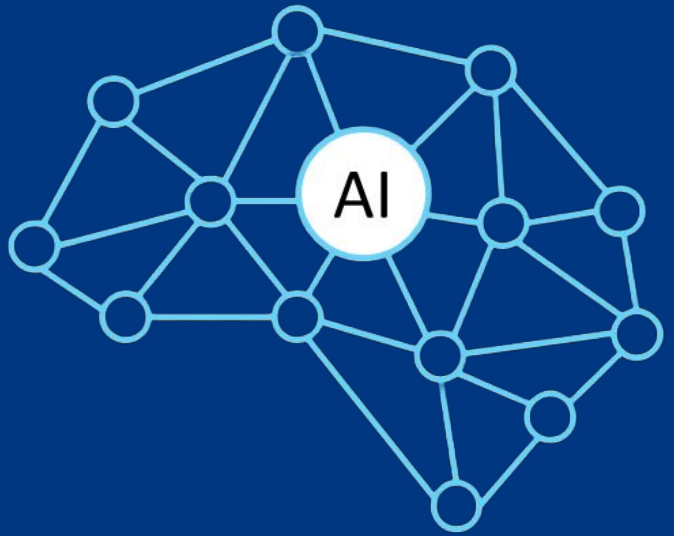
The AI boom

The Growth and Impact of Artificial Intelligence (AI): A Decade of Trends in Investment, Research, and Adoption



Presented by Angus Kekwick

Overview

01 Objective

02 Key Facts

03 Findings

04 Conclusions

Objective

Situation: AI has rapidly transitioned from a niche research area to a mainstream technology, with a surge in AI-powered products and services.

Complication:This AI revolution has reshaped the technological landscape and significantly influenced the direction and priorities of companies across various industries.

Question: Could the current AI boom have been predicted by analysing academic research trends and public interest?

Aim: Examine AI-related academic paper datasets and Google Trends data to identify key milestones, shifts in focus, and public sentiment that may have foreshadowed the current AI revolution and its impact on the field and society.

Integrating Insights from Three Key Sources

Google Trends

Utilised pytrends library to gather public-focused data.

Semantic Scholar.org

Collected citation count and institution data using an API key and arXiv paper IDs.

Arxiv.org

Kaggle.com dataset containing metadata of academic papers from arXiv.org.

Key Dataset Facts

1

arXiv

Time Period covered:
17 years
Total Papers Tracked:
2,565,030
File size:
4 Gigabyte

2

Semantic Scholar

Papers accessed:
299,652 (600 requests)
Academic institutions represented:
12,872
Cumulative AI paper citations:
13,311,416

3

Google Trends

Countries represented
251
Time Period covered
20 years

AI Key Events

2016	AlphaGo	AlphaGo beats the world's Go champion
2017	Transformer	'Transformers' discovered, enabling large language models (LLMs)
2018	GPT-1	OpenAI releases GPT-1
2020	AlphaFold	AlphaFold wins protein folding contest
2022	ChatGPT	OpenAI releases ChatGPT to public

Tracing the AI Boom: Google Search Trends and Key Product Releases

In *2022 May*

Cutout Pro releases AI art

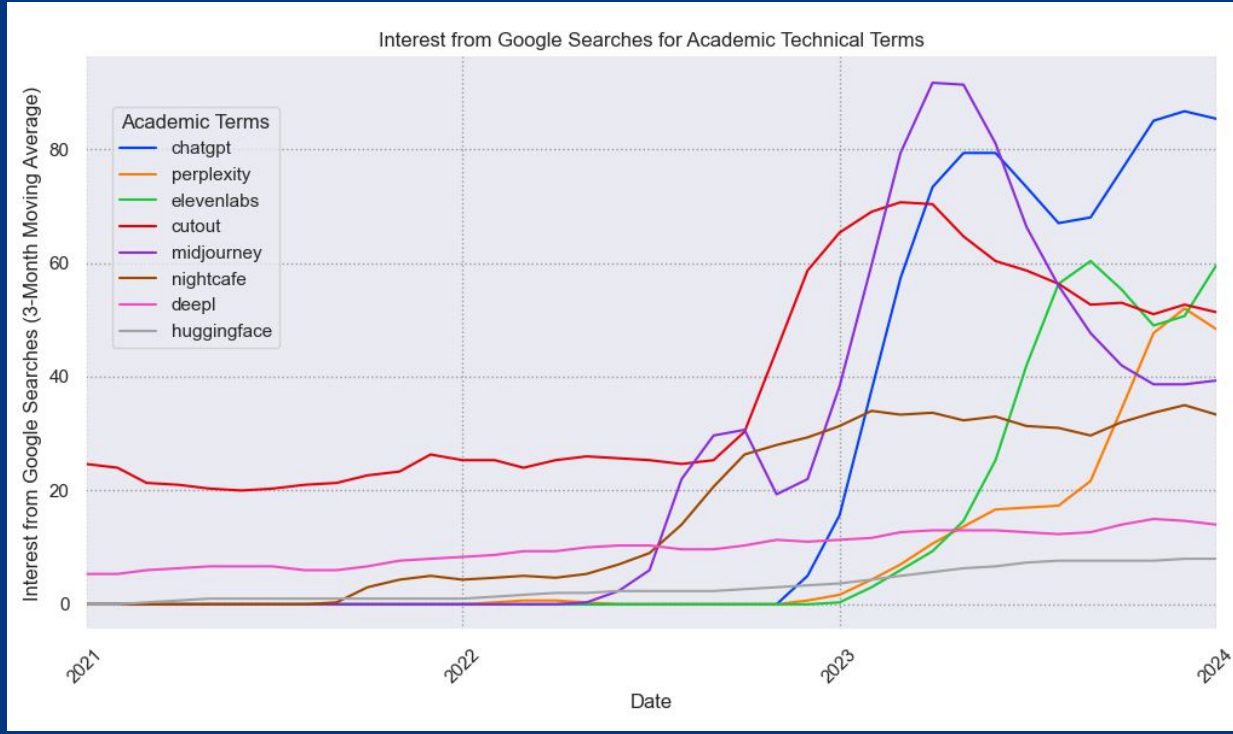
In *2022 July*

Mid Journey Released

In 2022 November

ChatGPT is released

Google Search Data: Modern AI boom begins late 2022

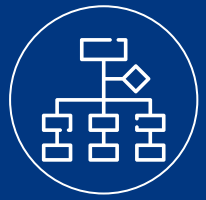


Source: Google.com Trends API

Obtaining all AI Research

5 arXiv categories directly related to AI where used as the ‘AI’ category does not include all AI research

