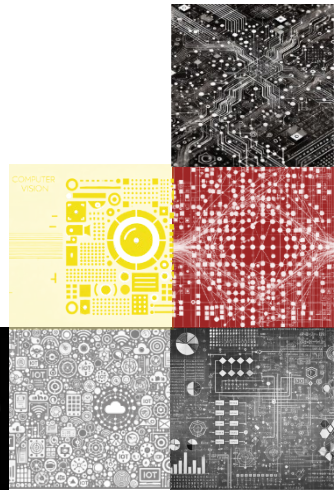


Research Centre for Computational Science & Mathematical Modelling

Prof. Matthew England
Coventry University
April 2025

What is AI and what can it do in 2025?
British Academy Early Career Researcher Network



This talk

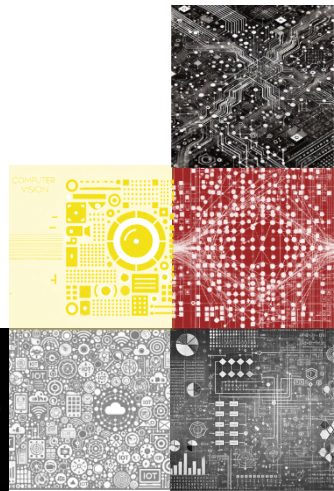
Audience: British Academy Early Career Researcher Network

= “An inclusive, researcher-led network for UK-based early career researchers working in the humanities and social sciences”.

I.e. smart people, interested to hear about AI (as they choose to attend), but unlikely to be studying the technical details.

This talk:

- Does not assume any prior knowledge of AI.
- Will not cover technical details.
- Will try to focus on what may be of interest to the audience.



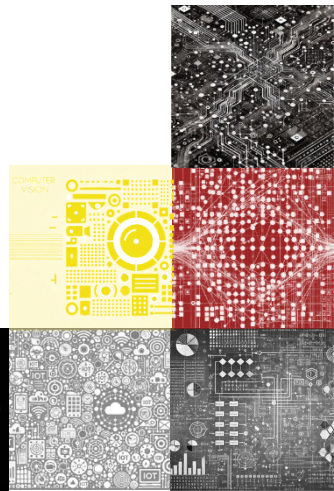
Contents

1. What is AI?

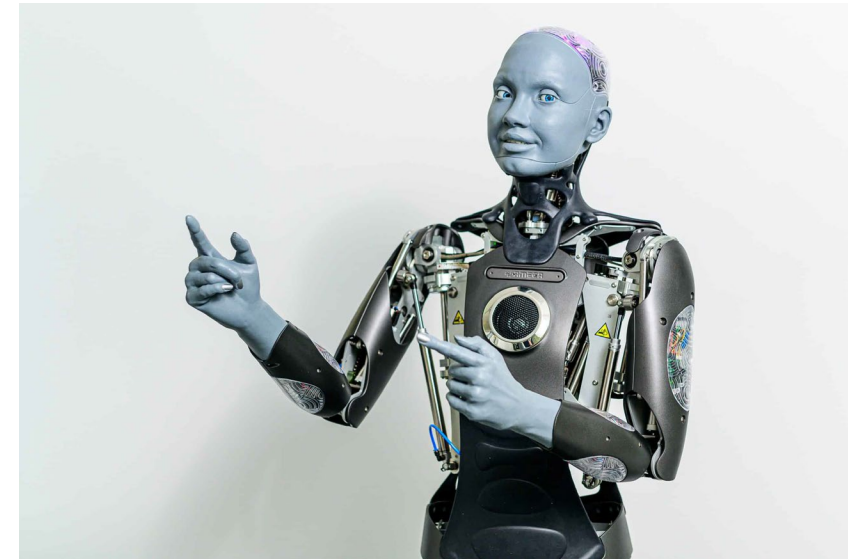
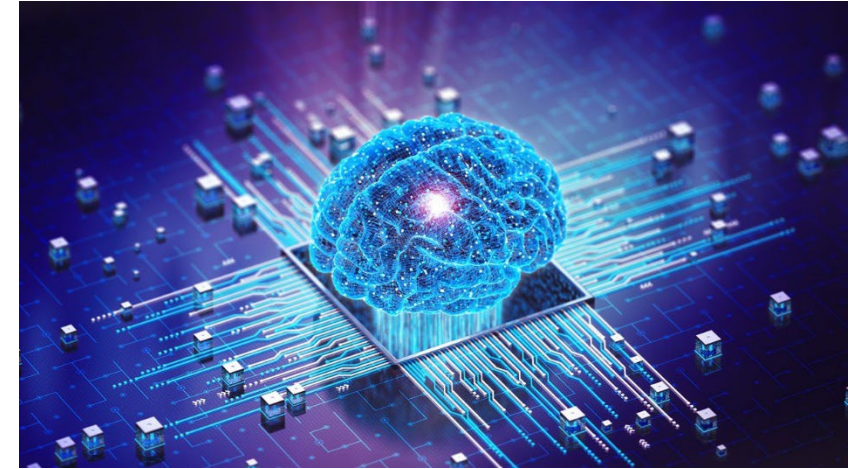
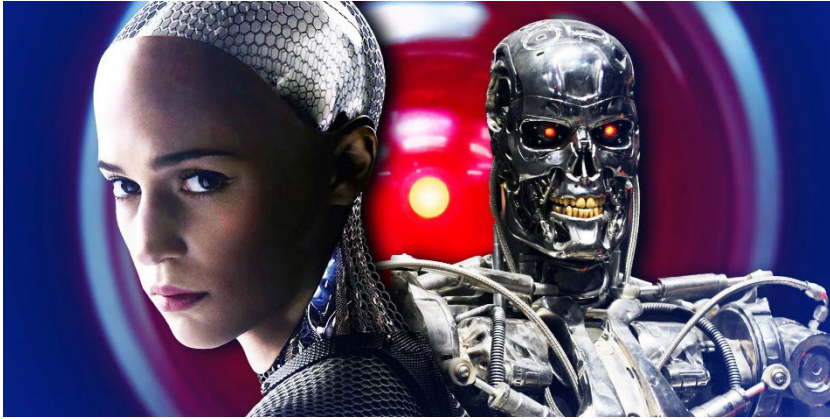
2. Key Questions for AI Beginners to Ask.

3. What's Changed in the Last Year?

4. AI for Researchers.



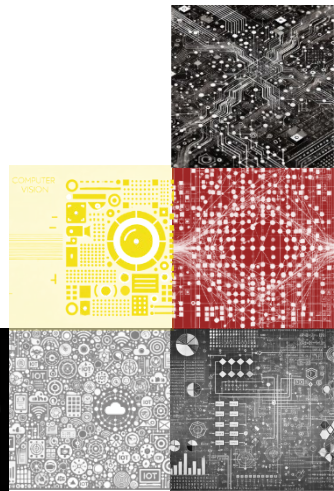
What is AI?

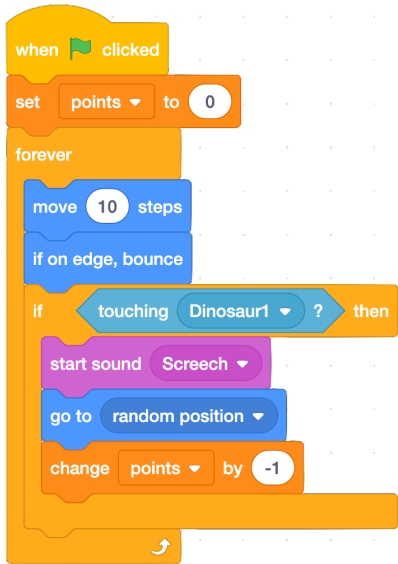


What is AI Technically?

Artificial Intelligence (AI): *methods for systems to perceive their environment, make decisions, and have the ability to learn and adapt over time.*

There are lots of very different approaches to AI.

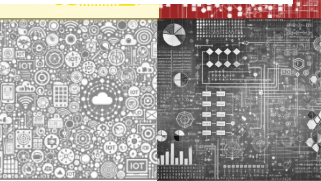
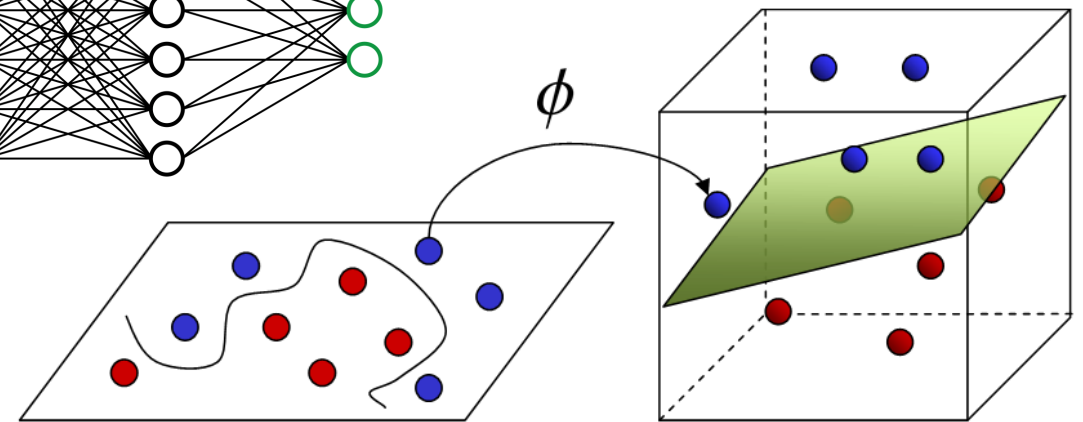
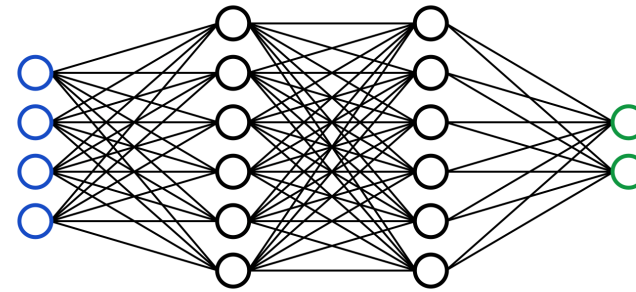
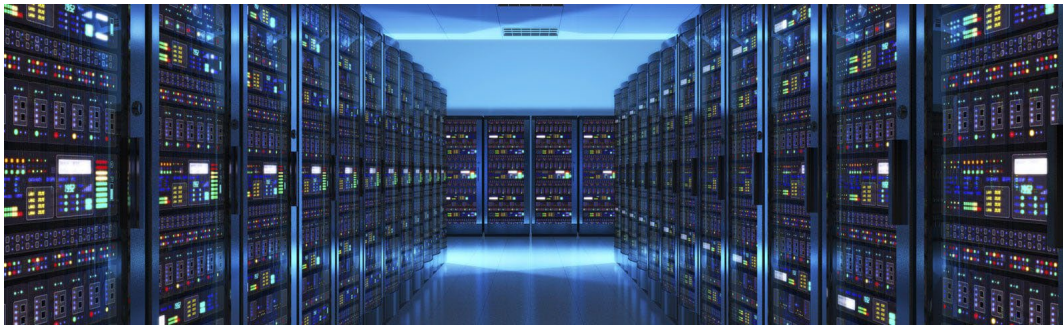
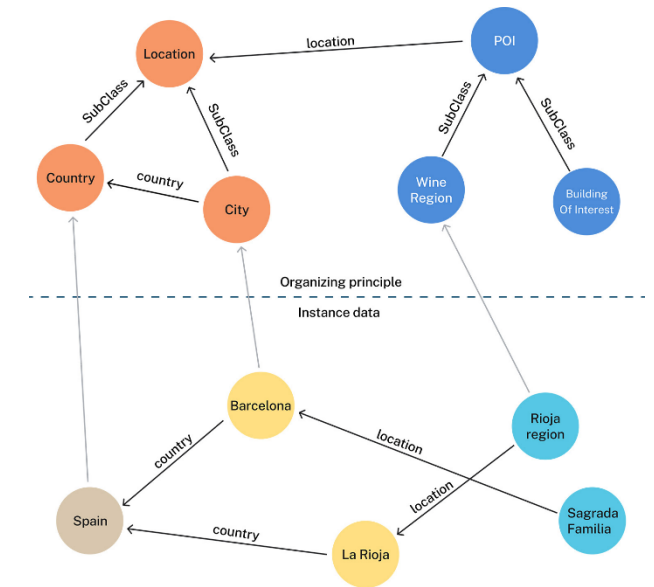




```

while (user != "bye"):
  keyword_found = False

  for index in range(len(keywords)):
    if (keywords[index] in user):
      print("Bot: " + responses[index])
      keyword_found = True
  
```



What is AI Technically?

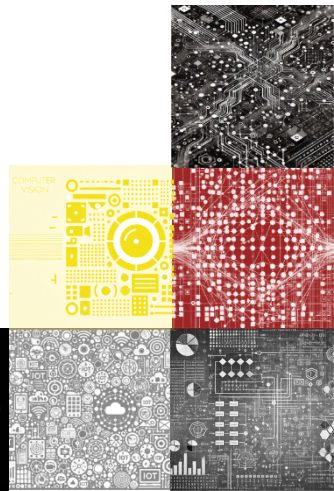
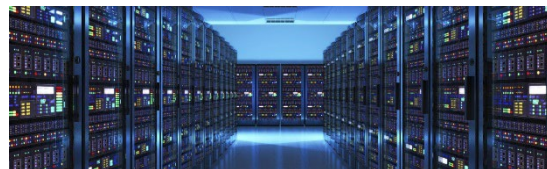
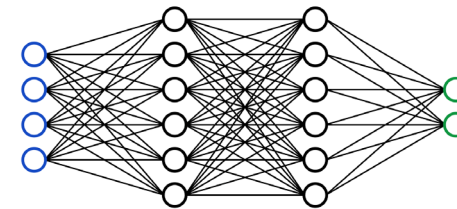
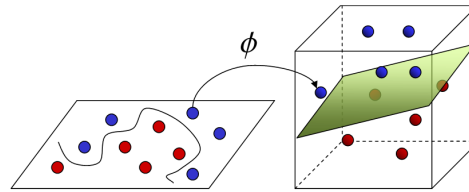
Artificial Intelligence (AI): *methods for systems to perceive their environment, make decisions, and have the ability to learn and adapt over time.*

There are lots of very different approaches to AI. But in 2025 it really means...

