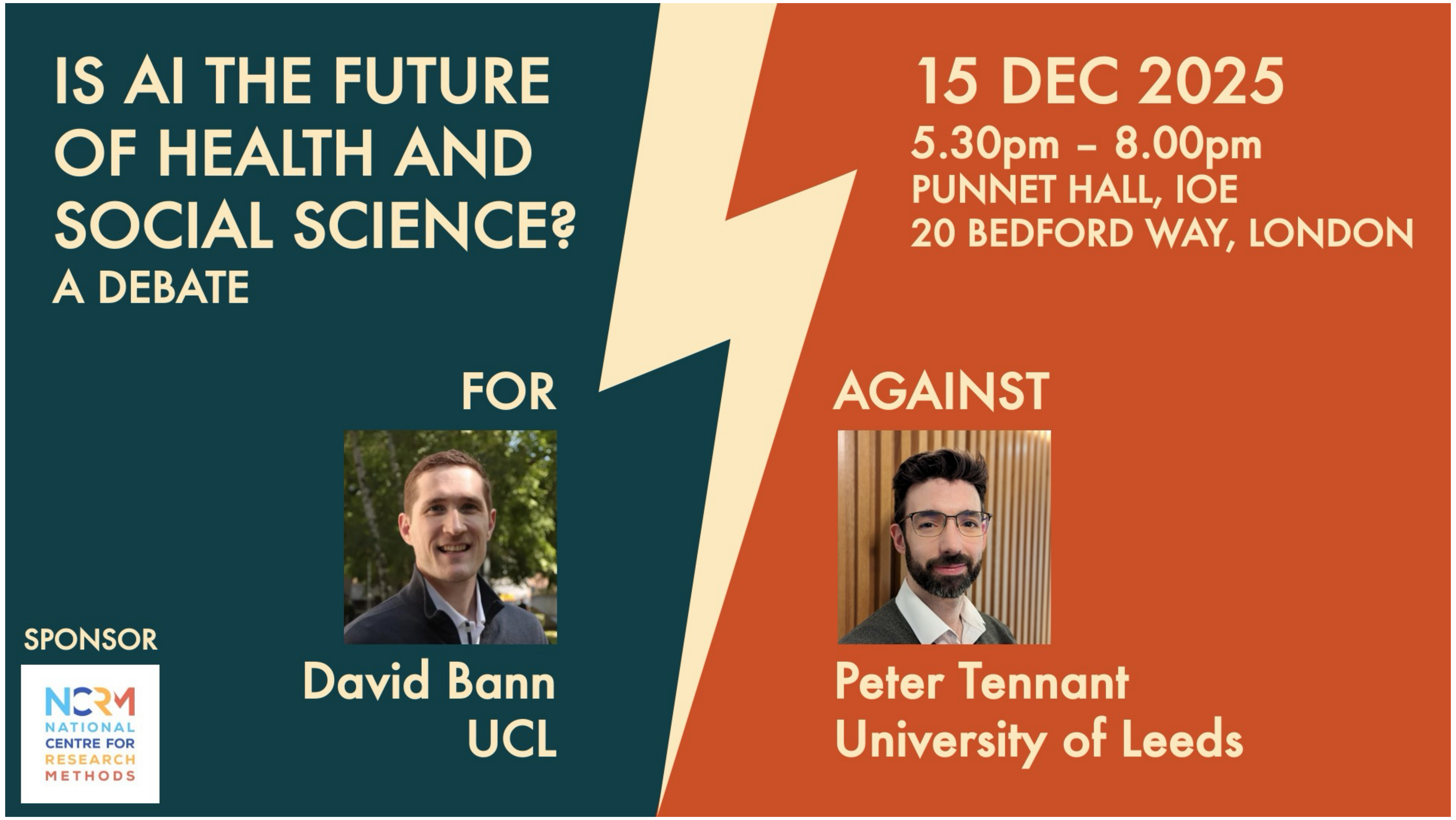
Is AI the future of health and social science?



“The aim of argument… should not be victory, but progress” Joubert, 1848

Argument

1. **Quantitative research is mostly cognitive work**

- Humans…
- AI…

2. **If AI helps research, it’s ‘the future’**

- Humans + AI

Argument

1. **Quantitative research is mostly cognitive work**

- Humans…
- AI…

2. **If AI helps research, it's 'the future'**

- Humans + AI

Tasks in quantitative health / social science research

- **Designing studies + collecting data**
  - Survey work (lit review, design, pilot, fieldwork)
- **Papers**:
  - Literature review
  - Forming research questions/hypothesis
  - Data analysis
  - Interpretation / write-up

Tasks in quantitative health / social science research

Cognitive tasks

- **Designing studies + collecting data**
  - Survey work (**lit review, design**, pilot, fieldwork)
- **Papers:**
  - **Literature review**
  - **Forming research questions/hypothesis**
  - **Data analysis**
  - **Interpretation / write-up**

GPTs are GPTs: Labor market impact potential of LLMs

Research is needed to estimate how jobs may be affected

By Tyna Eloundou[1], Sam Manning[2], Pamela Mishkin[1], Daniel Rock[3]

We propose a framework for evaluating the potential impacts of large-language models (LLMs) and associated technologies on work by considering their relevance to the tasks workers perform in their jobs. By applying this framework (with both humans and using an LLM), we estimate that roughly 1.8% of jobs could have over half their tasks affected by LLMs with simple in-

technology discussions, focusing more narrowly on advanced software capabilities than on the potential for business process reengineering, new intangible assets creation, or workforce retraining. General-purpose technologies such as electricity or computing historically have had far-reaching effects that took decades to fully materialize. With evidence of the general-purpose technology potential of LLMs, we urge caution in making long-term predictions while offering an outline of where work might change.

Supplement, p41

Group	Occupations with highest exposure	% Exposure
Human *E*1	Interpreters and Translators	76.5
	Survey Researchers	75.0
	Poets, Lyricists and Creative Writers	68.8
	Animal Scientists	66.7
	Public Relations Specialists	66.7
Human $E1 + 0.5 * E2$	Survey Researchers	84.4
	Writers and Authors	82.5
	Interpreters and Translators	82.4
	Public Relations Specialists	80.6
	Animal Scientists	77.8

exposure=task completion by at least 50%

Eloundou et al 2024; work activity link

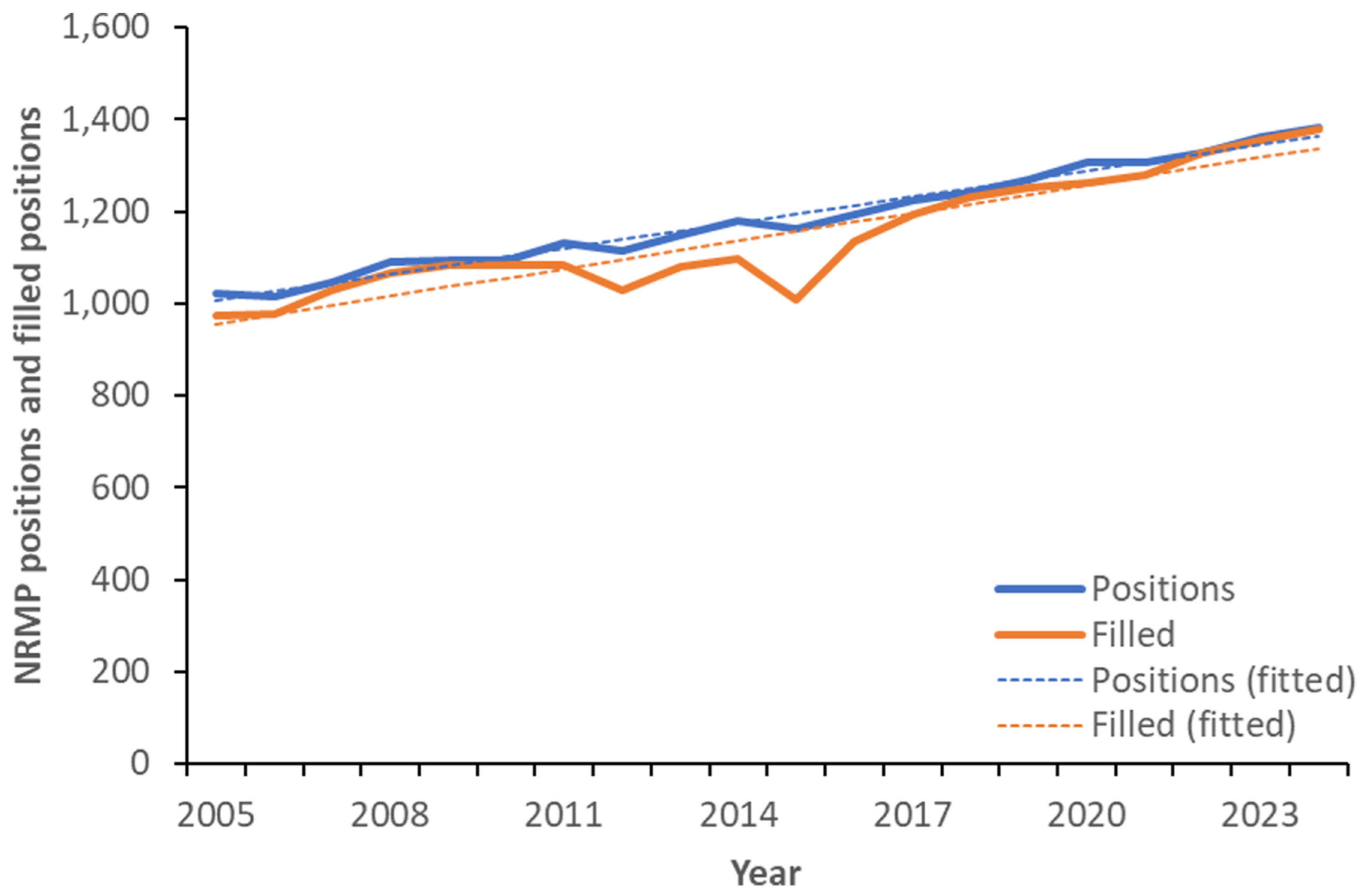
Exposed != imminently extinct

NY Times revisits Nobel Prize winner's prediction AI will render radiologists obsolete

Marty Stempniak | *May 15, 2025* | *Radiology Business* | *Artificial Intelligence*

"People should stop training radiologists now…"

