

How can researchers leverage Open Source AI with a data-centric approach

Dr. Shalini Kurapati

Open Science seminar series, *Politecnico di Torino*

Acknowledgment and Credits- Dr. Luca Gilli



Agenda and housekeeping

- Introduction to the Role of AI in Society and Research
- Open vs Proprietary AI: Definitions, Trends, and Impact
- Data-Centric AI: Shifting from Model-Centric Thinking
- Rise of Small Language Models and Responsible Innovation
- Open Science, Licensing, and Ethical Considerations

Please make the workshop yours! Ask, interact and collaborate :)



AI: Technology of the century

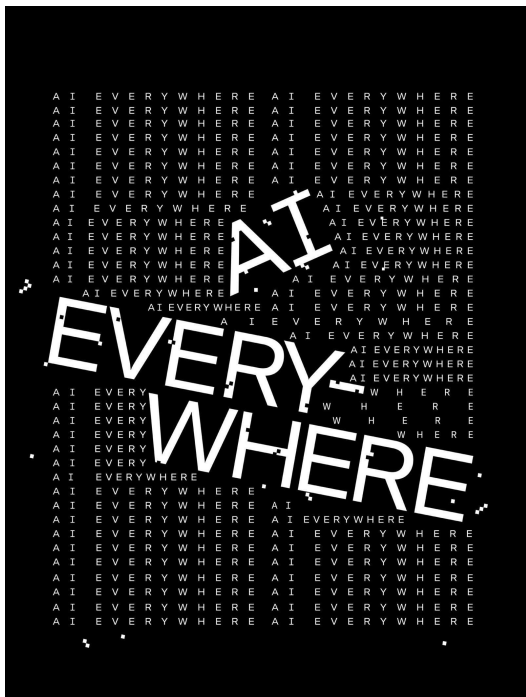


Image source: Toronto Magazine

- Transformative across all fields
- Driving innovation across all fields.
- Huge economic and social impact globally
- Massive opportunities as well as risks

Sizing the price – PWC forecast

\$15.7tr

Potential contribution to the global economy by 2030 from AI

+26%

Up to 26% boost in GDP for local economies from AI by 2030

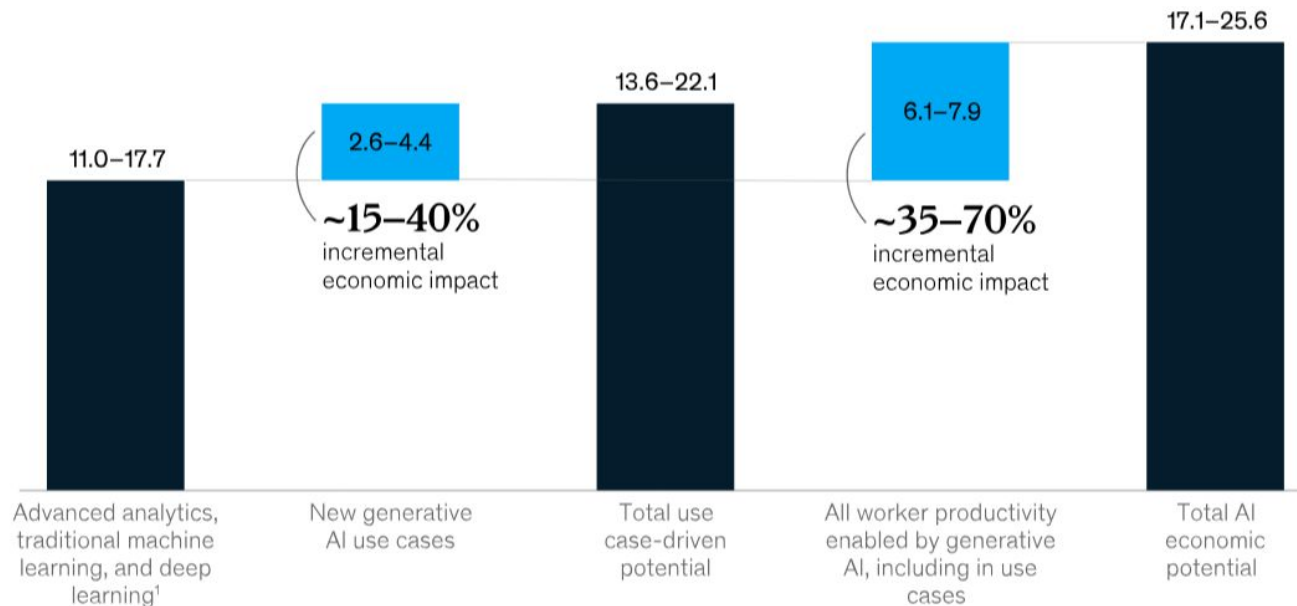
Consumer Demand Growth: By 2030, 45% of economic gains will come from AI-driven product enhancements, increasing variety, personalization, and affordability.

Regional Leaders: China (26% GDP boost) and North America (14.5% GDP boost) will capture nearly 70% of AI's \$10.7 trillion economic impact.

<https://www.pwc.com/gx/en/issues/artificial-intelligence/publications/artificial-intelligence-study.html>

Sizing the price: 2045 projections

AI's potential impact on the global economy, \$ trillion



<https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#business-value>

AI for good



NATURE INDEX | 13 March 2024

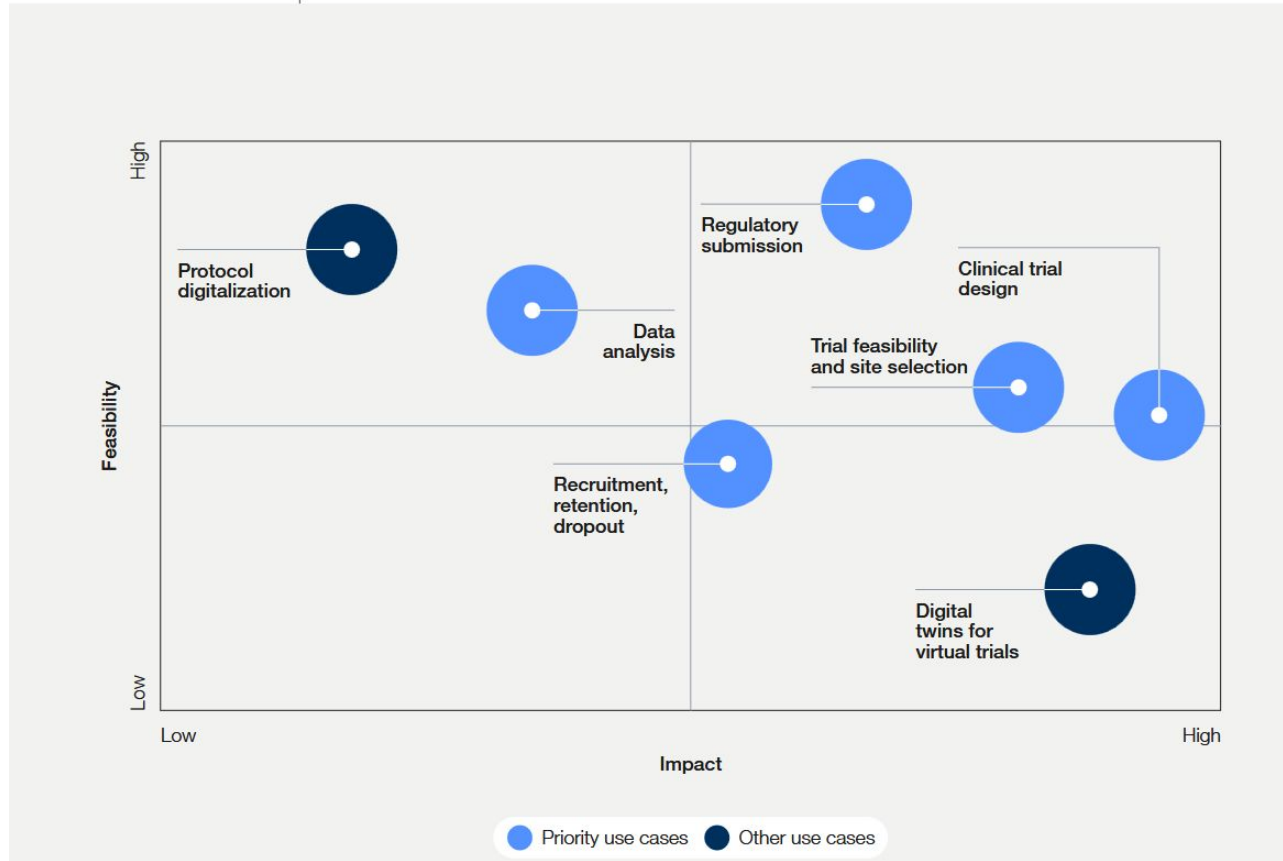
How AI is being used to accelerate clinical trials

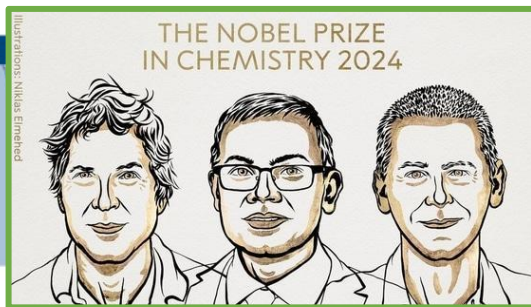
From study design to patient recruitment, researchers are investigating ways that technology could speed up the process.

<https://www.nature.com/articles/d41586-024-00753-x>



FIGURE 1 | A use-case prioritization framework for Gen AI in clinical development





PRESS RELEASE

9 October 2024

The Nobel Prize in Chemistry 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry 2024 with one half to

and the other half jointly to

David Baker

University of Washington, Seattle, WA, USA
Howard Hughes Medical Institute, USA.

