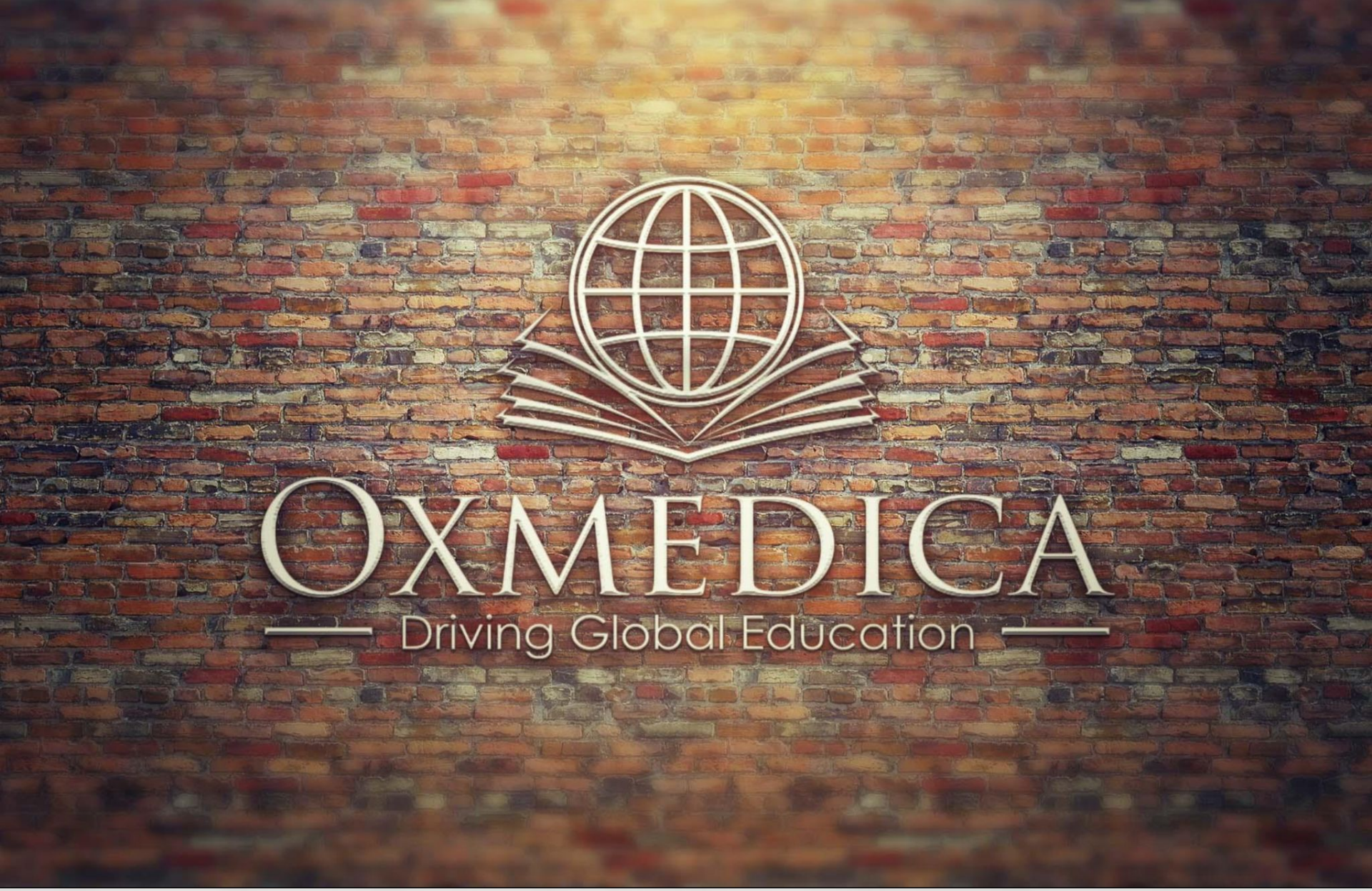
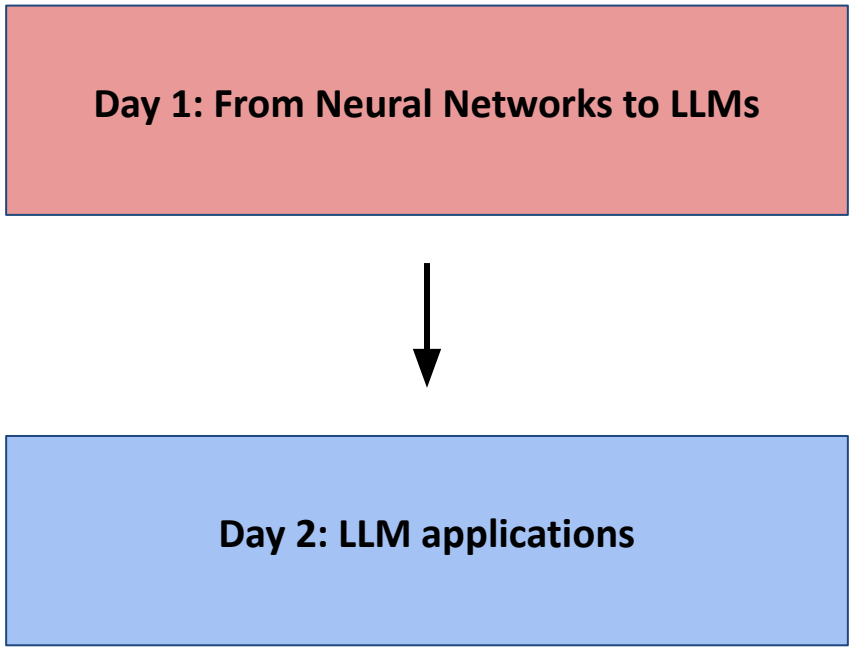