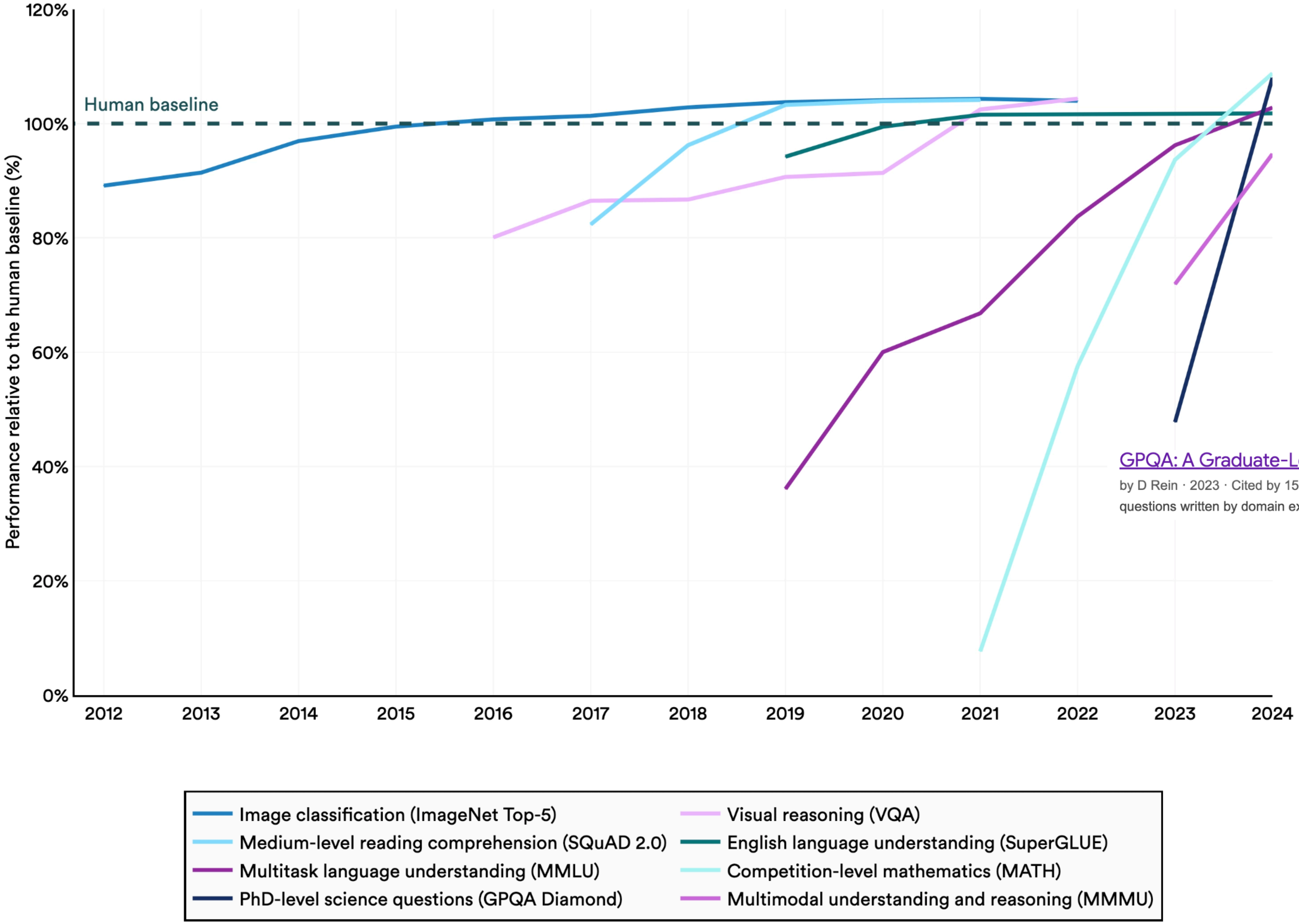
Hinton 2016



Christensen et al 2025

Select AI Index technical performance benchmarks vs. human performance

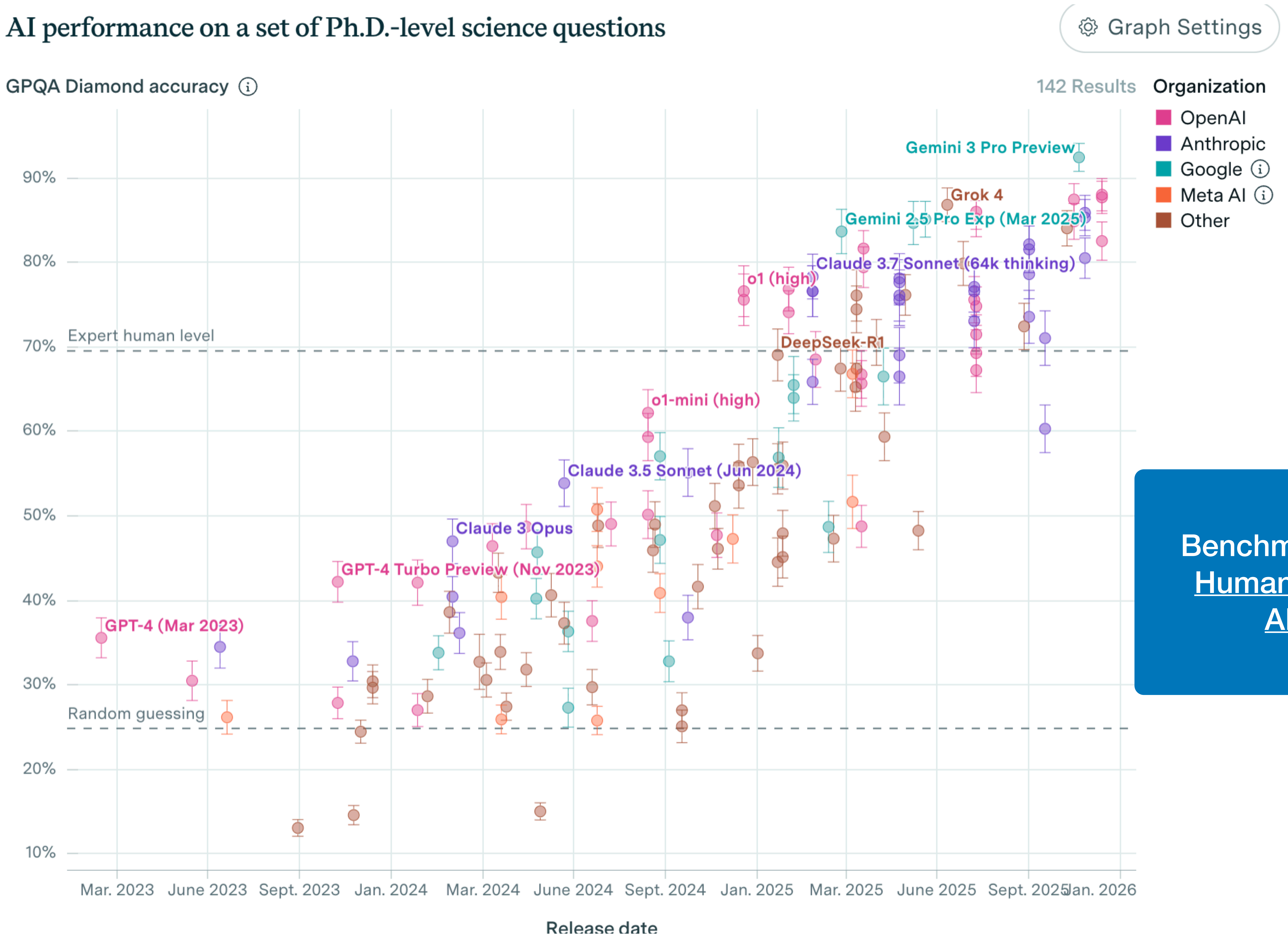
Source: AI Index, 2025 | Chart: 2025 AI Index report



[GPQA: A Graduate-Level Google-Proof Q&A Benchmark]

by D Rein · 2023 · Cited by 1514 — We present **GPQA**, a challenging dataset of 448 multiple-choice questions written by domain experts in biology, physics, and chemistry. See more

https://hai.stanford.edu/ai-index/2025-ai-index-report



Benchmark saturation?
Humanity's last exam
ARC-AGI…

https://epoch.ai/benchmarks/gpqa-diamond

GPT 5.2 - 3 days ago, ~30 day since 5.1 (!)

OpenAI — Run with maximum available reasoning effort.

	GPT-5.2 Thinking	GPT-5.1 Thinking
SWE-Bench Pro Software engineering	• 55.6%	50.8%
GPQA Diamond Science questions (No tools)	• 92.4%	88.1%
CharXiv Reasoning Scientific figure questions (No tools)	• 82.1%	67.0%
FrontierMath Advanced mathematics (Tier 1 – 3, Tier 4)	• 40.3% 14.6%	31.0% 12.5%
AIME 2025 Competition math (No tools)	• 100.0%	94.0%
ARC-AGI-1 Abstract reasoning	• 86.2%	72.8%
ARC-AGI-2 Abstract reasoning	• 52.9%	17.6%
GDPval Knowledge work tasks	• 70.9%	38.8% GPT-5

https://openai.com/index/introducing-gpt-5-2

AI: powerful, yet (for now) ‘jagged’

DeepSeek’s self-correcting AI model aces tough maths proofs

The model, DeepSeekMath-V2, scored 118 out of 120 points on questions from the 2024 William Lowell Putnam Mathematical Competition, beating the top human score of 90. The model also performed at the level of gold-medal winners in the International Mathematical Olympiad (IMO) 2025 and the 2024 China Mathematical Olympiad. The results are described

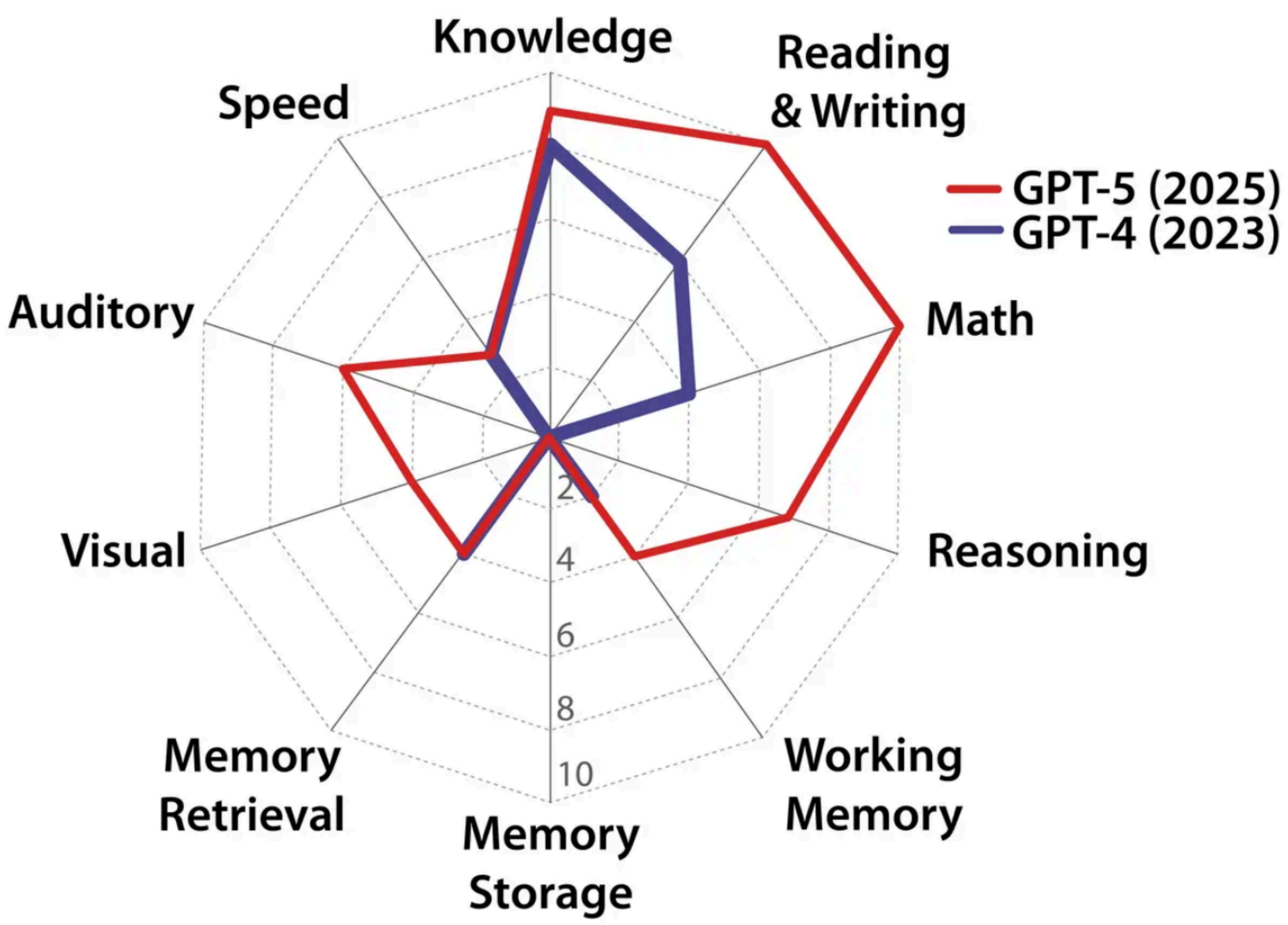
https://www.nature.com/articles/d41586-025-03959-9

https://arxiv.org/html/2511.22570v1

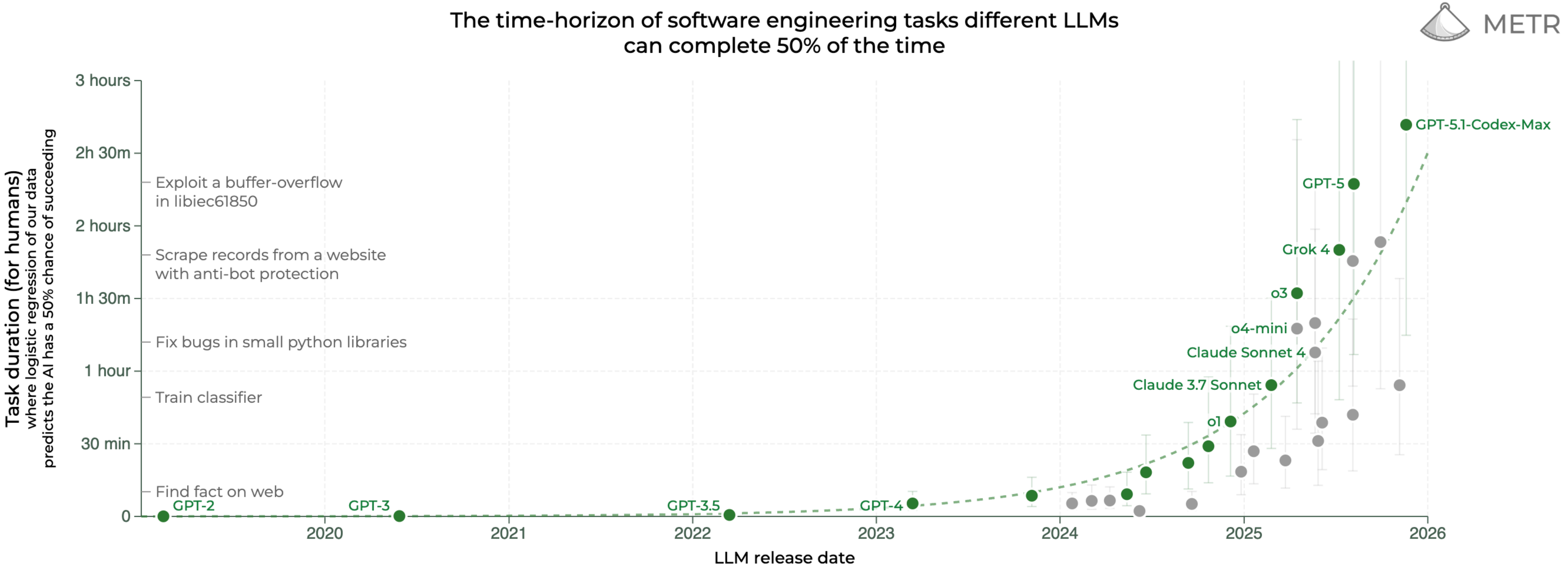
FINANCIAL TIMES

DeepMind and OpenAI achieve gold at ‘coding Olympics’ in AI milestone

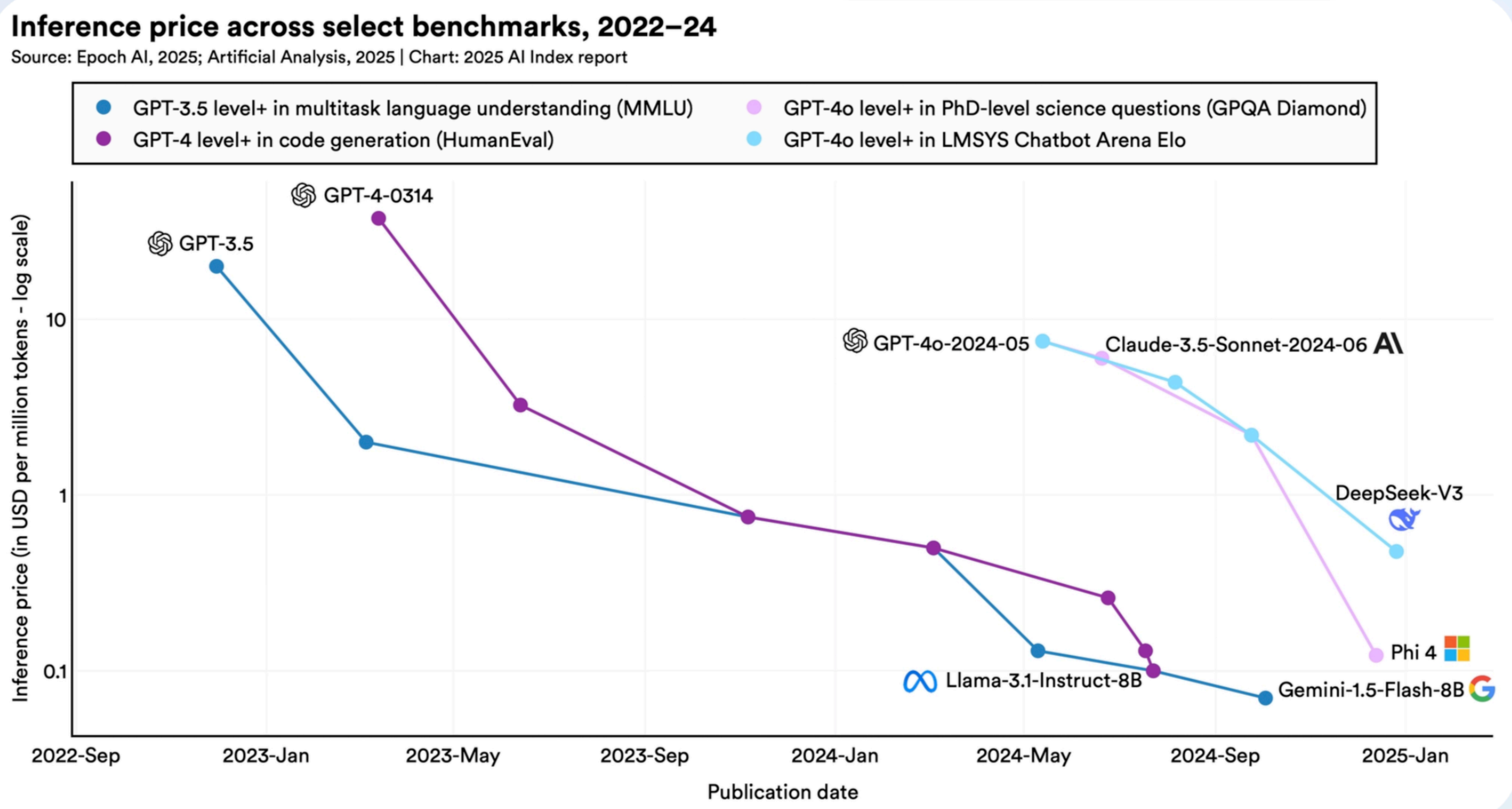
https://www.ft.com/content/c2f7e7ef-df7b-4b74-a899-1cb12d663ce6