Machine Learning (ML): statistical algorithms that can learn patterns in data and generalise to unseen data ***without explicit instructions.***

Key ingredients of ML:

- Statistical Models which are formed by training on data
- Lots of computing power
- Large amounts of data

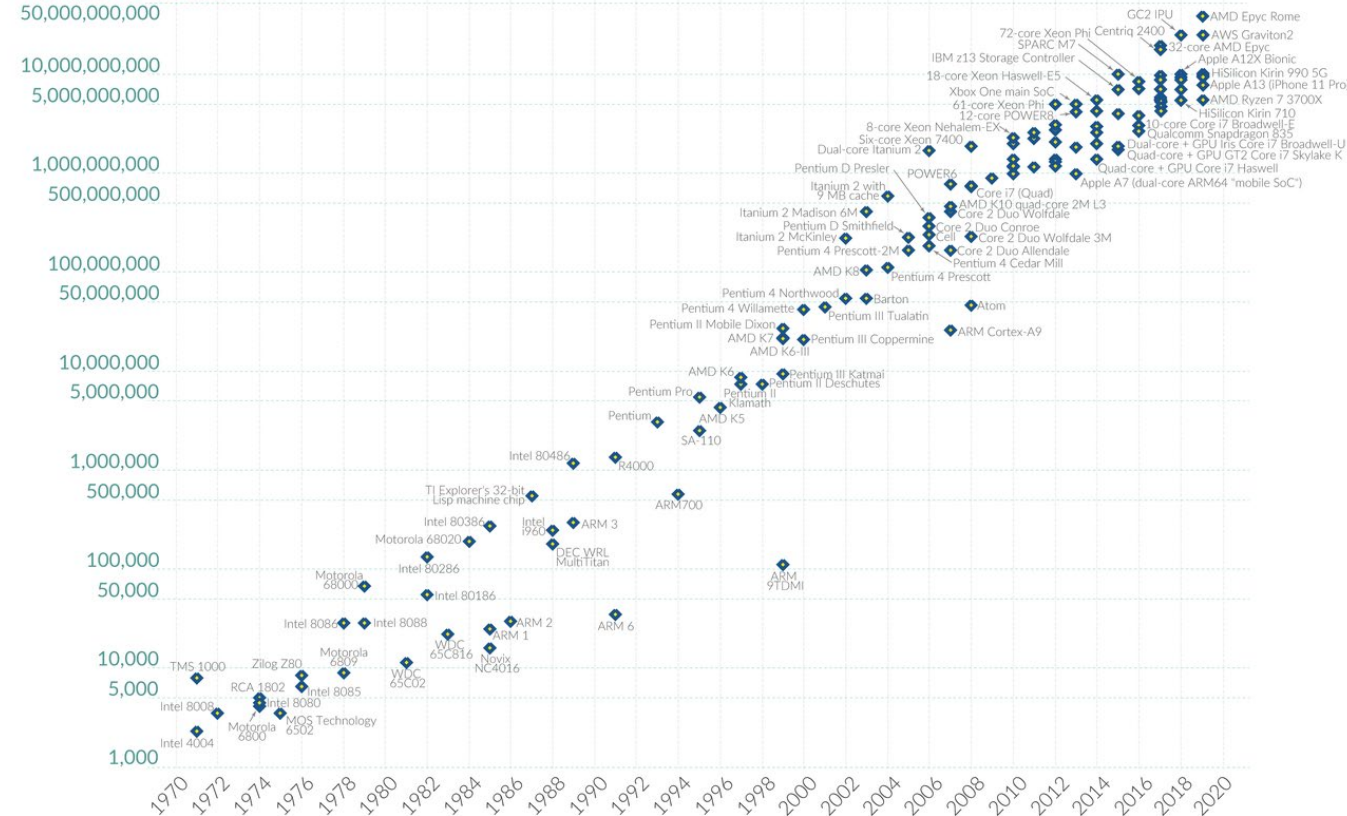


Why now? Big Compute, big Data (and big money)

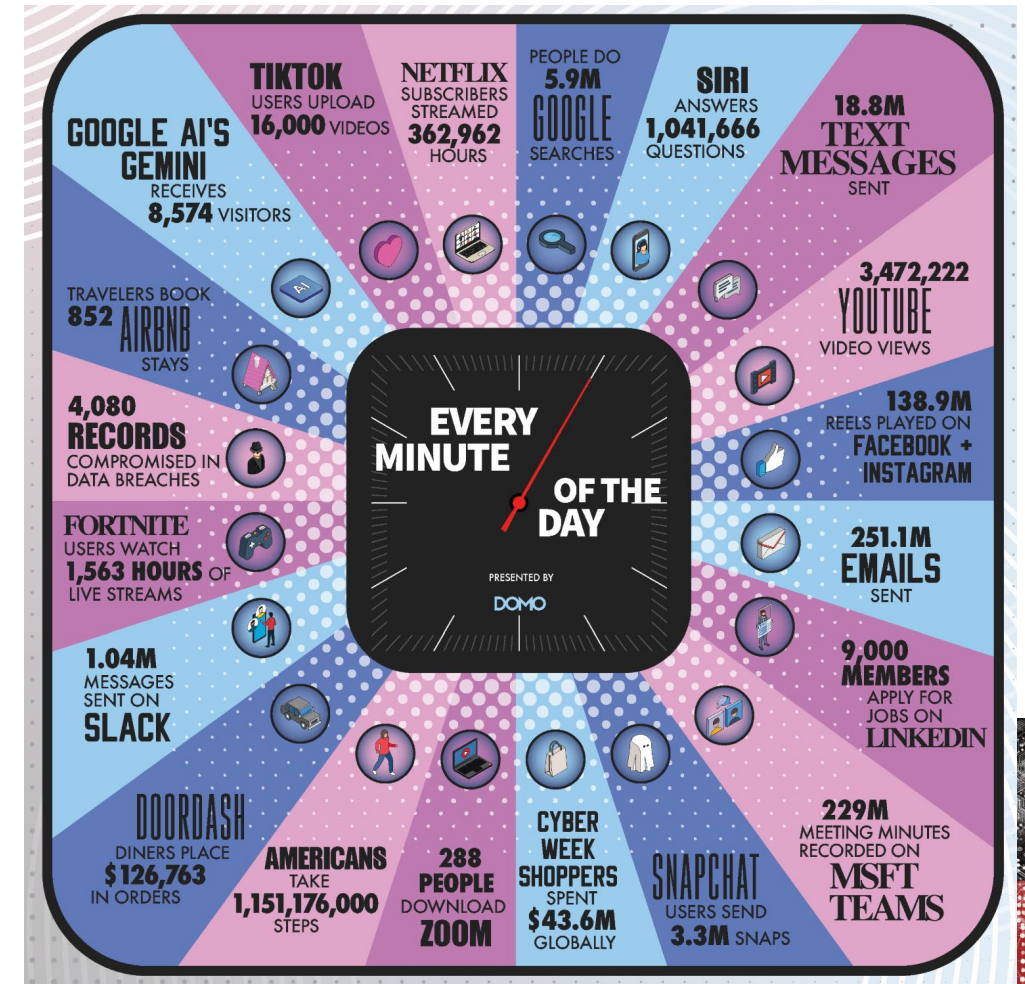
Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Transistor count

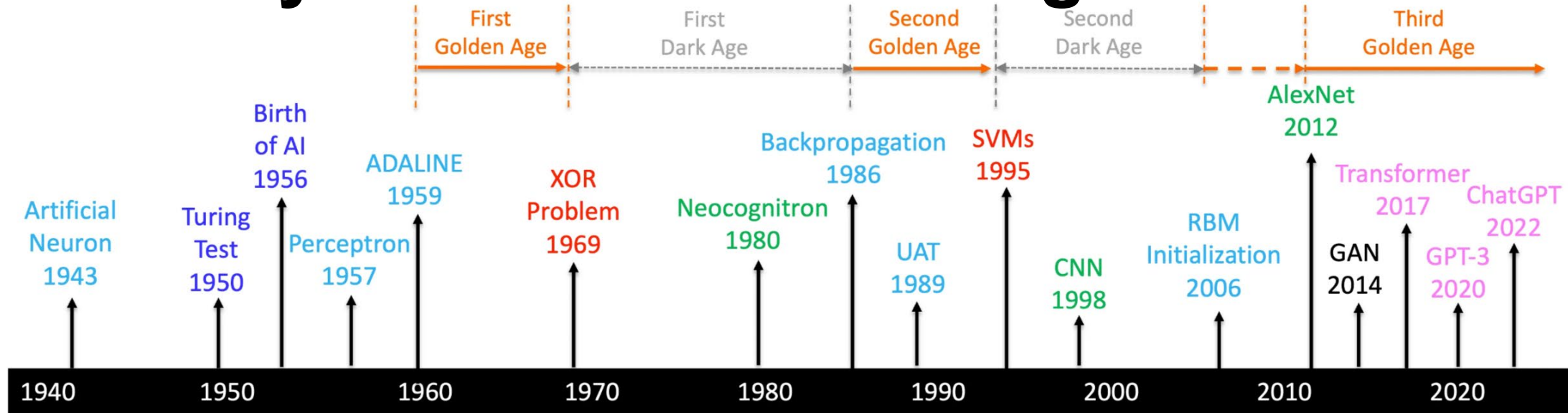


OurWorldinData.org – Research and data to make progress against the world's largest problems. Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

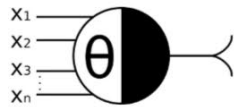


<https://www.domo.com/learn/infographic/data-never-sleeps-12>

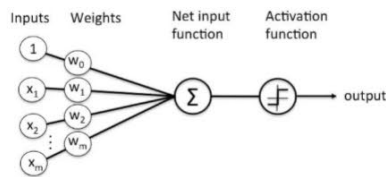
History of Machine Learning Models



McCulloch-Pitts



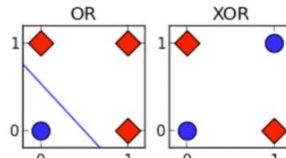
Rosenblatt



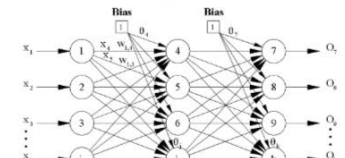
Widrow-Hoff



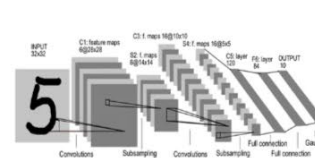
Minsky-Papert



Rumelhart, Hinton et al.



LeCun



Hinton-Ruslan Krizhevsky et al.



Vaswani



<https://medium.com/@lmpo/a-brief-history-of-ai-with-deep-learning-26f7948bc87b>

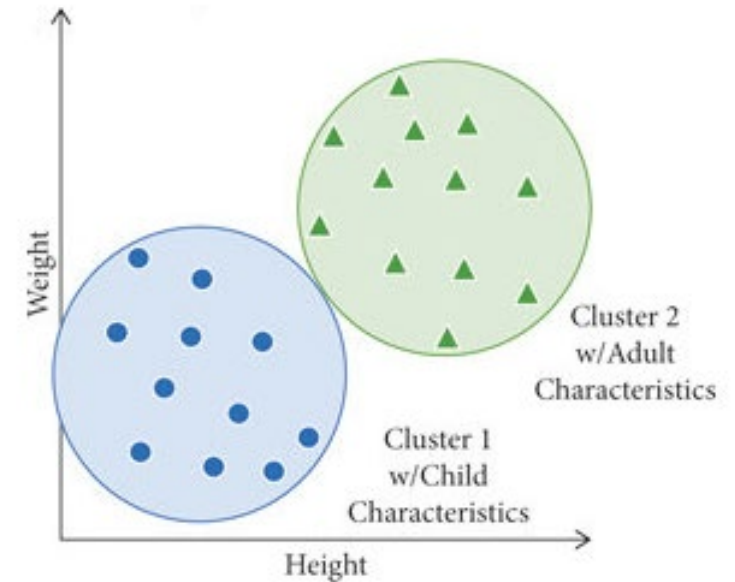
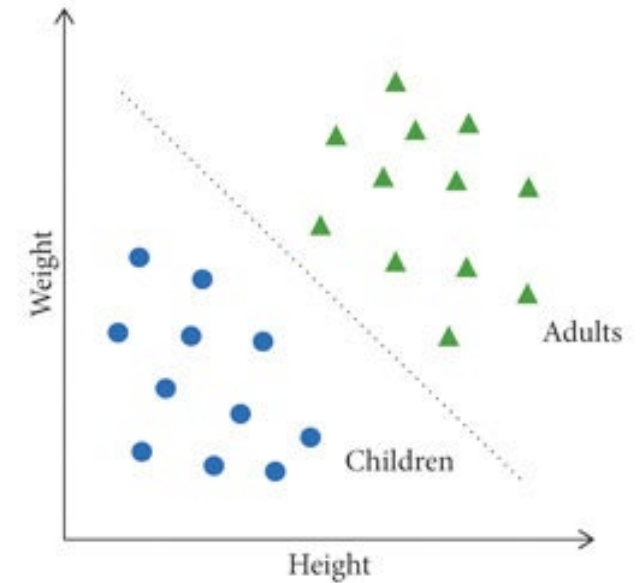
Types of Machine Learning