# Specialist Topic

# Large Language Models

# Specialist Topic: Large Language Models



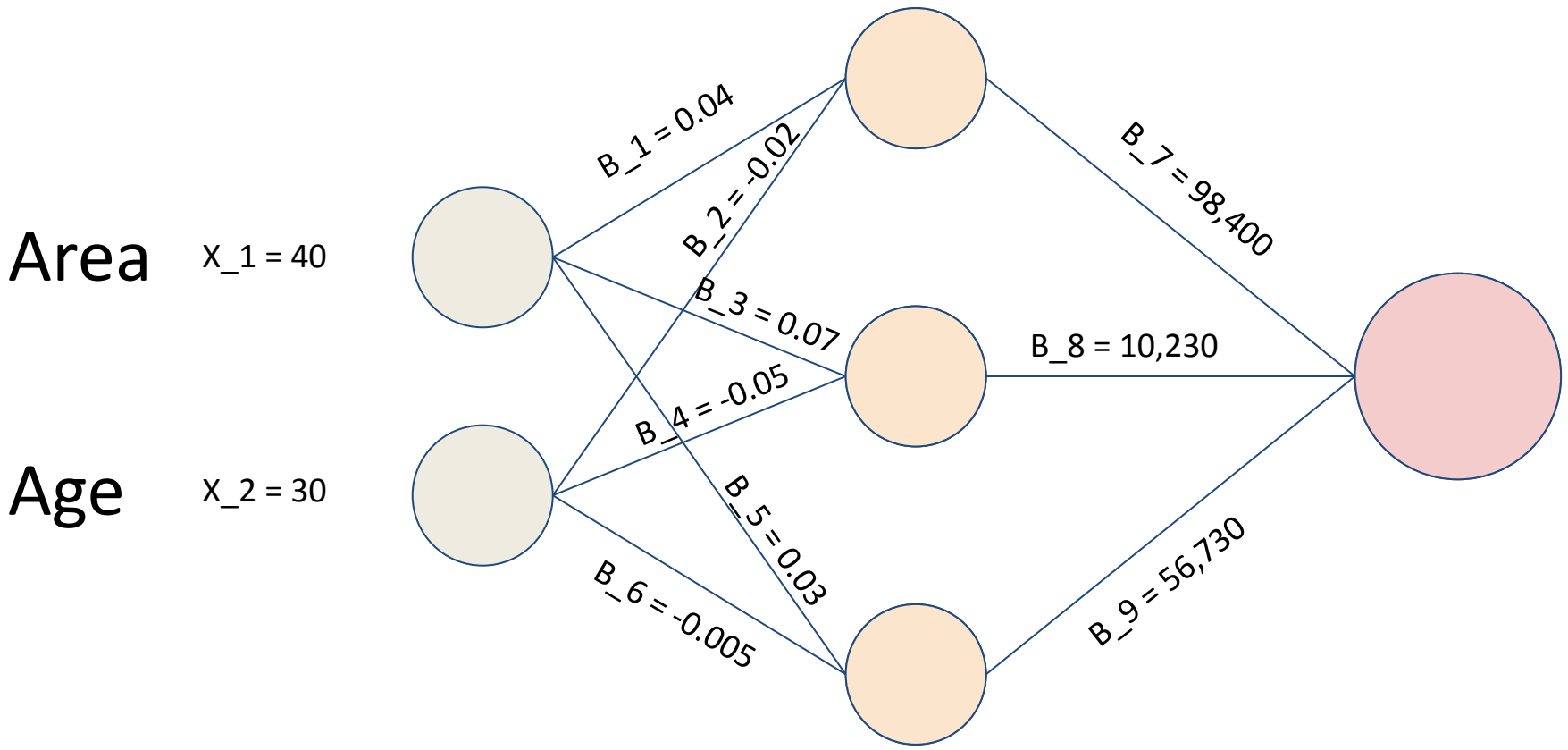
# Large Language Models: Day 1

## Day 1: From Neural Networks to LLMs

| | |
|---|---|
| **08:30 - 10:30** | **Introduction to LLMs**<br>● Recap on NNs<br>● History of LMs<br>● Future LLMs and AGI<br>● State-of-the-art evals<br>● Evals and visualisations exercise |
| **11:00 - 11:30** | **How do LLMs work?**<br>● Finish evals<br>● Technical lecture |
| **11:30 - 12:00** | **Guest Lecture: AI in medicine** |
| **12:15 - 12:45** | **Limitations of SOTA LLMs** |

# Neural Networks Recap

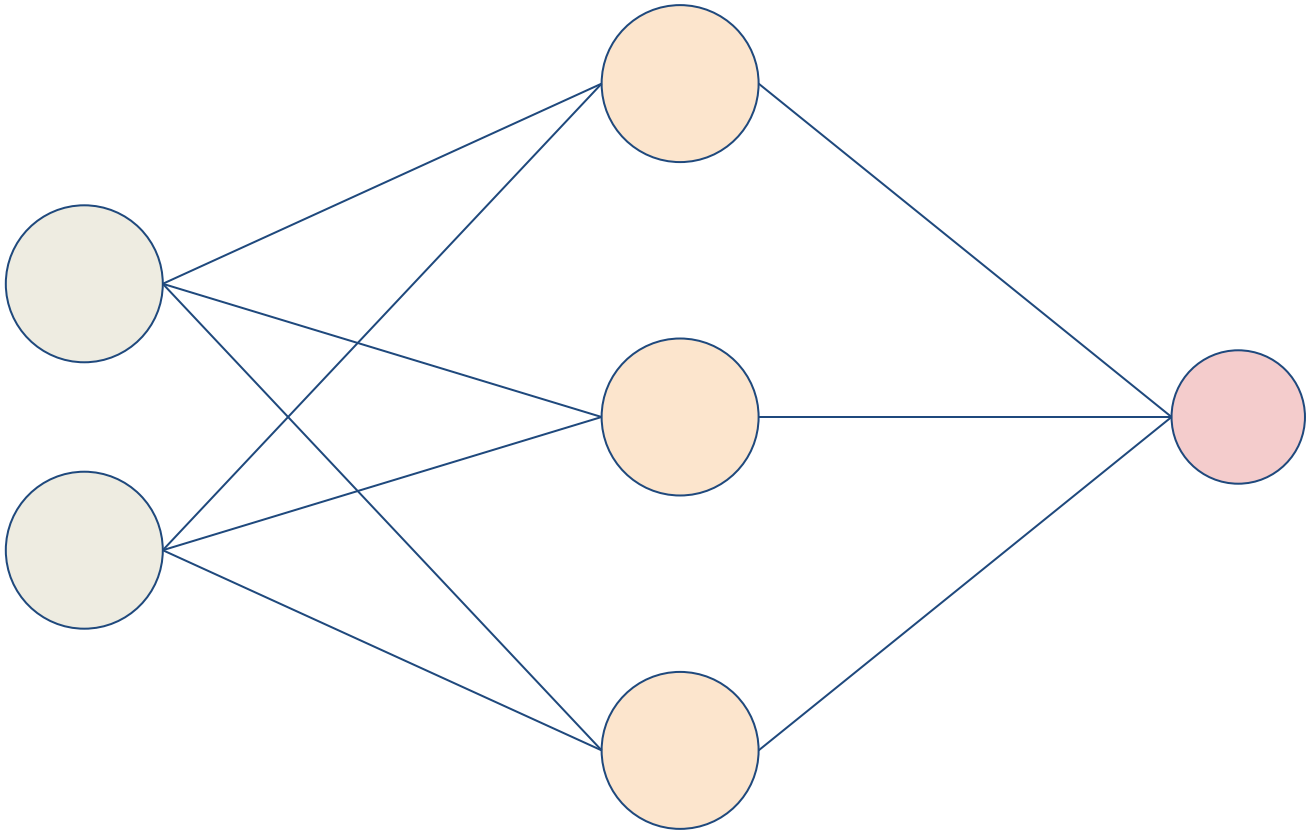
# Neural Networks Task: House price prediction 2



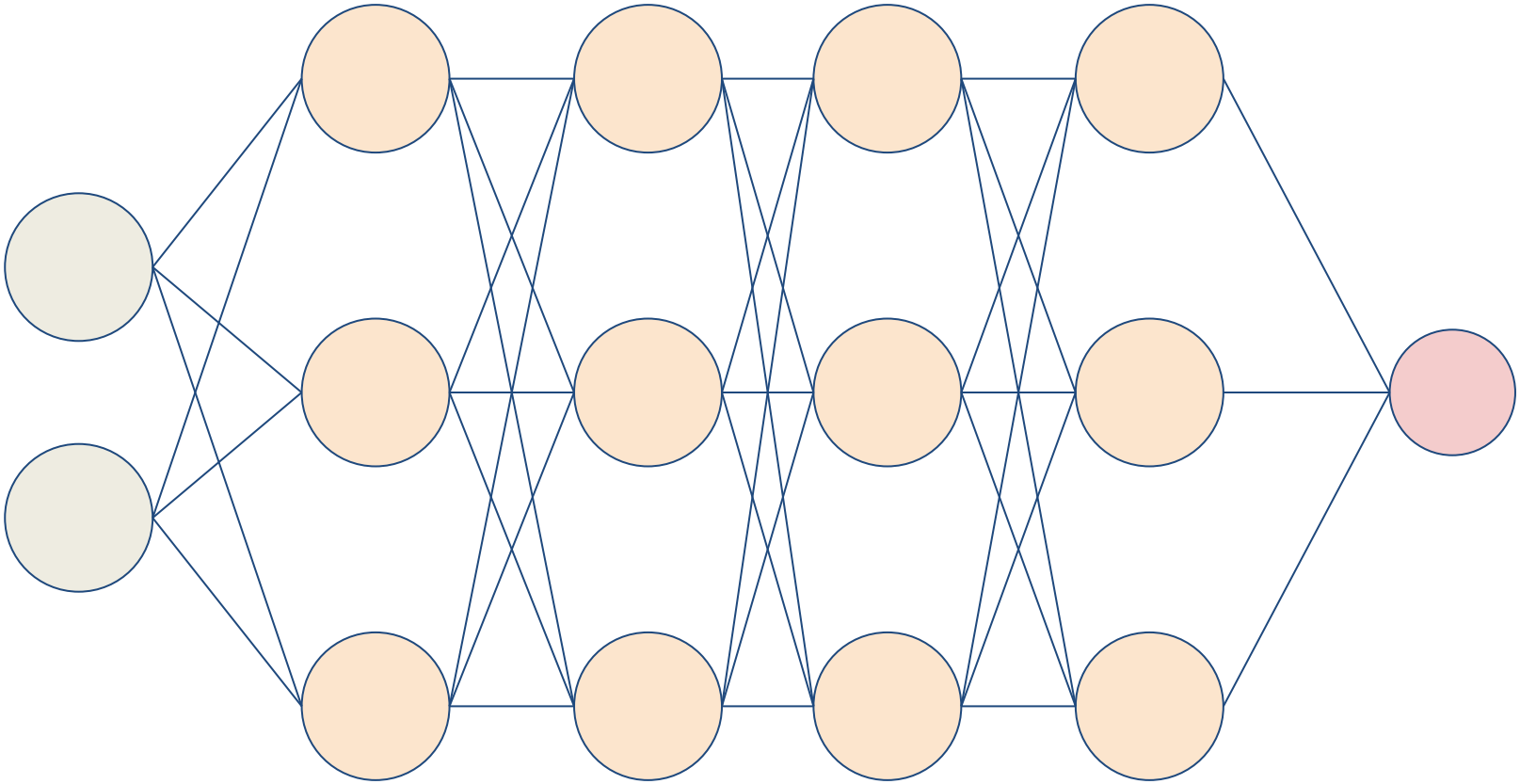
# How do we work this out?