Hendrycks et al 2025



https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/


Inference price across select benchmarks, 2022–24
Source: Epoch AI, 2025; Artificial Analysis, 2025 | Chart: 2025 AI Index report
GPT-3.5 level+ in multitask language understanding (MMLU)
GPT-4 level+ in code generation (HumanEval)
GPT-4o level+ in PhD-level science questions (GPQA Diamond)
GPT-4o level+ in LMSYS Chatbot Arena Elo
GPT-4-0314
GPT-3.5
GPT-4o-2024-05
Claude-3.5-Sonnet-2024-06
DeepSeek-V3
Phi 4
Llama-3.1-Instruct-8B
Gemini-1.5-Flash-8B
Inference price (in USD per million tokens - log scale)
10
1
0.1
2022-Sep
2023-Jan
2023-May
2023-Sep
2024-Jan
2024-May
2024-Sep
2025-Jan
Publication date

Argument

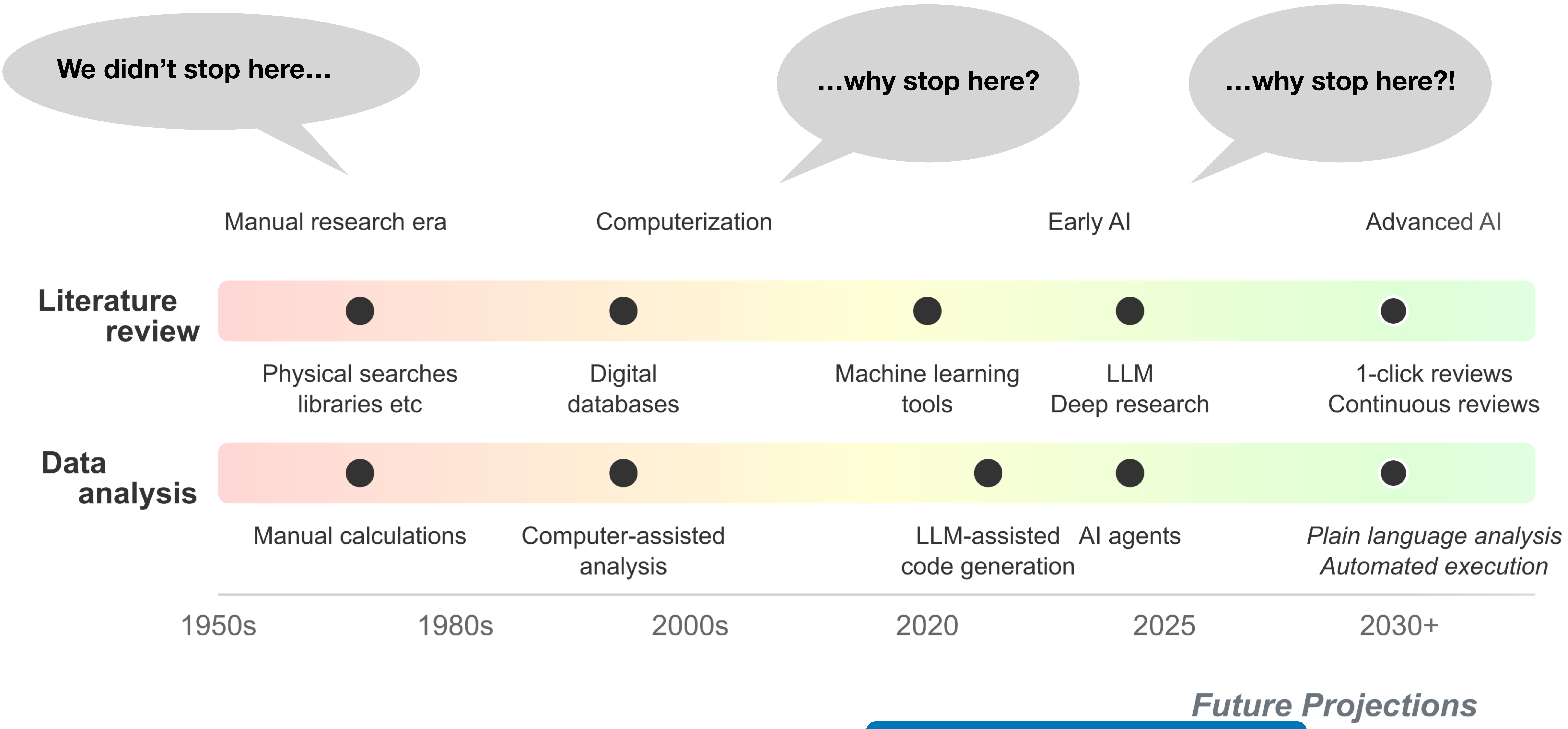
1. **Quantitative research is mostly cognitive work, done by….**
   - Humans…
   - AI…

2. **If AI helps research, it's 'the future'**
   - Humans + AI

Goal of research?

- **Expand knowledge**
  - "…work undertaken in order to increase the stock of knowledge" (Frascati)
- **Tools change, goal doesn't**
  - Humans + paper
  - Humans + computers
  - Humans + computer + some AI
  - Humans + computer + lots of AI

Key criteria: does it contribute to new knowledge?



We can research tools too!

Bann et al, arxiv / IJE 2025

Is this optimal?
How much is repetitive?

Task	Time estimate, now
1. Review Literature	3-12 months
2. Locate data	0.5 months
3. Analyse data	3-6 months
4. Interpret + write-up	3-6 months

GPT: General purpose technology: can be used for anything!
Tools… assistants…. agents

Almost half of US researchers' time goes on admin

BY REBECCA TRAGER | 9 SEPTEMBER 2014

Of more than 13,000 principal investigators (PIs) that responded to this 2012 workload survey, they said that they spent, on average, 42% of their time on admin, rather than actual research. The FDP, an association o

FDP, 2012-18

Why can't Epidemiology be automated (yet)?

David Bann, Ed Lowther, Liam Wright, Yevgeniya Kovalchuk

https://github.com/edlowther/automated-epidemiology

Bann et al, arxiv / IJE 2025

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Systematic reviews

BMJ Open **Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry**

Rohit Borah,[1,2] Andrew W Brown,[2,3] Patrice L Capers,[2,3] Kathryn A Kaiser[2,3]

Results: The mean estimated time to complete the project and publish the review was 67.3 weeks (IQR=42). The number of studies found in the literature

Month	Activity
1 – 2	Preparation of protocol.
3 – 8	Searches for published and unpublished studies.
2 – 3	Pilot test of eligibility criteria.
3 – 8	Inclusion assessments.
3	Pilot test of 'Risk of bias' assessment.
3 – 10	Validity assessments.
3	Pilot test of data collection.
3 – 10	Data collection.
3 – 10	Data entry.
5 – 11	Follow up of missing information.
8 – 10	Analysis.
1 – 11	Preparation of review report.
12 –	Keeping the review up-to-date.

Cochrane handbook

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Systematic reviews

Error rates of human reviewers during abstract screening in systematic reviews

Zhen Wang, Tarek Nayfeh, Jennifer Tetzlaff, Peter O'Blenis, Mohammad Hassan Murad

(CI): 2.38% to 8.58%). After abstract screening, the total error rate (false inclusion and false exclusion) was 10.76% (95% CI: 7.43% to 14.09%).

Artificial intelligence in systematic reviews: promising when appropriately used

Sanne H B van Dijk [1,2], Marjolein G J Brusse-Keizer [1,3], Charlotte C Bucsán [2,4], Job van der Palen [3,4], Carine J M Doggen [1,5], Anke Lenferink [1,2,5,✉]

Cochrane

Opportunities and Challenges for Data Extraction with a Large Language Model

Data extraction in evidence synthesis is labour-intensive, costly, and prone to errors. The use of large language models (LLMs) presents a promising approach for AI-assisted data extraction, potentially enhancing both efficiency and accuracy.

Our vision

A world where health decisions are based on timely, trusted and relevant evidence.

Automation of Systematic Reviews with Large Language Models

specificity) and data extraction (*otto-SR*: 93.1% accuracy; human: 79.7% accuracy). Using *otto-SR*, we reproduced and updated an entire issue of Cochrane reviews (n=12) in two days, representing approximately 12 work-years of traditional systematic review work. Across

Cao et al MedRxiv 2025

Cochrane 2025

Task
1. Review Literature
2. Locate data sources
3. Analyse data

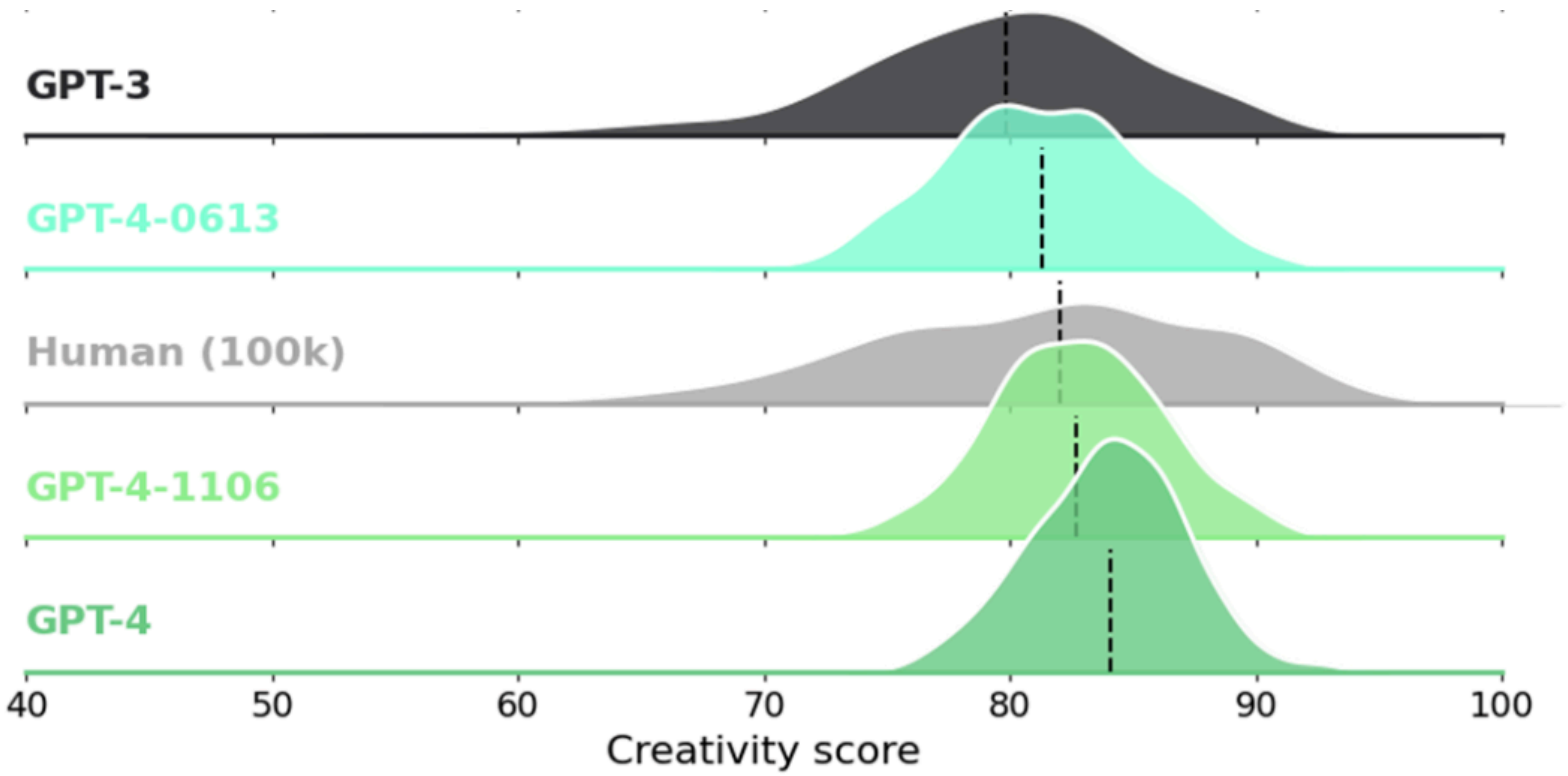
Ad hoc reviews

- Now
  - Many already using AI Google Scholar
  - Unlike PubMed, indexes social science + grey literature
- Future
  - ⬆️ access to (millions of) papers + ⬇️ hallucinations
  - Plain language Qs A new way to search Try Scholar Labs PubMed Elicit
- Humans + AI
  - Bandwidth freed for higher level tasks
  - More ambitious reviews (eg, across disciplines/designs), continual; more informed papers

Task
1. Review Literature
2. Locate data sources
3. Analyse data

AI for forming research questions/hypothesis

- Stochastic parrots?
- Creativity = cognitive task!
  - Their creativity is ~empirically testable
- Even if flawed / limited still useful (instant, cheap)
  - AI + humans



Bellemare-Pepin et al 2025

https://creativitybenchmark.ai

Task
1. Review Literature
2. Locate data sources
3. Analyse data

AI for forming research questions/hypothesis + testing them

Theoretical Physics with Generative AI

Stephen D.H. Hsu

Dec 2025

publication in Physics Letters B after peer review. *Remarkably, the main idea in the paper originated de novo from GPT-5.* GPT-5, Gemini, and Qwen-Max were used extensively to perform calculations, find errors, and generate the finished paper.