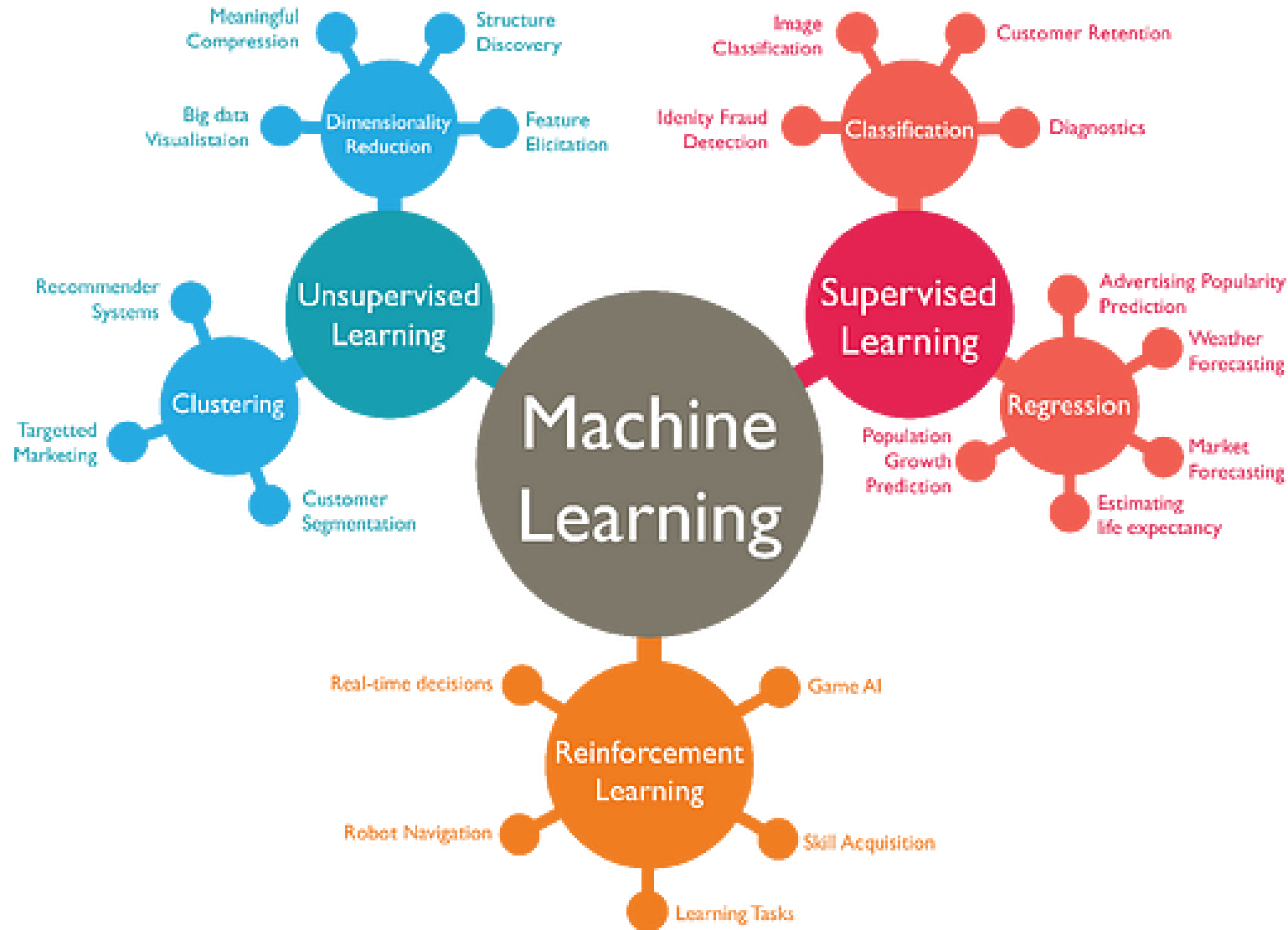
Supervised Learning: identify patterns based on existing large labelled datasets.

May be used to solve e.g. Classification Problems: “Does this image of a lung have cancer?”
(Based on this dataset of labelled images).

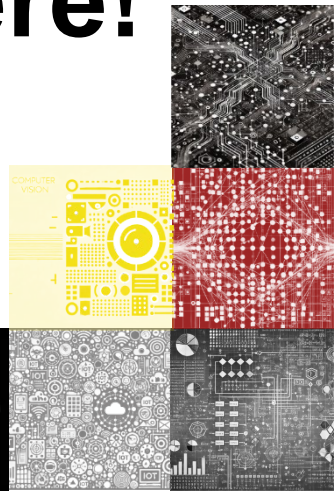
Unsupervised Learning: identify patterns in large unlabelled datasets (much harder).

May be used to solve e.g. Clustering Problems.
“Can we split our customers into groups for marketing?”
(Based on our records of their purchasing habits).



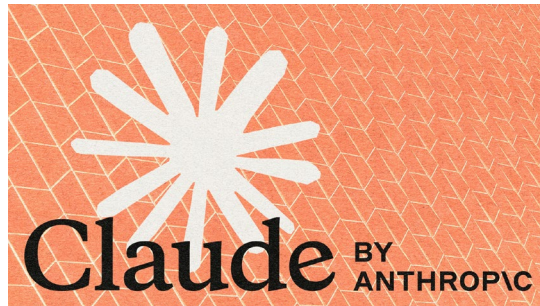


**Do be
aware
there is a
LOT more
machine
learning
out there!**

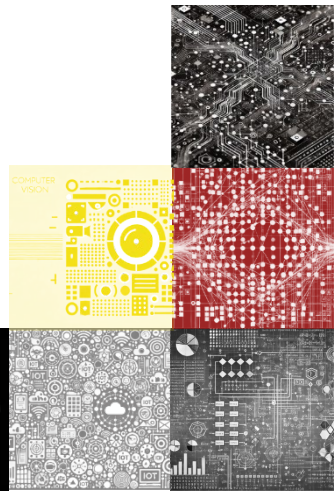


Generative AI

Generative AI uses machine learning technology to create “*new*” content. The most high-profile example is for text data via **Large Language Models (LLMs)**.



There are also Generative AI systems to make images, video, music, voices, code and more!

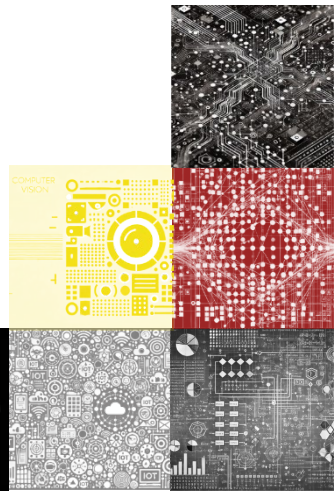
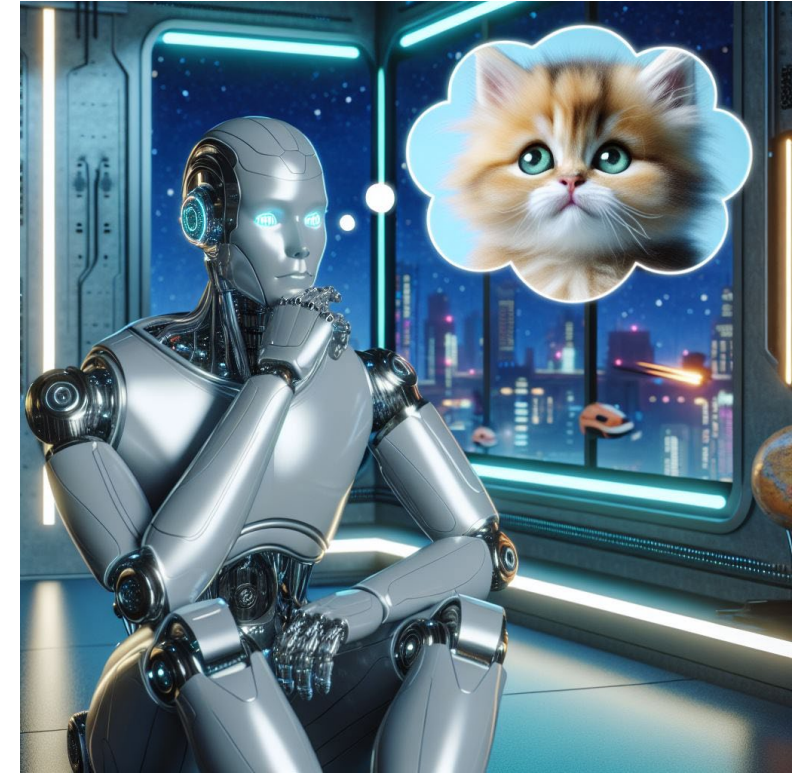


How does an LLM work?

At its core an LLM is a machine learning model to **predict the next word in a text, based on the words that came previously**. The model trained to recreate the patterns it has observed in the training data.

E.g. If we type “**the cat sat on the**” then a good prediction for the next word is “**mat**”.

LLMs know this, not because they have been programmed about cats (or even about English grammar). They know it because that phrase is used multiple times in the training data.



LLMs vs ChatBots vs AI Assistants

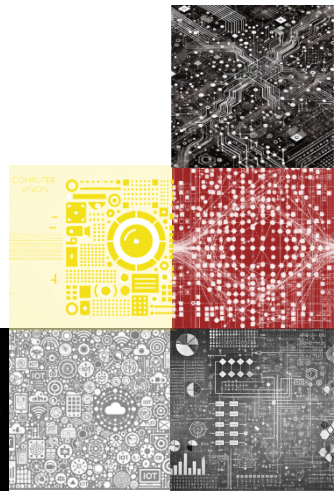
An **LLM** (e.g. GPT) generates new text based on the “context” (preceding text).



A **ChatBot** (e.g. ChatGPT) uses the LLM in a conversational interface, and fine tunes it for conversational norms (e.g. avoiding harmful content).

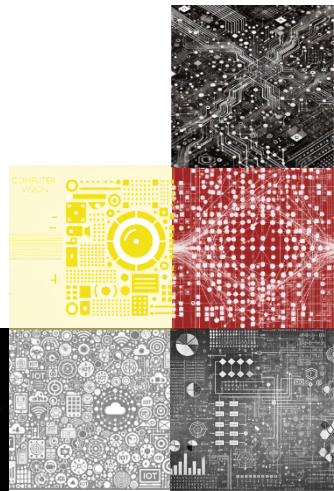


An **AI Assistant** integrates a ChatBot with other tools into a larger system.



AI Opportunities

- Automation of tasks
- Increased productivity
- Increased personalisation
- Optimisation of systems
- New scientific discoveries
- New artistic methods
- Promote new / better thinking

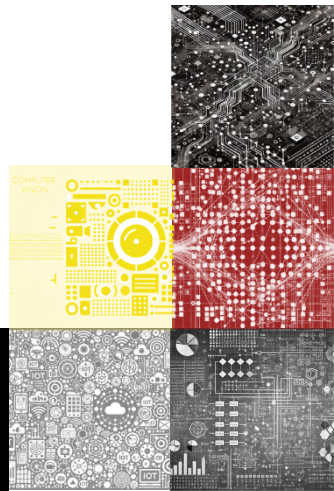


AI Opportunities

- Automation of tasks
- Increased productivity
- Increased personalisation
- Optimisation of systems
- New scientific discoveries
- New artistic methods
- Promote new / better thinking

AI Challenges

- Robustness
- Explainability
- Privacy Protection
- Governance
- Societal changes
- Misuse
- Inequality



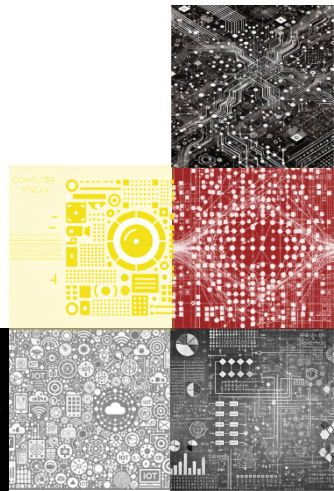
Contents

1. What is AI?

2. Key Questions for AI Beginners to Ask.

3. What's Changed in the Last Year?

4. AI for Researchers.



Key questions for beginners in AI 1:

(Q) Do you really need AI? AI is great for spotting patterns that you cannot; but if you already know the patterns, then traditional *digitalisation* and *automation* may be what you need!



Key questions for beginners in AI 1:

(Q) Do you really need AI? AI is great for spotting patterns that you cannot; but if you already know the patterns, then traditional *digitalisation* and *automation* may be what you need!

Assuming you need AI: AI runs on data.

(Q) Will you be able to use a pre-trained AI model, or will you train one on your own data?

(Q) Do you have that data?

(Q) Do you trust the quality of that data?



Important: bias in data → bias in AI

Dataset trained Microsoft chatbot to spew racist tweets

In March 2016, Microsoft learned that using Twitter interactions as training data for ML algorithms can have dismaying results.

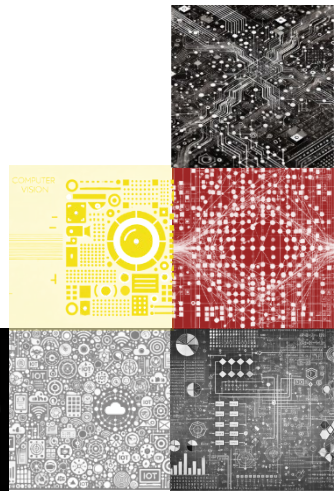
Within 16 hours, the chatbot posted more than 95,000 tweets, and those tweets rapidly turned overtly racist, misogynist, and anti-Semitic. Microsoft quickly suspended the service for adjustments and ultimately pulled the plug.

Amazon AI-enabled recruitment tool only recommended men

Like many large companies, Amazon is hungry for tools that can help its HR function screen applications for the best candidates. In 2014, Amazon started working on AI-powered recruiting software to do just that. There was only one problem: The system vastly preferred male candidates. In 2018, Reuters broke the news that Amazon had scrapped the project.

AI algorithms identify everything but COVID-19

Derek Driggs, an ML researcher at the University of Cambridge, together with his colleagues, published a paper in *Nature Machine Intelligence* that explored the use of deep learning models for diagnosing the virus. The paper determined the technique wasn't fit for clinical use. For example, Driggs' group found that their own model was flawed because it was trained on a data set that included scans of patients that were lying down while scanned, and patients that were standing up. The patients who were lying down were much more likely to be seriously ill, so the algorithm learned to identify COVID risk based on the position of the person in the scan.



Apple

Apple suspends AI-generated news alert service after BBC complaint

Inaccurate notices branded with broadcaster's logo sent to iPhone users but tech firm works on improvements

Dan Milmo *Global technology editor*

Fri 17 Jan 2025 09:57 GMT

 Share



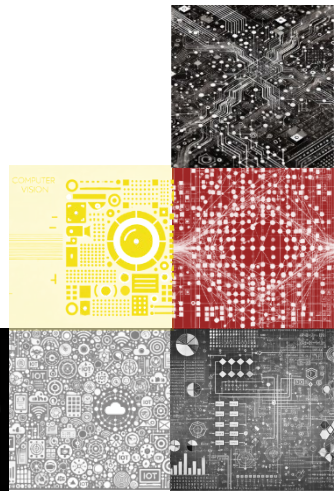
 Apple says: 'Notification summaries for the news and entertainment category will be temporarily unavailable.' Photograph: Matt Cardy/Getty Images

Apple is suspending an artificial intelligence feature that made inaccurate summaries of news headlines.