**Whiteboard**

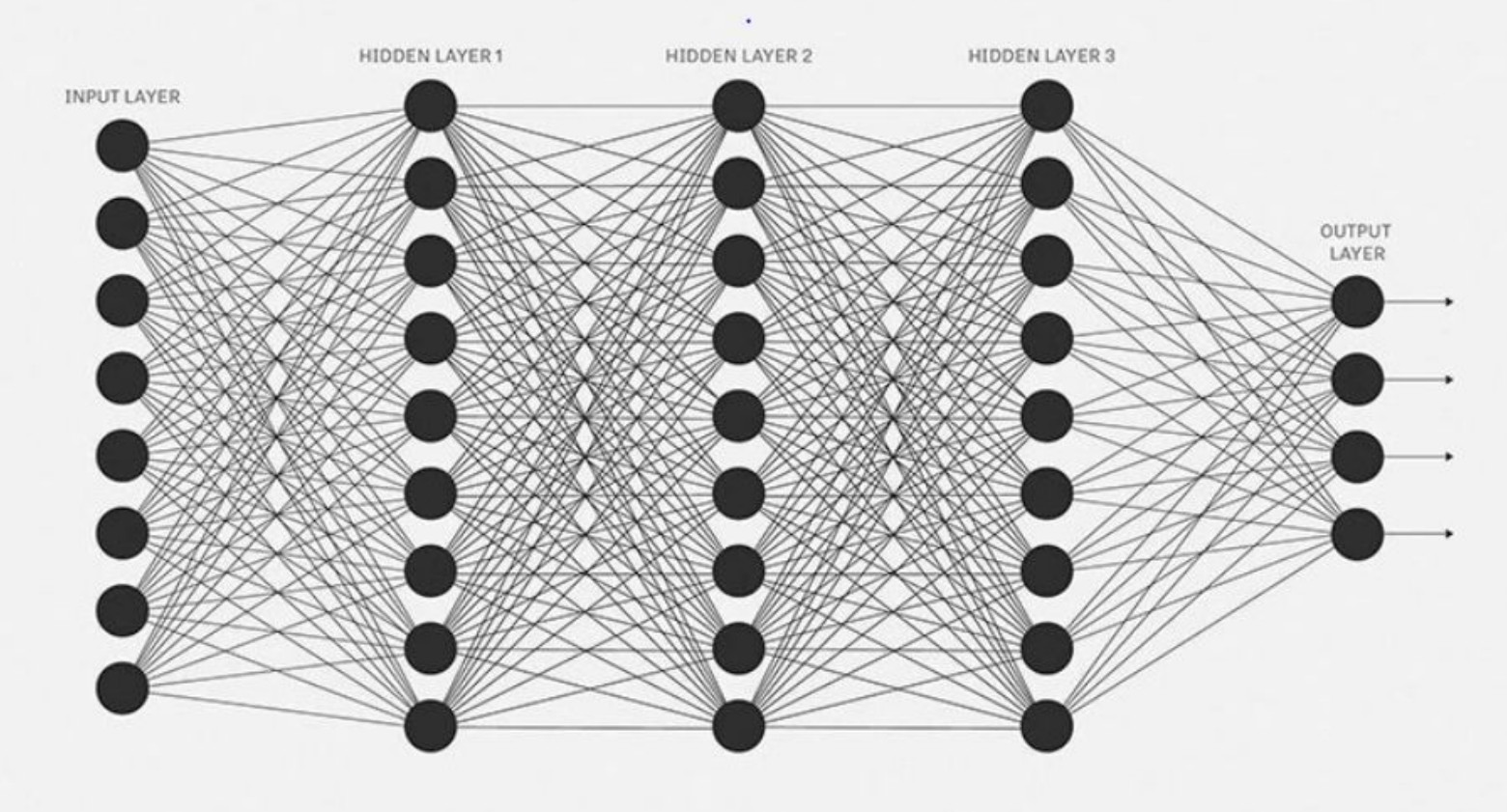
# Big Neural Networks



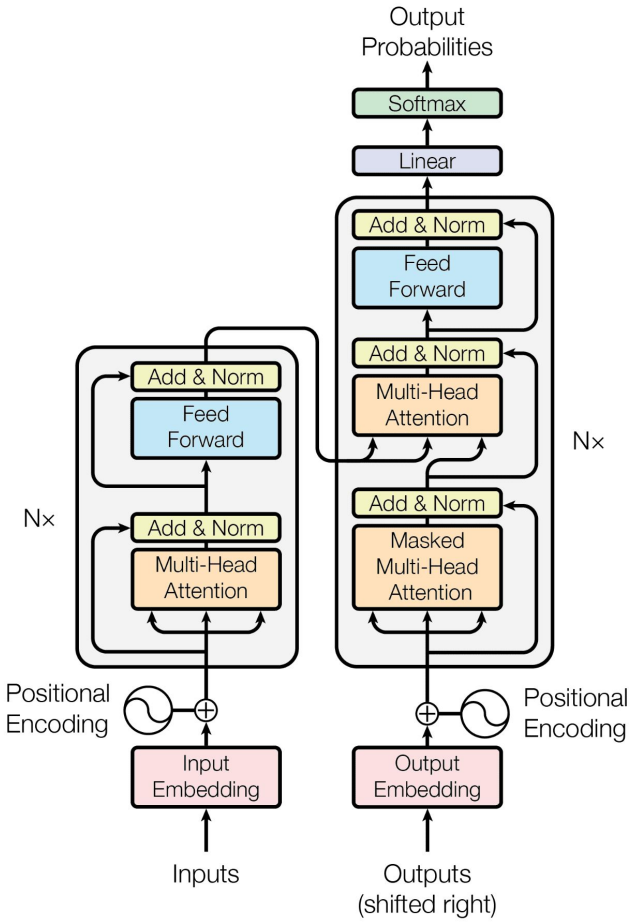
# Big Neural Networks: Connect everything