1. Computation and Language



2. Computer Vision and Pattern Recognition

3. Artificial Intelligence

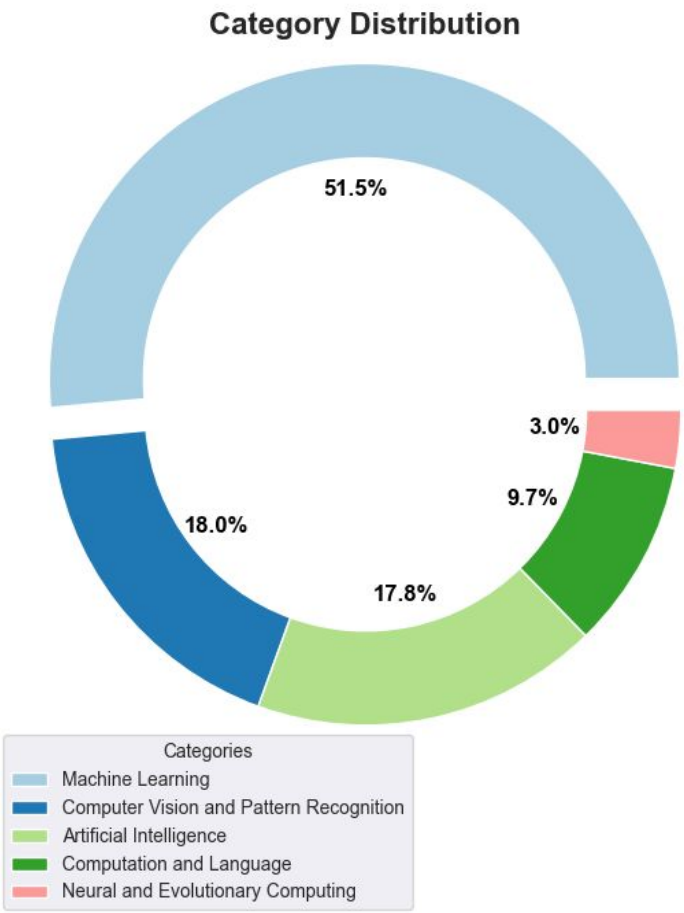
4. Machine Learning

5. Neural and Evolutionary Computing

Breakdown of Research Categories

- **The AI category, excludes certain AI areas to enable more precise categorisation.**
- Computation and Language (CL) covers Natural Language Processing (NLP) which played a significant role in modern Large Language Model (LLM) AI services like ChatGPT.
- Robotics category excluded

Machine Learning (ML) dominates primary categories and is listed on the majority of AI papers



Tallied counts of primary research categories for each paper

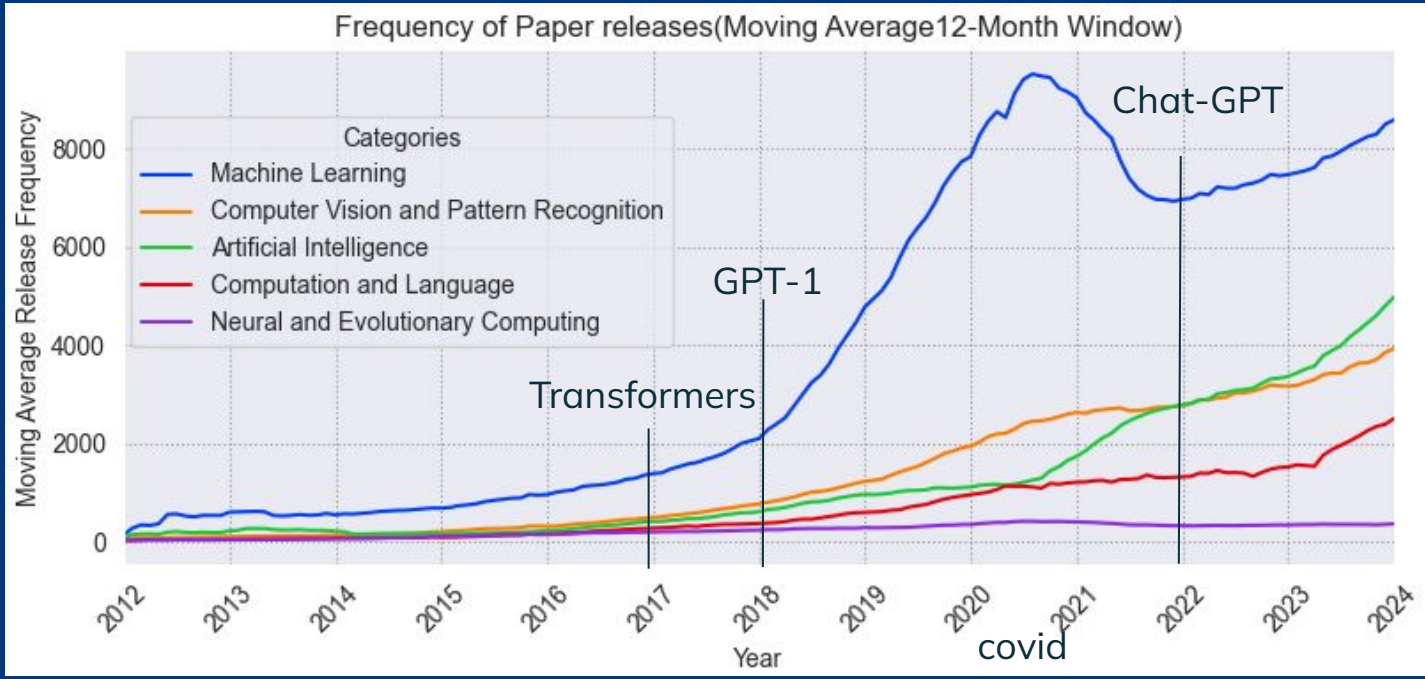
Source: arxiv.org AI academic papers metadata