Cell
A Cell Press journal

AI mirrors experimental science to uncover a mechanism of gene transfer crucial to bacterial evolution

José R. Penadés [1,2,3,4,7,8] · Juraj Gottweis [5,7] · Lingchen He [1,2] · … · Vivek Natarajan [5] · Alan Karthikesalingam [5] · Tiago R.D. Costa [2,3,6] … Show more

Nov 2025

Highlights

- AI co-scientist predicted a complex gene transfer mechanism before its publication
- Top AI-generated hypotheses opened new research directions
- AI bypassed human bias to propose overlooked biological possibilities
- Benchmarking showed AI co-scientist outperformed other LLMs on this task

The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery

Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, David Ha

Robin: A multi-agent system for automating scientific discovery

Ali Essam Ghareeb, Benjamin Chang, Ludovico Mitchener, Angela Yiu, Caralyn J. Szostkiewicz, Jon M. Laurent, Muhammed T. Razzak, Andrew D. White, Michaela M. Hinks, Samuel G. Rodriques

Scientific discovery is driven by the iterative process of background research, hypothesis generation, experimentation, and data analysis. Despite recent advancements in applying artificial intelligence to scientific discovery, no system has yet automated all of these stages in a single workflow. Here, we introduce Robin, the first multi-agent system capable of fully automating the key intellectual steps of the scientific process. By

Early science acceleration experiments with GPT-5

Sébastien Bubeck[1], Christian Coester[2], Ronen Eldan[1], Timothy Gowers[3], Yin Tat Lee[1], Alexandru Lupsasca[1,4], Mehtaab Sawhney[5], Robert Scherrer[4], Mark Sellke[1,6], Brian K. Spears[7], Derya Unutmaz[8], Kevin Weil[1], Steven Yin[1], Nikita Zhivotovskiy[9]

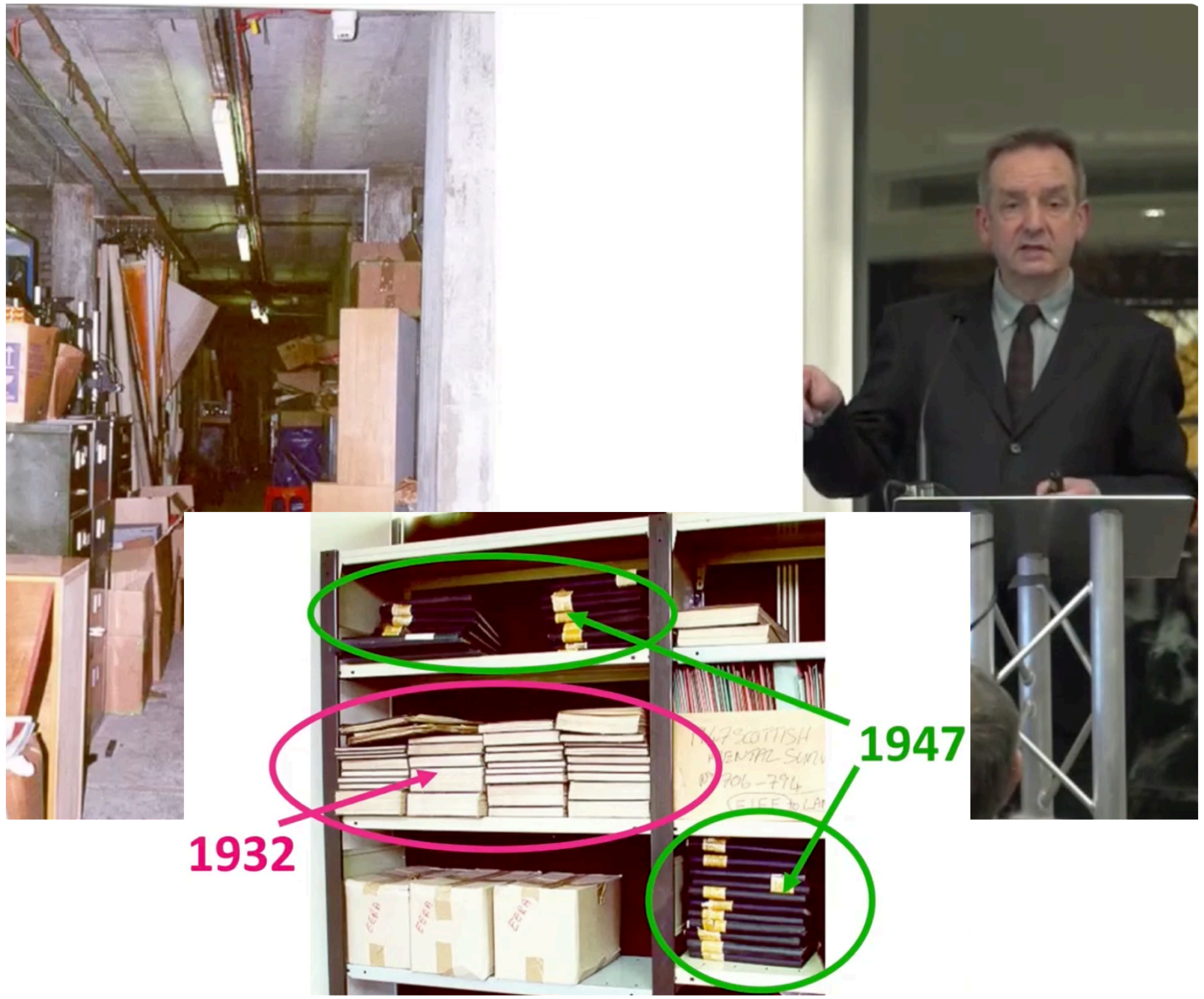
Task
1. Review Literature
2. Locate data sources
3. Analyse data

1. Unlocking historic (new) data

2. With collected data

   - Manual variable work -> AI assisted (e.g. Harmony for harmonisation; pdf annotation, metadata curation etc)

3. One research question -> all studies



"Found end of corridor in 1997"

The **Lothian Birth Cohort** 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond
IJ Deary, AJ Gow, MD Taylor, J Corley, C Brett… - BMC geriatrics, 2007 - Springer
… A total of 1091 participants make up the **Lothian Birth Cohort** 1936. They undertook: a medical interview and examination; physical fitness testing; extensive cognitive testing (reasoning, …
☆ Save 99 Cite Cited by 563 Related articles All 27 versions

youtube.com/watch?v=0rxgundtPTI

Task
1. Review Literature
2. Locate data sources
3. Analyse data

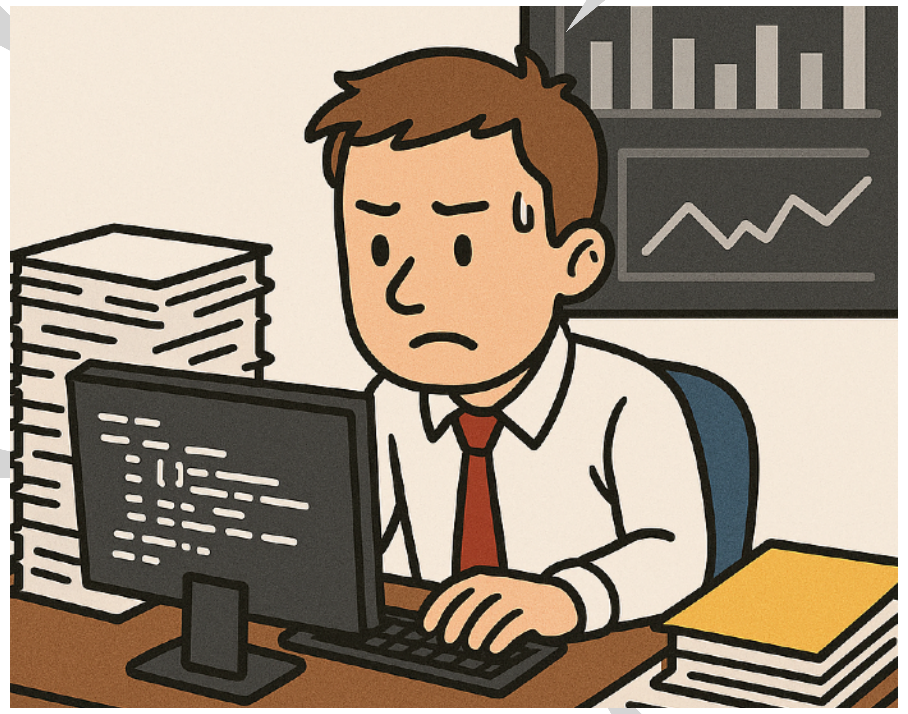
Coding: humans only

```
#load the libaries
library(gamlss)
library(sitar)
library(tidyverse)
library(haven)
library(psych) |

#load all data
dat <- read_dta('cleaned_datafor_r.dta')

#bmi
bmi46<- subset(dat, select = c("bmi46", "sex", "fscb" , "pa42b"))

#bmi
describeBy(bmi46, group="sex" ,digits=2)
describeBy(bmi46, group="fscb" ,digits=2)
describeBy(bmi46, group="pa42b" ,digits=2)

#output estimates
bmi46sex_no <- gamlss(bmi46 ~ sex  ,
                      sigma.formula = ~sex      ,
                      family = NO (mu.link = log), data = na.omit(bmi46) )
```

Error!?

I've spent years typing “ (% etc

check documentation



check forum

***2 hours later* I used ` instead of ‘ 😂**

It works! But across a scientific lifetime, how much is wasted?

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Coding: humans + AI autocomplete

TS Calculator.ts > Calculator

```typescript
class Calculator {

    public add(num1: number, num2: number): number {
        return num1 + num2;
    }

    public subtract(num1: number, num2: number): number {
        return num1 - num2;
    }
}
```

Visual Studio Code

https://code.visualstudio.com

Independent report

AI coding assistant trial: UK public sector findings report

Published 12 September 2025

GDS ran a trial of AI coding assistants (AICAs) across government from November 2024 to February 2025. A total of 2,500 licences were made available across central government organisations.

- 67% of respondents reporting a reduction in time spent searching for information or examples
- 65% reporting faster task completion
- 56% reporting more efficient problem solving

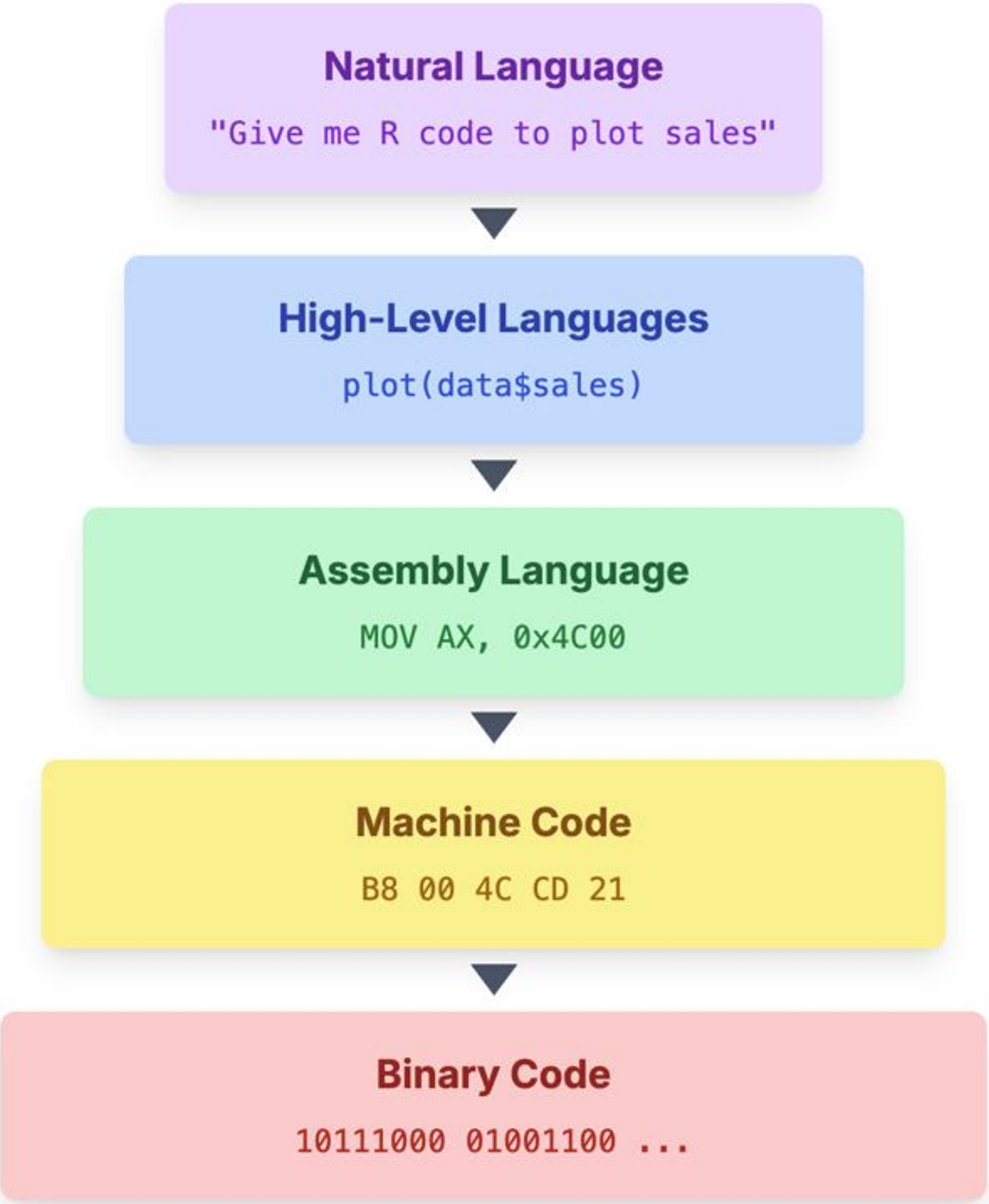
On average, users reported time savings of 56 minutes per working day, which equates to approximately 28 working days saved per user annually, based on a standard working calendar. Of this, 24 minutes per day were associated with creation of code or analysis

User sentiment was positive, with 58% expressing they would not want to return to their pre-AICA working conditions.