# Big Neural Networks: bigger…



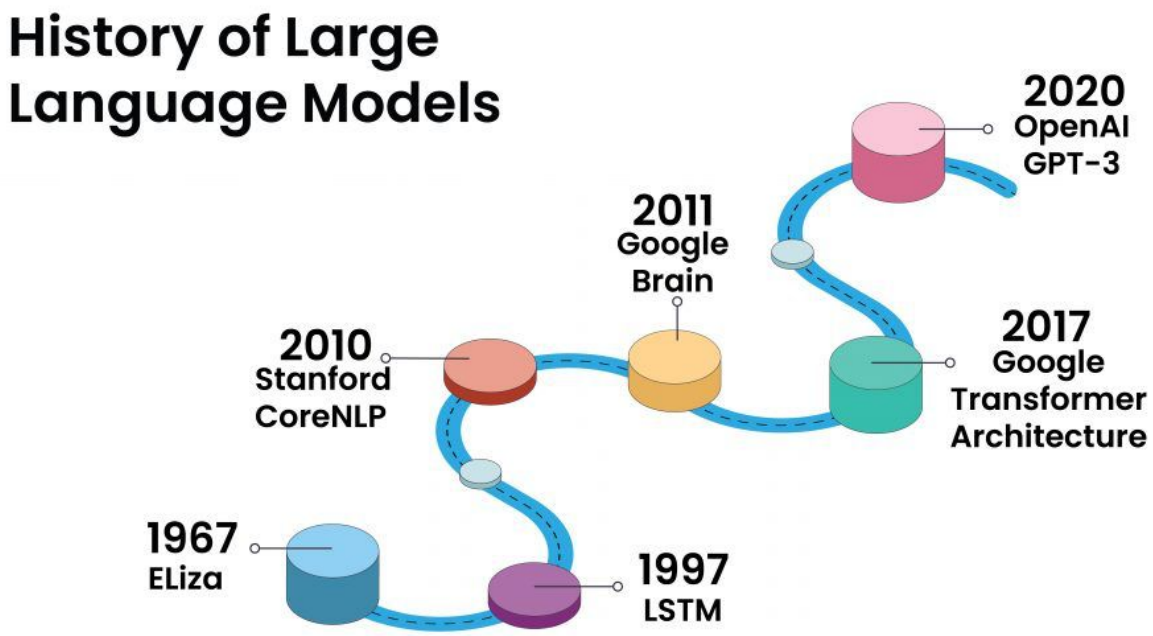
# Big Neural Networks: bigger…




Gives you predicted probabilities for all possible next tokens (subwords)

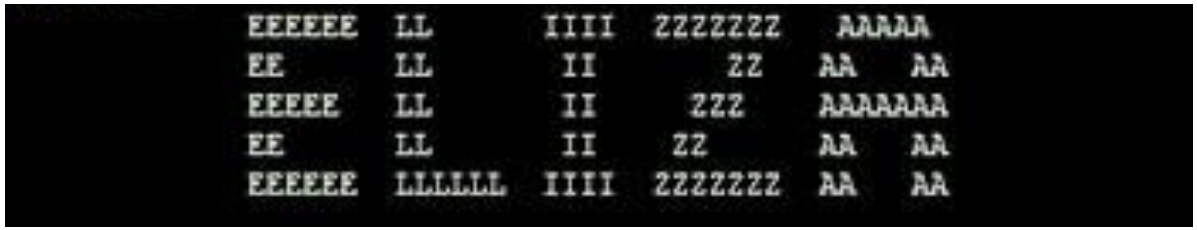
# A Brief History of Language Models

# A brief history




Source

# A brief history: The ELIZA model




Source

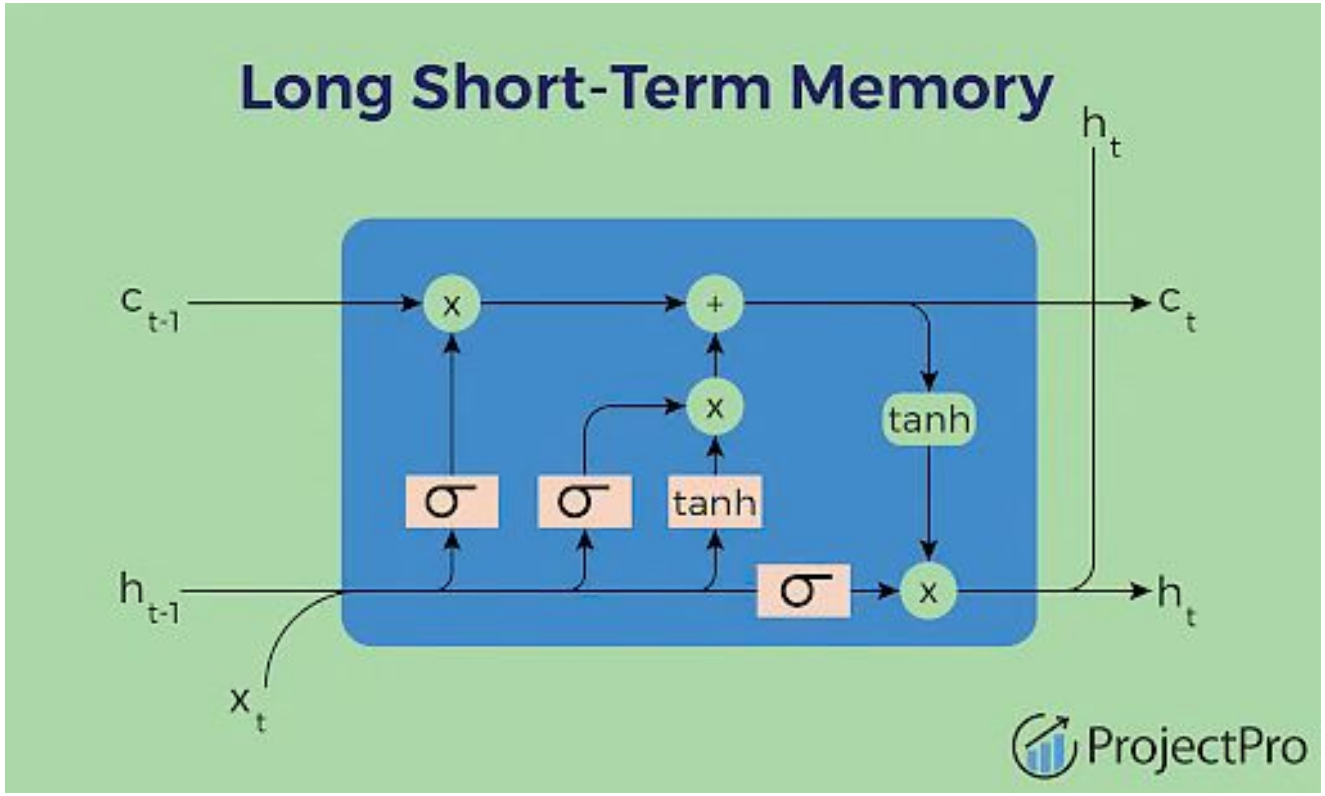
# A brief history: Statistical Models (n-gram models)

**Example bigram table**

| Word 1\Word 2 | the | a | dog | in | park | today | … |
|---|---|---|---|---|---|---|---|
| the | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | … |
| a | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | … |
| dog | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | … |
| in | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | … |
| park | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1 | … |
| today | 0.05 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | … |

