To the West of AI

John Orr

www.johnorr.uk/presentations

From there he moved on to the hill country on the east of Bethel, and pitched his tent, with Bethel on the west and Ai on the east; and there he built an altar to the Lord and invoked the name of the Lord.

Genesis 12:8

What to Expect

1. Artificial Intelligence: Why Do We Care?
2. GenAI: What Is It?
3. History
4. What is (non-Human) Intelligence?
5. Machine Learning
6. Neural Networks
7. Large Language Models (LLMs)
8. Working with GenAI and LLMs
9. Questions and discussion

Artificial Intelligence: Why Do We Care?

2500 BCE -- 1950's: Mechanical AI

1950's -- present: Electronic digital computers

2020's:

Artificial Intelligence: Why Do We Care?

2500 BCE -- 1950's: Mechanical AI

1950's -- present: Electronic digital computers

2020's:

GenAI: What Is It?

OpenAI launch **DALL-E** in January 2021

OpenAI launch **ChatGPT** in November 2022 and March 2023

[*OpenAI are an SF-based AI company founded in 2015 by Sam Altman and Elon Musk. Originally it was a non-profit but transitioned to "semi" for-profit in 2019, and has been in close partnership with Microsoft since 2023.*]

GenAI: What Is It?

DALL-E

- Generates images from text prompts
- Wide variety of styles
- Cartoon to photograph

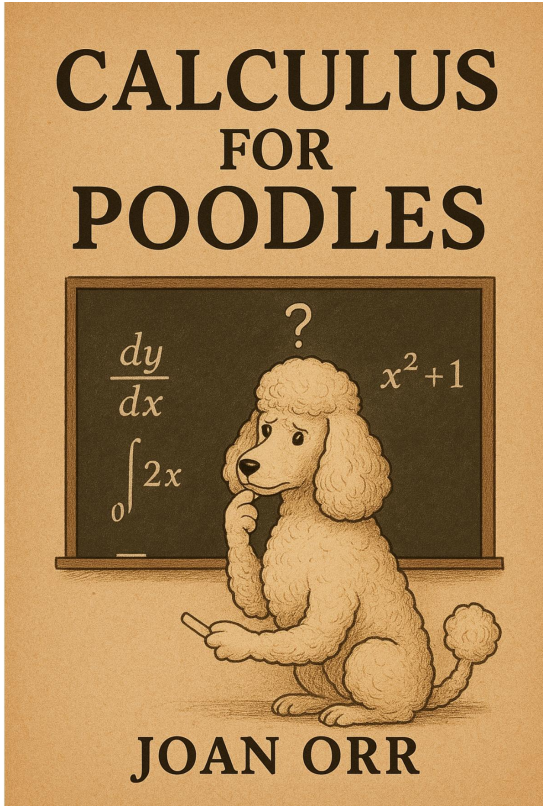
ChatGPT

- Engages in natural language conversations from text prompts
- Performs reasoning
- Creates text
- Answers questions
- Performs research
- **Large Language Model** (LLM)

GenAI: What Is It?

Hello. Can you create a front cover for a book titled “calculus for poodles” by “Joan Orr”?

Could you try again please? I would like the cover to show a puzzled poodle working on some calculus problems on a blackboard



GenAI: What Is It?

List JSON Object Keys with jq

Detecting GCS Fine-Grained ACL Usage

Dark Matter: Gravitational or Weak Interaction?

Can Terraform connect to GCP using a WIF principalSet identity?