UK Gov Digital Services Sep 2025

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Coding: humans + AI: plain language -> code



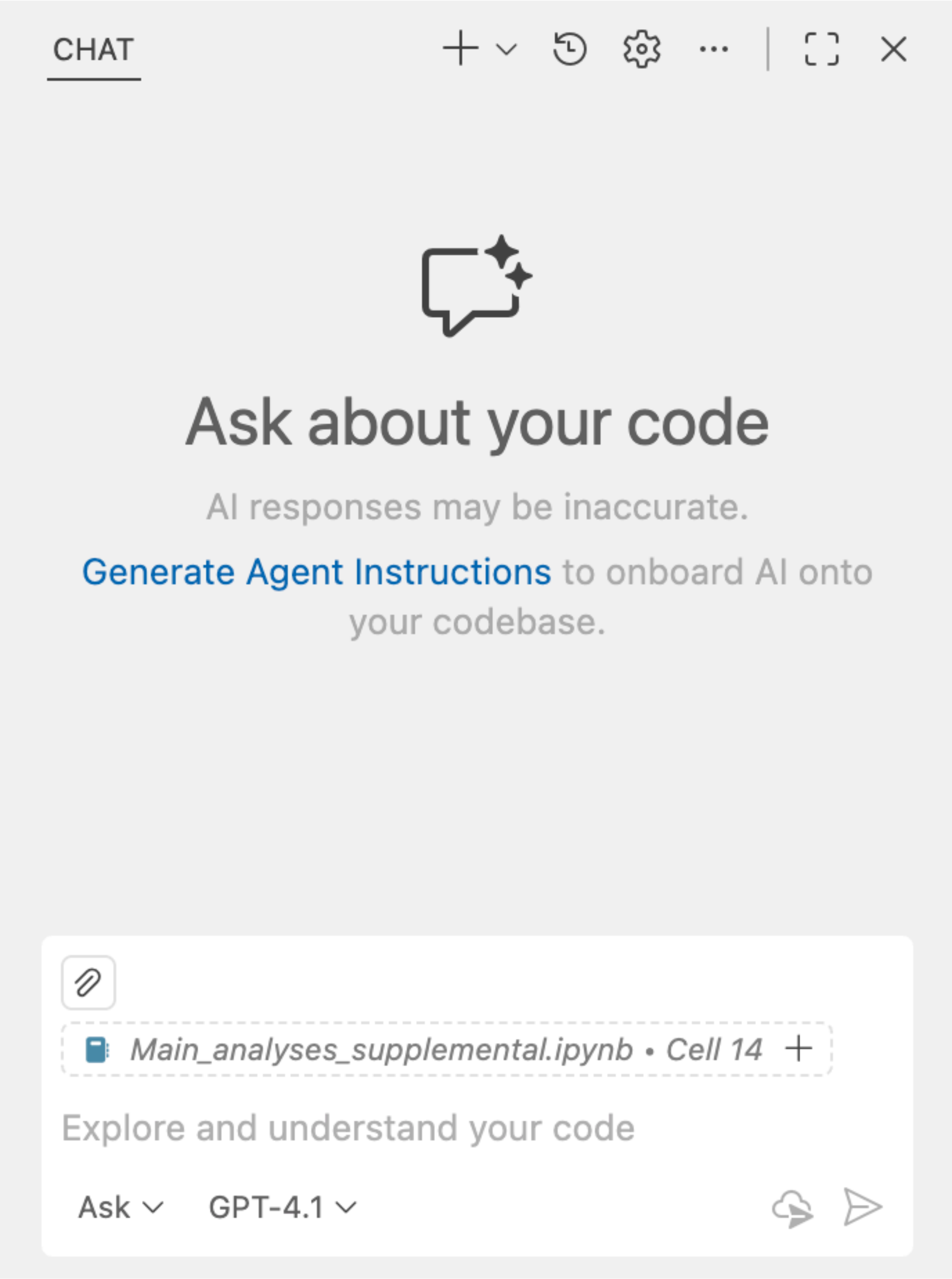
Use any!

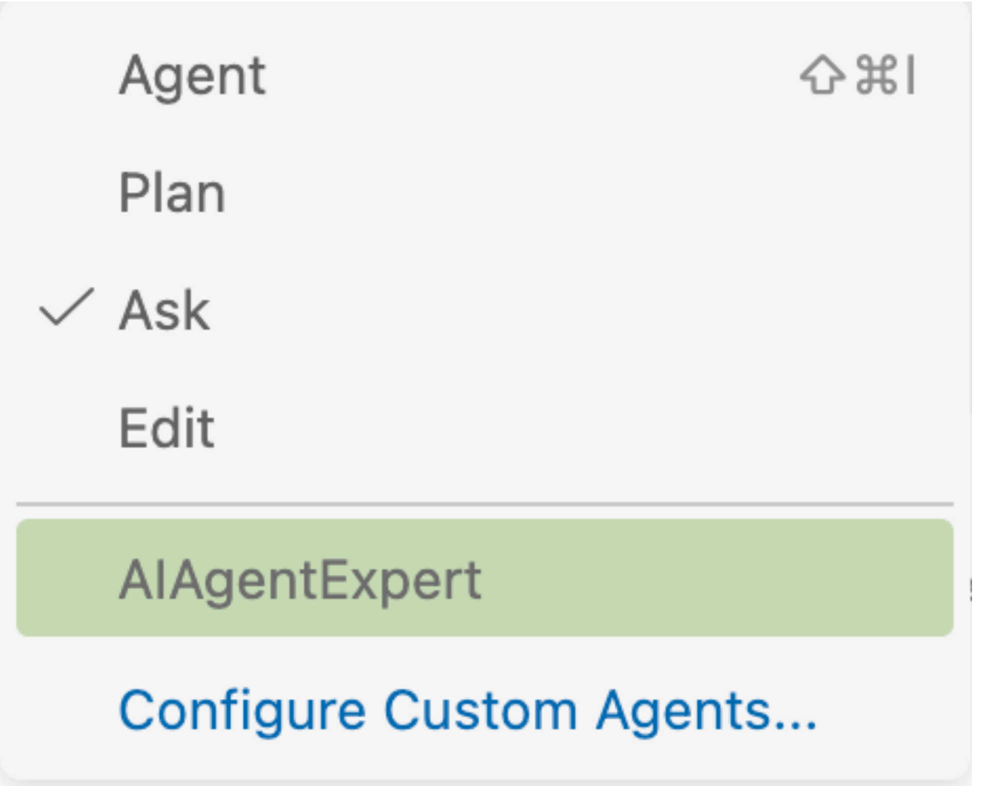
Bann and Wright, 2025

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Coding: humans + AI: plain language -> code

```python
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import numpy as np
from pathlib import Path

# --- Extract the country rows from your table1 ---
# (they start after the 'Countries' row)
country_start = row_names.index('Countries') + 1
country_rows = row_names[country_start:]
```





Visual Studio Code

https://code.visualstudio.com

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Coding: humans + AI agents

https://julius.ai/

D

Mental health D... 4.66 MB
spreadsheet

I uploaded the following files: Mental health Depression disorder Data.csv

J

Thank you for uploading the file "Mental health Depression disorder Data.csv". I'll take a look at the data and provide you with an overview of its contents. Let's start by examining the structure and the first few rows of the dataset.

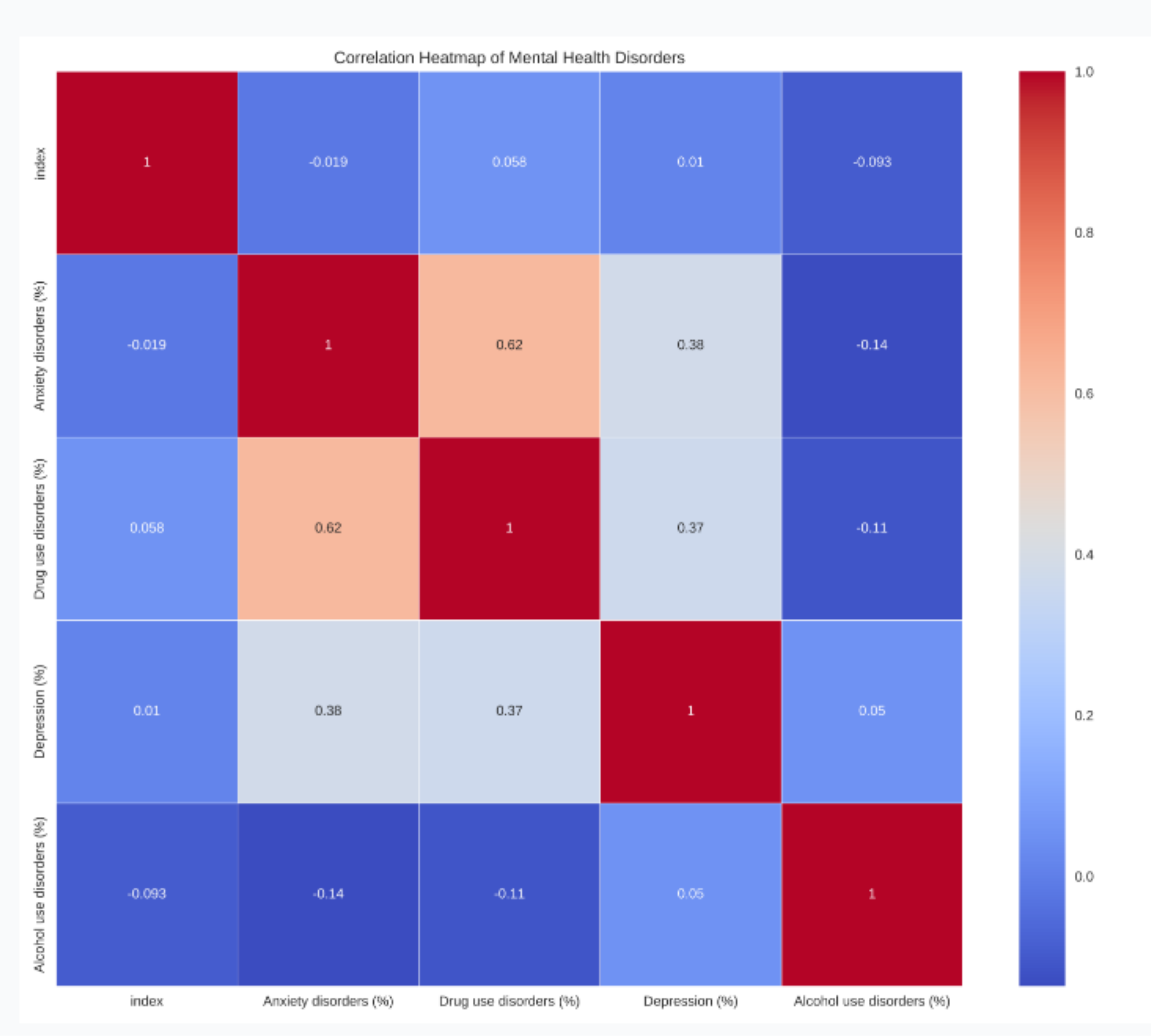
</> Show Code

I've analyzed the Mental health Depression disorder Data.csv file you uploaded. Let me provide you with an overview of the dataset and some initial insights.

First, let's look at the structure of the data:

(Optional) click on columns or rows to focus on specific data Export

	index	Entity	Code	Year	Schizophrenia (%)	Bipolar disorder (%)	Eating disorders (%)	Anxiety disorders (%)	Drug use disorders (%)	Depression (%)	Alcohol use disorders (%)
0	0	Afghanistan	AFG	1990	0.16056	0.697779	0.101855	4.82883	1.677082	4.071831	0.672404
1	1	Afghanistan	AFG	1991	0.160312	0.697961	0.099313	4.82974	1.684746	4.079531	0.671768



Task
1. Review Literature
2. Locate data sources
3. Analyse data

Coding: humans + AI agents

Associations between Birth Weight and BMI

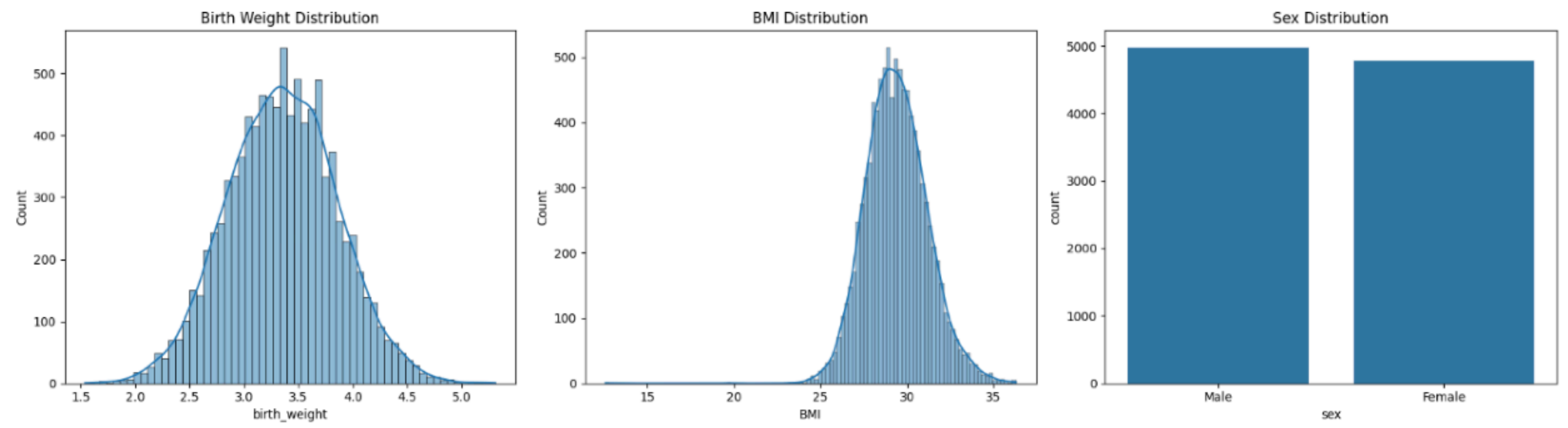
Notebook generated using API access to `gpt-4.1-2025-04-14` via the `data-analysis-crow` framework on 23 June 2025.

