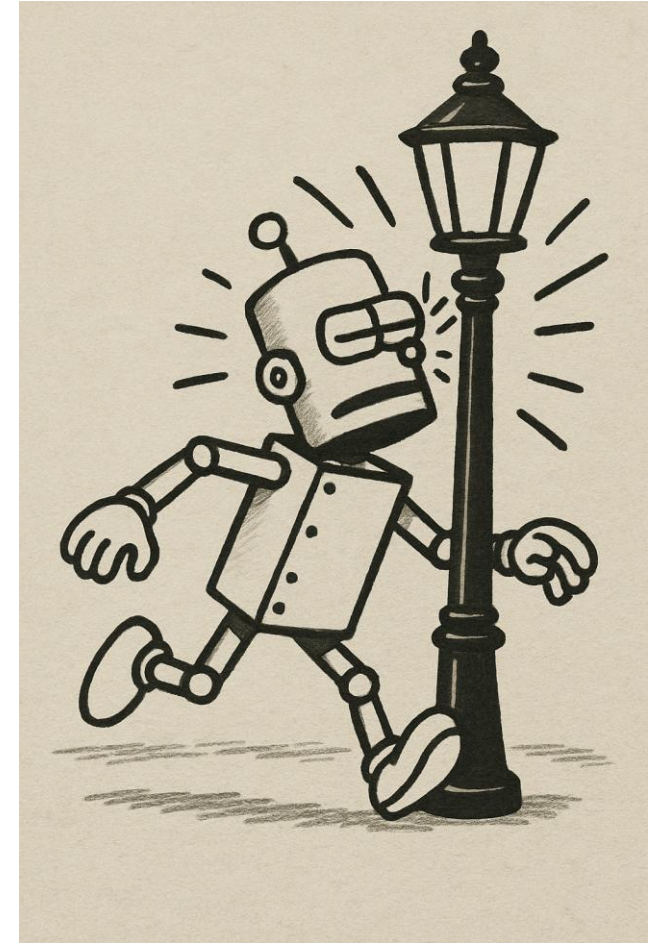


WHY ARE THEY TERRIBLE?

EXPLORE
PD25
New Depths

WHY ARE THEY TERRIBLE?

And what can we do to help improve them?



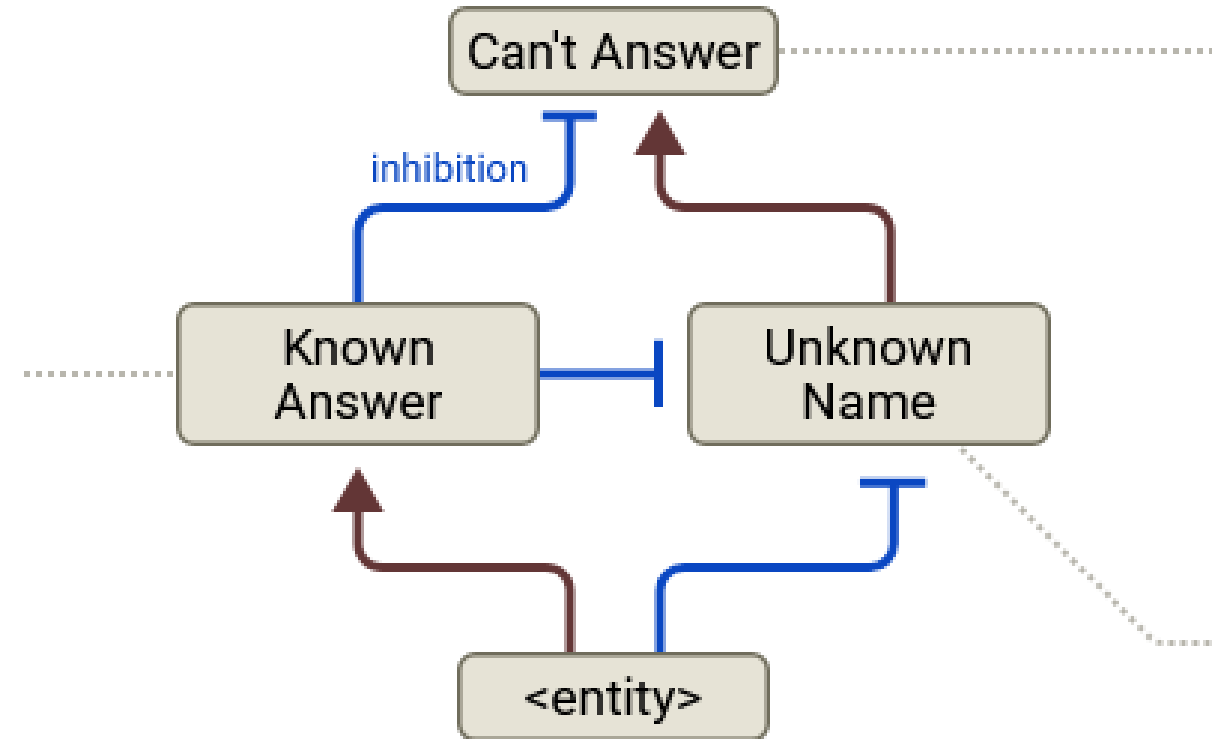
THEY DON'T KNOW EVERYTHING

- Training data can be out of date
- Niche domains might not be in the training data
- Can be improved by giving the models more data at “inference” time:
 - Retrieval augmented generation (RAG)
 - Browsing tools
 - Tool calling



WHAT THEY DON'T KNOW, THEY MAKE UP

- Models are trained to produce completions for blocks of text
- Nothing in the pre-training step to incentivize truth, that comes later
- Happens most often when asked about obscure facts or topics



STOP HALLUCINATING

- Same solutions as knowledge cutoffs:
 - RAG
 - Browsing
 - Tool calling
- Always check your sources!



DEFICIENCIES INTO STRENGTHS: REASONING & RESEARCH

- New “reasoning” models spend time “thinking” about how to answer questions
- Research models search the web and “reason” about the results



PRIVACY, SECURITY, ETHICS, OH MY!

EXPLORE
PD25
New Depths

WHO'S GOT YOUR DATA?

- Where is your data going?
- What is it being used for?
- What data points are you allowed to share?



AN LLM OF ONE'S OWN

- Open source tools and open weight models provide an alternative to private models and companies
- Run an LLM on your laptop
 - Transformer Lab
 - Jan
 - AnythingLLM
- Run an LLM in the cloud (or on your private network)



ETHICAL CONSIDERATIONS

- How will this AI process/output be used in the future?
 - Like transportation?
 - Like social media?
- Try to anticipate risks early
 - Bias
 - Audience
 - Transparency
 - Authenticity



WHERE TO START

- Identify resources/stakeholders already in place
- Start small
- Try something!



THANK YOU!

matthew@charitycan.ca

<https://charitycan.ca/ai>

<https://www.linkedin.com/in/matthew-charters-a1894639/>



THANK YOU!

Please complete your session
evaluations in the mobile app.