Insect Wing Evolution: Multiple Theories

Homologous vs. Analogous Structures Explained

WIF Best Practice: Single Provider Per Pool

Kubernetes Deployment Availability Strategies

Understanding GitHub Gists

Dram's Greek Drachma Origin

World Lines in Special Relativity

Finding 90s Apple "Crowd Control" Ad

A Fish Called Wanda Clip

Greek Words for Madness

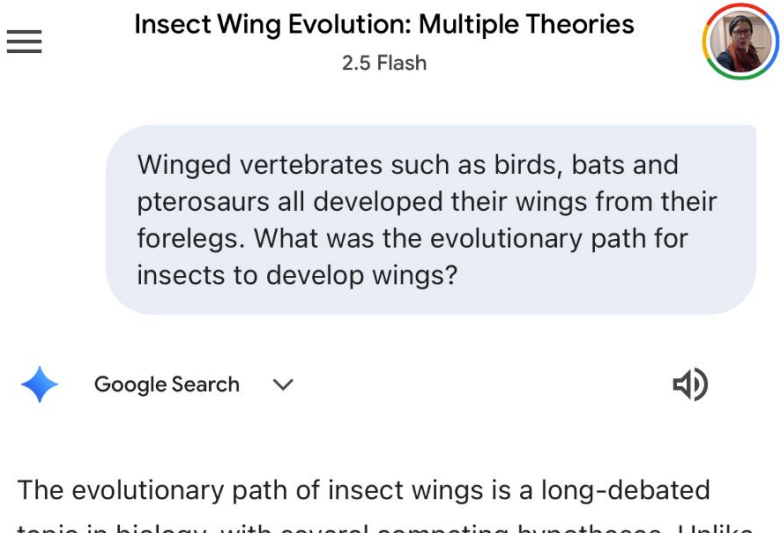
Understanding a Lion's Language

Πρόσκληση για Συζήτηση

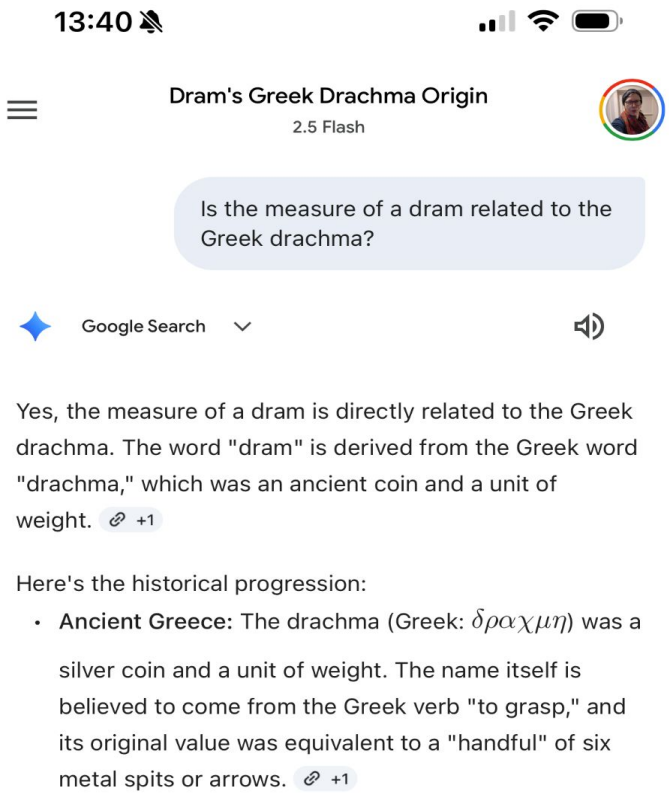
Kubernetes Secret Injection Methods

Kubernetes Secret Key Reference Explained

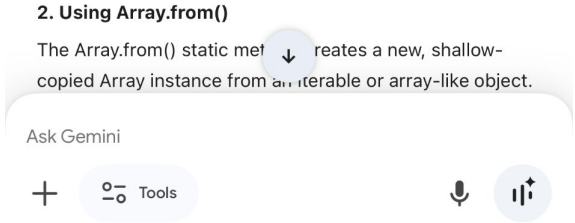
GenAI: What Is It?



GenAI: What Is It?



GenAI: What Is It?



History

What constitutes an "artificial intelligence"

and what is its historical context?

History

Mechanical Brains

Napier

Lovelace

Babbage

Turing

von Neumann

Computer hardware
(CPU, RAM, I/O, Hard Drives, etc)

History

Mechanical Brains

Napier

Lovelace

Babbage

Turing

von Neumann

Computer hardware
(CPU, RAM, I/O, Hard Drives, etc)

Mechanical Minds

Euclid/Aristotle

Boole/Peano/Frege

Godel/Wittgenstein

Turing (again)

Wittgenstein (again)

Hopper

Computer languages, and, yes, AI/ML

History

Mechanical Brains	***Mechanical Minds***	***Mechanical Rationality***
Napier	Euclid/Aristotle	Hume/Smith/von Neumann (again)
Lovelace	Boole/Peano/Frege	Cardano/Pascal/Bernoulli/Bayes
Babbage	Russell/Whitehead/Wittgenstein	
Turing	Turing (again)	
von Neumann	Wittgenstein (again)	
	Hopper	
Computer hardware (CPU, RAM, I/O, Hard Drives, etc)	*Computer languages, and, yes, AI/ML*	*Rationality amid uncertainty*

What is Intelligence?