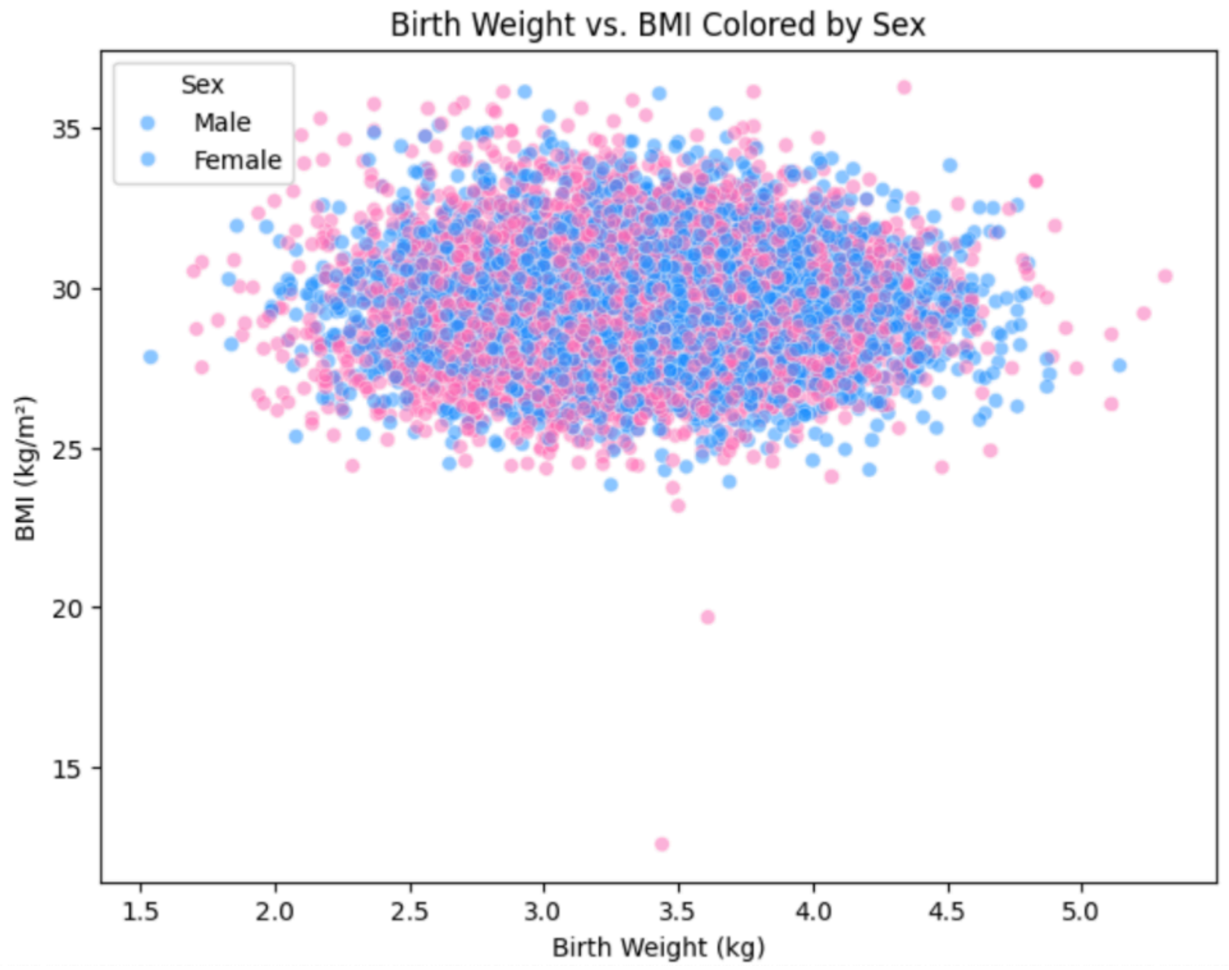
Prompt:

The attached dataset contains information about participants in a longitudinal study. Investigate whether there is an association between birth weight and BMI using linear regression anlayses adjusted for sex.

```
# Step 1: Load Data and Perform Descriptive Statistics
#
# Analysis Planning:
#
# - Relevant file: 'cohort-data.csv' looks like the main dataset. We will load this file with pandas.
# - We will inspect the file's shape and column names with .head() and .info() to understand the structure and
# - Use .describe() for numerical summaries (mean, std, min, max, etc.) for birth weight and BMI.
# - Check for missing data with .isnull().sum().
# - If there are missing/strange values (e.g., negative birth weights or BMI), we will note, and consider handl
# - Plan to present summaries as printed outputs and initial visual inspection if columns allow.
# - Key statistics: distribution and range of birth weight, BMI; balance of sexes.
# - Data quality issues: unexpected values, missing data, unexpected data types; will address as found.

import pandas as pd

# Load the data file
cohort = pd.read_csv('cohort-data.csv')

# Examine the structure and first few rows
display(cohort.head())
print(cohort.info())
print(cohort.describe(include='all'))
# Check for missing data
print('Missing values per column:')
print(cohort.isnull().sum())
```

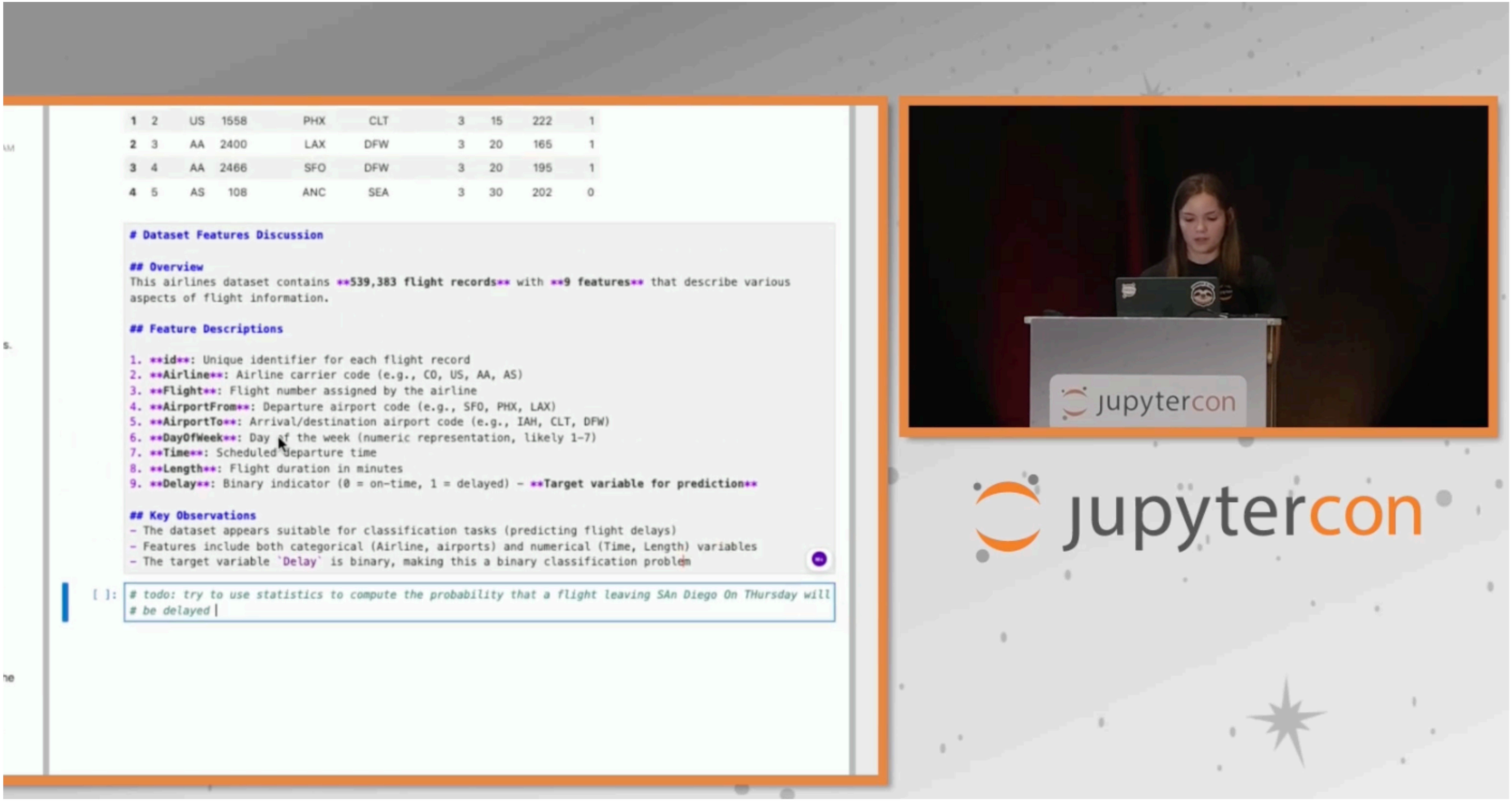
	ID	sex	education	birth_weight	adult_body_weight	height	salary	depression_symptoms
0	1	0	4	3.29	91.882406	1.791942	70603.442728	20.433706
1	2	1	3	3.48	81.210334	1.619915	53981.525954	13.147941





Human baseline (code sharing) 2% papers (Hamilton et al, 2023)

Bann et al, arxiv / IJE 2025



Nov 2025. Zach Sailer & Abigayle Mercer youtube.com/watch?v=l5DhqN3su0o

Day-to-day tasks ...	Repetitive admin tasks Inbox management	Scientific thinking Paper / infrastructure design
Literature Reviews	Manual screening Data extraction	Designing AI-assisted reviews
Analysing data	Line-by-line writing code Single study / designs	Orchestrating agents to code Multiple studies / study designs

Current time allocation → **AI Augmented time allocation**

Bann et al, arxiv / IJE 2025

Centre for Longitudinal Studies (CLS)

Code repositories from CLS

11 followers · United Kingdom · https://cls.ucl.ac.uk/ · @clscohorts · company/ucl-centre-for-longitudin...

🔗 Harmonisation / Derived variable code

Core datasets (useful variables across domains e.g., health, socioeconomic, demographic)

- MCS Core
- Next Steps Core
- BCS70 + NCDS forthcoming in 2026

Cross-cohort harmonisation:

- Fertility Histories
- Smoking
- Asthma

Code for single cohorts:

- Time use data in MCS

🧬 Genomic Projects

- CLS Polygenic Index Repository - Tools and scripts for calculating polygenic scores in longitudinal studies

💻 Data Management Code

Tools and resources for longitudinal data management:

- See CLS Data Management Resources - code for common data management tasks

https://github.com/cls-data

CLS PGI Pipeline (v1.0)

Contributors

Tim Morris - Designed the pipline and led development and implementation.
Gemma Shireby - Designed the pipline and led curation of genotype data.
Georg Otto - Contributed to pipeline design including feature suggestions.
David Bann - Provided code annotations and documentation support.
Liam Wright - Checked PGIs and designed PGI visualations.

https://github.com/CLS-Data/CLS_PGI_repository

CLS Data Handling Guide

Search CLS Data Handling Guide

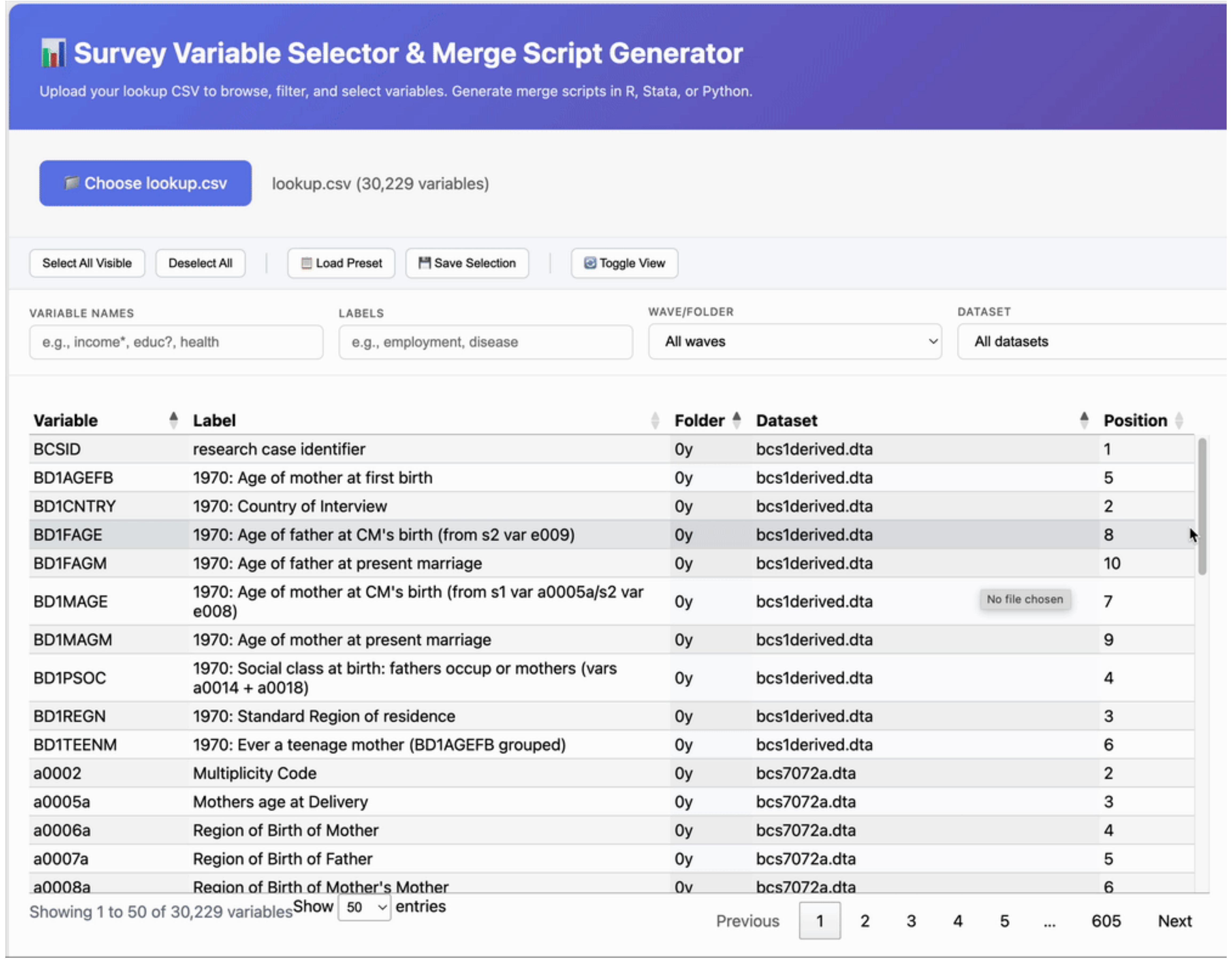
- Introduction
- Downloading Documentation
- NCDS
- BCS70
- Next Steps
- MCS
  - Creating a Simple Folder Structure
  - Data Discovery
  - Data Structures
  - Working with the Household Grid
  - Combining Data Across Sweeps
  - Combining Data Within A Sweep
  - Reshaping Data from Long to Wide (or Wide to Long)

This section presents code to clean and handle data from the Millennium Cohort Study (MCS). The MCS has the most complex data structures of CLS's cohorts, so we provide an introduction to the data as necessary background.