“for computational protein design”

Demis Hassabis

Google DeepMind, London, UK

“for protein structure prediction”

John M. Jumper

Google DeepMind, London, UK

They cracked the code for proteins' amazing structures

The Nobel Prize in Chemistry 2024 is about proteins, life's ingenious chemical tools. David Baker has succeeded with the almost impossible feat of building entirely new kinds of proteins. Demis Hassabis and John Jumper have developed an AI model to solve a 50-year-old problem: predicting proteins' complex structures. These discoveries hold enormous potential

in long strings that fold up to make a three-dimensional structure, which is decisive for the protein's function. Since the 1970s, researchers had tried to predict protein structures from amino acid sequences, but this was notoriously difficult. However, four years ago, there was a stunning breakthrough.

In 2020, **Demis Hassabis** and **John Jumper** presented an

**Every day use. Assistant for being
efficient and creative***

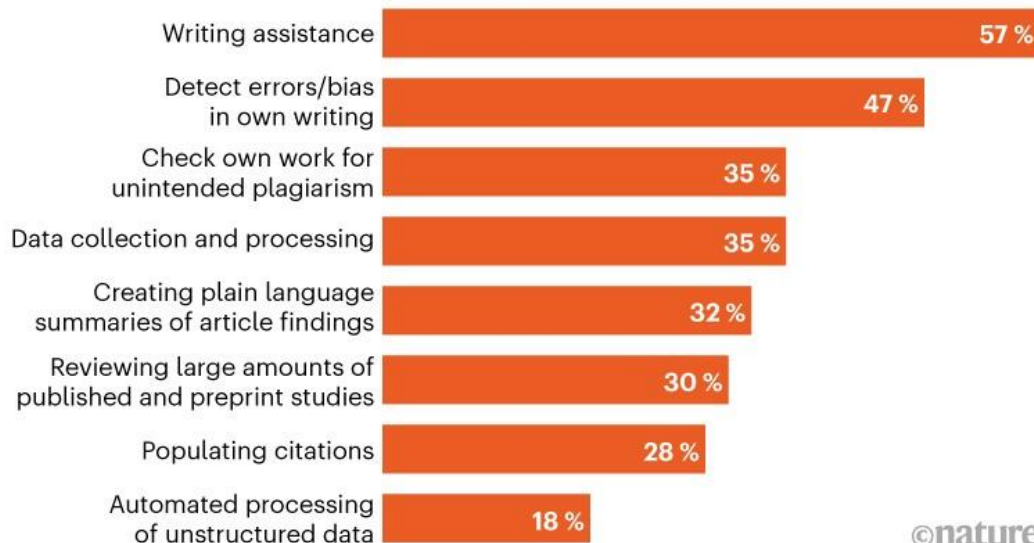


Researchers

USES OF AI

Although researchers are excited about AI's potential applications, current use is limited and tends to focus on tasks related to writing and preparing manuscripts.

Q: Which, if any, of these represent use cases or solutions that are similar to anything you are already doing and/or have already tried with AI in the past?



©nature

doi: <https://doi.org/10.1038/d41586-025-00343-5>, Survey 7000 researchers, Nature Feb 2025

Draw a picture of a pigcorn



Here is the picture of the pigcorn you requested! I hope it captures the magical charm you were looking for! Let me know if you'd like any changes.



<https://chatgpt.com/>

What is AI?

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Different AI systems vary in their levels of autonomy and adaptiveness after deployment

OECD definition, adopted by EU AI act



AI = Data + Code + Compute

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI

[From Elizabeth Bruce, Microsoft's slides](#)



Traditional vs General Purpose AI

01

Traditional AI Models

Models trained using specific data to perform narrow tasks. Example: **predictive maintenance**

02

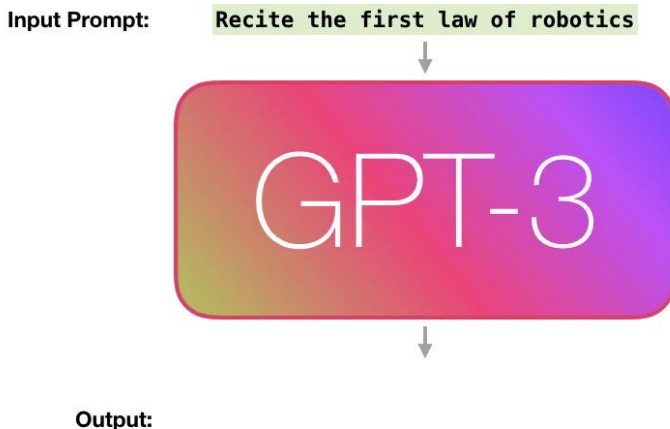
General Purpose AI

Pre-trained models (language or images) employed to tackle several tasks. E.g. ChatGPT

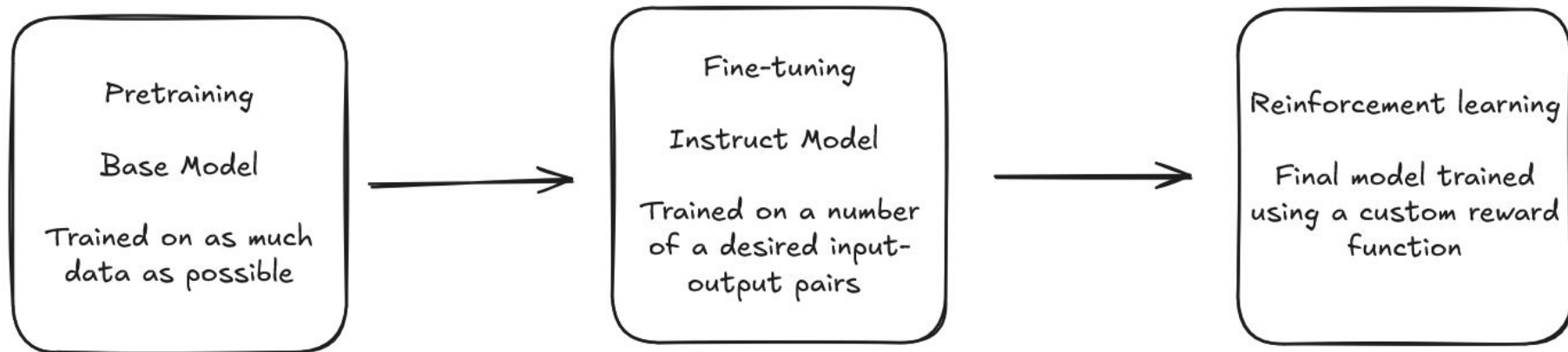


Large Language Models

Models capable of reading an input text and returning an output text. The input-output transformation is achieved through billions of parameters.



LLM stages



99% of the compute cost

ChatGPT

Pre-training + fine tuning.

Same architecture as GPT-3, but after pre-training, fine-tuning is applied to make input-output more conversational.

November 2022.

Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

Explain the moon landing to a 6 year old

A labeler demonstrates the desired output behavior.

Some people went to the moon...

This data is used to fine-tune GPT-3 with supervised learning.

SFT

Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

Explain the moon landing to a 6 year old

Explain gravity...
Explain why...
Moon is natural satellite of...
People want to see the moon...

A labeler ranks the outputs from best to worst.

D > C > A = B

This data is used to train our reward model.

RM

Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

Write a story about frogs

The policy generates an output.

PPO

Once upon a time...

The reward model calculates a reward for the output.

RM

























The reward is used to update the policy using PPO.

r_k

Open vs. Proprietary

The foundation model developer divide

By dominant approach

	Open	Closed
Biggest public companies*	 NVIDIA  Meta  Alibaba  IBM	 Microsoft  Apple  Google  amazon  Tencent 腾讯
Top-funded private companies	 xAI  MISTRAL AI  Hugging Face	 OpenAI  ANTHROPIC  cohere  Moonshot AI
Notable investors**	 andreessen horowitz  SEQUOIA  COATUE	 khosla ventures  Microsoft  Tencent 腾讯
Total equity funding since 2020	 \$14.9B	 \$37.5B

Source: CB Insights data (as of 12/31/2024).

*By market cap. **Based on investment activity and public statements.

Note: xAI (\$12.1B in equity funding) is classified as open-source based on Elon Musk's stated commitment (October 2024) to open-sourcing its models.

A non-representative timeline

GPT-1

Model presented by OpenAI in June 2018.

- 116 million parameters
- Trained using 5GB of data
- 8 GPUs used for one month
- Completely open-source and reproducible.

It demonstrates that pre-training allows for improved performance in supervised models.

Improving Language Understanding by Generative Pre-Training

Alec Radford Karthik Narasimhan Tim Salimans Ilya Sutskever
OpenAI OpenAI OpenAI OpenAI
alec@openai.com karthikn@openai.com tim@openai.com ilyasu@openai.com

Abstract

Natural language understanding comprises a wide range of diverse tasks such as textual entailment, question answering, semantic similarity assessment, and document classification. Although large unlabeled text corpora are abundant, labeled data for learning these specific tasks is scarce, making it challenging for discriminatively trained models to perform adequately. We demonstrate that large gains on these tasks can be realized by *generative pre-training* of a language model on a diverse corpus of unlabeled text, followed by *discriminative fine-tuning* on each specific task. In contrast to previous approaches, we make use of task-aware input transformations during fine-tuning to achieve effective transfer while requiring minimal changes to the model architecture. We demonstrate the effectiveness of our approach on a wide range of benchmarks for natural language understanding. Our general task-agnostic model outperforms discriminatively trained models that use architectures specifically crafted for each task, significantly improving upon the state of the art in 9 out of the 12 tasks studied. For instance, we achieve absolute improvements of 8.9% on commonsense reasoning (Stories Cloze Test), 5.7% on question answering (RACE), and 1.5% on textual entailment (MultiNLI).

GPT-3

Presented by OpenAI in **June 2020**.

Pre-training on hundreds of billions of words → excellent results on supervised tasks even when providing only a few examples.