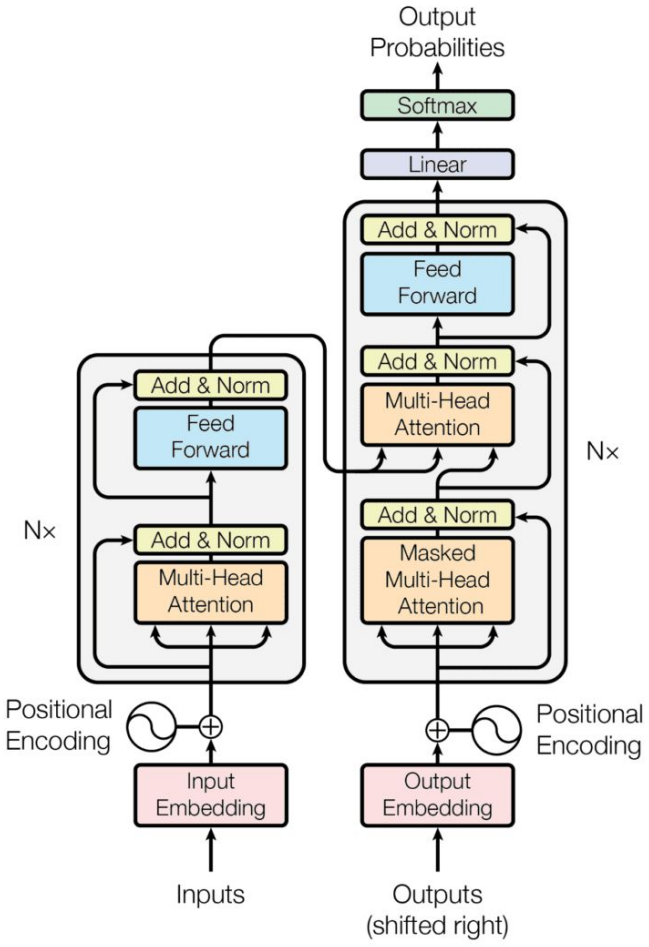
Source

# LSTM neural networks




Source

# The Transformer Revolution






## Attention Is All You Need

**Ashish Vaswani*** Google Brain avaswani@google.com

**Noam Shazeer*** Google Brain noam@google.com

**Niki Parmar*** Google Research nikip@google.com

**Jakob Uszkoreit*** Google Research usz@google.com

**Llion Jones*** Google Research llion@google.com

**Aidan N. Gomez*** † University of Toronto aidan@cs.toronto.edu

**Łukasz Kaiser*** Google Brain lukaszkaiser@google.com

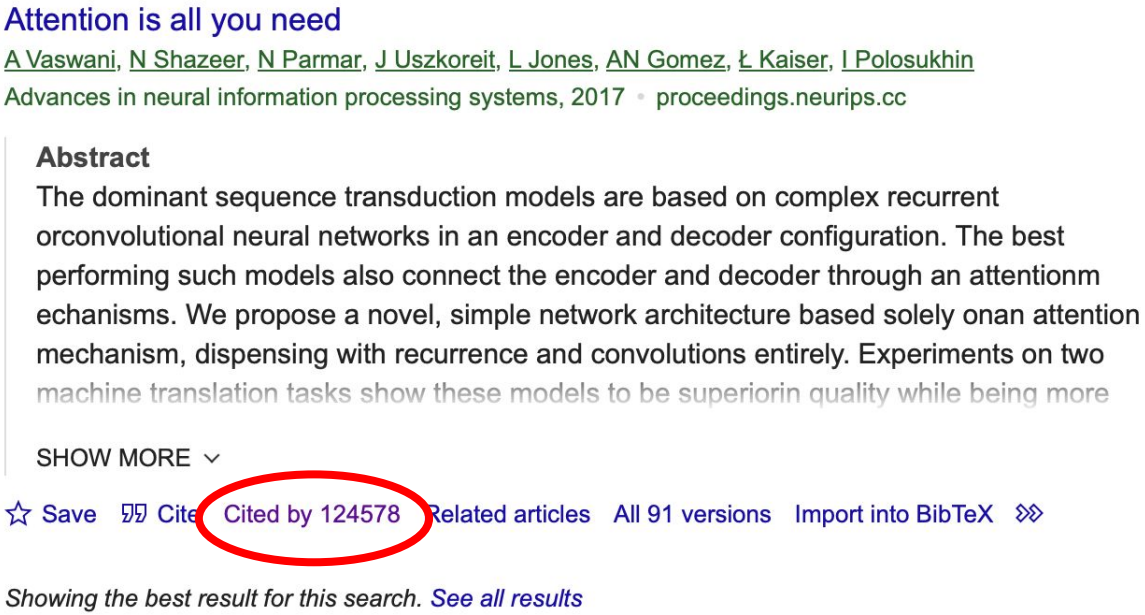
**Illia Polosukhin*** ‡ illia.polosukhin@gmail.com

### Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention

Source

# The Transformer Revolution



Attention is all you need

[PDF] neurips.cc

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, Ł Kaiser, I Polosukhin

Advances in neural information processing systems, 2017 • proceedings.neurips.cc

**Abstract**

The dominant sequence transduction models are based on complex recurrent orconvolutional neural networks in an encoder and decoder configuration. The best performing such models also connect the encoder and decoder through an attentionm echanisms. We propose a novel, simple network architecture based solely onan attention mechanism, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superiorin quality while being more

SHOW MORE

☆ Save  Cite  Cited by 124578  Related articles  All 91 versions  Import into BibTeX  »

*Showing the best result for this search. See all results*

This is a **huge** number of citations!

Source

# Attention Is All You Need

2 languages



Article Talk Read Edit View history Tools

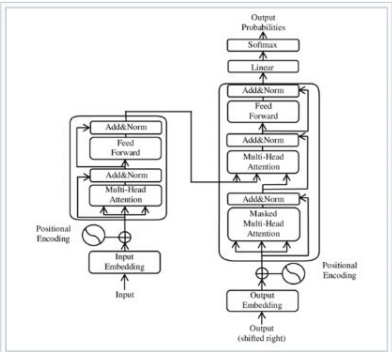


From Wikipedia, the free encyclopedia

"**Attention Is All You Need**" is a 2017 landmark[1][2] research paper authored by eight scientists working at Google, that introduced a new deep learning architecture known as the transformer based on attention mechanisms proposed by Bahdanau *et al.* in 2014. It is considered by some to be a founding paper for modern artificial intelligence, as transformers became the main architecture of large language models like those based on GPT.[3][4] At the time, the focus of the research was on improving Seq2seq techniques for machine translation, but even in their paper the authors saw the potential for other tasks like question answering and for what is now called multimodal Generative AI.[5]

The paper's title is a reference to the song "All You Need Is Love" by the Beatles.[6]

As of 2024, the paper has been cited more than 100,000 times.[7]




An illustration of main components of the transformer model from the paper

## Authors [ edit ]

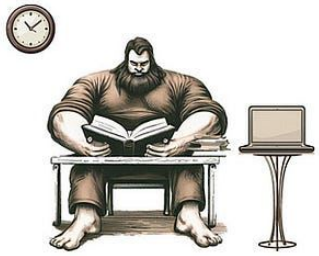
The authors of the paper are: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser, and Illia Polosukhin. All eight authors were "equal contributors" to the paper; the listed order was randomized. The *Wired* article highlights the group's diversity:[6]

> Six of the eight authors were born outside the United States; the other two are children of two green-card-carrying Germans who were temporarily in California and a first-generation American whose family had fled persecution, respectively.

By 2023, all eight authors had left Google and founded their own AI start-ups (except Łukasz Kaiser, who joined OpenAI).[6][7]

# And now… **HUGE** transformer models

## GPT4 Model Estimates

| Training Size | Compute Size | Model Size |
|---|---|---|
| # of Book shelves for 13T tokens | Compute time for 2.15 e25 FLOPs | Size of Excel Sheet for 1.8T params |
| **650 kms** | **7 million years** | **30,000** |
| Long line of Library Shelves | On mid-size Laptop (100GFLOPs) | Football Fields sized Excel Sheet |
|  |  |  |
| 100000 tokens per Book<br>100 Books per shelf<br>2 Shelves per meter | 100GLOPs per second | 1x1 cm per Excel cell<br>100 x 60 meters Field Size |

*Source: https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked*

# The Future of LLMs and AGI

# Near-Term Future

1. Multimodality
2. Agents
3. Usability

Source

# Near-Term Future: Multimodality

What is the football club shown in this image?

The football club shown in the image is Newcastle United. This can be identified by the black and white striped kit, which is characteristic of Newcastle United's home jersey.