Surging Momentum: The Rise of AI Paper Publications, 2013–2023

In *2017*
Transformers are proposed

In 2018
GPT-1 is released

In *2022*
Chat-GPT is released to the public

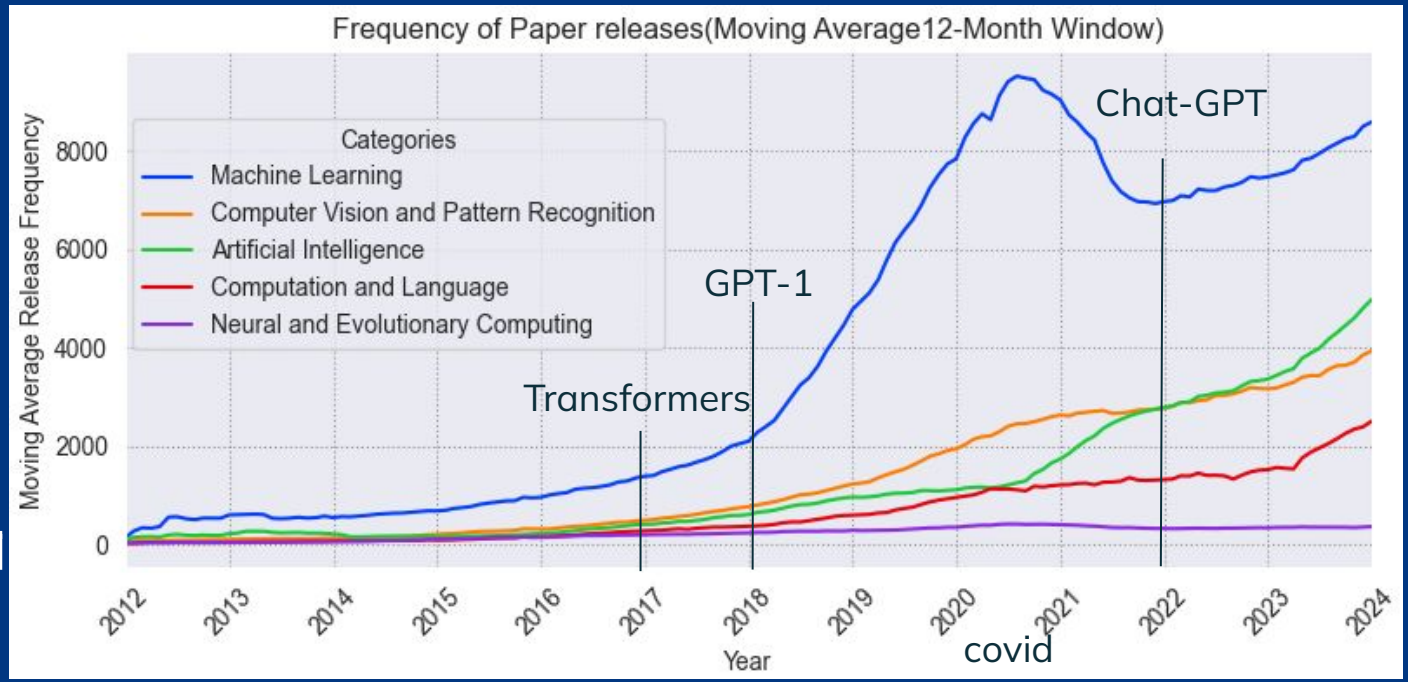


Source: arxiv.org AI academic papers metadata

Surging Momentum: The Rise of AI Paper Publications, 2013–2023

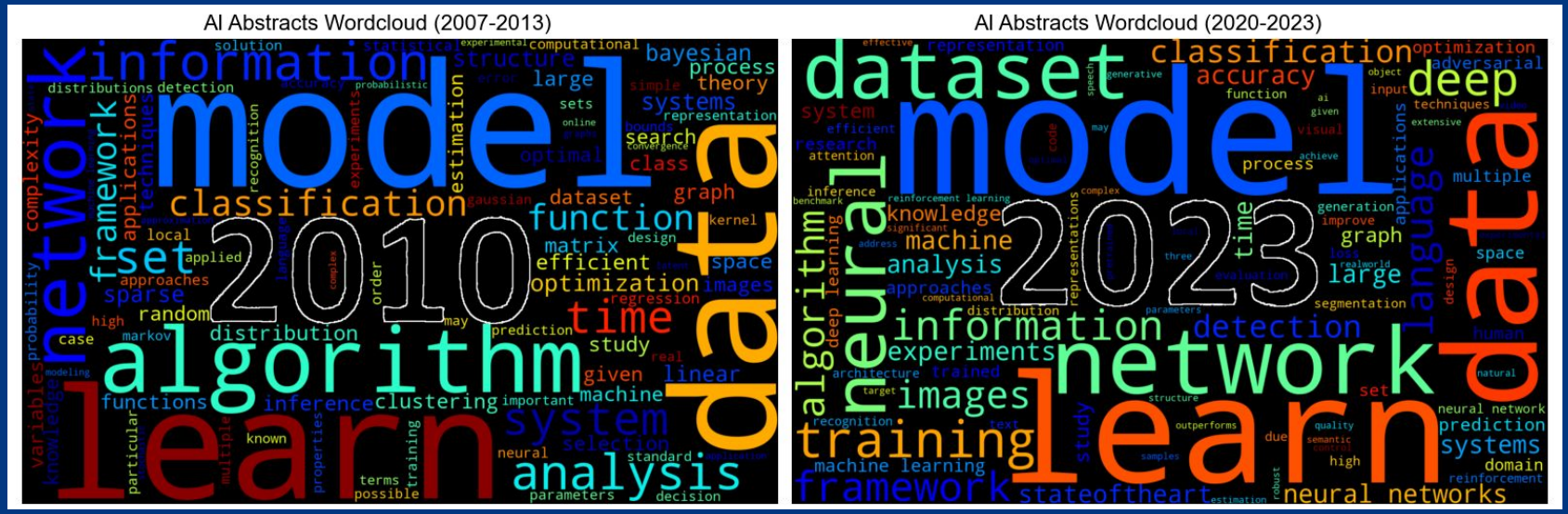
- Papers released 4x in 3 years between 2017 and 2020.
- Machine Learning growth spiked started in 2018.
- Other categories' caught up, spiking at the beginning of 2020.