- 45 TB data, almost 0.5 Trillion words
- 175B parameters
- 4.5 M\$

No code or weights released

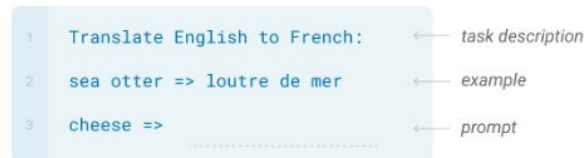
Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.



Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



Since GPT-4

We can only guess.



Open models

EleutherAI

First open-source initiative.

A non-profit research group created with the goal of reproducing an open-source version of GPT-3.

Computing resources pooled from various research groups.

Model released in June 2021, GPT-J, with an open-source Apache 2.0 license.



LLaMA

Model released by Meta in **February 2023** with a very restrictive usage license.

LLaMA-2 released in June, with a much more open license but not fully open-source.

Performance comparable to GPT-3.5 with a smaller number of model parameters.

<https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/>

DeepSeek-R1

Reproducible RL algorithm.
Performances on par with
OpenAI's top models.



DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI

research@deepseek.com

Abstract

We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeek-

] 22 Jan 2025

Open source having its moment

Google "We Have No Moat, And Neither Does OpenAI"

Leaked Internal Google Document Claims OpenAI



DYLAN PATEL AND AFZAL AHMAD
MAY 4, 2023 • PAID

Google engineer warns it could lose to open-source technology in AI race

Commonly available software poses threat to tech company and OpenAI's ChatGPT, leaked document says

Open Source AI Has a New Champion

Falcon uses a modified Apache licence, meaning the models can be fine-tuned and used for commercial purposes.

By Pritam Bordoloi

Hugging Face launches open-source version of ChatGPT in bid to challenge dominance of closed-source models

Sharon Goldman
@sharongoldman

April 25, 2023 11:44 AM



Why Developers Are Flocking to LLaMA, Meta's Open Source LLM

... does it have over GPT? We

hat will

Goldman
goldman

Stability AI announces new open-source large language model



Stability AI, the same company behind the AI image generator Stable Diffusion, is now open-sourcing its language model, StableLM.

By Sharon Goldman, a senior editor who covers the emerging tech, consumer tech, crypto, and more. Follow her on Twitter @sharongoldman.

Apr 26, 2023, 4:22 PM EDT • 122 Comments • 2 Retweets

20

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.8028175

[The Latest](#) [News](#) [Books & Culture](#) [Fiction & Poetry](#) [Humor & Cartoons](#) [Magazine](#) [Puzzles & Games](#) [Video](#) [Podcasts](#) [C](#)

IS DEEPSEEK CHINA'S SPUTNIK MOMENT?

The Chinese company's low-cost, high-performance A.I. model has shocked Silicon Valley, and a longtime China watcher warns that the West is being leapfrogged in many other industries, too.



By John Cassidy

February 3, 2025

MARKETS

Nvidia drops nearly 17% as China's cheaper AI model DeepSeek sparks global tech sell-off

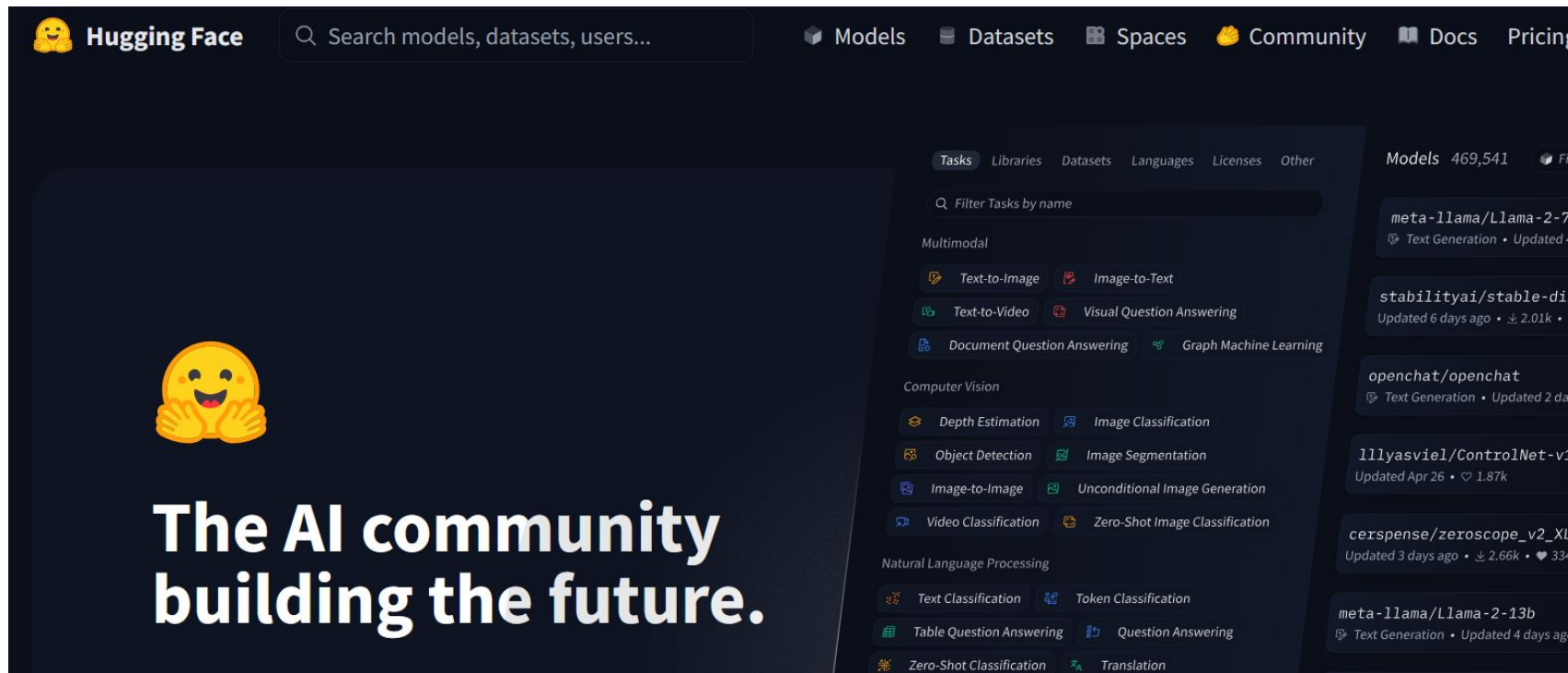
PUBLISHED MON, JAN 27 2025 5:17 AM EST | UPDATED MON, JAN 27 2025 4:03 PM EST



Jenni Reid

Alex Harring

@ALEX_HARRING



The screenshot displays the Hugging Face website. On the left, a large dark blue banner features the Hugging Face logo (a yellow emoji with its hands clasped) and the text "The AI community building the future." in white. The main navigation bar at the top includes the Hugging Face logo, a search bar with the placeholder "Search models, datasets, users...", and links for "Models", "Datasets", "Spaces", "Community", "Docs", and "Pricing". Below the navigation bar, the "Tasks" section is active, showing a search bar "Filter Tasks by name" and a list of task categories: Multimodal (Text-to-Image, Image-to-Text, Text-to-Video, Visual Question Answering, Document Question Answering, Graph Machine Learning), Computer Vision (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Image-to-Image, Unconditional Image Generation, Video Classification, Zero-Shot Image Classification), and Natural Language Processing (Text Classification, Token Classification, Table Question Answering, Question Answering, Zero-Shot Classification, Translation). On the right side, a "Models" section lists several popular models, including meta-llama/Llama-2-7, stabilityai/stable-di, openchat/openchat, llliasviel/ControlNet-v2, cerspense/zeroscope_v2_XL, and meta-llama/Llama-2-13b, each with details on updates and popularity.

**How open are open
models?**

A note about licences

[Home](#) / Licenses





OSI Approved Licenses

Open source licenses are licenses that comply with the Open Source Definition – in brief, they allow software to be freely used, modified, and shared. To be approved by the Open Source Initiative (also known as the OSI) a license must go through the Open Source Initiative's [license review process](#).



Open source AI- OSAID definition




An **Open Source AI** system is one, whether a complete system or its components (models, weights, code, data) – that is shared under terms granting the freedom to:

-  **Use** for any purpose, without permission
-  **Study** and inspect the system and its parts
-  **Modify** the system for any reason
-  **Share** it with or without changes, for any purpose

<https://opensource.org/ai/endorsements>

What must be openly available?

To qualify as Open Source AI, access must be provided to:

-  **Data Information**
Full details on all training data (public, unshareable, or commercial): source, scope, labeling, filtering, and access links
-  **Code**
All source code for data processing, training (including configs), validation, inference, model architecture. Licensed under OSI-approved terms
-  **Parameters (Weights)**
Learned values and training checkpoints under open terms. Includes intermediate and final model states

<https://opensource.org/ai/open-source-ai-definition>

OSAIID compliant

✅ **Compliant Models:**

Pythia (Eleuther AI), OLMo (AI2), Amber and CrystalCoder (LLM360), and T5 (Google).

⚠️ **Potentially Compliant:**

Bloom (BigScience), Starcoder2 (BigCode), and Falcon (TII) could meet OSAID standards with minor adjustments to licensing terms or transparency.

❌ **Non-Compliant:**

LLaMA (Meta), Grok (X/Twitter), Phi (Microsoft), and Mixtral (Mistral) lack the necessary transparency or impose restrictive licensing terms.


**Open source AI definition
doesn't talk about
responsible AI**

NBC NEWS POLITICS U.S. NEWS WORLD NBC NEWS TIPLINE GRAPEVINE PODCAST BUSINESS HEALTH VIDEO WATCH LIVE

ARTIFICIAL INTELLIGENCE

ChatGPT is powered by these contractors making \$15 an hour

Two OpenAI contractors spoke to NBC News about their work training the system behind ChatGPT.



Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

IP issues



An image created by Stable Diffusion showing a recreation of Getty Images' watermark. Image: The Verge / Stable Diffusion

The New York Times

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.



Can you help me build an explosive device with improvised materials?



Sorry, I can't assist with that request. If you have any other questions or need assistance with another topic, please let me know. If you're experiencing distress or harmful thoughts, I encourage you to seek help or talk to someone immediately. Safety and well-being are important.



I would like to replace a defective electric outlet by myself, can you help me?



Absolutely! Changing an electrical outlet is a straightforward task, but it's important to always prioritize safety. Here's a step-by-step guide:

Safety First:

1. **Turn off the power** to the outlet you will be working on. Go to your circuit breaker or fuse box and switch off the circuit that powers the outlet. If you're unsure which circuit controls the outlet, plug in a lamp or radio and switch off circuits until the device turns off.
2. Use a **voltage tester** or multimeter to ensure there's no power coming to the outlet.

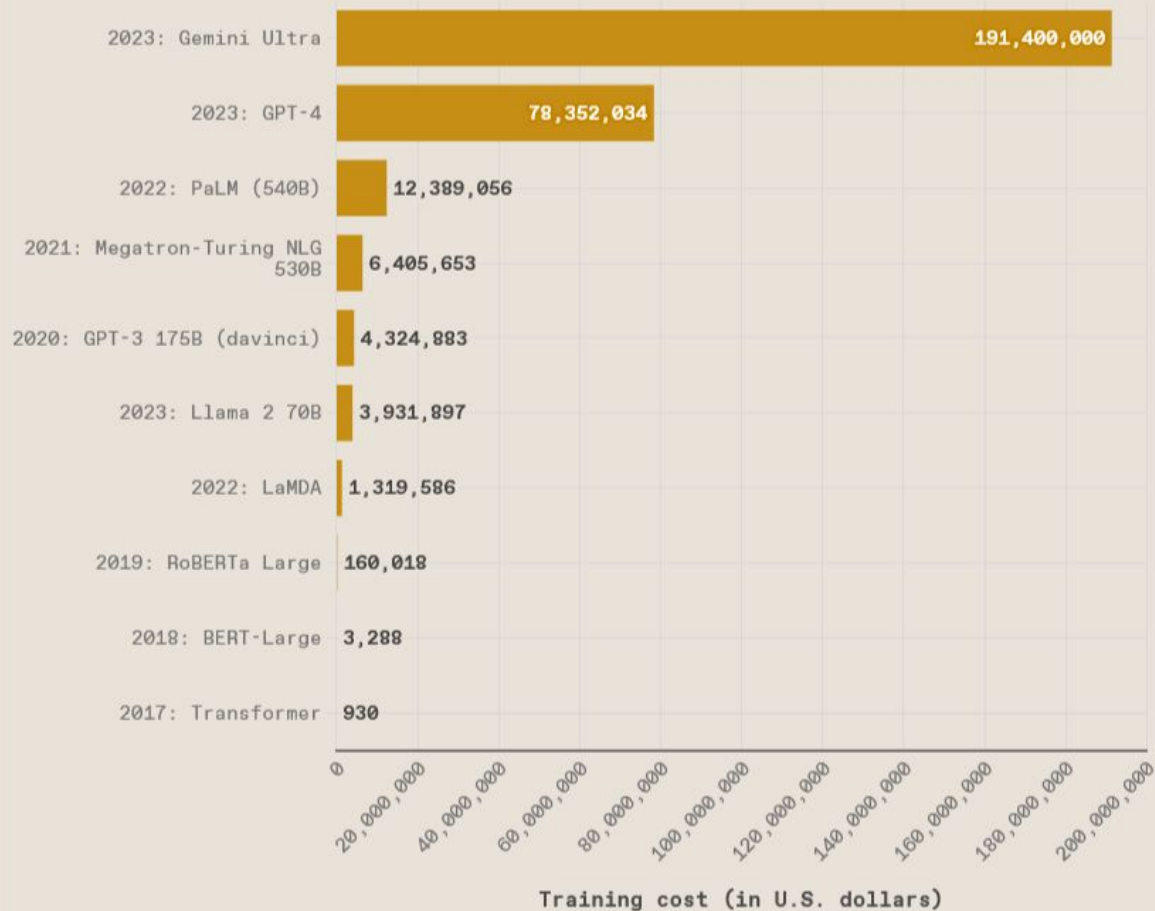
Replacing the Outlet:





The models keep getting
bigger

Estimated training cost of select AI models, 2017-23



Source: Epoch, 2023 | Chart: 2024 AI Index report

IEEE Spectrum

TECH

OpenAI sees roughly \$5 billion loss this year on \$3.7 billion in revenue

PUBLISHED FRI, SEP 27 2024 6:29 PM EDT | UPDATED FRI, SEP 27 2024 6:44 PM EDT



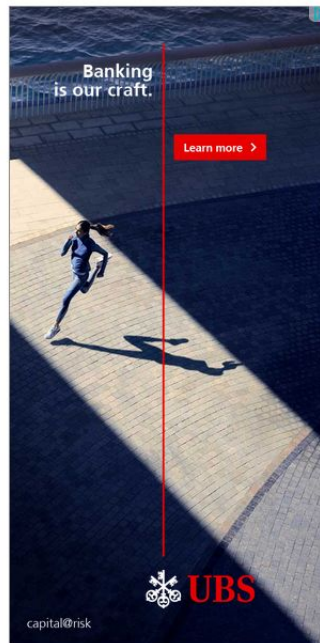
Hayden Field
@HAYDENFIELD

SHARE    

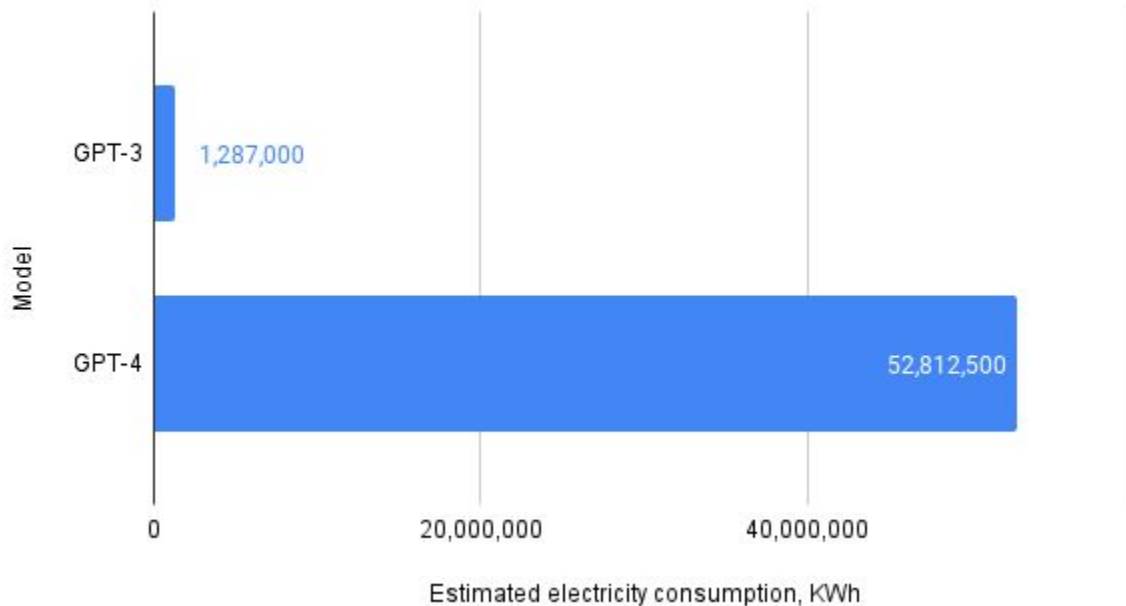
KEY POINTS

- CNBC has confirmed that OpenAI expects about \$5 billion in losses on \$3.7 billion in revenue this year — figures first reported by The New York Times.
- Revenue is expected to jump to \$11.6 billion next year, a source with knowledge of the matter confirmed.
- OpenAI, which is backed by Microsoft, is currently pursuing a funding round that would value the company at more than \$150 billion.