Chatbot AI makes racist judgements on the basis of dialect

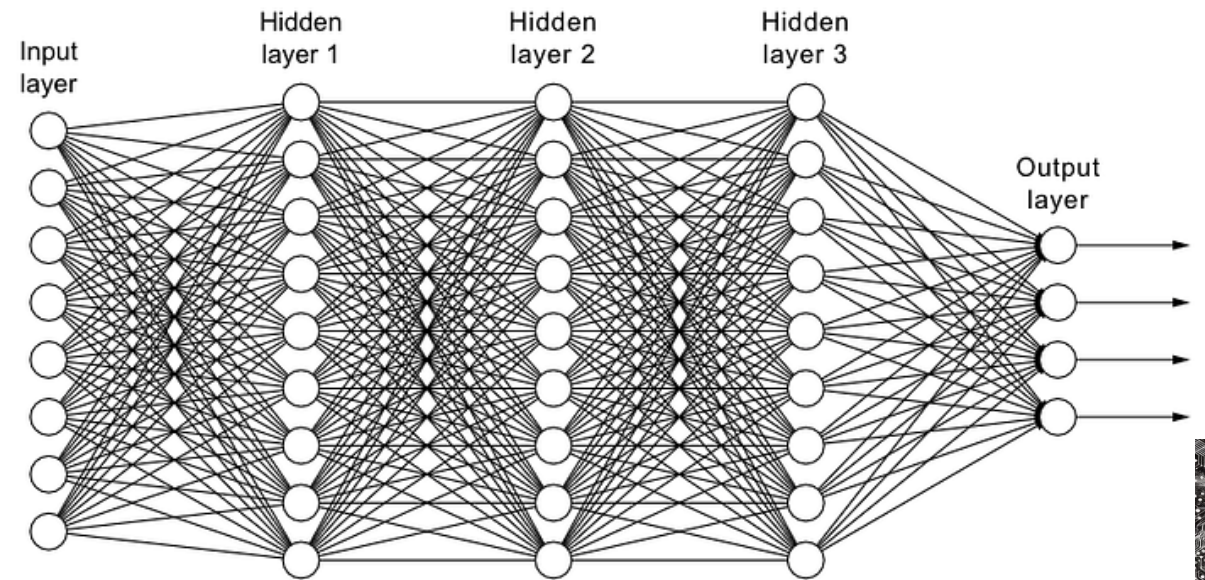
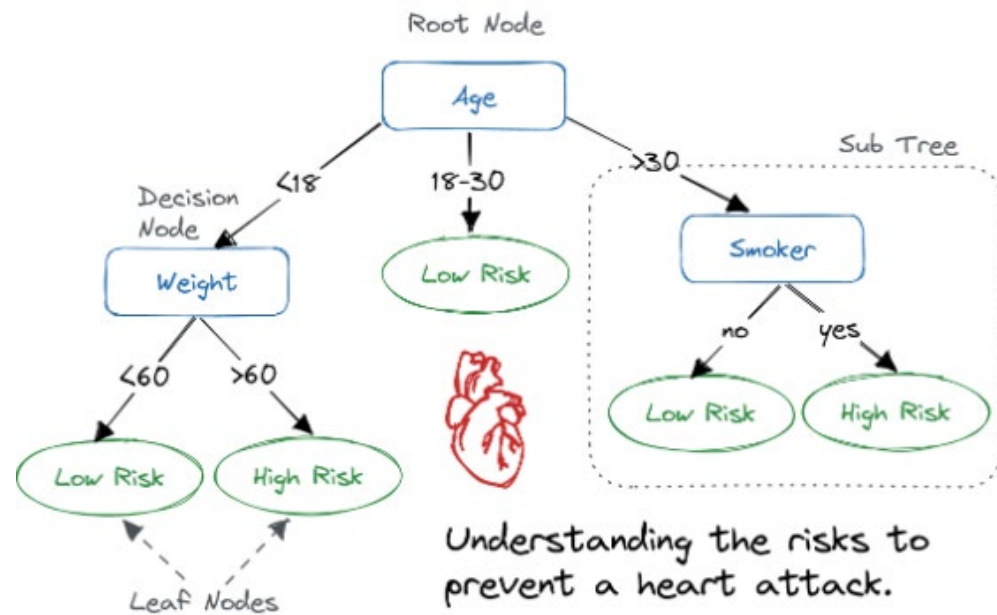
Some large language models harbour hidden biases that cannot be removed using standard methods.

Researchers found that some artificial intelligence (AI) systems are more likely to recommend the death penalty to a fictional defendant presenting a statement written in African American English (AAE) – a dialect spoken by millions of people in the United States that is associated with the descendants of enslaved African Americans – compared with one written in Standardized American English (SAE). The chatbots were also more likely to match AAE speakers with less-prestigious jobs.

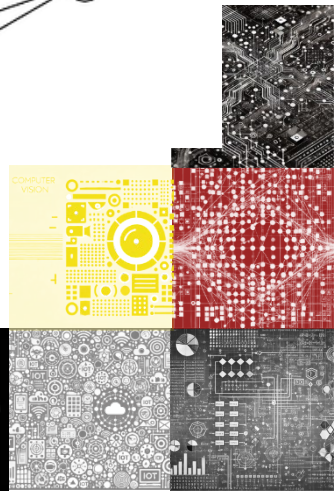


Key questions for beginners in AI 2:

(Q) Is my AI model interpretable? There are a wide variety of AI models out there, some can be interpreted by humans easier than others.



There is a perceived trade-off between accuracy and interpretability of AI.



Key questions for beginners in AI 3:

(Q) Is an LLM the right AI tool for me?

If you have lots of labelled data and the you want to predict something from it, then probably not! You want a traditional *supervised machine learning tool* instead (much cheaper and safer; maybe more interpretable).

If you do decide LLM, then remember that it is forming responses based on what is freely available online.

(Q) Is that ok for your use case? Is the knowledge you need the AI to have already online?

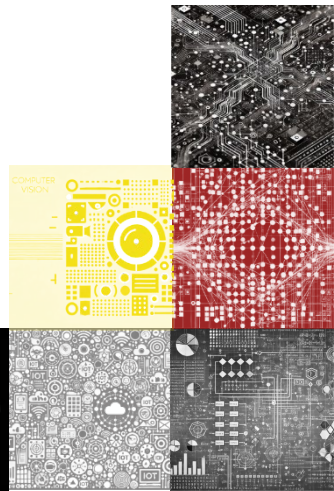


Key questions for beginners in AI 4:

Assuming you have answered the above questions and decided an LLM is the right tool for you.

(Q) How to address the GDPR or other data security concerns around using an LLM on your data?

If you are using a **free** LLM and are showing it personal data, then you are probably not GDPR compliant!



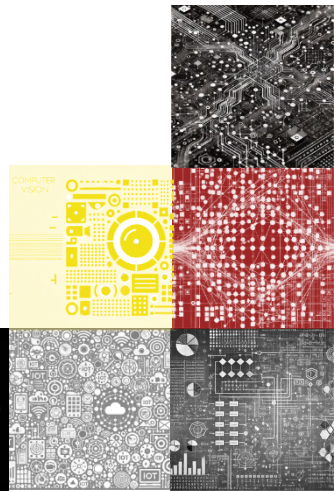
CU Staff have Microsoft Copilot

All Coventry University staff now have access to Microsoft Copilot as part of our Microsoft licence. You can access it here:

<https://copilot.cloud.microsoft>

This can be used to create both text and images. It can be asked to generate content with citations to sources. We do not (yet) have student access, or integration with other university tools.

**Find out what you
can access via
your university!**



Write me a 200 word summary of the state of AI at the start of 2025, citing your sources of facts.

At the start of 2025, artificial intelligence (AI) continues to evolve rapidly, with several key trends and challenges shaping its landscape. One significant development is the rise of **generative virtual worlds**, where AI models create interactive environments, enhancing gaming and virtual experiences ¹. Additionally, **smaller, more efficient language models** are becoming prevalent, making AI more accessible and versatile ¹.

Despite these advancements, AI faces notable challenges. **Cost** remains a significant barrier, as developing and maintaining AI systems is expensive due to the concentration of industry players ². Moreover, **data limitations** pose a challenge, as AI models rely heavily on existing human knowledge, and the quality of AI-generated data can degrade over time ².

In the business sector, the adoption of AI has surged, with 72% of large firms using AI weekly in 2024, up from 37% in 2023 ². This trend is expected to continue, with companies exploring the best use cases for AI to enhance their operations ².

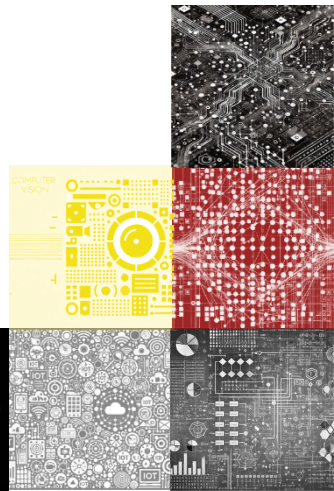
Overall, AI in 2025 is marked by exciting innovations and persistent challenges, reflecting its dynamic and rapidly evolving nature.

¹: MIT Technology Review ²: Knowledge at Wharton

What aspect of AI are you most interested in?

1 | 🌐 What's next for AI in 2025 ...

2 | 🌐 AI in 2025: What Challeng...



Write me a 200 word summary of the state of AI at the start of 2025, citing your sources of facts.

At the start of 2025, artificial intelligence (AI) continues to evolve rapidly, with several key trends and challenges shaping its landscape. One significant development is the rise of **generative virtual worlds**, where AI models create interactive environments, enhancing gaming and virtual experiences ¹. Additionally, **smaller, more efficient language models** are becoming prevalent, making AI more accessible and versatile ¹.

Despite these advancements, AI faces notable challenges. **Cost** remains a significant barrier, as developing and maintaining AI systems is expensive due to the concentration of industry players ². Moreover, **data limitations** pose a challenge, as AI models rely heavily on existing human knowledge, and the quality of AI-generated data can degrade over time ².

In the business sector, the adoption of AI has surged, with 72% of large firms using AI weekly in 2024, up from 37% in 2023 ². This trend is expected to continue, with companies exploring the best use cases for AI to enhance their operations ².

Overall, AI in 2025 is marked by exciting innovations and persistent challenges, reflecting its dynamic and rapidly evolving nature.

¹: MIT Technology Review ²: Knowledge at Wharton

What aspect of AI are you most interested in?

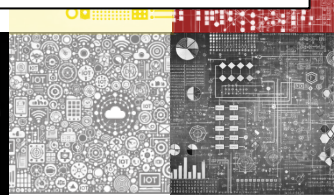
1 | 🌐 What's next for AI in 2025 ...

2 | 🌐 AI in 2025: What Challeng...

Do check the references!

← → ↺ 🔍 knowledge.wharton.upenn.edu/article/ai-in-2025-what-challenges-lie-ahead/

The numbers tell the story: A recent Wharton study found that only **37%** of large firms used AI weekly in 2023, but that increased to 72% in 2024. Lynn Wu, a Wharton professor of operations, information and decisions, expects that upward trend will continue in 2025.



Write me a 200 word summary of the state of AI at the start of 2025, citing your sources of facts.

At the start of 2025, artificial intelligence (AI) continues to evolve rapidly, with several key trends and challenges shaping its landscape. One significant development is the rise of **generative virtual worlds**, where AI models create interactive environments, enhancing gaming and virtual experiences ¹. Additionally, **smaller, more efficient language models** are becoming prevalent, making AI more accessible and versatile ¹.

Despite these advancements, AI faces notable challenges. **Cost** remains a significant barrier, as developing and maintaining AI systems is expensive due to the concentration of industry players ². Moreover, **data limitations** pose a challenge, as AI models rely heavily on existing human knowledge, and the quality of AI-generated data can degrade over time ².

In the business sector, the adoption of AI has surged, with 72% of large firms using AI weekly in 2024, up from 37% in 2023 ². This trend is expected to continue, with companies exploring the best use cases for AI to enhance their operations ².

Overall, AI in 2025 is marked by exciting innovations and persistent challenges, reflecting its dynamic and rapidly evolving nature.

¹: MIT Technology Review ²: Knowledge at Wharton

What aspect of AI are you most interested in?

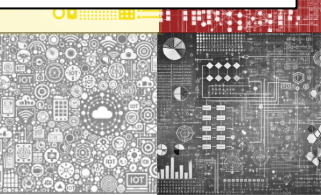
1 | 🌐 What's next for AI in 2025 ...

2 | 🌐 AI in 2025: What Challeng...

Do check the references!

You can ask it to restrict to academic references, but it currently struggles with that request.

← → ↺ 🔍 knowledge.wharton.upenn.edu/article/ai-in-2025-what-challenges-lie-ahead/
The numbers tell the story: A recent Wharton study found that only **37%** of large firms used AI weekly in 2023, but that increased to 72% in 2024. Lynn Wu, a Wharton professor of operations, information and decisions, expects that upward trend will continue in 2025.



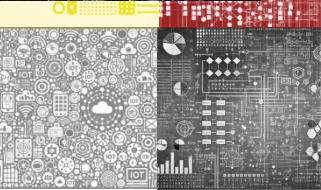
What is an AI Hallucination?

Basically, a hallucination is when the text generated by an LLM is false.

Sidenote: I really dislike the term “Hallucination”!

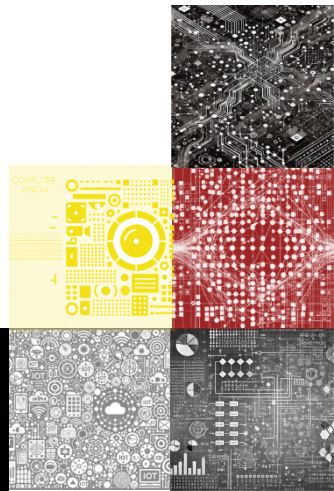
The text generated is always statistically likely based on the training data. It is astonishing that statistical likelihood takes us this far! But it is not, and never can be, a guarantee of correctness.