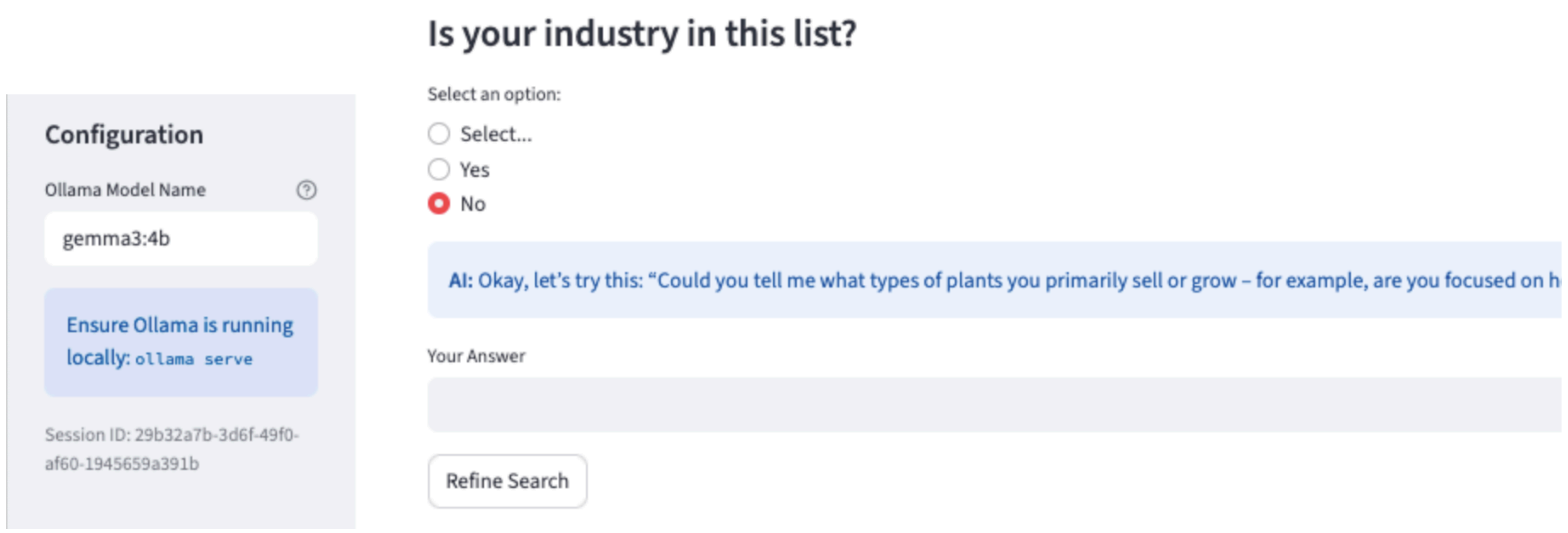
TABLE OF CONTENTS

https://cls-data.github.io/

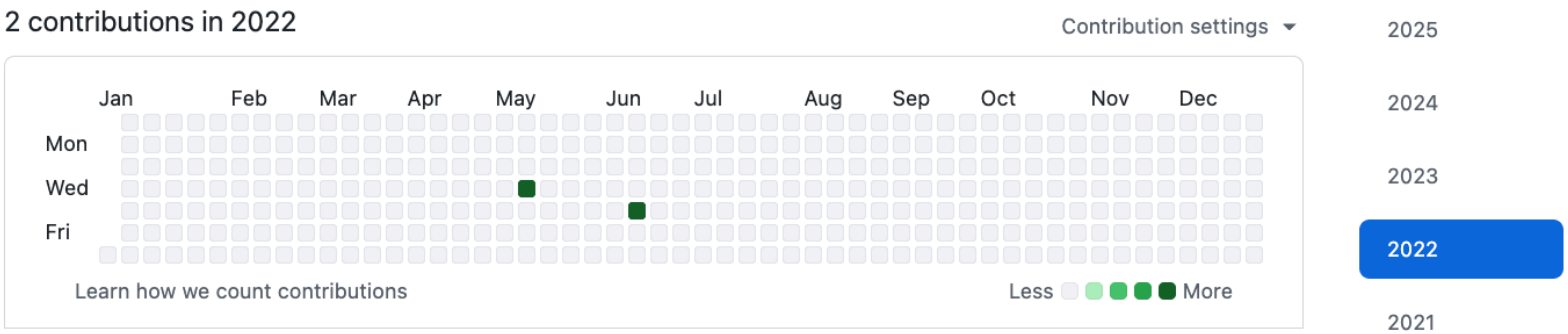
NCRM hackathon Nov 2025

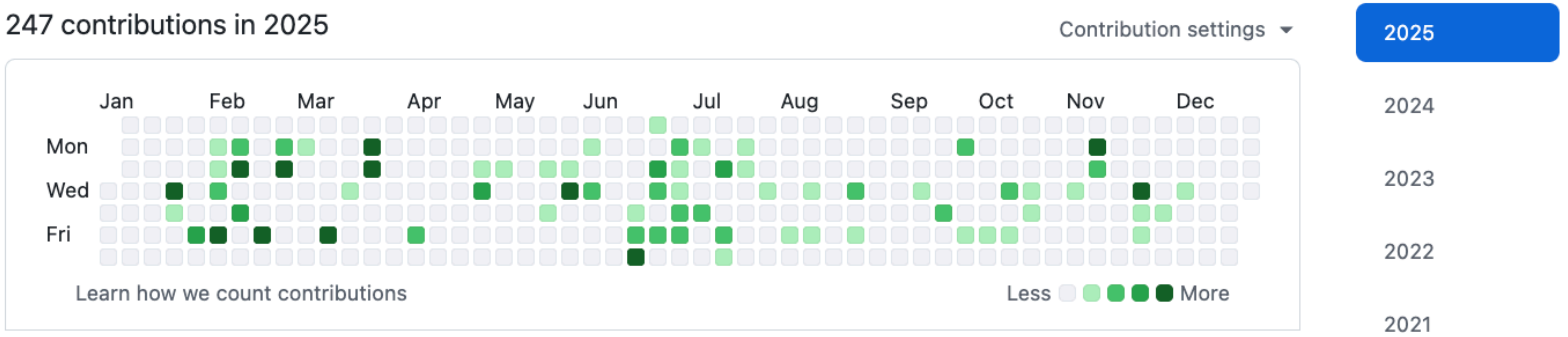


https://github.com/dbann/surveymerge



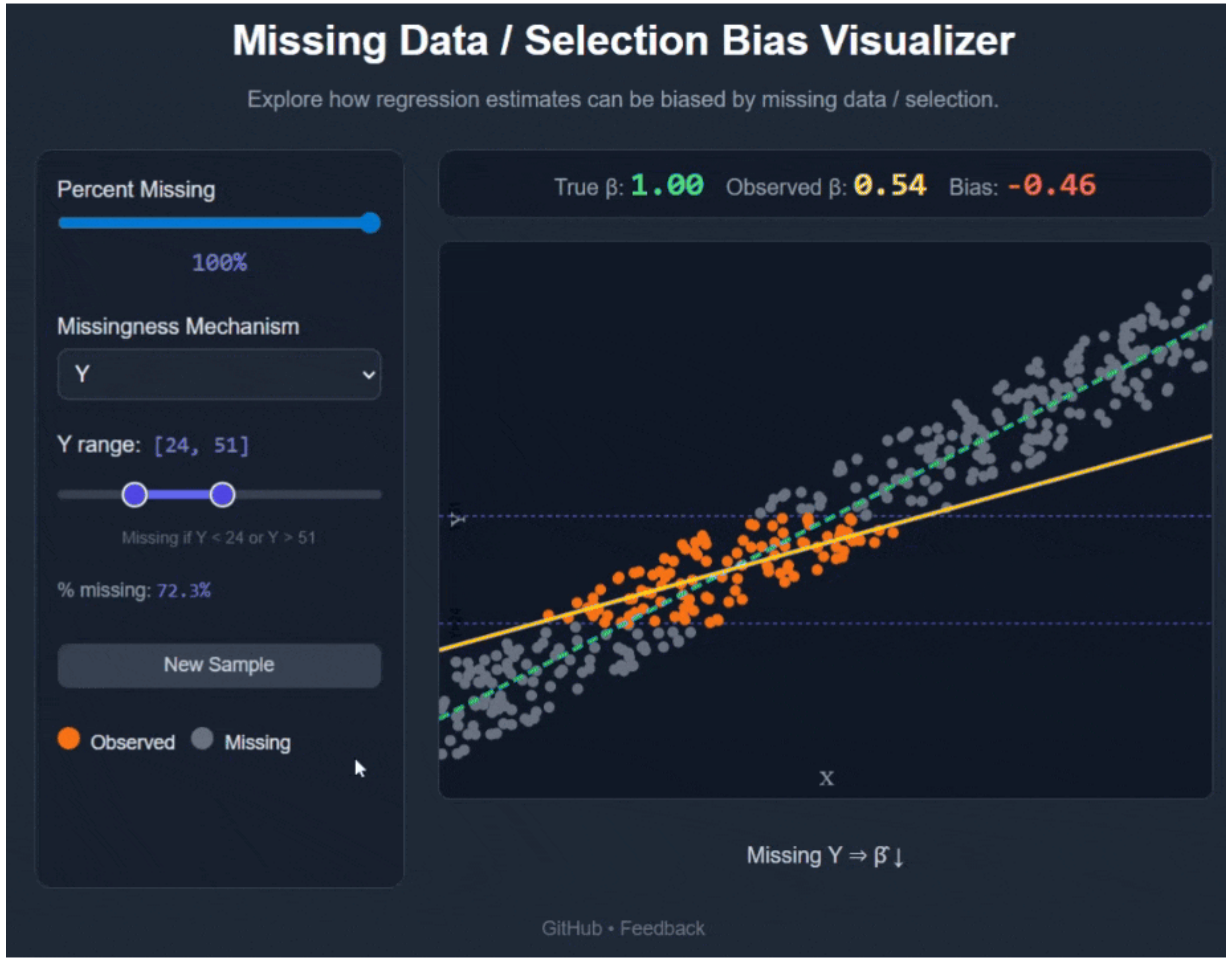
Liam Wright liam.wright@ucl.ac.uk



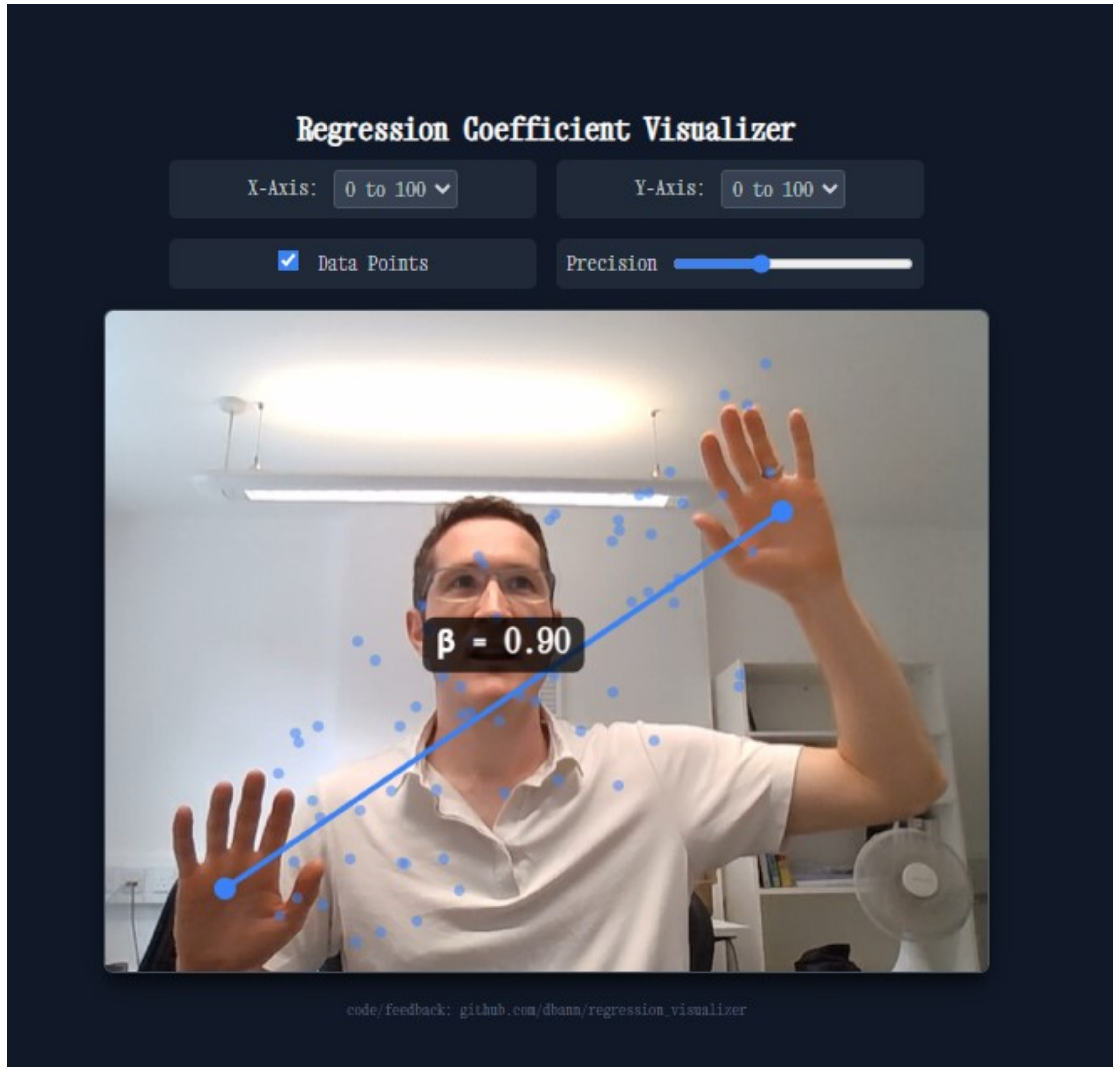


https://github.com/dbann

Teaching, making visualisers



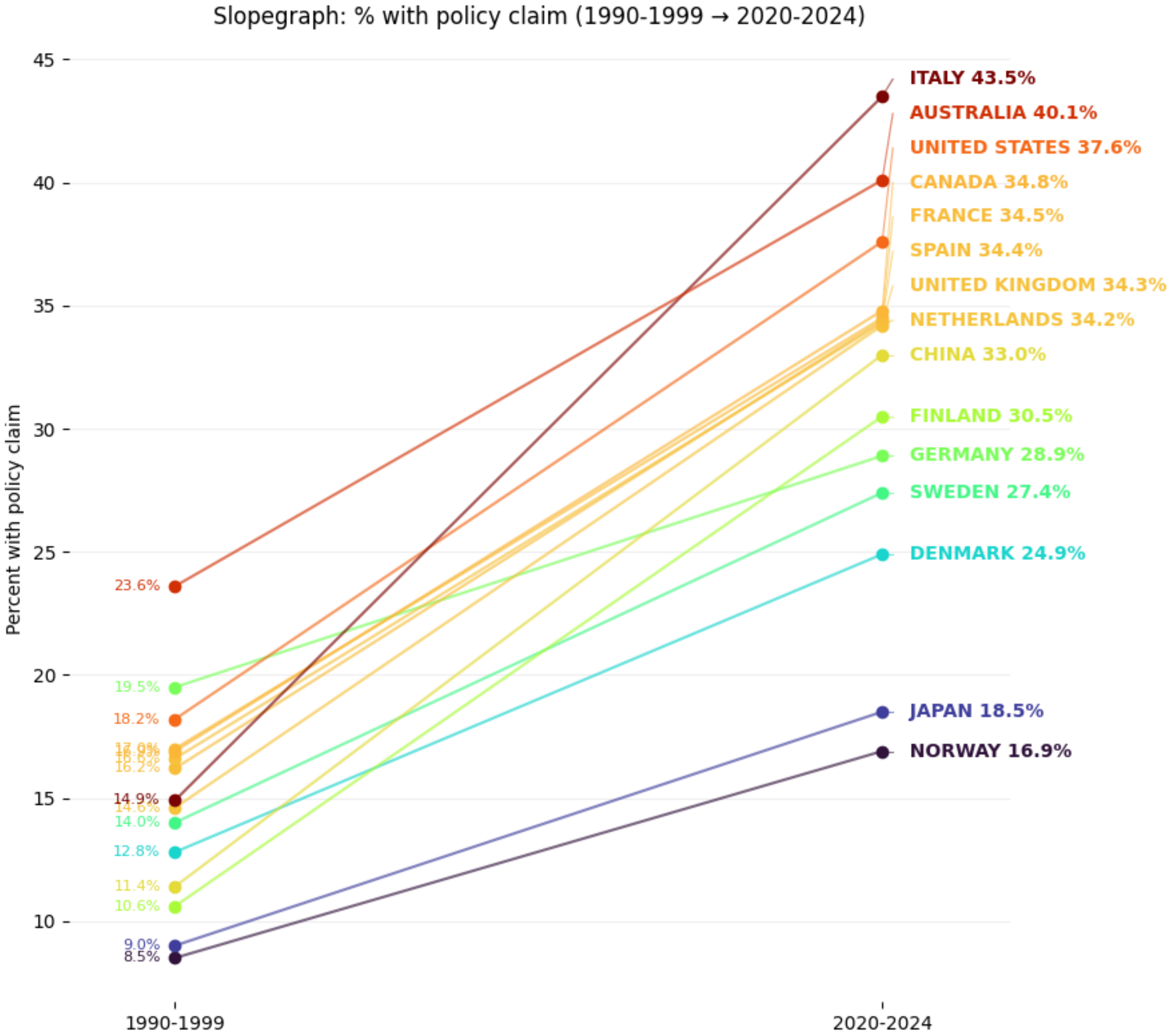
https://github.com/dbann/selection



https://github.com/dbann/regression_visualizer

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Example of a human + AI paper