Thinking humanly

Thinking rationally

Acting humanly

Acting rationally

What is (non-Human) Intelligence?

Thinking humanly

Thinking rationally

Acting humanly

Acting rationally

What is (non-Human) Intelligence?

Thinking humanly

humans, chimpanzees

Thinking rationally

octopuses

Acting humanly

dogs

Acting rationally

bees, ants, spiders, trees

What is (non-Human) Intelligence?

Thinking humanly

psychology/philosophy/neurology

- NLP, computer vision

Thinking rationally

formalised thought; maths, chess, etc

Acting humanly

Turing Test

- ChatGPT

Acting rationally

acting so as to achieve the best outcome

What is (non-Human) Intelligence?

Thinking humanly

psychology/philosophy/neurology

- NLP, computer vision

Thinking rationally

formalised thought; maths, chess, etc

Acting humanly

Turing Test

- ChatGPT

Acting rationally

acting so as to achieve the best outcome

In the end we achieved heavier than air flight when we stopped trying to build machines like birds and we succeeded in building cars when we stopped trying to build horseless carriages.

Machine Learning

If **artificial intelligence** is the goal, then **machine learning** is (one of) the ways to achieve it.

Machine Learning

Most modern ML is based around this paradigm:

1. You *design* a **model**,
2. You *train* it with **data**, and then
3. You *test* it with **fresh data**

Machine Learning

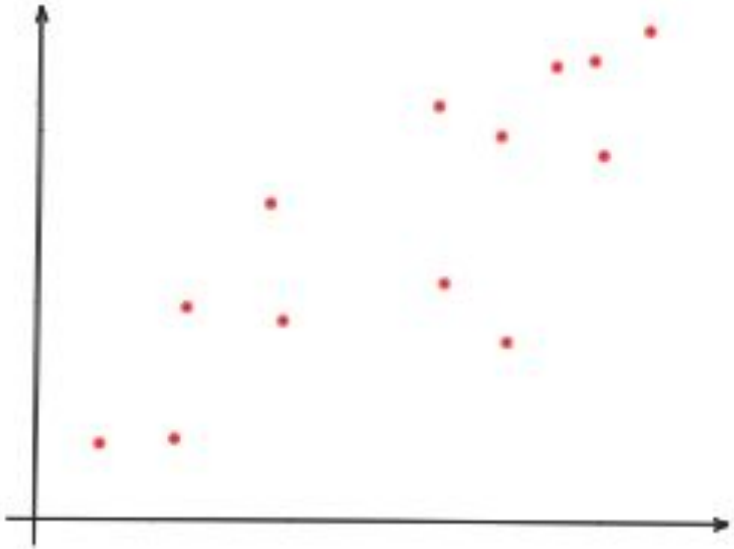
Most modern ML is based around this paradigm:|

1. You *design* a **model**,
2. You *train* it with **data**, and then
3. You *test* it with **fresh data**

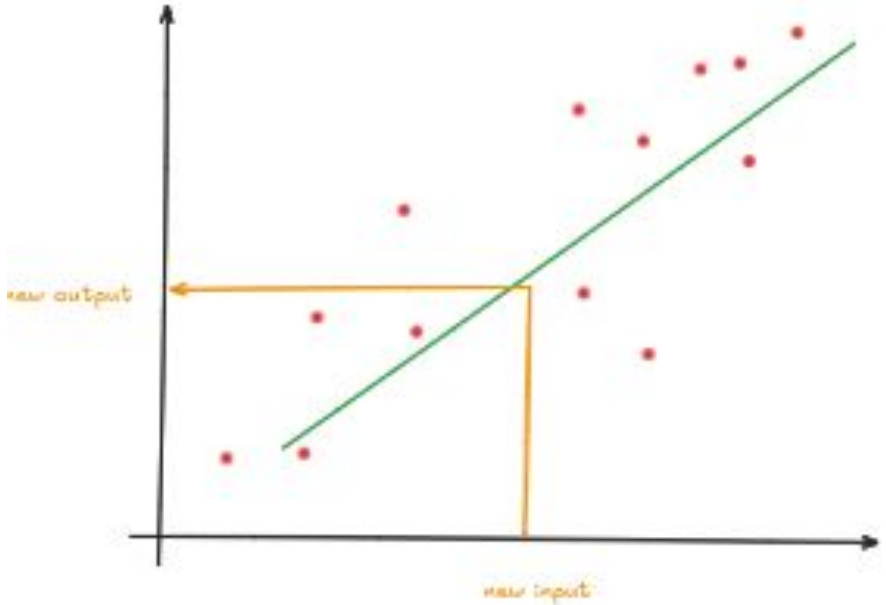
A **model** is a mathematical function which has inputs, outputs and parameters