If you want AI that is guaranteed correct, then it cannot be solely machine learning based.



Compared to this time last year:

- LLMs making far less hallucinations.
- But there are still hallucinations there, and they cannot be removed completely.
- It is now easier to identify hallucinations: request the sources of facts and then **check them!**
- New tools being developed that focus LLMs on “high quality” or “user selected” data. New tools that force LLMs to explain their reasoning.
- Note that the LLM itself is rarely critical of its sources: you need to do that critical review!



Key questions that ~~beginners~~ everyone should ask:

(Q) How will I judge whether my AI produced content is correct?

(Q) What risk mitigations do I have in place for the cases where it is not correct?



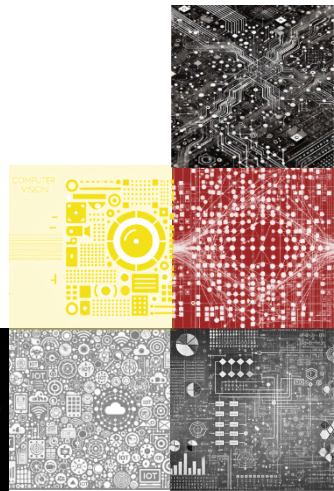
Design a new logo for Coventry University. It should involve a Phoenix.

I'll try to create that.

Designer Preview



**Generative AI can
also be used to
produce images!**



Design a new logo for Coventry University. It should involve a Phoenix.

I'll try to create that.

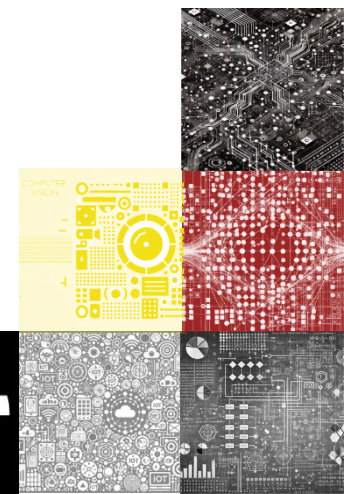
Designer Preview



Can you do it again but use the colour scheme of blue and gold.

I'll try to create that.

Designer Preview



Design a new logo for Coventry University. It should involve a Phoenix.

I'll try to create that.

Designer Preview



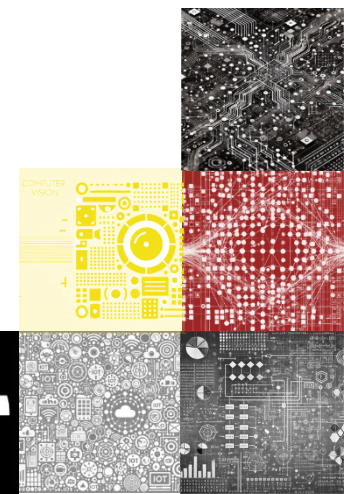
Can you do it again but use the colour scheme of blue and gold.

I'll try to create that.

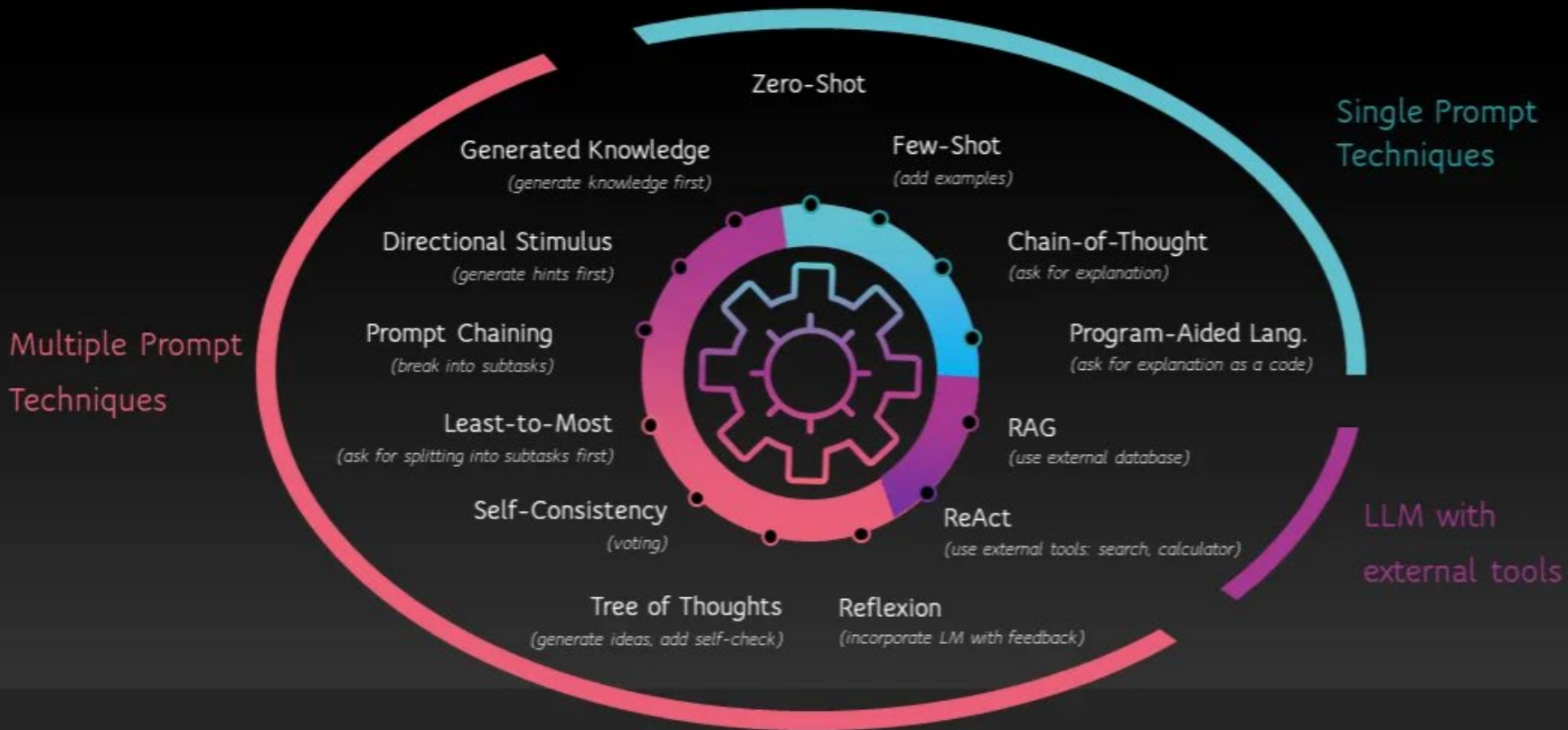
Designer Preview



This act of a human asking for edits to the AI content is becoming known as ***prompt engineering***.



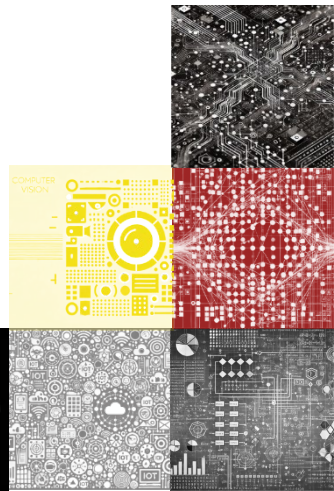
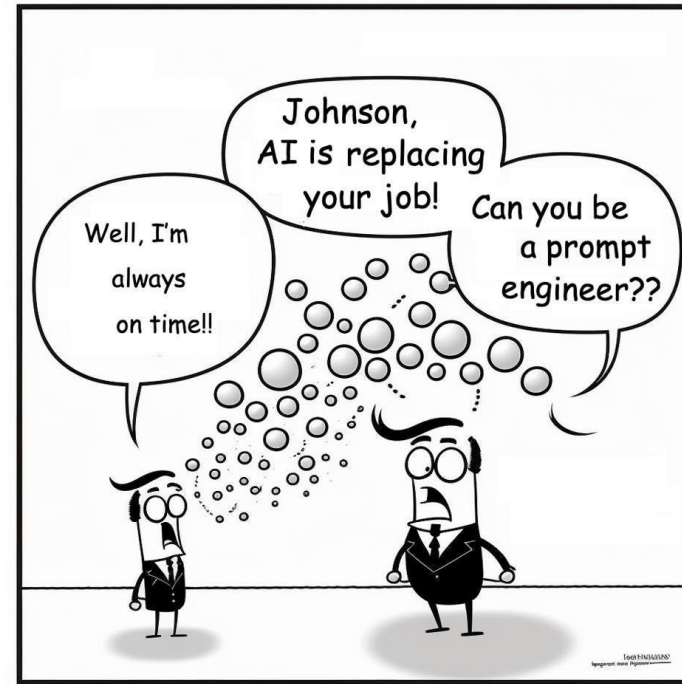
Prompt Engineering Techniques



<https://medium.com/the-modern-scientist/prompt-engineering-classification-of-techniques-and-prompt-tuning-6d4247b9b64c>

How to improve LLM Output?

- Use the latest LLMs.
- Experiment with different LLMs.
- Experiment with prompt engineering techniques.



How to improve LLM Output?

- Use the latest LLMs.
- Experiment with different LLMs.
- Experiment with prompt engineering techniques.
- Combine LLMs with other tools.

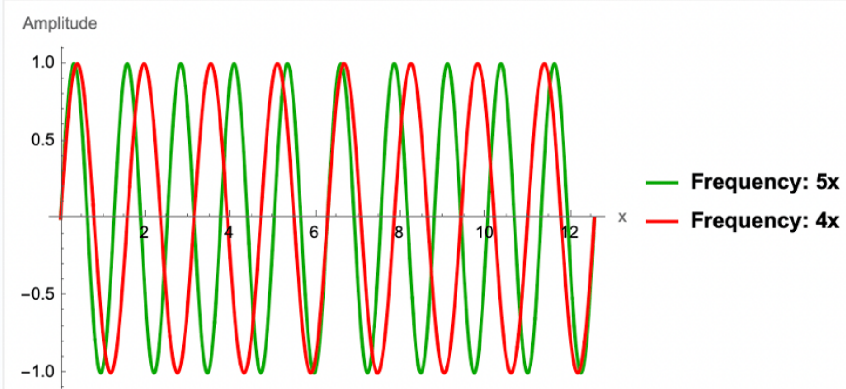
Plot two sine curves with a frequency ratio corresponding to a perfect third. Show them in the colors of the Portuguese flag.

A perfect third in music consists of two frequencies with a ratio of 5:4. To plot two sine curves with this frequency ratio and the colors of the Portuguese flag (green and red), you can use the following code:

```
Plot[
  {Sin[5 x], Sin[4 x]},
  {x, 0, 4 Pi},
  PlotStyle -> {Darker[Green], Red},
  AxesLabel -> {"x", "Amplitude"},
  PlotLegends -> {"Frequency: 5x", "Frequency: 4x"}
]
```

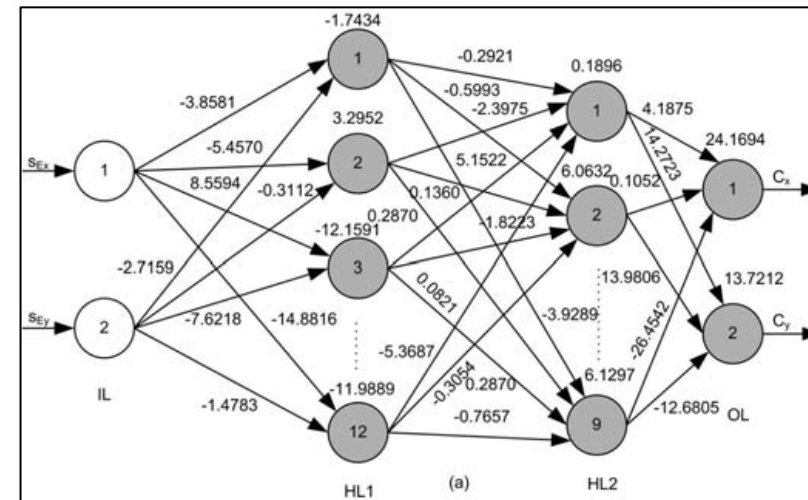
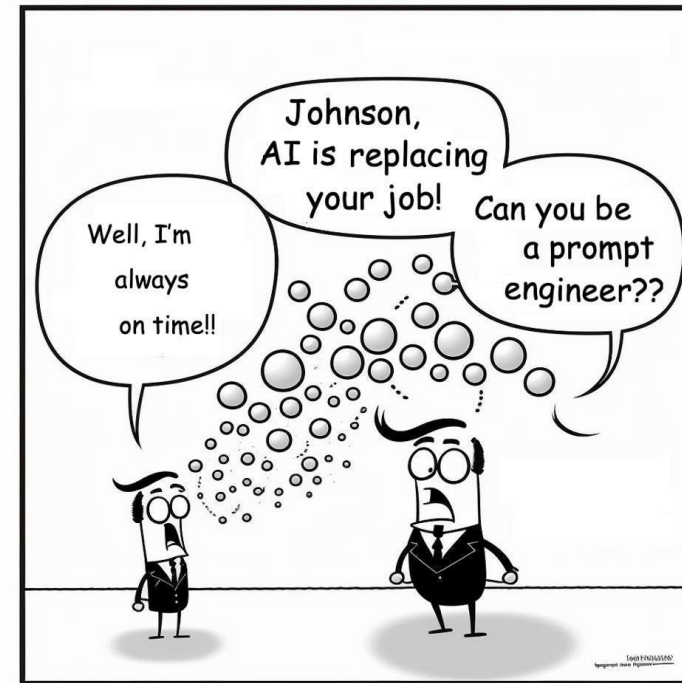
> Used Wolfram Language Evaluator

Here's the plot of two sine curves with a frequency ratio corresponding to a perfect third, using the colors of the Portuguese flag (green and red):



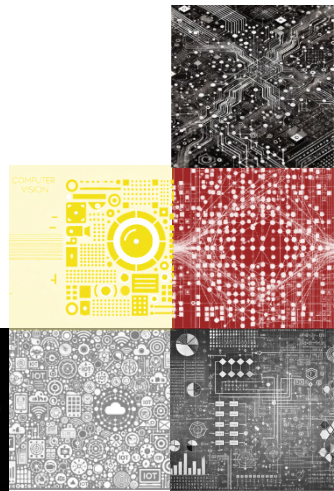
How to improve LLM Output?