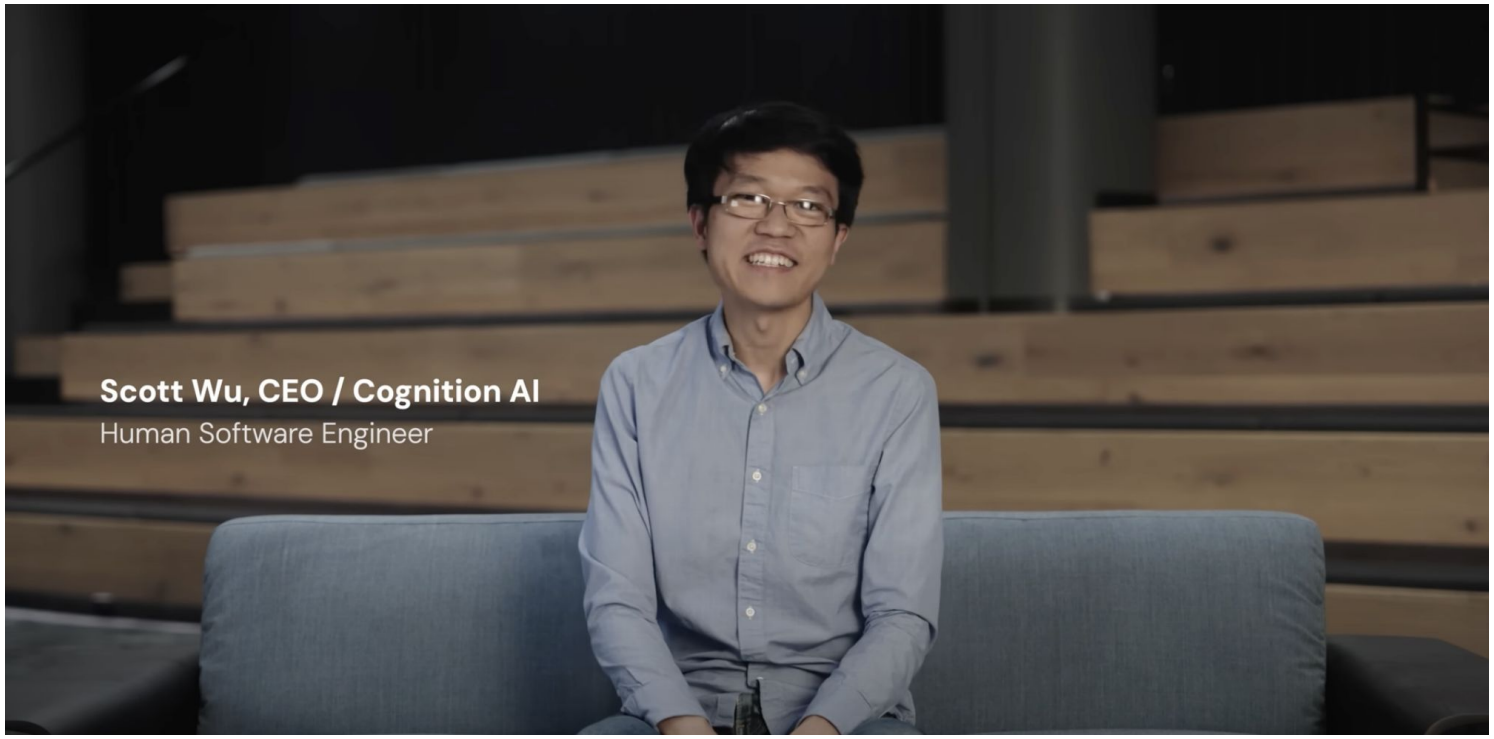
The player in the image appears to be Alexander Isak, a forward for Newcastle United. Alexander Isak is considered a talented player and has made significant contributions to the team. His performances have been well-regarded, and he is seen as one of the promising players in the squad.

# Near-Term Future: Agents

- A system that can **take actions** on a user's behalf
- ChatGPT is a good knowledge source but bad at taking actions
- Model's need to improve to get good at multi-step, agents work