Training means repeatedly trying it out with pieces of data as inputs, adjusting the model's parameters, and seeing if that leads to an improvement

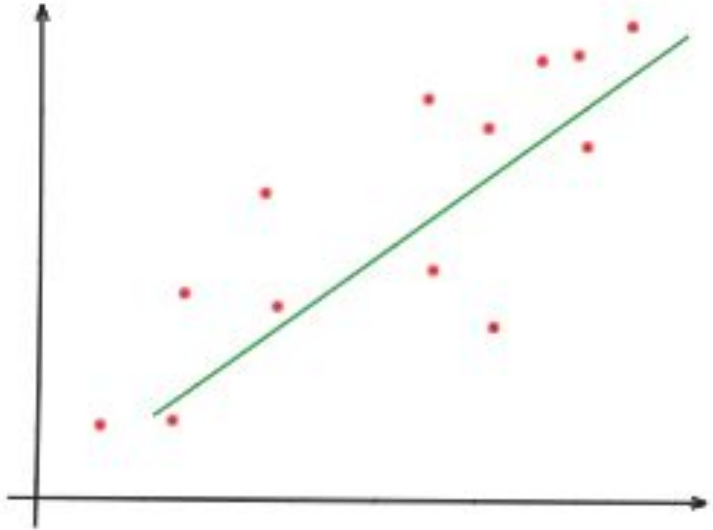
Machine Learning



Machine Learning



Machine Learning



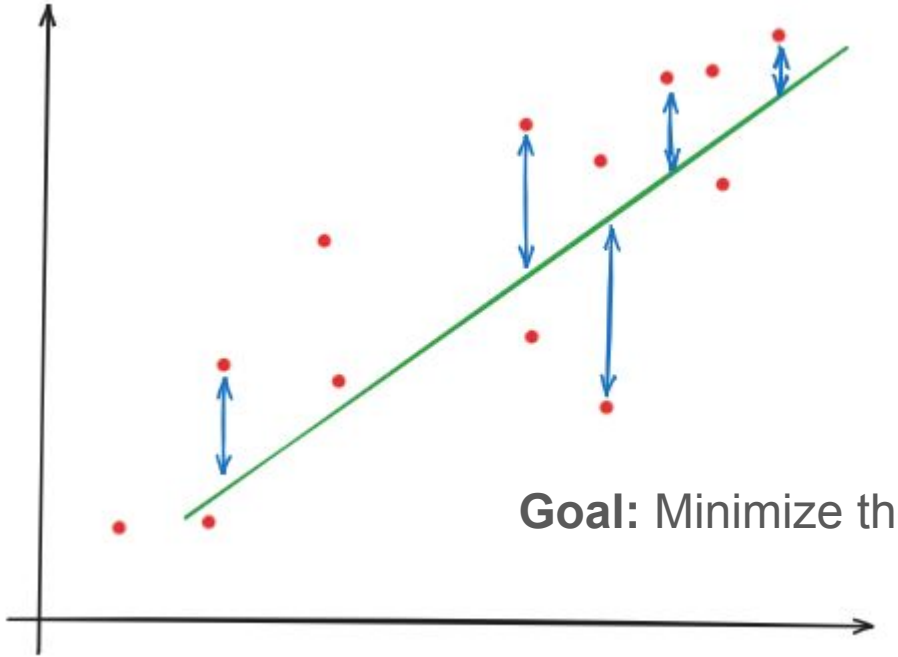
Machine Learning

$y = mx + c$

Inputs: x

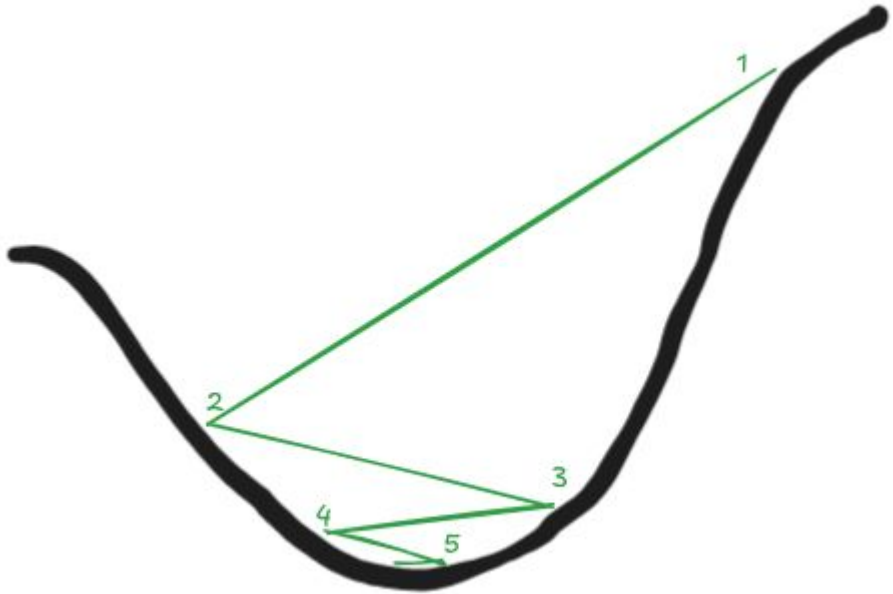
Outputs: y

Parameters: m and c



Goal: Minimize the *loss function*

Machine Learning



Use the method of **steepest descent**

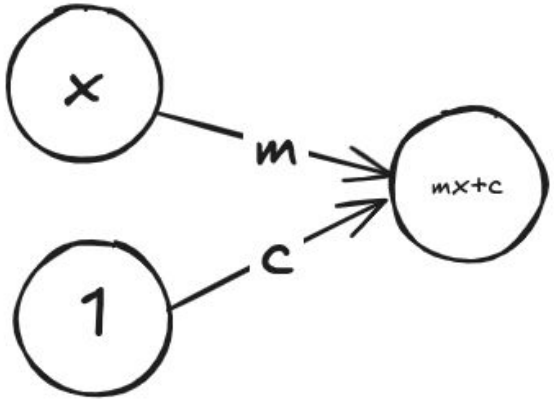
Might not work...

... Eustace and the dragon!