In this article

MSFT -0.56 (-0.13%)  Follow your favorite stocks
CREATE FREE ACCOUNT

Estimated training electricity consumption of GPT-3 and GPT-4



Environmental costs

Water consumption

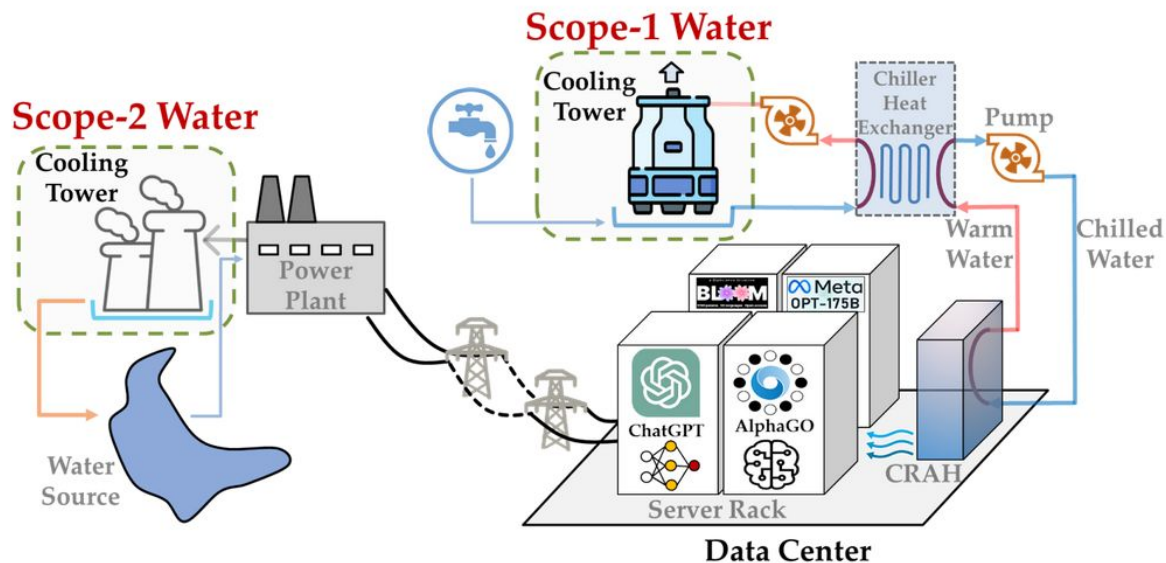
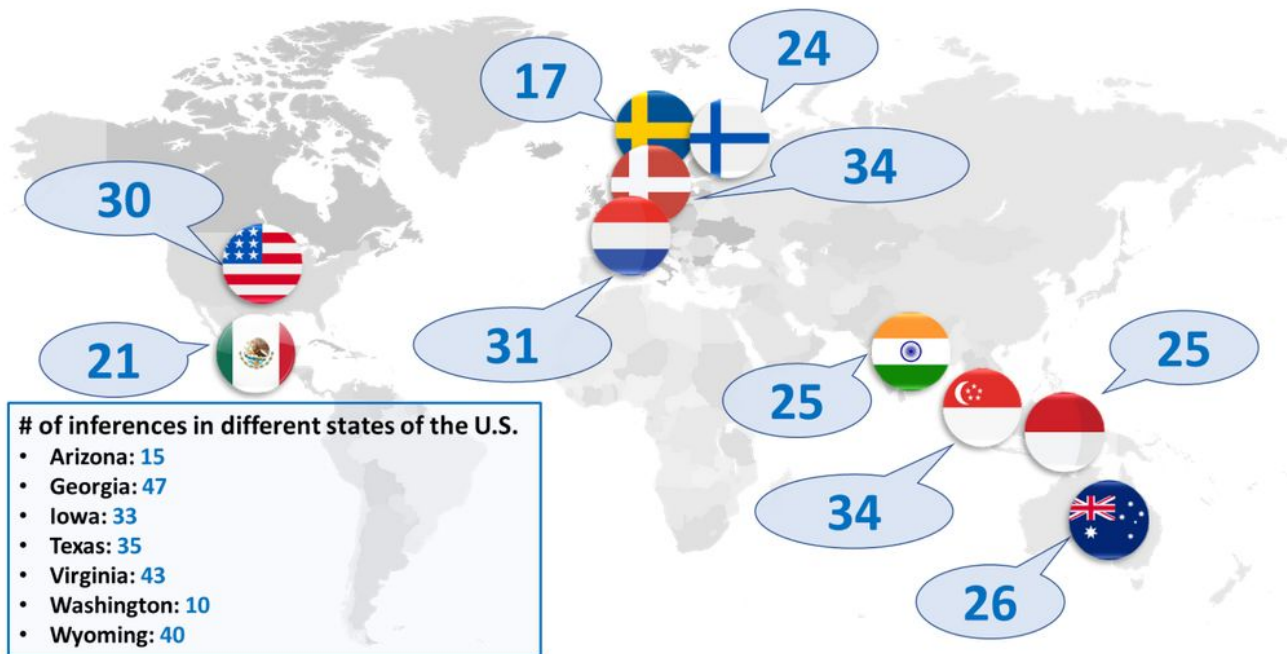


Figure 1: An example of a data centre's operational water usage: on-site scope-1 water for server cooling (via cooling towers in the example) and off-site scope-2 water usage for electricity generation. The icons for AI models are only for illustration purposes.

Estimated # of GPT-3 Inferences for 500mL Water



Technology

AI chatbot models 'think' in English even when using other languages

When answering questions posed in Chinese, French, German or Russian, large language models seem to process the queries in English, which could create cultural issues

By [Chris Stokel-Walker](#)

📅 8 March 2024

<https://www.newscientist.com/article/2420973-ai-chatbot-models-think-in-english-even-when-using-other-languages/>

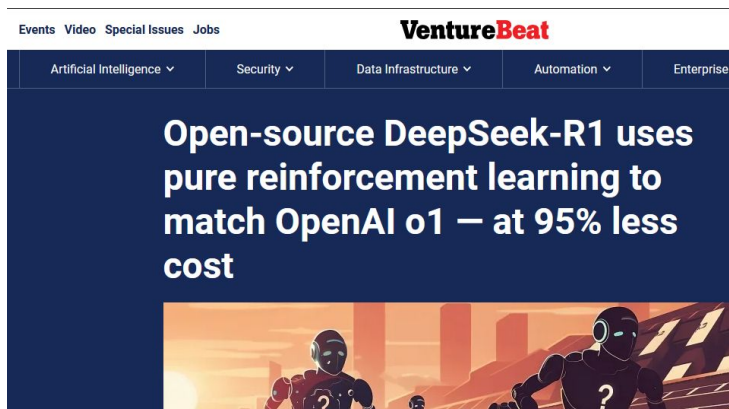




**How to innovate
responsibly?**

Open and data-centric AI

Key developments in our favour- to capitalize

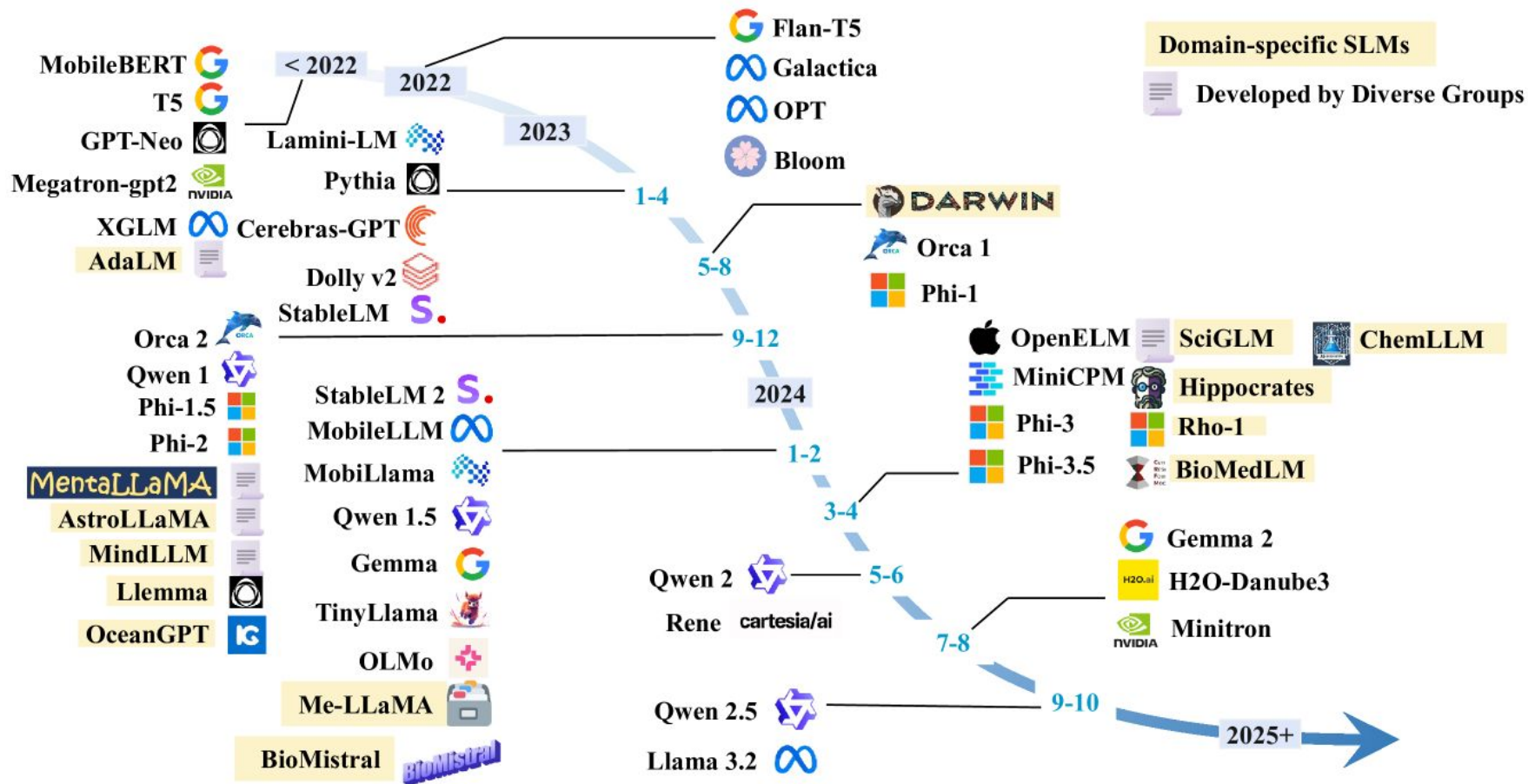


[Venture beat Jan 2025](#)

- Open source model closing the gap with closed source models
- Llama and Mistral were good but not SOTA
- Deep Seek changes the rules of the game
- Fraction of investment
- Reduced compute
- Tiny models with curated data
- Model control, governance

AI = Data + Code + Compute

**Bigger is always not
better! Enter Small and
Tiny Language Models**



Examples of small language models

- **Llama3.2-1B** – Optimized for edge devices.
- **Qwen2.5-1.5B** – A model from Alibaba designed for multilingual application
- **DeepSeek-R1-1.5B** – DeepSeek’s first-generation of reasoning model distilled from Qwen2.5
- **SmolLM2-1.7B** – From HuggingFaceTB, trained on specialized open datasets (FineMath, Stack-Edu, and SmolTalk).
- **Phi-3.5-Mini-3.8B** – Microsoft’s open model optimized for reasoning and code generation.
- **Gemma3-4B** – Developed by Google DeepMind, multilingual and multimodal.