# Example agent: **Devin**

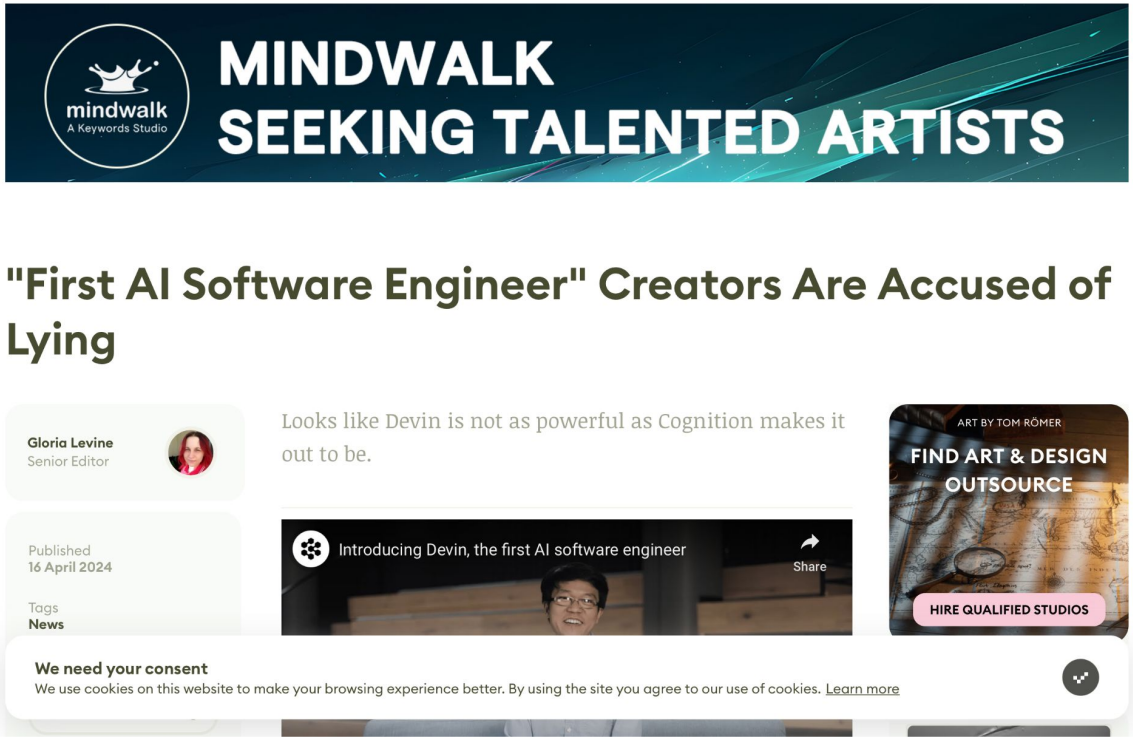
Cognition




https://www.youtube.com/watch?v=fjHtjT7GO1c

# Near-Term Future: Agents

But they’re not that good yet…




Source

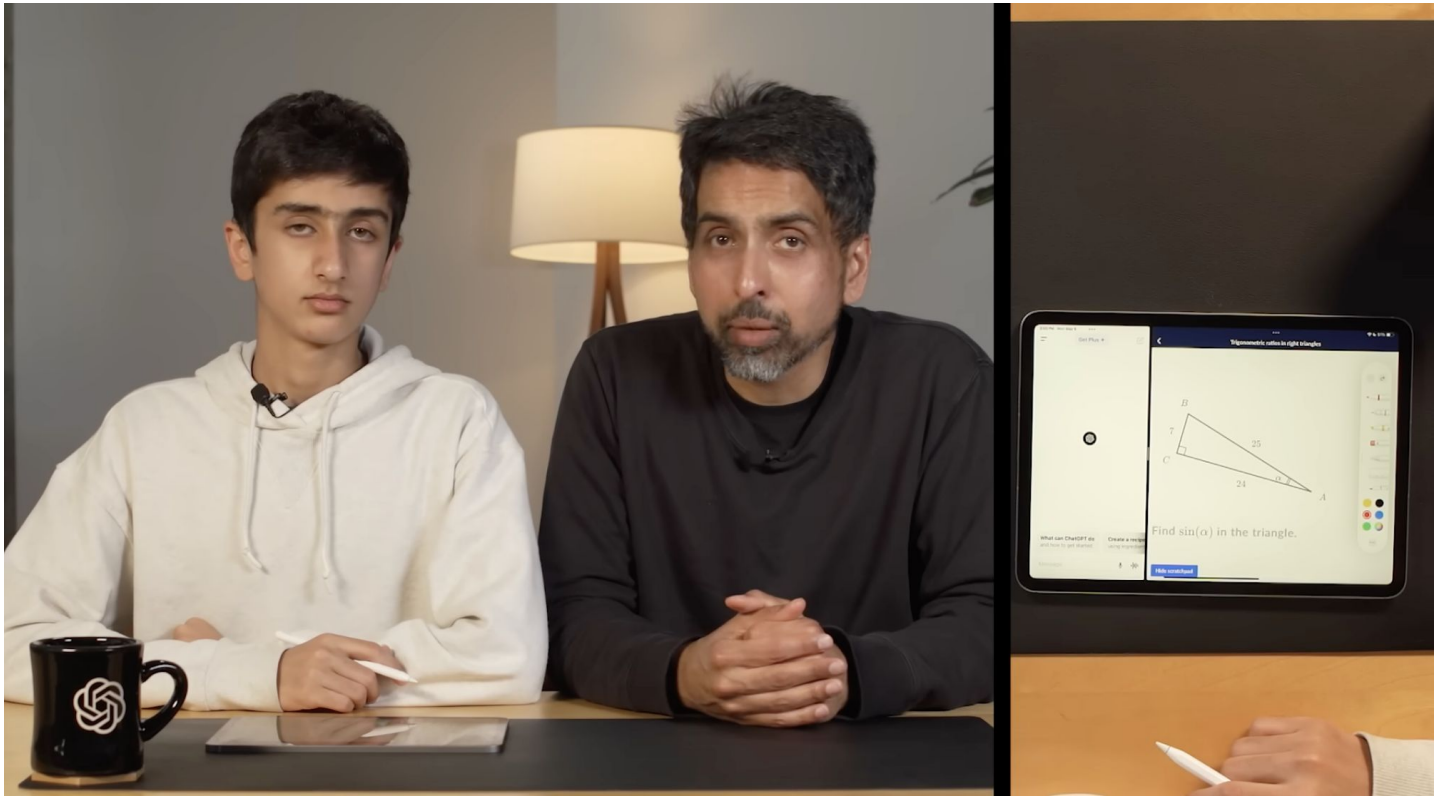
# Near-Term Future: Usability

- Current models aren't very usable!
- You have to type in text… very slow and painful!

# Near-Term Future: Usability


https://www.youtube.com/watch?v=c2DFg53Zhvw

# Near-Term Future: Good but far from great…




https://www.youtube.com/watch?v=_nSmkyDNulk

# Long-term future: Some definitions

**Artificial General Intelligence (AGI):** No agreed definition but an intelligence that is roughly human-level across the board.