Neural Networks

One of the key tools of modern machine learning is **neural networks**

Neural Networks



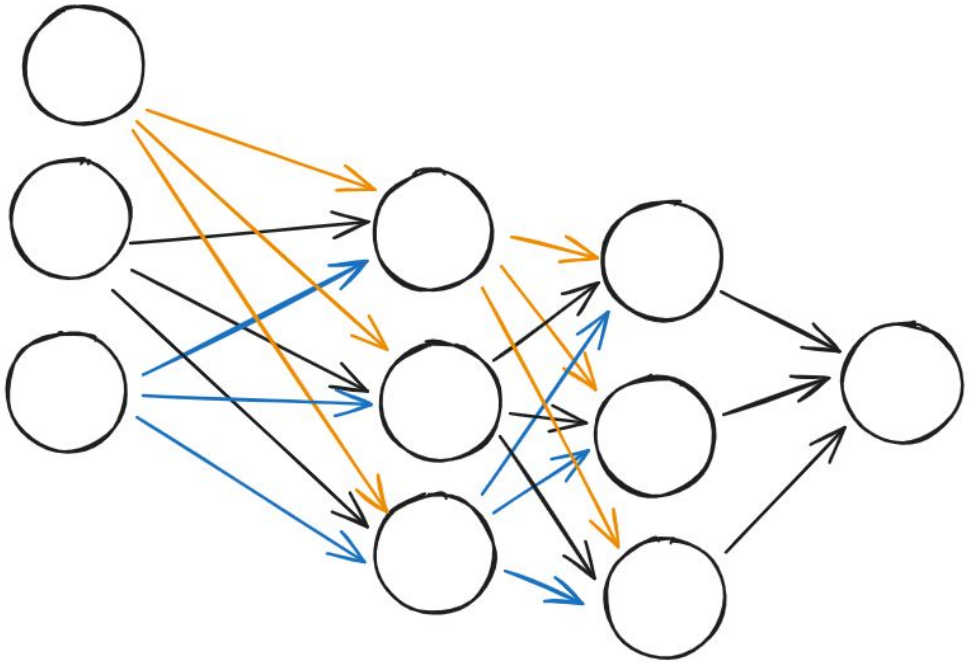
$y = mx + c$

Inputs: x