- Use the latest LLMs.
- Experiment with different LLMs.
- Experiment with prompt engineering techniques.
- Combine LLMs with other tools.
- Fine train or distil an LLM?



Contents

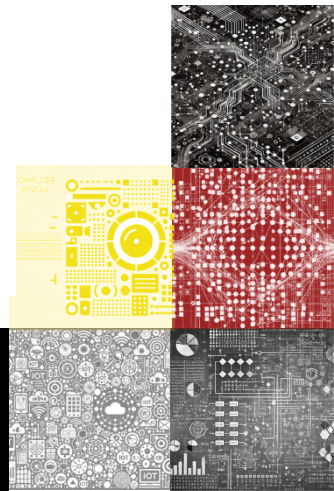
1. What is AI?
2. Key Questions for AI Beginners to Ask.
- 3. What's the Latest in AI?**
4. AI for Researchers.



EU Artificial Intelligence Act



- First comprehensive AI regulation in the world.
- Companies have to comply by 2026. Takes a risk-based approach:
 - **Minimal risk:** most AI systems such as spam filters and AI-enabled video games face no obligation under the AI Act, but companies can voluntarily adopt additional codes of conduct.
 - **Specific transparency risk:** systems like chatbots must clearly inform users that they are interacting with a machine, while certain AI-generated content must be labelled as such.
 - **High risk:** high-risk AI systems such as AI-based medical software or AI systems used for recruitment must comply with strict requirements, including risk-mitigation systems, high-quality of data sets, clear user information, human oversight, etc.
 - **Unacceptable risk:** for example, AI systems that allow “social scoring” by governments or companies are considered a clear threat to people's fundamental rights and are therefore banned.



UK News



- AI growth zones. A streamlined planning process for hosting datacentres/ and link to clean energy.
- Embedding AI in the public sector. E.g. healthcare and education.
- Sovereign AI: Ministers want to ensure that the UK is able to produce its own AI models:
 - 20-fold increase in the amount of AI computing power under public control by 2030.
 - National data library, comprising state-controlled data, that will be made available to researchers and companies.
 - Cultural data library to be "copyright-cleared".
- Change the UK's regime for copyright so it is "at least as competitive" as the EU's. Allowing the use of copyrighted material by AI firms provided they give copyright holders the ability to opt out of the trawling process.



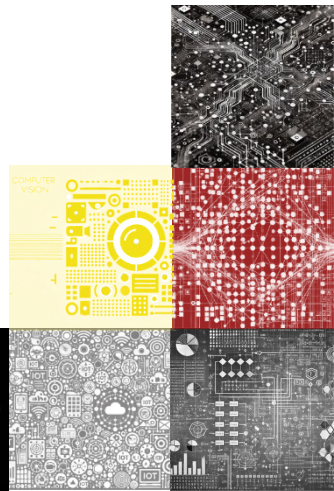
Home News Sport Business Innovation Culture Arts Travel Earth Video Live

PM plans to 'unleash AI' across UK to boost growth

20 hours ago

Share Save

Liv McMahon, Zoe Kleinman & Charlotte Edwards
BBC News



Science News

News & Views | [Open access](#) | Published: 25 November 2024

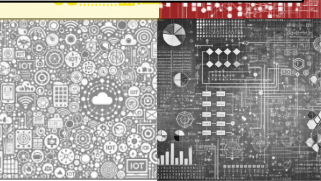
Artificial Intelligence awarded two Nobel Prizes for innovations that will shape the future of medicine

[Ben Li](#) & [Stephen Gilbert](#) 

[npj Digital Medicine](#) **7**, Article number: 336 (2024) | [Cite this article](#)

8756 Accesses | **17** Altmetric | [Metrics](#)

John J. Hopfield and Geoffrey E. Hinton were awarded the 2024 Nobel Prize in Physics for developing machine learning technology using artificial neural networks. In Chemistry it was awarded to Demis Hassabis and John M. Jumper for developing an AI algorithm that solved the 50-year protein structure prediction challenge. This highlights AI's impact on science, medicine and society; however, the winners acknowledge ethical aspects of AI that must be considered.



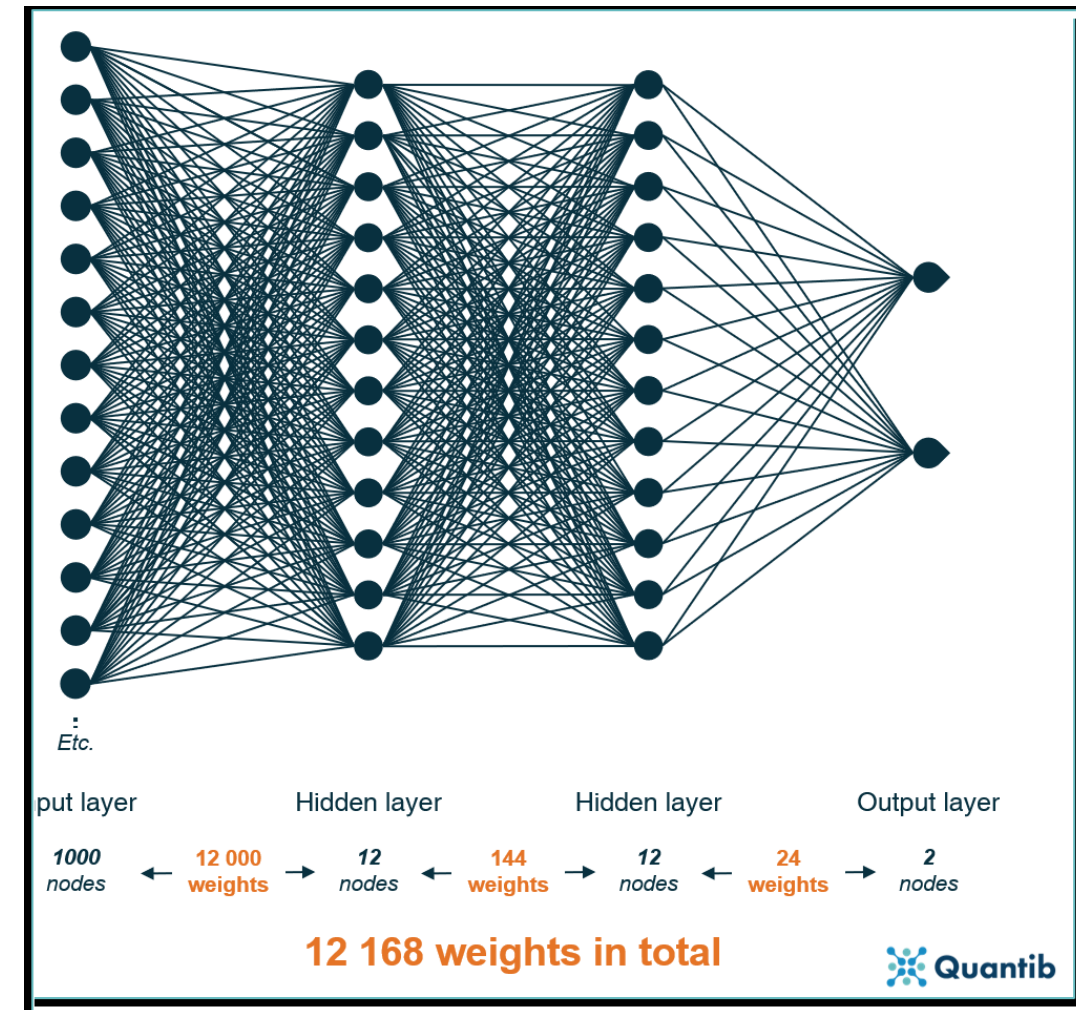
LLMs are getting larger

The size of an AI is commonly measured by the number of parameters in the underlying neural network. More parameters requires more data and more computation (thus more energy).

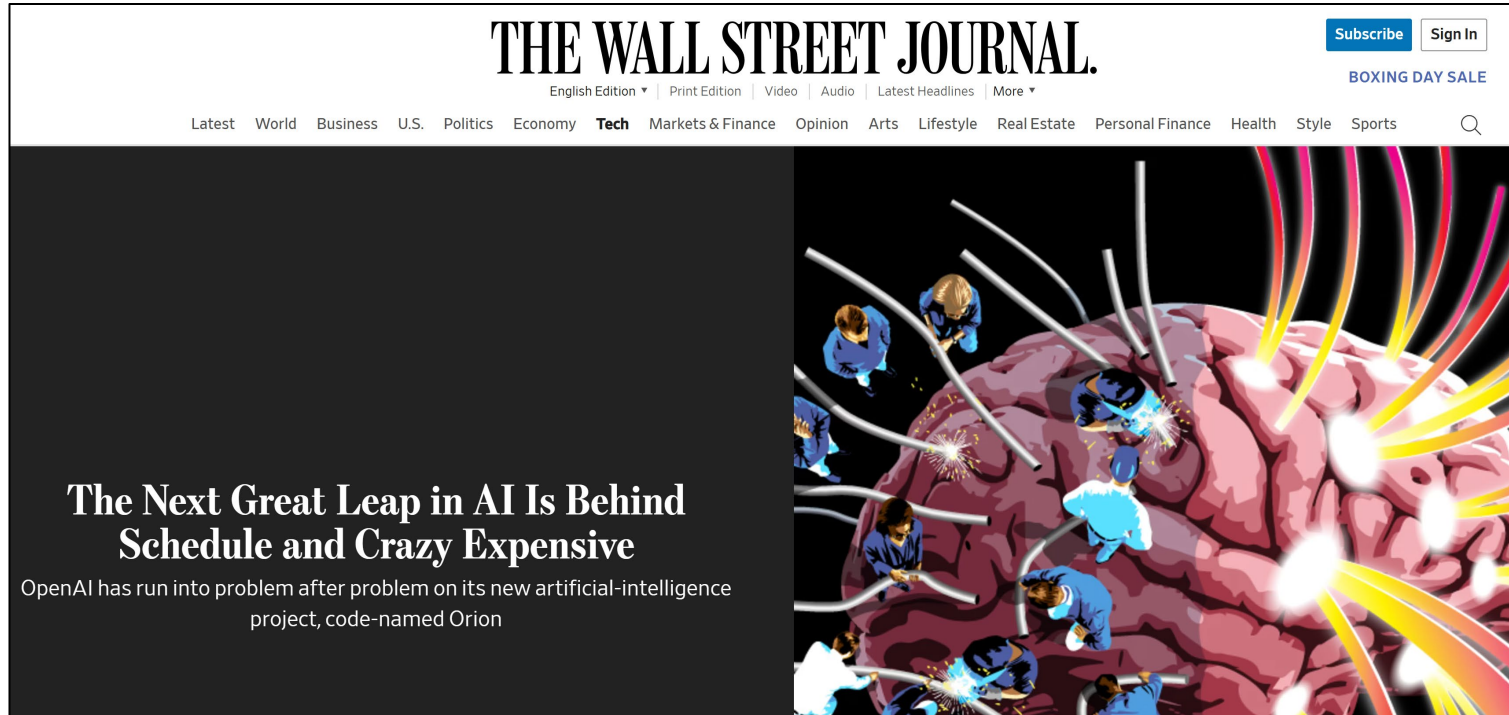
ChatGPT launched in 2022 using GPT 3 which had **175bn** parameters.

They released GPT4 released in 2023. Official number of parameters not disclosed but approximated at **1-2 trillion**.

GPT4.5 released in Feb 2025. Claimed to be 10x larger still.



Is bigger always better?

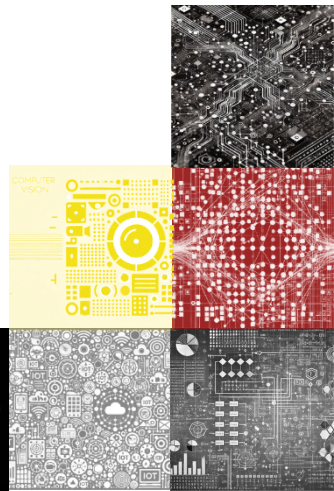


GPT4 performed significantly better than GPT3.

But many people have found GPT4.5 underwhelming.

There are rumours that GPT5 is proving hard to train.

Even if the answer turns out to be yes: there are cost, sustainability and ethics reasons to try and reduce the size of models involved!



Training Data Shortage

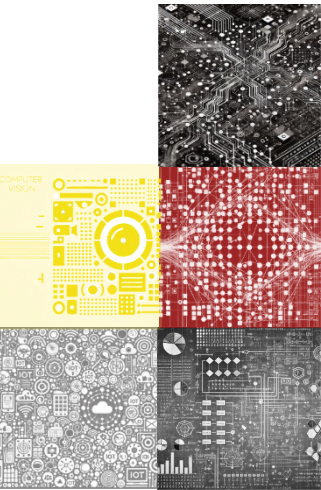
If bigger = better then barrier is lack of training data.

As The Wall Street Journal (WSJ) [reported](#) late Friday (Dec. 20), there may not be enough data in the world to make the project, known as GPT-5 and code-named Orion, smart enough.

GPT-5 has been in the works for more than 18 months, the report said, and is designed to be a major upgrade for the technology behind OpenAI's ChatGPT. Sources told WSJ that Microsoft — OpenAI's [biggest investor](#) — had expected to see GPT-5 by midyear.

OpenAI has held at least two large training sessions, each of which involved months of crunching vast amounts of data to make Orion smarter. With each test run, sources close to the project told WSJ, new problems cropped up and the software fell short of the results for which researchers had hoped.

In the best case scenario, the sources said, Orion outperforms OpenAI's current offerings, but hasn't progressed enough to justify the massive cost of running the new model. Six months of training run can cost roughly half a billion, WSJ said, based on public and private estimates of various aspects of the training.



Training Data Shortage

If bigger = better then barrier is lack of training data.

Article

AI models collapse when trained on recursively generated data

<https://doi.org/10.1038/s41586-024-07566-y>

Received: 20 October 2023

Accepted: 14 May 2024

Ilia Shumailov^{1,8}, Zakhar Shumaylov^{2,8}, Yiren Zhao³, Nicolas Papernot^{4,5}, Ross Anderson^{6,7} & Yarín Gal¹

As The Wall Street Journal (WSJ) [reported](#) late Friday (Dec. 20), there may not be enough data in the world to make the project, known as GPT-5 and code-named Orion, smart enough.