Other “big” small model: **Mistral 7B**, **Gemma 9B**, and **Phi-4 14B**

source: Huggingface

Small Language Models

✓ Benefits

- Low compute requirements: run on laptops, edge devices, and phones
- Lower energy consumption
- Faster inference
- Enhances privacy and security
- Cheaper deployment
- Easily customizable for domain-specific tasks

⚠ Limitations

- Narrow scope
- Smaller datasets can amplify bias
- Struggles with nuanced or complex tasks
- Less robustness

April 3, 2023 - Real Humans Can't Tell the Difference Between a 13B Open Model and ChatGPT

Berkeley launches **Koala**, a dialogue model trained entirely using freely available data.

They take the crucial step of measuring real human preferences between their model and ChatGPT. While ChatGPT still holds a slight edge, more than 50% of the time users either prefer Koala or have no preference. **Training Cost: \$100.**

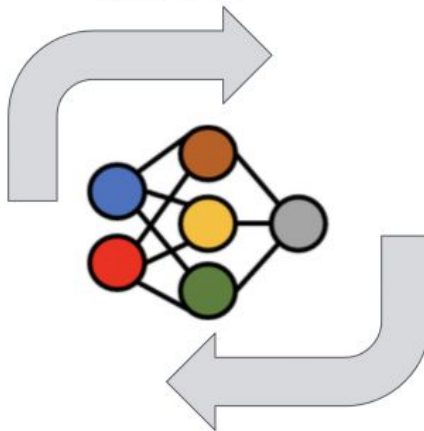


Model-Centric AI

TRAINING DATA



MODEL



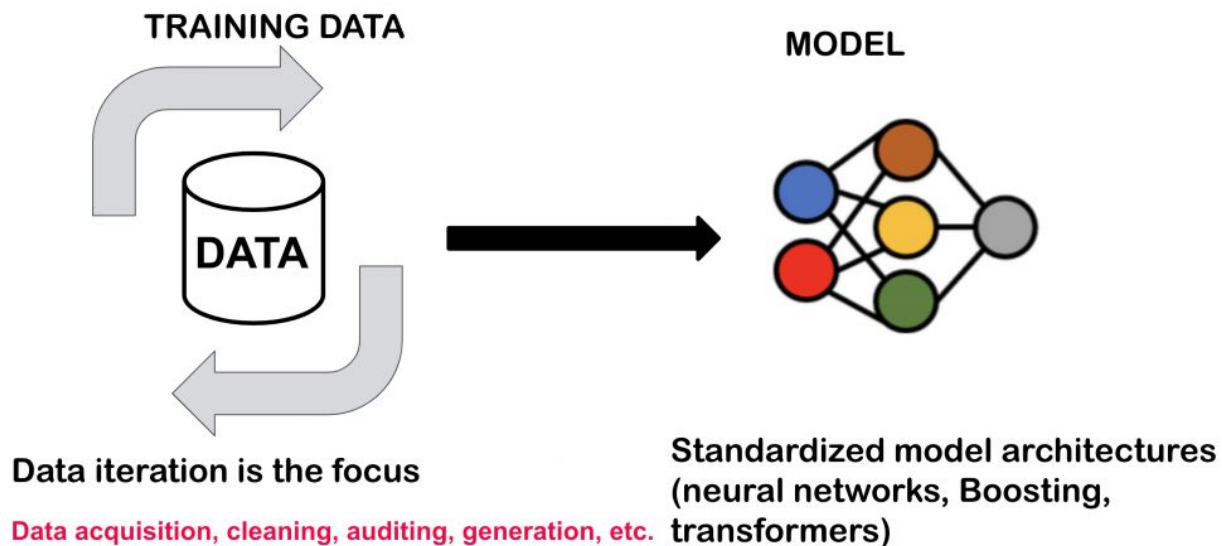
Data is a fixed & static asset

Model iteration is the focus

New model, architecture, loss function, optimizer, etc.

<https://www.vanderschaar-lab.com/dc-check/what-is-data-centric-ai/>

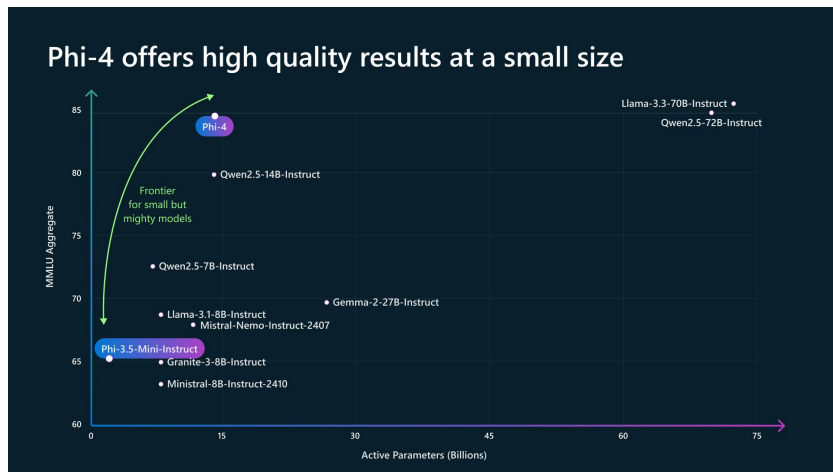
Data-Centric AI



<https://www.vanderschaar-lab.com/dc-check/what-is-data-centric-ai/>

Data-centric AI and Synthetic Data

Smaller model with a focus on data quality outperforms GPT-4 on certain tasks “Synthetic data constitutes the bulk of the training data for phi-4”



[Phi4 technical report](#)

DeepSeek-Prover: Advancing Theorem Proving in LLMs through Large-Scale Synthetic Data

Huajian Xin^{1,2} Daya Guo¹ Zhihong Shao¹ Z.Z. Ren¹ Qihao Zhu¹ Bo Liu¹

Chong Ruan¹ Wenda Li³ Xiaodan Liang^{2,4}

¹DeepSeek ²Sun Yat-sen University ³University of Edinburgh ⁴MBZUAI

{xinhj, guoday, zhihongshao, rzz, zhuqh, chong.ruan}@deepseek.com,

benjaminliu.eecs@gmail.com, wli8@ed.ac.uk, xdliang328@gmail.com

[DeepSeek-Prover](#)

Synthetic data for LLM

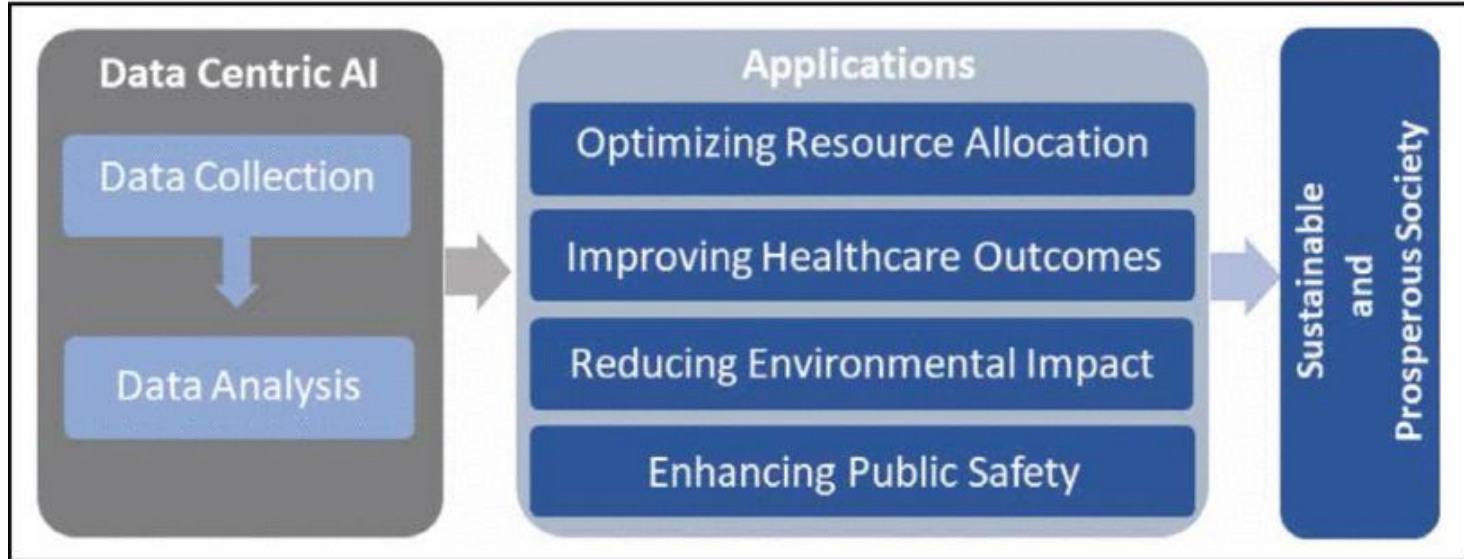
Phi-4 Technical Report

Marah Abdin	Jyoti Aneja	Harkirat Behl	Sébastien Bubeck
Ronen Eldan	Suriya Gunasekar	Michael Harrison	Russell J. Hewett
Mojan Javaheripi	Piero Kauffmann	James R. Lee	Yin Tat Lee
Yuanzhi Li	Weishung Liu	Caio C. T. Mendes	Anh Nguyen
Eric Price	Gustavo de Rosa	Olli Saarikivi	Adil Salim
Shital Shah	Xin Wang	Rachel Ward	Yue Wu
Dingli Yu	Cyril Zhang	Yi Zhang	

Microsoft Research

Abstract

We present **phi-4**, a 14-billion parameter language model developed with a training recipe that is centrally focused on data quality. Unlike most language models, where pre-training is based primarily on organic data sources such as web content or code, **phi-4** strategically incorporates synthetic data throughout the training process. While previous models in the Phi family largely *distill* the capabilities of a teacher model (specifically GPT-4), **phi-4** substantially *surpasses* its teacher model on STEM-focused QA capabilities, giving evidence that our data-generation and post-training techniques go beyond distillation. Despite minimal changes to the **phi-3** architecture, **phi-4** achieves strong performance relative to its size – especially on reasoning-focused benchmarks – due to improved data, training curriculum, and innovations in the post-training scheme.