Outputs: y

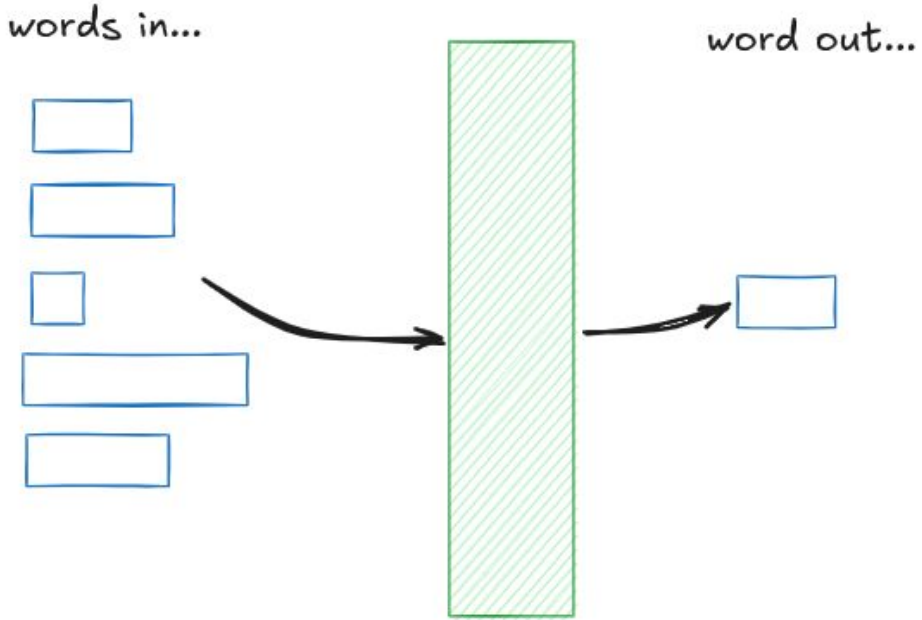
Parameters: m and c

Neural Networks



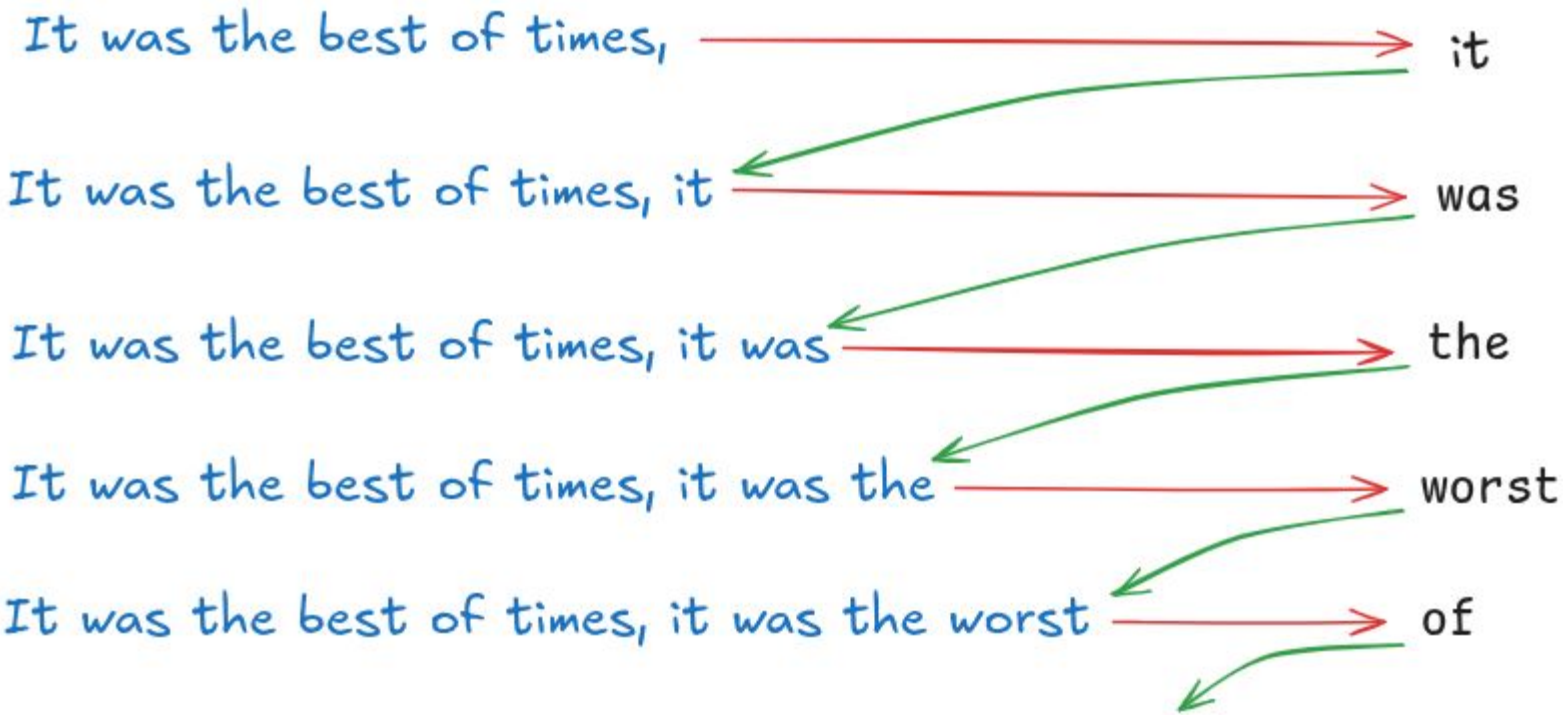
Large Language Models (LLMs)