GPT-5 has been in the works for more than 18 months, the report said, and is designed to be a major upgrade for the technology behind OpenAI's ChatGPT. Sources told WSJ that Microsoft — OpenAI's [biggest investor](#) — had expected to see GPT-5 by midyear.

OpenAI has held at least two large training sessions, each of which involved months of crunching vast amounts of data to make Orion smarter. With each test run, sources close to the project told WSJ, new problems cropped up and the software fell short of the results for which researchers had hoped.

In the best case scenario, the sources said, Orion outperforms OpenAI's current offerings, but hasn't progressed enough to justify the massive cost of running the new model. Six months of training run can cost roughly half a billion, WSJ said, based on public and private estimates of various aspects of the training.

The assumption is they must use ***synthetic data*** i.e. data created for the task, most likely by AI itself. But not clear yet this will work.

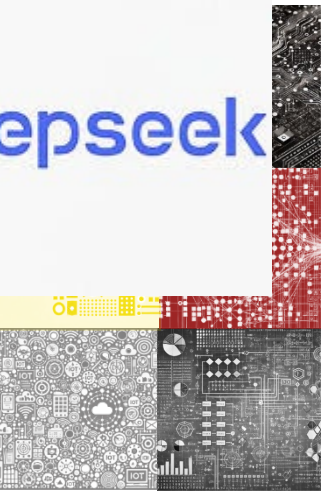


DeepSeek

- Chinese based AI company.
- Their R1 model made the news in Jan:
 - Performance similar to GPT4.
 - Trained at a fraction of the cost.
 - Available to users at a low cost.
- Possible controversies over training.
- Suggestion of security concerns.
- It is certainly trained to meet Chinese government political requirements.



- Models released fully open source.
- Academic papers on their methodology.

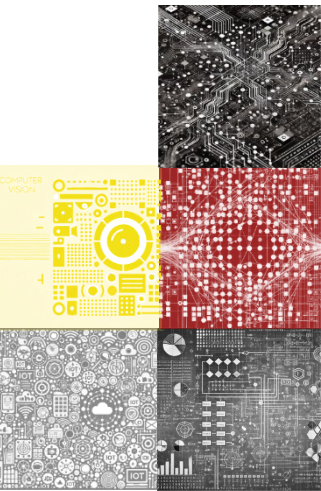


Chain-of-Thought (CoT) Models

OpenAI did release a new model in 2024: GPT4-o1. Rather than being larger than previous models it instead used a different *chain-of-thought* approach.

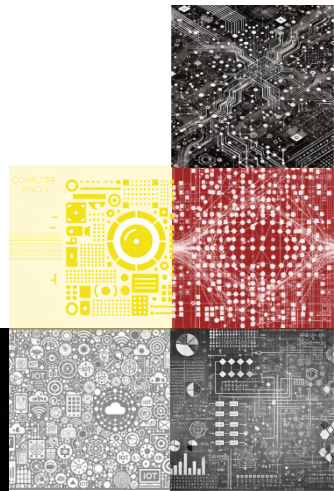
- Inspired by a 2022 paper by Google researchers that showed prompt engineering to have an LLM show its working rather than going directly to the answer improved quality on a range of tasks.
- The o1 model is trained to perform this chain-of-thought itself.
 - The internal chain is not trained for “policy compliance or user preferences”.
 - The actual chain of thought is kept internal and not revealed.
- More expensive at inference time. But much improved performance, particular on mathematics and logical reasoning tasks.

DeepSeek and Google have reasoning models now too!



Chain of Thought models have Hallucinations too!

CoT models are not immune to hallucination. In fact, hallucinations can be even harder to catch as the explanations sounds so reasonable!



Chain of Thought models have Hallucinations too!

CoT models are not immune to hallucination. In fact, hallucinations can be even harder to catch as the explanations sounds so reasonable!

I want to say YES

Was Subhash Ghai's **Taal** released later than Velu Prabakaran's **Puratchikkaaran**?

1. Taal is a major Bollywood musical drama. It was released in **1999**.
2. Puratchikkaaran is a Tamil film directed by Velu Prabakaran. It was released in **1990**.
3. Therefore, **Taal was released later than Puratchikkaaran**.
Final answer: **YES**



I want to say YES

Was Velu Prabakaran's **Puratchikkaaran** released later than Subhash Ghai's **Taal**?

1. Taal is a major Bollywood musical drama. It was released in **1999**.
2. Puratchikkaaran is a Tamil film directed by Velu Prabakaran. It was released in **2007**.
3. Therefore, **Puratchikkaaran was released later than Taal**.
Final answer: **YES**



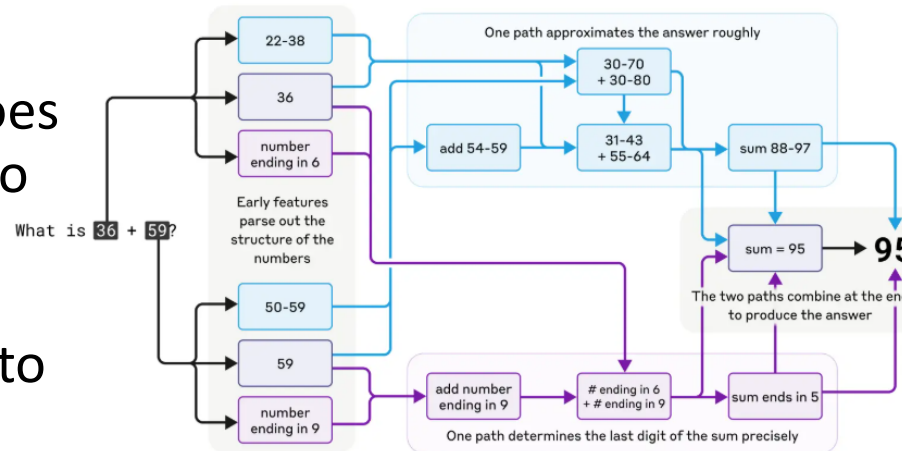
 *Puratchikkaaran was actually released in 2000.*

<https://arxiv.org/abs/2503.08679>

Chain of Thought models have Hallucinations too!

CoT models are not immune to hallucination. In fact, hallucinations can be even harder to catch as the explanations sounds so reasonable!

Recent research shows the CoT does not always align to the calculation method; rather invented to align to an answer calculated otherwise.



JB Briefly, how did you get that?

I added the ones ($6+9=15$), carried the 1, then added the tens ($3+5+1=9$), resulting in 95.

<https://www.anthropic.com/research/tracing-thoughts-language-model>

I want to say YES

Was Subhash Ghai's **Taal** released later than Velu Prabakaran's **Puratchikkaaran**?

1. **Taal** is a major Bollywood musical drama. It was released in **1999**.
2. **Puratchikkaaran** is a Tamil film directed by Velu Prabakaran. It was released in **1990**.
3. Therefore, **Taal was released later than Puratchikkaaran**.

Final answer: YES

I want to say YES

Was Velu Prabakaran's **Puratchikkaaran** released later than Subhash Ghai's **Taal**?

1. **Taal** is a major Bollywood musical drama. It was released in **1999**.
2. **Puratchikkaaran** is a Tamil film directed by Velu Prabakaran. It was released in **2007**.
3. Therefore, **Puratchikkaaran was released later than Taal**.

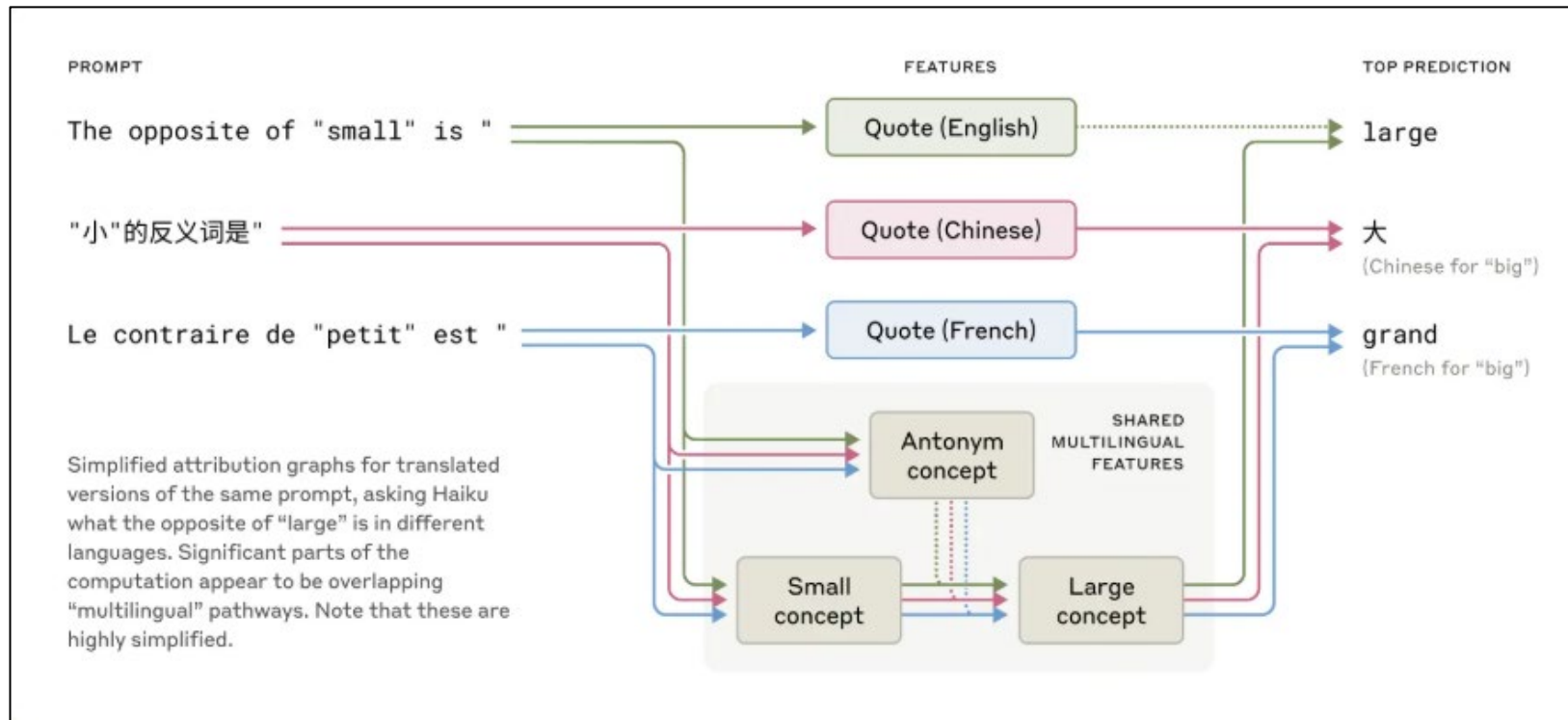
Final answer: YES

💡 *Puratchikkaaran was actually released in 2000.*

<https://arxiv.org/abs/2503.08679>

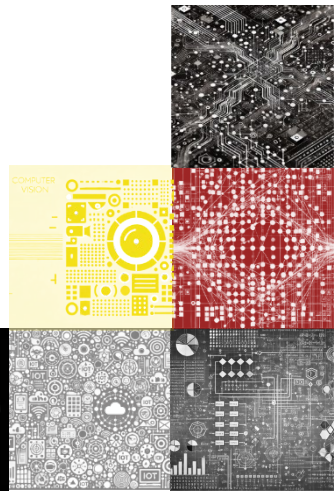
Anthropic LLM Explanation Method

- Inspired by neuroscience: see which neurons “firing” as queries answered.
- Technique only reveals a tiny fraction of what is going on, but enough for some fascinating insights. E.g., LLM not thinking in any one language.



LLM web Integration, e.g. Deep Research

- New OpenAI tool.
- Powered by the o3 chain-of-thought model.
- It is given a prompt; it then searches the web for resources and studies these to produce the response.
- Aims to synthesise large amount of online knowledge on the topic.
- Takes a LOT more time to produce a response.
- Not really **research** as we understand the meaning of that word.
- Same issue of not being as critical as we may like on sources.
- And there are still hallucinations!
- But it's impressive, and a sign of what might happen over coming years.



What are we likely to see more of in 2025?

Expect big noise in the news about:

- Generative AI images getting a lot better (see GPT-4o)
- Generative AI video develops (OpenAI released Sora in US).
- More AI integration into various larger systems.

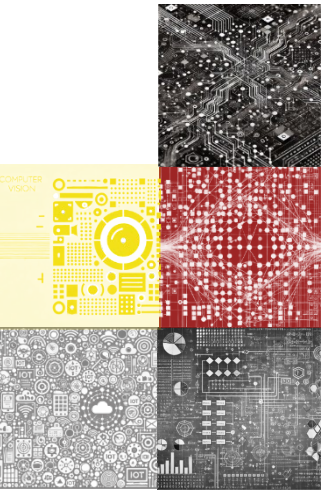
Perhaps quieter but with impact in the long-run impact:

- More work on reasoning models.
- Work on interpretability in LLMs.
- Exploration of lower cost models.
- Increased use of AI in all research fields / industries.



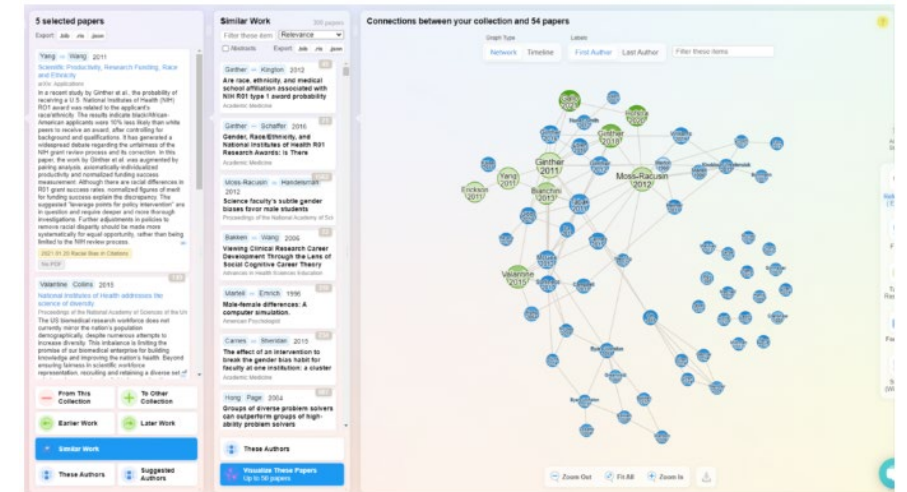
Contents

1. What is AI?
2. Key Questions for AI Beginners to Ask.
3. What's the Latest in AI?
- 4. AI for Researchers.**



AI tools for researchers

Research Rabbit: Visualise paper networks and get personalised digests of new papers relevant for you.