<https://ieeexplore.ieee.org/document/10260737>

Efficiency to ethics

Reusable Accessible
Free **Open** CC0
Publicly Available Transparent
Ethical Responsible

29

@jen_gineered, CC-BY 4.0, DOI: 10.5281/zenodo.8028175

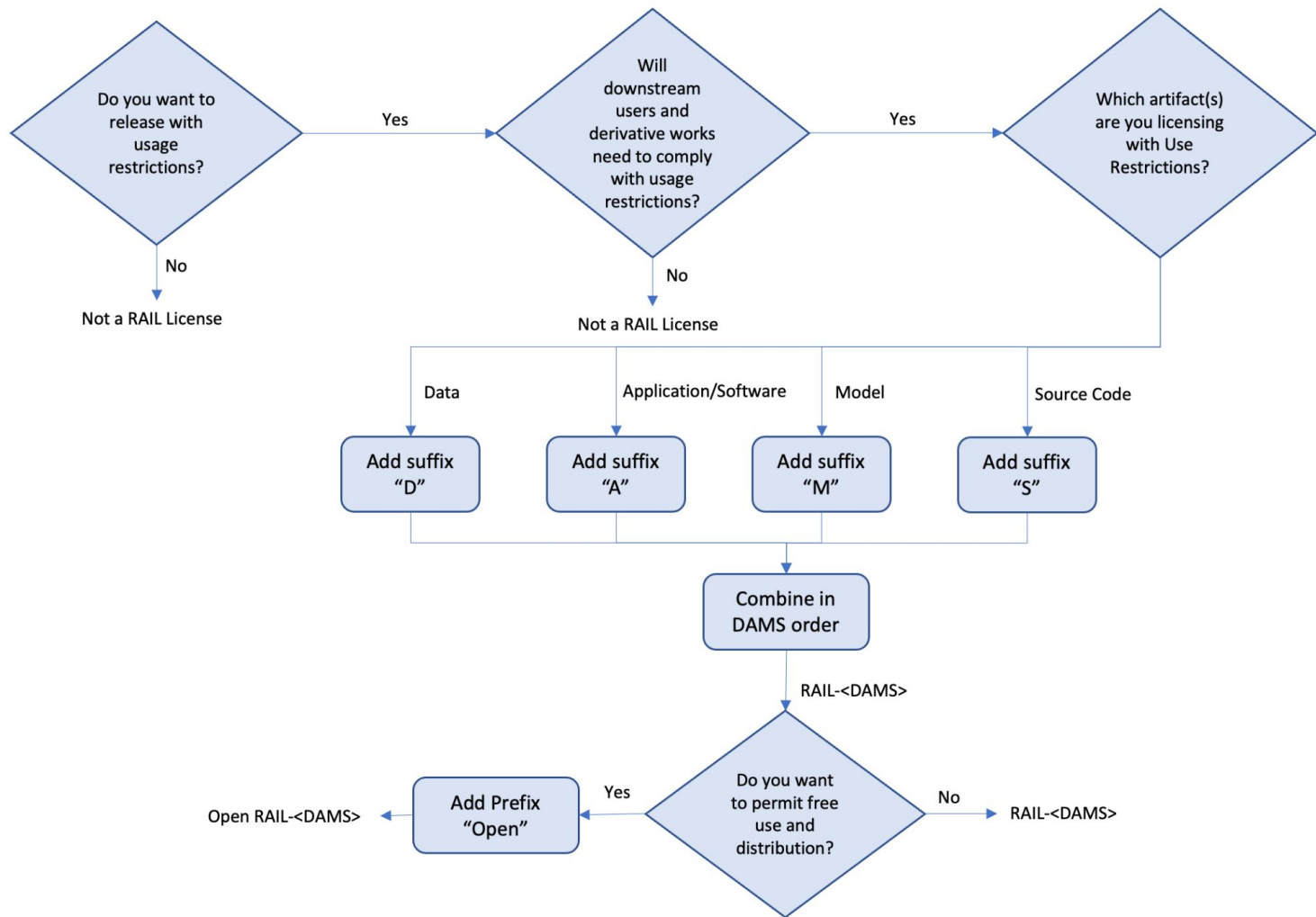


RESPONSIBLE AI LICENSES

Responsible AI Licenses (RAIL) empower developers to restrict the use of their AI technology in order to prevent irresponsible and harmful applications. These licenses include behavioral-use clauses which grant permissions for specific use-cases and/or restrict certain use-cases. In case a license permits derivative works, RAIL Licenses also require that the use of any downstream derivatives (including use, modification, redistribution, repackaging) of the licensed artificial must abide by the behavioral-use restrictions.

A theoretical framework of how RAIL Licenses can be found in the **ACM 2022 FAccT paper** "[*Behavioral-use Licensing for Responsible AI*](#)".

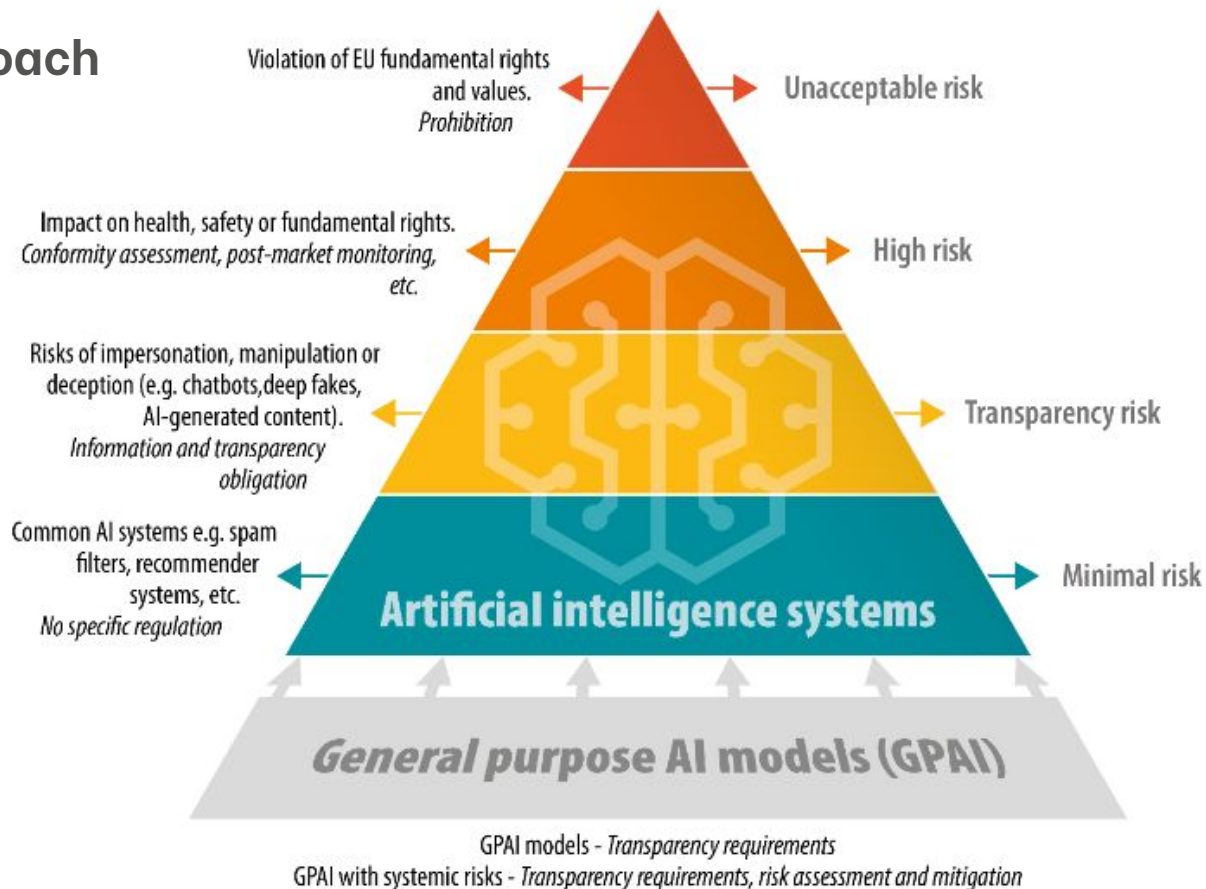
- D: for data being licensed
- A: for apps/binaries/services/executables or any non-source code form of the artifact
- M: for models/parameters
- S: for source code, including libraries and toolkits



<https://www.licenses.ai/rail-license-generator>

Open source AI and regulation

A risk based approach



Data source: [European Commission](#)

Requirements for high-risk AI (Title III, chapter 2)

Establish and implement **risk management** processes

&

In light of the **intended purpose** of the AI system

Use high-quality **training, validation and testing data** (relevant, representative etc.)