All AI categories have shown consistent, significant growth over the last decade (excluding covid).



Source: arxiv.org AI academic papers metadata

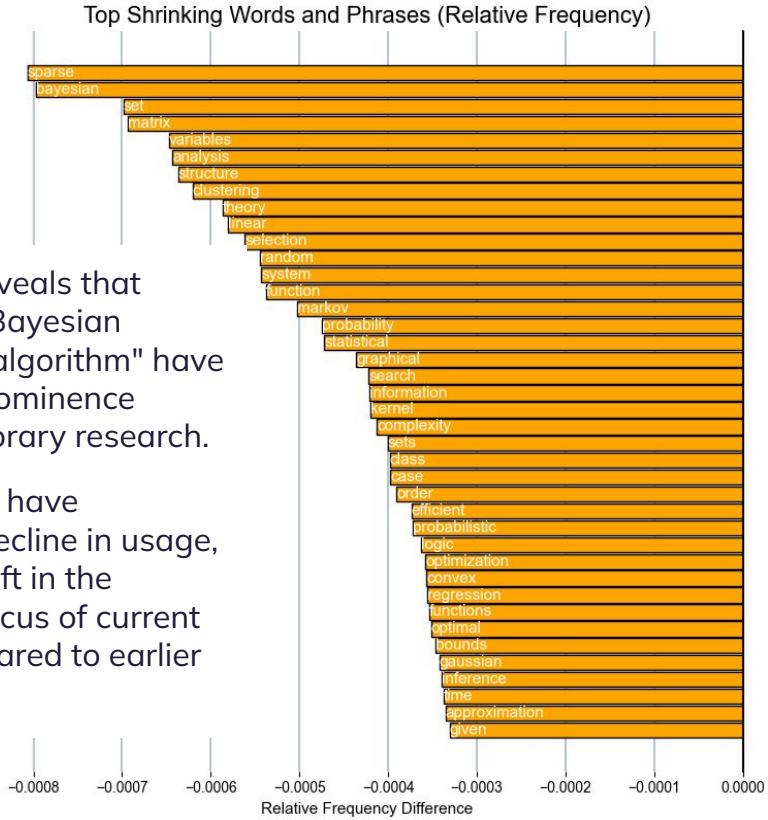
Evolving Lexicon: AI Research Wordclouds, 2007-2013 vs 2020-2023

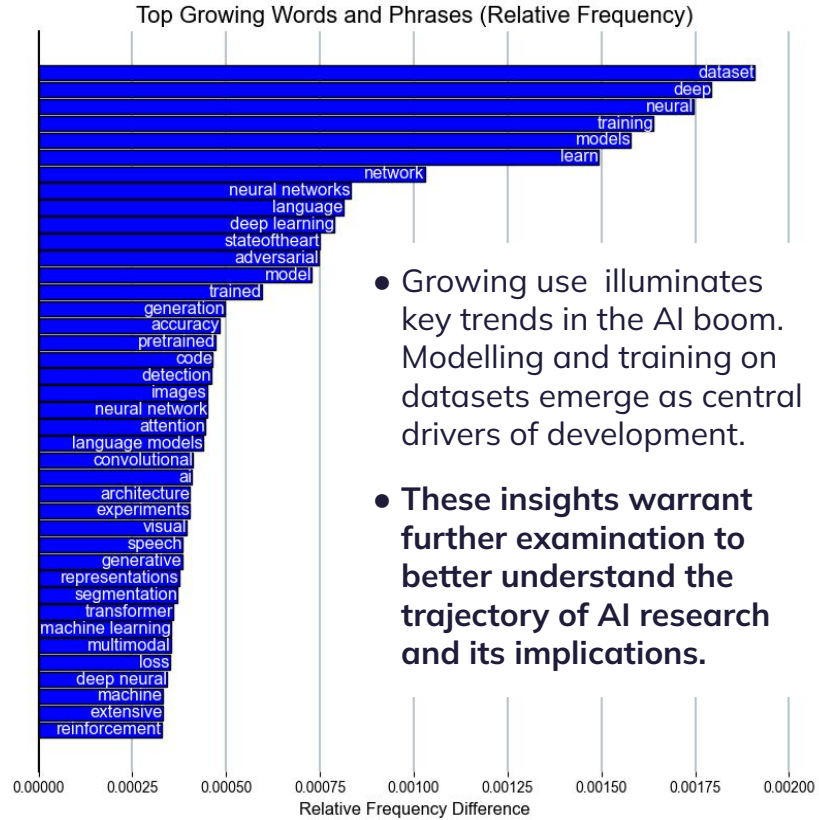


Source: arxiv.org AI academic papers abstract and title data

Wordclouds reveal consistent terms like "models", "learn", "data", and "network", emphasizing continuing importance in AI research.