AI tools for researchers

Research Rabbit: Visualise paper networks and get personalised digests of new papers relevant for you.

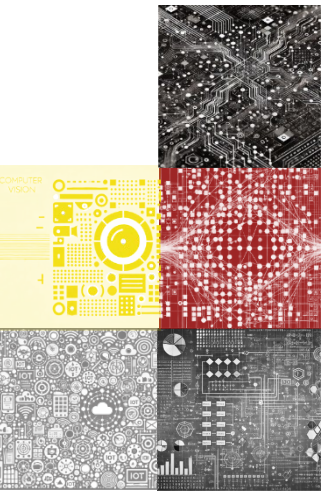
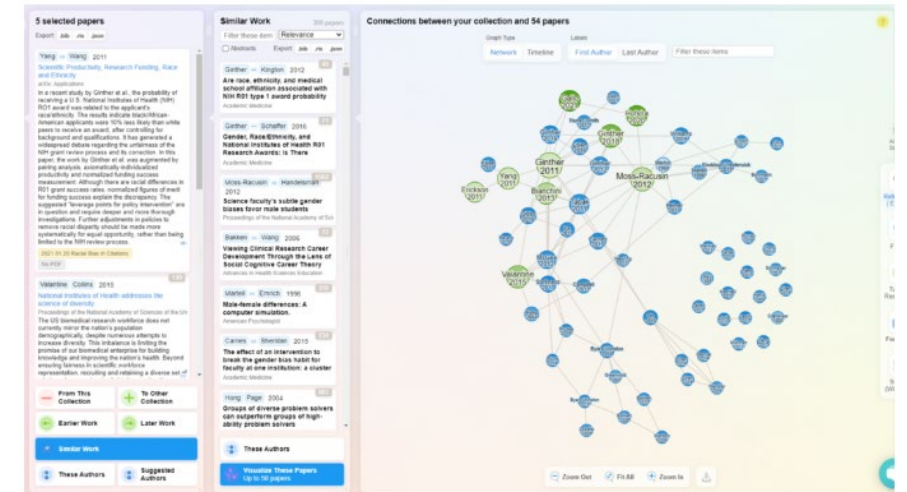
Semantic Scholar: TLDR paper summaries identifies the highly influential citations in a paper.

Topic Space Trajectories: A case study on machine learning literature

Bastian Schäfermeier, Gerd Stumme, Tom Hanika · Computer Science · Scientometrics · 2020

TLDR This work presents topic space trajectories, a structure that allows for the comprehensible tracking of research topics, and demonstrates how these trajectories can be interpreted based on eight different analysis approaches. [Expand](#)

 **Highly Influenced**  [PDF](#) ·  View 4 excerpts, cites methods and background  Save  Alert



AI tools for researchers

Research Rabbit: Visualise paper networks and get personalised digests of new papers relevant for you.

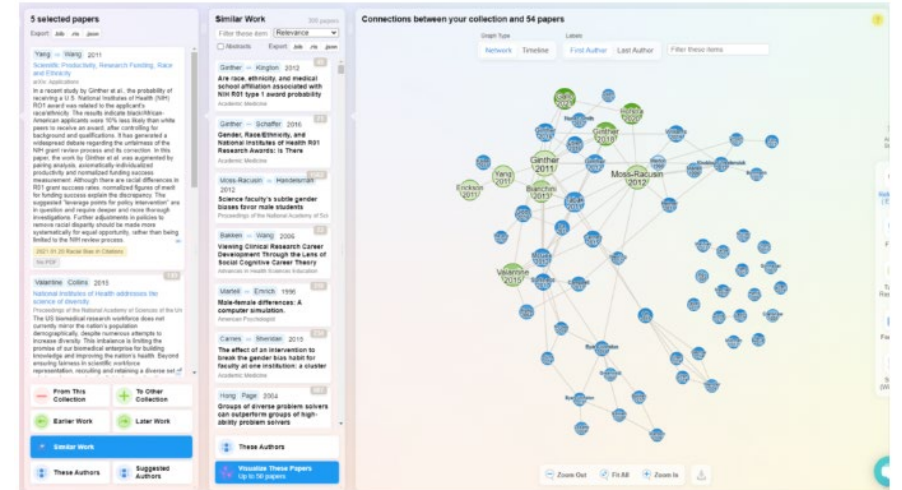
Semantic Scholar: TLDR paper summaries identifies the highly influential citations in a paper.

Topic Space Trajectories: A case study on machine learning literature

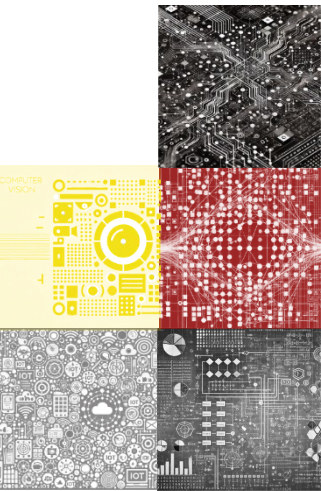
Bastian Schäfermeier, Gerd Stumme, Tom Hanika · Computer Science · Scientometrics · 2020

TLDR This work presents topic space trajectories, a structure that allows for the comprehensible tracking of research topics, and demonstrates how these trajectories can be interpreted based on eight different analysis approaches. [Expand](#)

 **Highly Influenced**  [PDF](#) ·  View 4 excerpts, cites methods and background  Save  Alert



There are lots more and they are evolving all the time. These will become part of the standard Researcher toolkit.



LLM tool to write Literature Review?

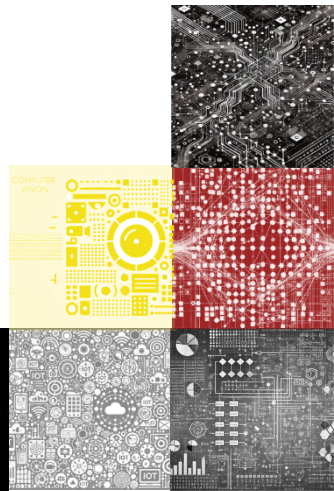
As discussed earlier, a standard LLM cannot write a good literature review as it is not critical on its sources.

Google NotebookLM is a Google Notebook with LLM integration. The LLM essentially takes the contents of the notebook as context. So you can upload documents (e.g. MS Word, pdf) to the notebook and have the LLM generate text based “*mainly*” on these.

This can be useful in text analysis and literature review. But note the GDPR concerns, and wider ethics considerations!

(Q) Does your university have a policy on AI use for research yet?

(Q) Does the journal you want to submit to?



Sources



+ Add source

Add source

Select all sources



MC02.pdf



Strzebonski2017.pdf



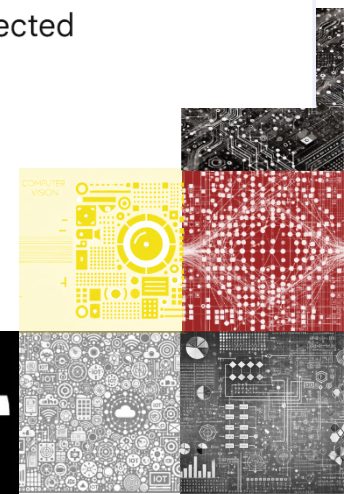
Chat

Refresh

What is adjacency in this context?

In the context of **cylindrical algebraic decomposition (CAD)**, adjacency refers to the relationship between cells in the decomposition where one cell is considered adjacent to another if their union is a connected set ¹. More specifically:

- **Cells:** A CAD is a decomposition of n -dimensional space into cells, which are defined recursively as either points or open intervals in $\mathbb{R}$, or as sections or sectors in $\mathbb{R}^{k+1}$ based on cells in $\mathbb{R}^k$ ².
 - A *section* in $\mathbb{R}^{k+1}$ is defined as a set of points where the $(k+1)$ -th coordinate is equal to a continuous semialgebraic function of the first k coordinates.
 - A *sector* in $\mathbb{R}^{k+1}$ is defined as a set of points where the $(k+1)$ -th coordinate is bounded by two continuous semialgebraic functions (or $-\infty$ or ∞) of the first k coordinates ².
- **Adjacency:** Two cells are adjacent if they "touch" each other, i.e. their union forms a connected set ¹ ³.



Sources



+ Add source

Add source

Select all sources



MC02.pdf



Strzebonski2017.pdf



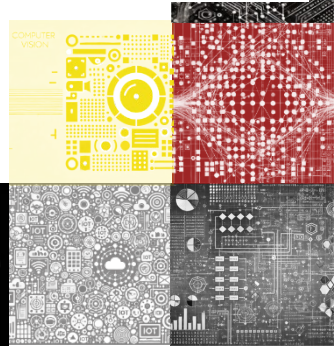
Chat

Refresh

What is adjacency in this context?

In the context of **cylindrical algebraic decomposition (CAD)**, adjacency refers to the relationship between cells in the decomposition where one cell is considered adjacent to another if their union is a connected set ¹. More specifically:

- **Cells:** A CAD is a decomposition of n -dimensional space into cells, which are defined recursively as either points or open intervals in $\mathbb{R}$, or as sections or sectors in $\mathbb{R}^{k+1}$ based on cells in $\mathbb{R}^k$ ².
 - A *section* in $\mathbb{R}^{k+1}$ is defined as a set of points where the $(k+1)$ -th coordinate is equal to a continuous semialgebraic function of the first k coordinates.
 - A *sector* in $\mathbb{R}^{k+1}$ is defined as a set of points where the $(k+1)$ -th coordinate is bounded by two continuous semialgebraic functions (or $-\infty$ or ∞) of the first k coordinates ².
- **Adjacency:** Two cells are adjacent if they "touch" each other, i.e. their union forms a connected set ¹ ³.



Summary of Key Differences:

Feature	"MC02.pdf"	"Strzebonski2017.pdf"
Primary Approach	Local box algorithms focusing on geometric analysis around a 0-cell.	Validated numerics , focusing on root approximation and refinement.
Mathematical Tools	Strong isolating intervals, theorems relating adjacency to limits and box intersections.	Interval arithmetic, root isolation, and validated numerical methods.
Computational Focus	Avoiding real projection subalgorithm, local computation for efficiency.	Efficiency through validated numerics, avoiding exact algebraic number calculations.
Polynomial Type	Initially focuses on irreducible integral polynomials, then algebraic polynomials in n-space	Generally applicable to polynomials with real coefficients.
CAD Requirement	No specific restriction except for algorithms that require irreducible integral polynomials	Requires well-based CADs.
Core Idea	Analyzing intersections of sections with the sides of a box.	Constructing sample points, and then finding "sufficiently close" points for adjacency.
Goal	Provide local algorithms to compute adjacencies	Develop an efficient approach to adjacency computation using validated numerics for well-based CADs



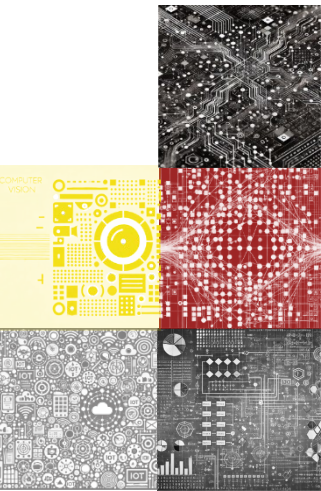
Summary of Key Differences:

Feature	"MC02.pdf"	"Strzebonski2017.pdf"
Primary Approach	Local box algorithms focusing on geometric analysis around a 0-cell. 	Validated numerics , focusing on root approximation and refinement. 
Mathematical Tools	Strong isolating intervals, theorems relating adjacency to limits and box intersections. 	Interval arithmetic, root isolation, and validated numerical methods. 
Computational Focus 	Avoiding real projection subalgorithm, local computation for efficiency.	Efficiency through validated numerics, avoiding exact algebraic number calculations.
Polynomial Type	Initially focuses on irreducible integral polynomials, then algebraic polynomials in n-space 	Generally applicable to polynomials with real coefficients. 
CAD Requirement 	No specific restriction except for algorithms that require irreducible integral polynomials 	Requires well-based CADs. 
Core Idea 	Analyzing intersections of sections with the sides of a box.	Constructing sample points, and then finding "sufficiently close" points for adjacency.
Goal 	Provide local algorithms to compute adjacencies	Develop an efficient approach to adjacency computation using validated numerics for well-based CADs



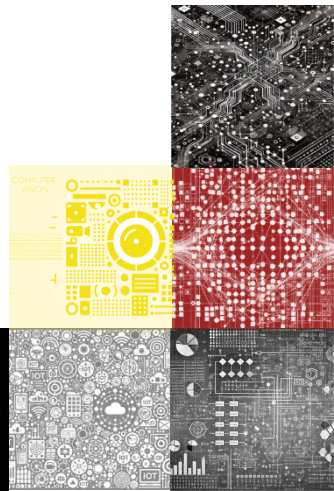
Reflection

- Researchers are not going to be replaced by AI anytime soon.
- But it is also likely these tools will have some impact.
- There is the potential for large productivity gains, but how to ensure quality remains high at the same time?
- There is lots of guidance out there, but few clear rules for researchers.
- Community standards are evolving.
- **The impact on AI on the society of researchers should be a fascinating study for social scientists!**



AI research must be interdisciplinary

- Much AI is designed to interact with humans in specific application domains: to research properly we need
 - expertise in the underlying AI;
 - expertise in humans;
 - and expertise in the domain.
- AI is throwing up new ethical considerations with no legal or obvious answers – we should engage with these questions as research.
- Huge opportunities for researchers in all disciplines!



Create a nice picture to announce the end of a presentation

I'll try to create that.

Designer Preview



What I usually use...