Sam Altman (OpenAI CEO) → here

**Artificial Superintelligence (ASI):** An intelligence which is far beyond human level. Far harder to achieve.

Source

# What is AI?

## ANI vs. AGI vs. ASI

**Artificial narrow intelligence (ANI)**

Designed to perform specific tasks

**Artificial general intelligence (AGI)**

Can behave in a human-like way across all tasks

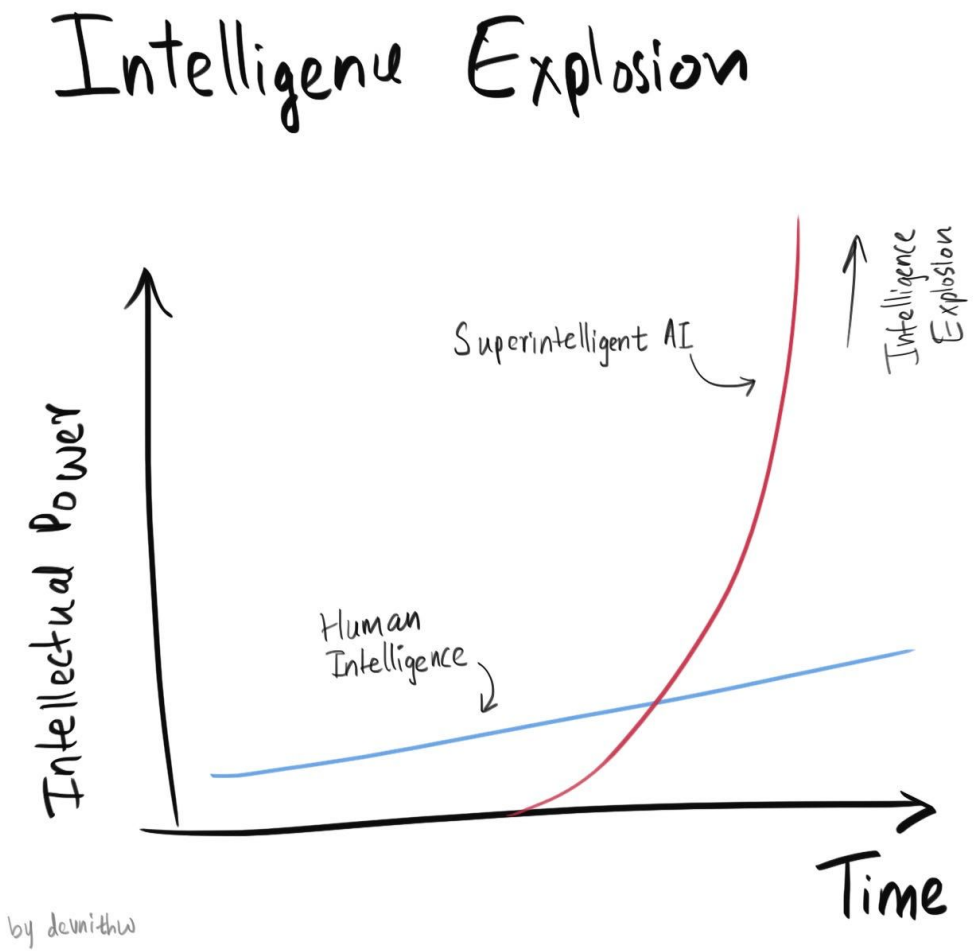
**Artificial super intelligence (ASI)**

Smarter than humans—the stuff of sci-fi

_zapier

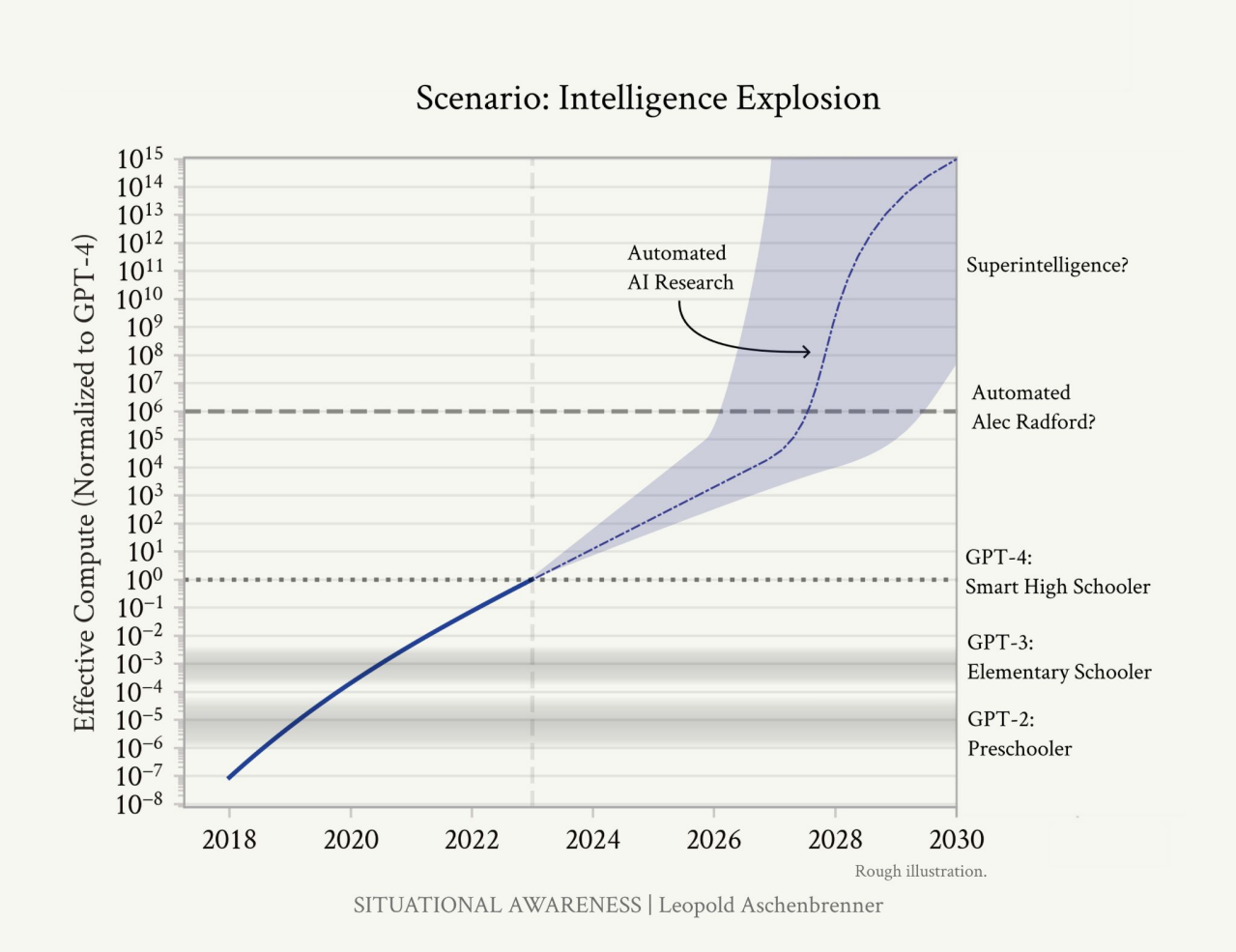
Source

# The intelligence explosion?




Source

# The intelligence explosion?




Source

# Sir Demis Hassabis' thoughts…

https://www.youtube.com/watch?v=BGxiufHVVd0

# Dario Amodei's thoughts…


https://www.youtube.com/watch?v=wo4o09lKAQQ

# What might be missing from the story of an intelligence explosion?

**Discussion**

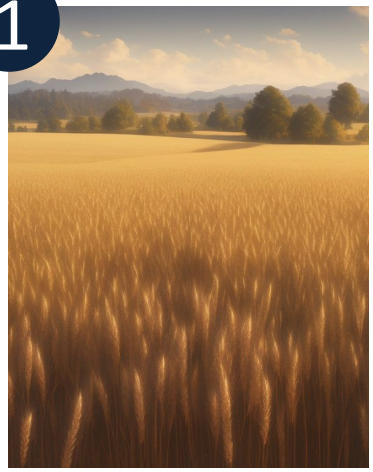
# State-of-the-art Evaluations

# What are Language Model Evaluations?

**Discussion**

# The Turing Test

1

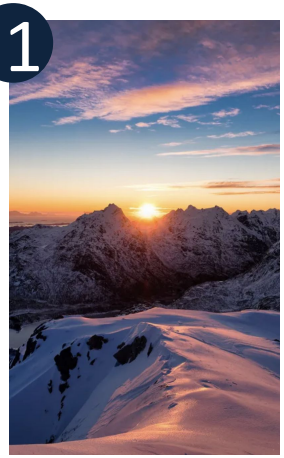


Link to original source

2

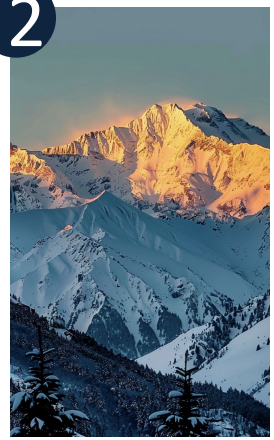
Link to original source

1



Link to original source

2



Link to original source

1

**Five English proverbs**

1. Many hands make light work
2. Strike while the iron is hot
3. The grass is always greener
4. Don’t judge a book by its cover
5. An apple a day keeps the doctor away

2

**The top five English proverbs:**

1. Actions speak louder than words.
2. A silent cat catches no mice.
3. A picture is worth a thousand words.
4. When in Rome, do as the Romans do.
5. The stone that rolls grows no moss.

# Modern Evals: What makes a good eval?

**Discussion**

# Modern Evals: Things to consider

1. Difficulty
2. Generalisability
3. Memorisation risk…
4. Superhuman abilities…?

# Modern Evals: Lessons from the trenches

Preprint. Under review.

## Lessons from the Trenches on Reproducible Evaluation of Language Models

Stella Biderman[1*], Hailey Schoelkopf[1*], Lintang Sutawika[1*], Leo Gao[1], Jonathan Tow[2], Baber Abbasi[1], Alham Fikri Aji[3], Pawan Sasanka Ammanamanchi[4], Sidney Black[1], Jordan Clive[5], Anthony DiPofi[1], Julen Etxaniz[6], Benjamin Fattori[1], Jessica Zosa Forde[7], Charles Foster[8], Jeffrey Hsu[9], Mimansa Jaiswal[10], Wilson Y. Lee[11], Haonan Li[3,12], Charles Lovering[13], Niklas Muennighoff[14], Ellie Pavlick[7], Jason Phang[1,15], Aviya Skowron[1], Samson Tan[16], Xiangru Tang[17], Kevin A. Wang[7], Genta Indra Winata[18], François Yvon[19], and Andy Zou[20]

[1]EleutherAI, [2]Stability AI, [3]MBZUAI, [4]IIIT Hyderabad, [5]Chattermill AI, [6]HiTZ Center - Ixa, UPV/EHU, [7]Brown University, [8]Finetune, [9]Ivy Natal, [10]University of Michigan, [11]HubSpot, [12]LibrAI, [13]Kensho, [14]Contextual AI, [15]New York University, [16]Amazon, [17]Yale University, [18]HKUST, [19]Sorbonne University, [20]CMU

[*]Equal Contribution

### Abstract

https://arxiv.org/pdf/2405.14782

# Modern Evals / Benchmarks

**Task in pairs:** 10 mins to research this eval/benchmark

- What does this eval test?
- Give an example problem - how hard is it?
- Overall is this benchmark good?

| | |
|---|---|
| **GSM8K** | **MGSM** |
| **MMLU** | **MMMU** |
| **BBH** | **MathVista** |
| **ARC / The ARC Challenge** | **WMDP Benchmark** |
| **HumanEval (Code)** | **GPQA** |

**Individual Task:** 10 minutes to read this paper

# LINGOLY: A Benchmark of Olympiad-Level Linguistic Reasoning Puzzles in Low-Resource and Extinct Languages

**Andrew Bean**[1*] **Simi Hellsten**[1,2]
**Harry Mayne**[1] **Jabez Magomere**[1] **Ethan A. Chi**[3] **Ryan Chi**[3]
**Scott A. Hale**[1,4] **Hannah Rose Kirk**[1]

[1]University of Oxford [2]United Kingdom Linguistics Olympiad
[3]Stanford University [4]Meedan

## Abstract

In this paper, we present the LINGOLY benchmark, a novel benchmark for advanced reasoning abilities in large language models. Using challenging Linguistic Olympiad puzzles, we evaluate (i) capabilities for in-context identification and

https://arxiv.org/pdf/2406.06196

# Evals and Visualisation Task

# **Evals Challenge:** 30 minutes

- Most modern evals are becoming memorised by the language models and are thus poor tests of true ability.
- Your task is to design your own eval. It can be really specific (only testing a certain type of behaviour) or really general but it should, in some way, test language model ‘intelligence’
- You need to design 7-10 questions for your eval and work out how to score the model responses.
- Benchmark 5 different language models on this benchmark and present the results in a table (remember how to make nice visualisations!)