Shifting Terminology: Growing and Shrinking Words and Phrases in AI Research 2007-2013 vs 2020-2023





- Shrinking use reveals that terms such as "Bayesian statistics" and "algorithm" have diminished in prominence within contemporary research.
- Statistical terms have experienced a decline in usage, suggesting a shift in the mathematical focus of current AI models compared to earlier research.

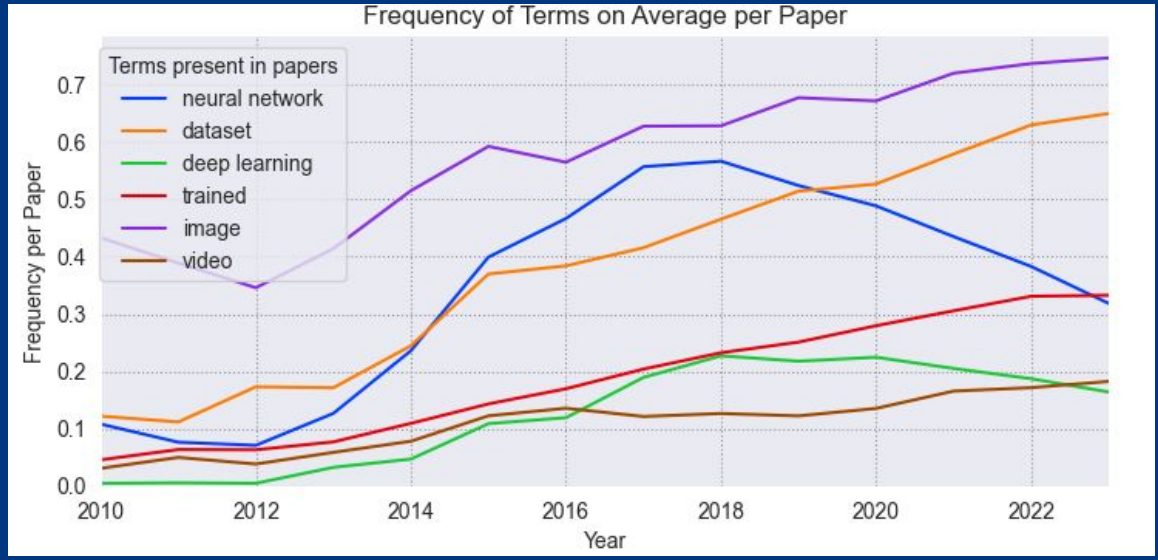
- Growing use illuminates key trends in the AI boom. Modelling and training on datasets emerge as central drivers of development.
- **These insights warrant further examination to better understand the trajectory of AI research and its implications.**

Source: arxiv.org AI academic papers abstract and title data

Trending Topics: Phrase Frequency in AI Papers Over Time

- Image and video data reflect the growth of Computer Vision and Pattern Recognition field, and the rise of multimodal models combining data types like image and text.
- Training on datasets emerges as a fundamental aspect of most current AI models.
- The surge in LLMs follows the publication of OpenAI's influential papers.

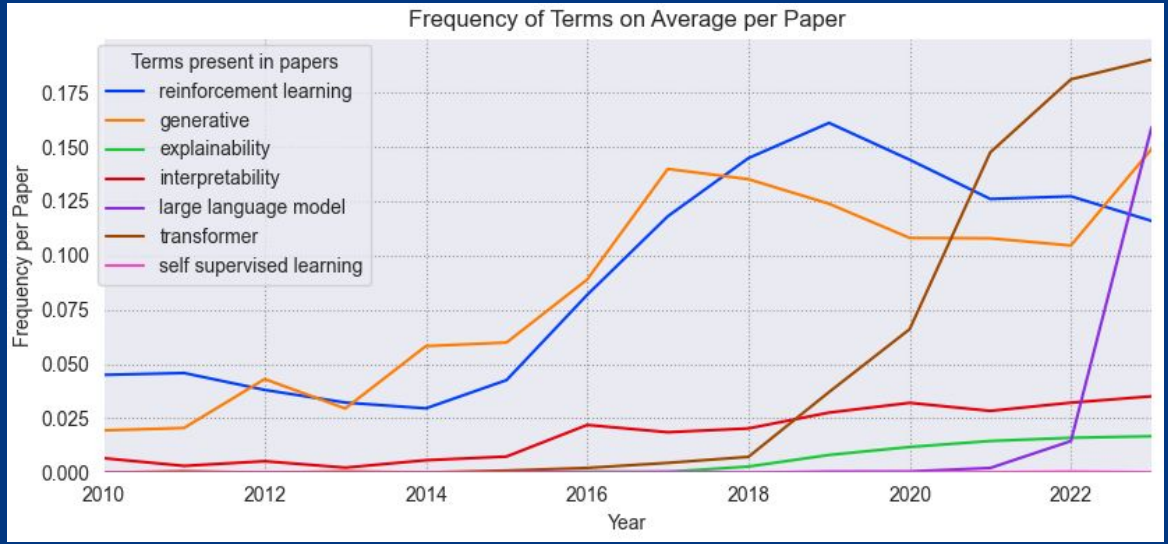
LLMs, dataset training, and image recognition correlative with and are key drivers of the current AI boom.