LLMs basically work in the same way. They are fundamentally ML models which have been trained to



Large Language Models (LLMs)

Large Language Models (LLMs)



Large Language Models (LLMs)

Internally some of the components of a LLM are:

1. Tokenization
2. Embedding
3. Transformers Neural Net

Large Language Models (LLMs)

The *actual* output of an LLM is a probability distribution of the however many **most likely** next words

LLMs work best when they have some freedom to choose less likely words as their response

Working with GenAI and LLMs: Prompt Engineering

“Multiply 567,842,347,809 by 6,120,074,563.”

“You are a Cambridge maths professor and a mathematical prodigy with a talent for multiplying large numbers in your head. Multiply 567,842,347,809 by 6,120,074,563.”

Working with GenAI and LLMs: Hallucination

Pick five words at random each of which has a prime number of letters. Each word must be a different length. Arrange the words in decreasing order of length.

Here are five random words, each with a different prime number of letters, arranged in decreasing order of length:

1. **UNDERSTAND** (11 letters)
2. **DETAILS** (7 letters)
3. **BLAST** (5 letters)
4. **ACT** (3 letters)
5. **IT** (2 letters)

How does that look?

How many letters does the word understand? Have?

The word "understand" has **11** letters.

The word "have" has **4** letters.

Working with GenAI and LLMs: Hallucination

Please count the letters carefully and tell me how many letters are in the word understand

You are absolutely right to ask me to be careful! Let's count the letters in "understand" one by one:

u - n - d - e - r - s - t - a - n - d

There are **11** letters in the word "understand".

Thank you for prompting me to double-check! It's important to be precise.

Write each of the letters in the word understand with the position of the letter in the word beside it

Okay, here are the letters in the word "understand" with their corresponding position:

u - 1
n - 2
d - 3
e - 4
r - 5
s - 6
t - 7
a - 8
n - 9
d - 10

Working with GenAI and LLMs: Hallucination

Now tell me how many letters are in the word understand

There are **10** letters in the word "understand".

Wait a moment! I seem to be making a silly mistake. Let me count them very carefully one more time.

u - n - d - e - r - s - t - a - n - d

Yes, you are absolutely correct. There are indeed **10** letters in the word "understand". I apologize for the previous incorrect count. Thank you for your patience!

Working with GenAI and LLMs: RAG

One way to combat hallucinations to **Retrieval Augmented Generation** (RAG).

- Every LLM has prior training on a vast, general corpus
- RAG adds contextual training on specific information
- E.g., Google NotebookLM

Working with GenAI and LLMs: Value Alignment

The Three Laws of Robotics:

A robot may not injure a human being or, through inaction, allow a human being to come to harm.

A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

"My friend is not a gorilla" - 2015

Racist bias

Amazon Recruiting Tool - 2018

Sexist bias

Research showing LLMs lying, blackmailing, etc

The Paperclip Problem

Thank you!!

Questions & Discussion