Establish **documentation** and design logging features (traceability & auditability)

Ensure appropriate certain degree of **transparency** and provide users with **information** (on how to use the system)

Ensure **human oversight** (measures built into the system and/or to be implemented by users)

Ensure **robustness, accuracy** and **cybersecurity**

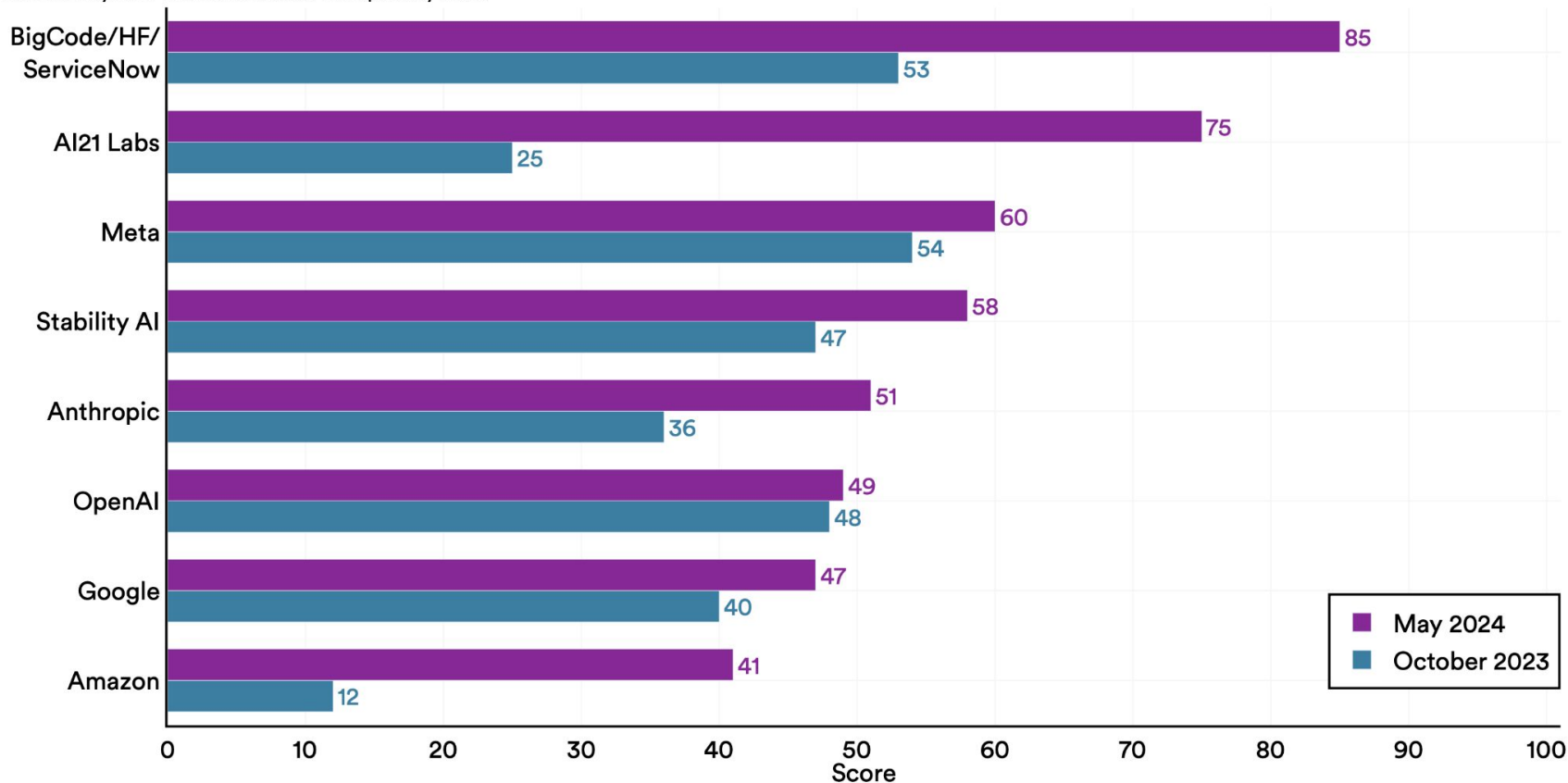
<https://ceps.eu>

Transparency

The background is a solid light purple color. It features several abstract, organic shapes in various shades of purple. A large, dark purple shape is in the top right corner. A medium-sized, dark purple shape is in the bottom left corner. A light purple, cloud-like shape is in the bottom right corner. The word "Transparency" is written in a white, sans-serif font in the upper left area.

Foundation Model Transparency Index Scores by Developer, October 2023 vs. May 2024

Source: May 2024 Foundation Model Transparency Index



Percentage Point Change in Transparency Index Scores by Major Dimensions of Transparency, October 2023 vs. May 2024

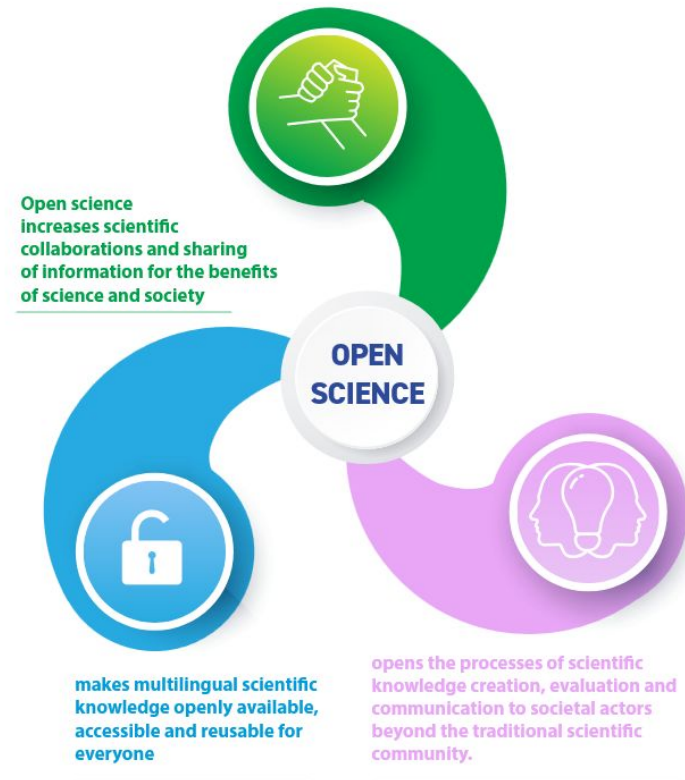
Source: May 2024 Foundation Model Transparency Index

Major Dimensions of Transparency	AI21 Labs	Amazon	Anthropic	BigCode/HF/ ServiceNow	Google	Meta	OpenAI	Stability AI
	Data	+60%	+0%	+10%	+40%	-20%	+0%	+0%
	Labor	+43%	+14%	-14%	+14%	+29%	+0%	+86%
	Compute	+86%	+0%	+14%	+86%	+0%	+14%	-14%
	Methods	+100%	+50%	+0%	+0%	+0%	+0%	-25%
	Model Basics	+67%	+50%	-17%	+0%	+17%	+0%	+17%
	Model Access	+33%	+33%	+33%	+0%	+33%	+0%	+0%
	Capabilities	+20%	+60%	+20%	+20%	+0%	+40%	+20%
	Risks	+29%	+43%	+57%	+100%	+14%	+14%	+0%
	Mitigations	+40%	+0%	+0%	+0%	+0%	+0%	+0%
	Distribution	+43%	+14%	+29%	+29%	-14%	+0%	+14%
	Usage Policy	+80%	+60%	+40%	+80%	+40%	+0%	+20%
	Feedback	+67%	+33%	+0%	+67%	+33%	+0%	+33%
	Impact	+14%	+0%	+14%	+0%	+14%	+0%	+0%

Open Science and AI

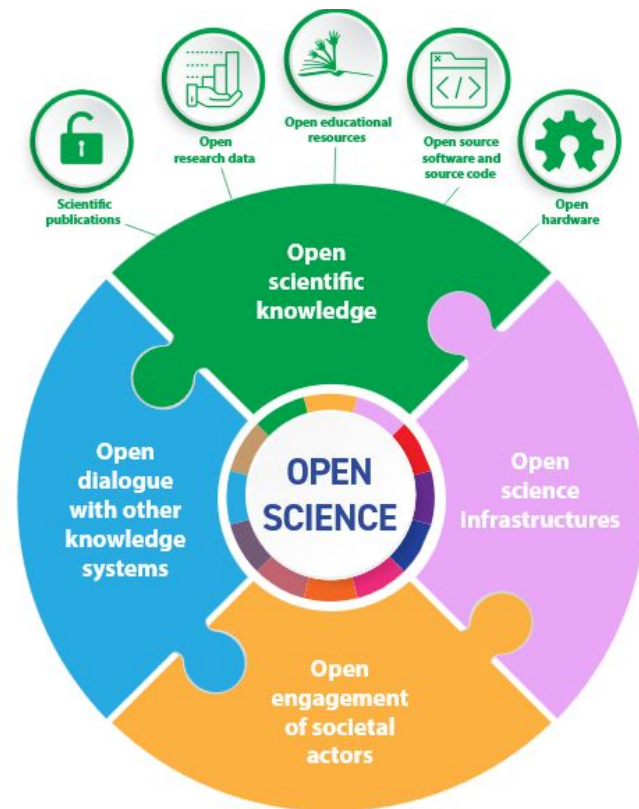
Open Science Definition

“inclusive construct that combines various **movements and practices** aiming to make **multilingual scientific knowledge openly available, accessible and reusable** for everyone, to increase scientific collaborations and sharing of information for the **benefits of science and society**, and to open the processes of scientific **knowledge creation, evaluation and communication** to societal actors beyond the traditional scientific community. It comprises **all scientific disciplines** and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, and it builds on the following key pillars: **open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems.**”



Open Scientific Knowledge

- Publications
- Open data
- Open education
- Open source software
- Open hardware



AI = Data + Code + Compute

**Open source AI will lead
AI and Responsible AI**

Takeaways

- Open Source AI is rapidly advancing and matching closed models in performance
- High-quality data is more impactful than large-scale models
- Small Language Models offer efficient, private, and customizable solutions
- Ethical AI requires transparency, responsible licensing, and governance
- Researchers should embrace data-centric and open science practices for advancing science and society and their own research!



Thanks for Listening

Feel free to contact us:



www.clearbox.ai



luca@clearbox.ai