Source: arxiv.org AI academic papers abstract and title data

Trending Topics: Phrase Frequency in AI Papers Over Time

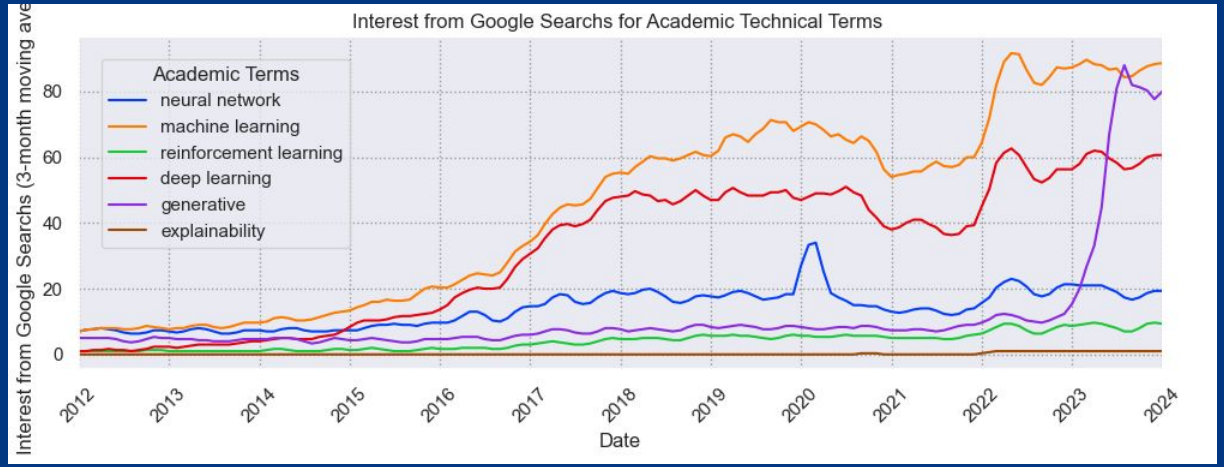
- The Transformer architecture exhibits significant growth, closely correlating with the AI boom.
- Generative models steadily rise, leading to products like Midjourney and Cutout Pro
- Self-supervised and reinforcement learning techniques grow in prominence, with
reinforcement learning as the favored approach
- Explainability and interpretability have grown since 2018, potentially indicating an emerging AI research trend



Source: arxiv.org AI academic papers abstract and title data

Trending Topics: Google Search Data

- AI academic keyword searches trend upward from 2017, with a second spike in 2022, coinciding with the rise of AI products.
- "Generative" grows consistently, surging dramatically in 2023 as it becomes a publicly recognized term for AI image generators services.



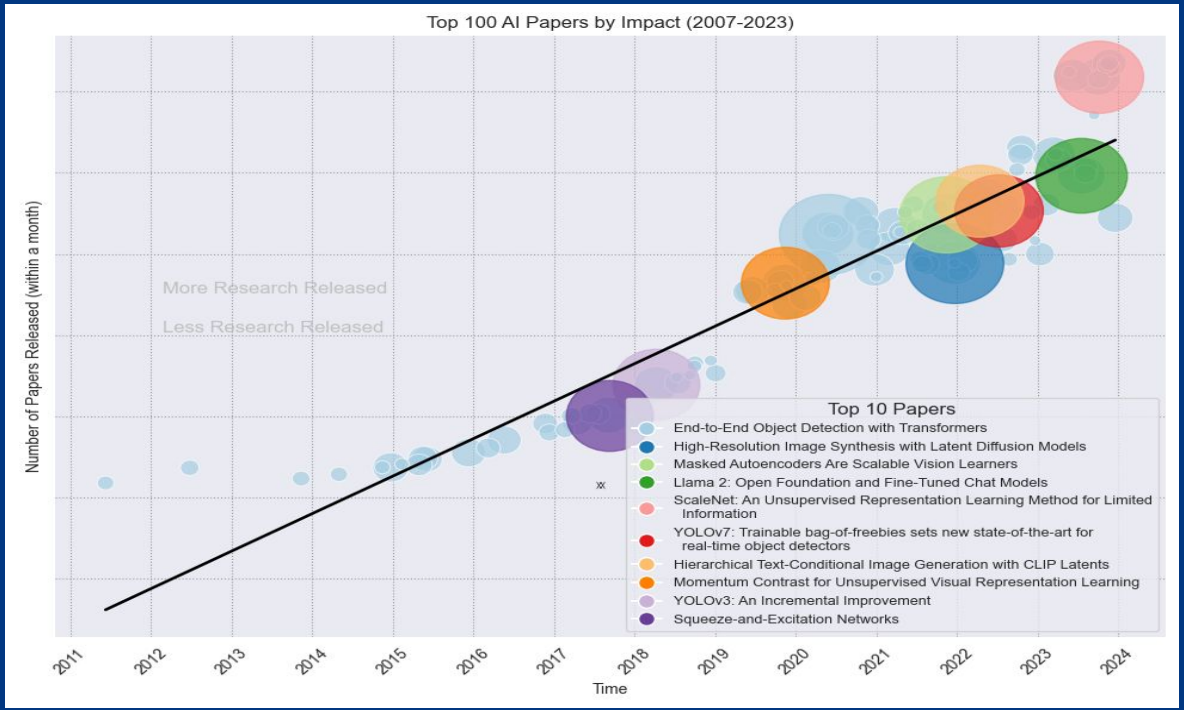
Source: Google.com Trends API

Google searches for AI academic keywords reflect research trends, with AI topics showing substantial growth in public interest and awareness

Groundbreaking Research: Top 100 Influential AI Papers, 2013-2023

- Linear regression model shows consistent upward trend in AI paper releases
- Top-ranked paper based on z-score involves Transformers, extensively used in current AI products like ChatGPT
- Computer vision and image detection are prominent, with 6 of the top 10 most impactful papers focusing on these areas.
- Add info on most recent impactful paper scalenet

Most of the impactful AI papers were published recently, with 6 of the all-time most influential papers in the last 2 years since the AI boom began



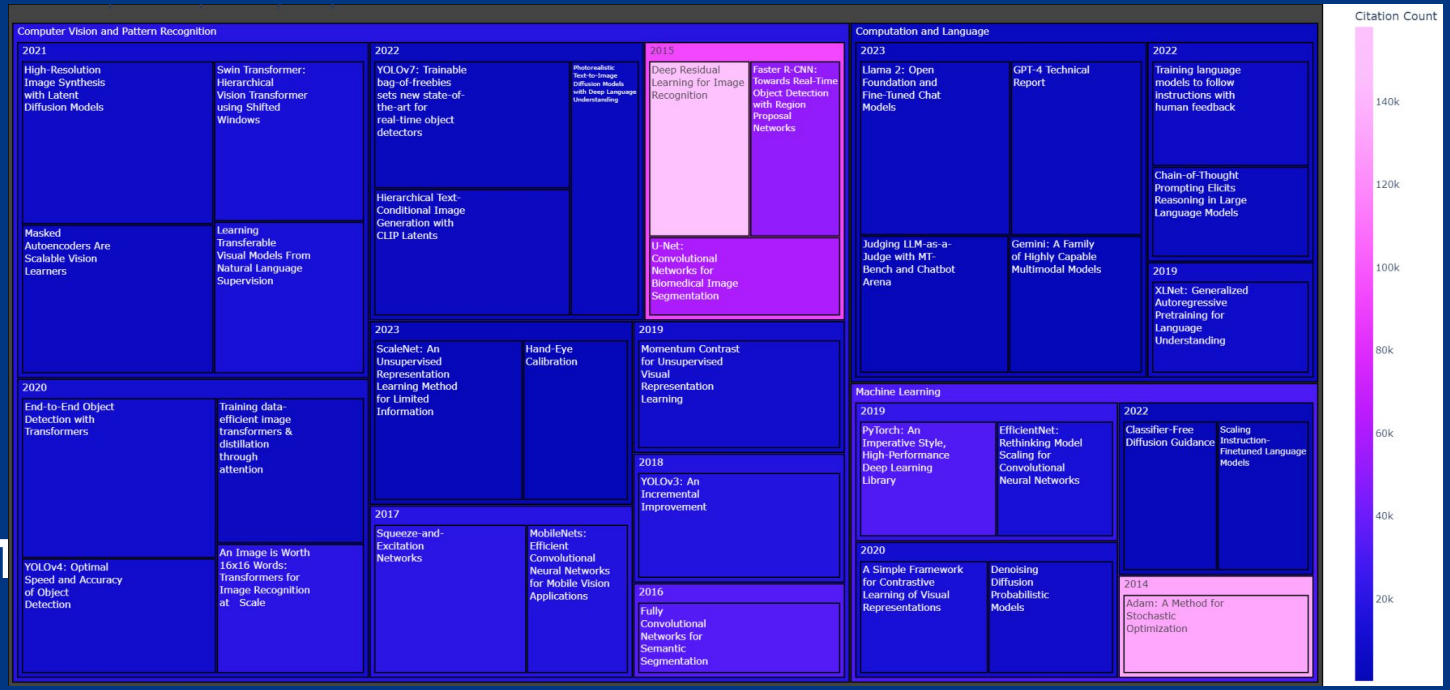
Bubbles represent citation counts relative to contemporary papers, mitigating bias favoring older papers

Source: semantic scholar API and arxiv.org AI academic papers metadata

Trailblazers: Dissecting the Top 35 AI Papers by Impact, Category, and Publication Date

- Significant proportion of influential papers have been published in the last 4 years, despite data spanning the previous decade
- Two papers from 2014-2015 have the highest citation counts, suggesting they have shaped AI research direction since publication

Computer vision leads in impactful research, but computation and language show influential papers in 2022-2023, potentially indicating a future trend



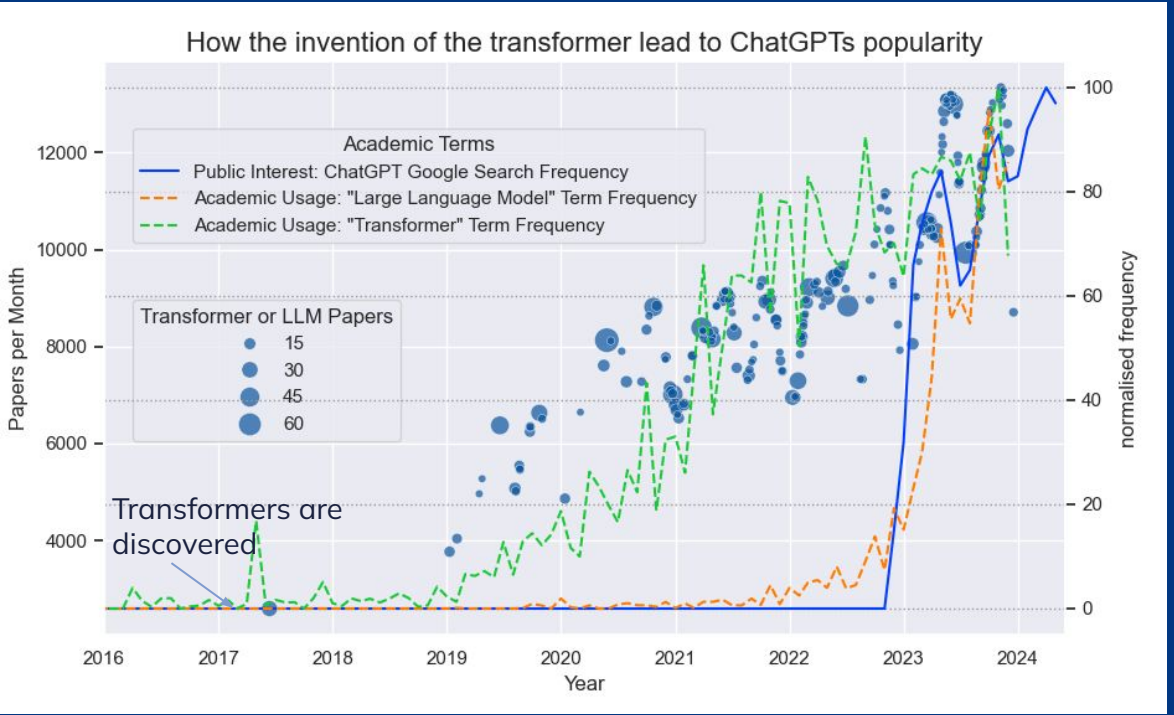
Size representing impact and colour denotes citation count

Source: semantic scholar API and arxiv.org AI academic papers metadata

‘Transformers’ Unraveled: Tracing the Impact of Pivotal Papers, Research Trends, and ChatGPT's Influence

- "Attention is All You Need" (2017), introduced ‘Transformer’ architecture, sparking research surge
- Numerous influential papers on Transformers were published.
- Academic usage of ‘Large Language Model’ correlates with Google Search of ‘ChatGPT’ trend, indicating research influenced by ChatGPT's commercial success
- LLMs emerged as rapidly growing research area during AI boom

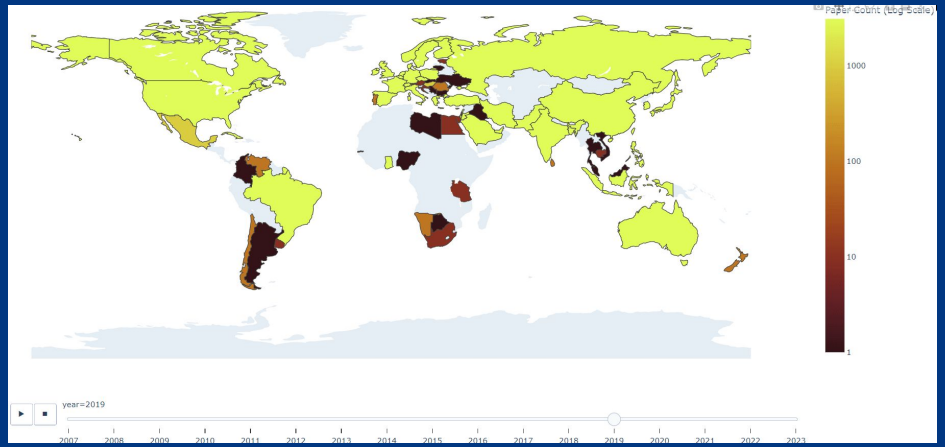
Transformer innovation reached the public through ChatGPT, igniting 2022 AI boom



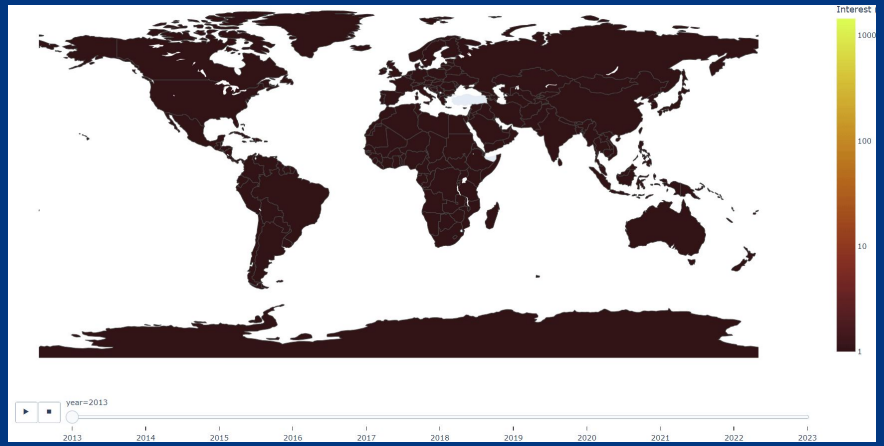
This analysis combines multiple graphs, plotting the impactful of influential Transformer papers, academic keyword usage in papers, and ChatGPT's Google search data.

Source: Semantic scholar API, Google.com Trends API and arxiv.org AI academic papers metadata

Geographical Disparity: AI Research Hubs vs AI Service Adoption



Source: Semantic scholar API and ‘Universities’ Python library



Source: Google.com Trends API

The map of AI academic papers' publication locations based on the primary author's affiliation shows a centralized distribution, with the USA leading, followed by China and the UK.

In contrast, the map of AI service Google Trends search data reveals a more dispersed user base, with slightly higher concentrations in developed countries with internet access and strong education systems.

Conclusions

The current AI boom could have been anticipated by analysing trends in academic research and public interest. AI paper releases began accelerating in 2017, preceding the product AI boom that started in mid-2022. The delay between research advancements and consumer products could have served as a predictor for the AI product boom.

Trending AI keywords and research areas have directly contributed to successful products. The academic data reveals consistent growth in the frequency of these keywords over the past 5 years, along with a significant number of impactful papers focusing on these topics.

The growth of these research areas may have signalled an impending product boom as the influential and popular research was incorporated into products.

Given the centralised nature of AI research, it is likely that most future AI growth will be concentrated in the USA, closely followed by the UK and China.

Despite machine learning's category majority, computer vision leads in impactful AI research, with computation and language showing promise in 2022-2023 impactful papers, potentially signaling a future trend.

Future AI products may emerge from currently growing research areas, such as scalenet, which has produced impactful papers using deep convolutional neural networks. Driveless cars

Next Steps

- Predictive Analysis: Utilise the established relationships between past AI academic paper metadata and subsequent AI industry growth to forecast future trends and growth areas based on AI research from the last 2 years.
- Data Modelling: Employ data modelling techniques to analyse current AI trends and leverage them for predictive analysis of future AI developments and their potential impact on the industry.