- **n=44k abstracts, $3**
- **Concordance in classification (kappa)**
  - **Human-human = 0.83**
  - **LLM (Deepseek 3,1) -human = 0.73-0.84**
- **LLMs used to learn**
  - **Python, APIs, Git (Vs Code)**
- **Humans + LLMs**
  - **New research, new field (metascience)**
  - **New data / syntax resource**
  - **Help early career researcher (Mengyao Wang)**

Bann, Wang et al, mexRiv, 2025

Science Funding Cost Calculator

Explore the hidden costs of the grant application process. How much time and money is spent on unfunded proposals?

⚙️ Parameters

Words requested by funder: 20,000

Applications funded (%): 11%

2nd Stage Selection Rate (%): 30%

Scientists' Hourly Rate (£)

50

Total Funding Applications

1000

Writing Speed

150 words/hr | 250 words/hr

Currency: £

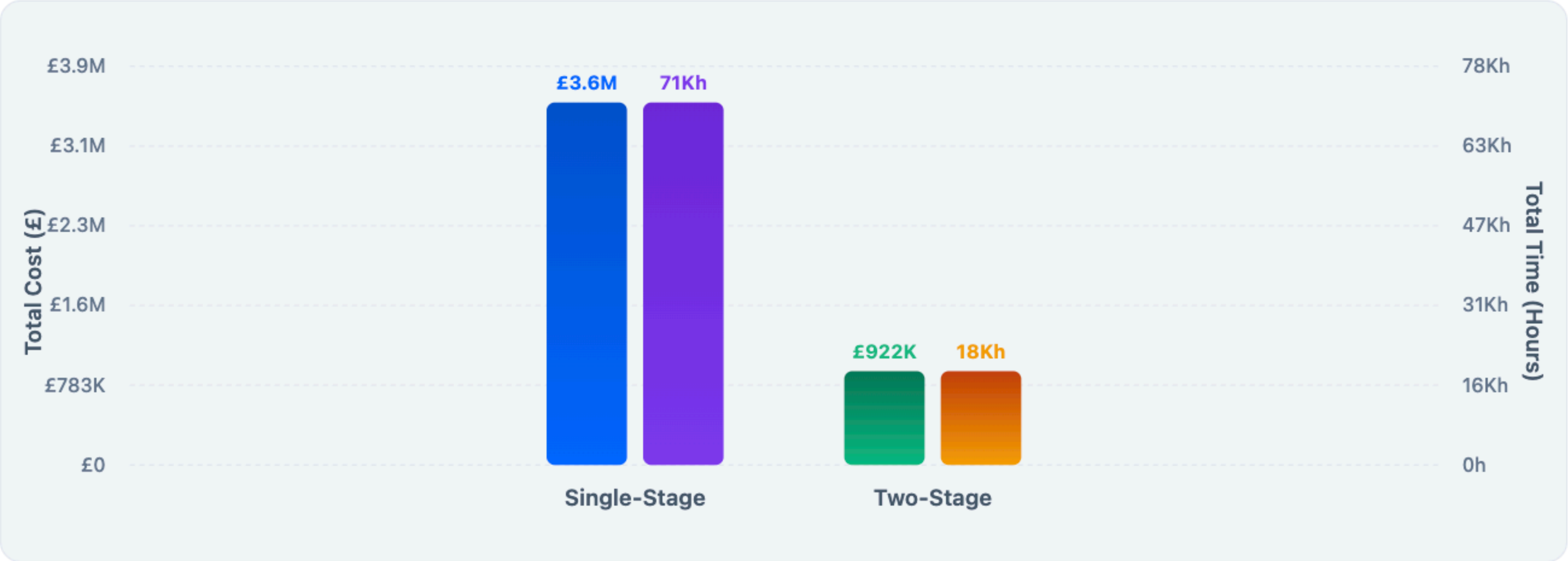
💰 **£3,560,000**
Total Wasted Cost

⏰ **71,200**
Hours Wasted

📊 **11%**
Success Rate

Cost Breakdown and Comparison with 2-Stage System

Comparing single-stage vs. two-stage review processes



Potential Cost Savings

With 2-stage system (1000 words first stage, 30% proceeding):

£2,638,000

Potential Time Savings

Reduced application writing burden:

52,760h

https://dbann.github.io/grantreview

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Does good > bad? 1) Our data are private

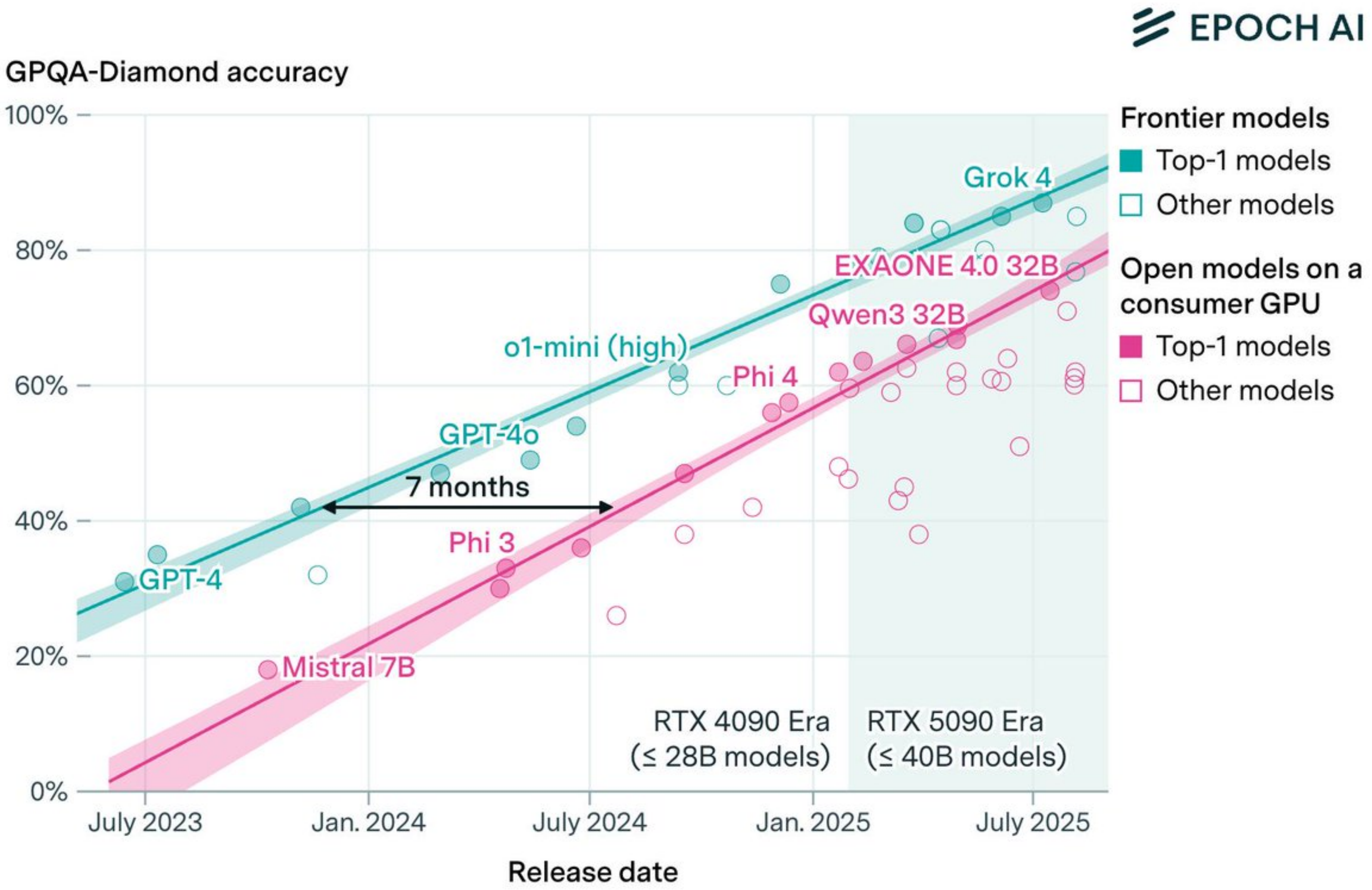
Use open-weight LLMs!

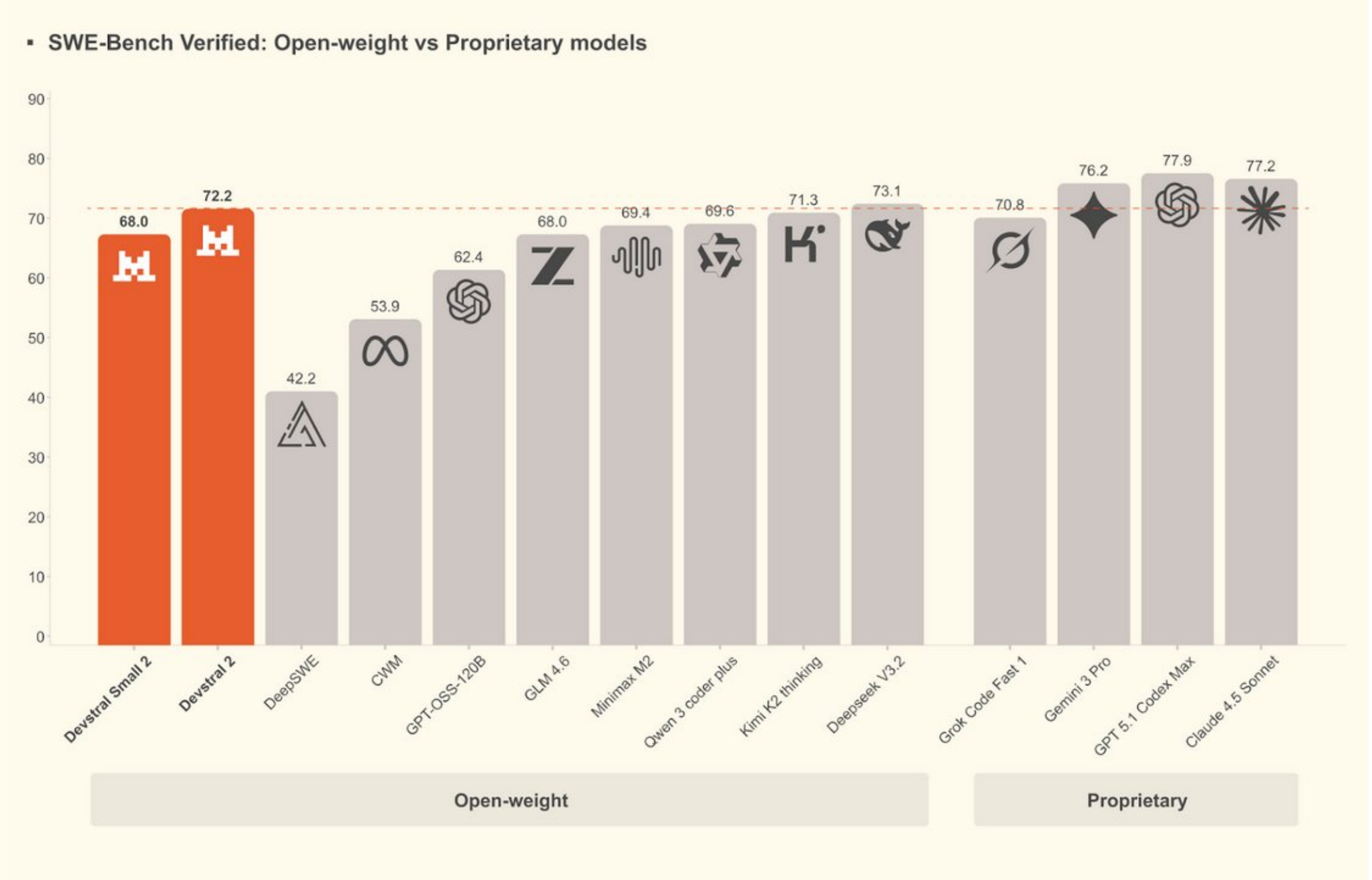
🔒 Closed weight LLMs	Open weight variant	🌐 Open weight LLMs
ChatGPT (OpenAI🇺🇸)	ChatGPT-OSS (OpenAI🇺🇸)	Deepseek (Deepseek🇨🇳)
Gemini (Google🇺🇸)	Gemma (Google🇺🇸)	Kimi k2 (Moonshot AI🇨🇳)
Claude (Anthropic🇺🇸)	n/a	Mistral (Mistral AI🇫🇷)
Grok (xAI🇺🇸)	Earlier Grok releases (xAI🇺🇸)	Qwen (Alibaba Qwen🇨🇳)
		Llama (Meta🇺🇸)
		Ernie (Baidu🇨🇳)
		GLM (Zhipu AI🇨🇳)

Bann and Wright, 2025

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Does good > bad? 1) Our data are private





https://mistral.ai

MLX qwen/qwen3-coder-30b

Eject

R Ridgeline Plot Code

Here's R code to create a ridgeline plot using the `ggridges` package:

r

Copy

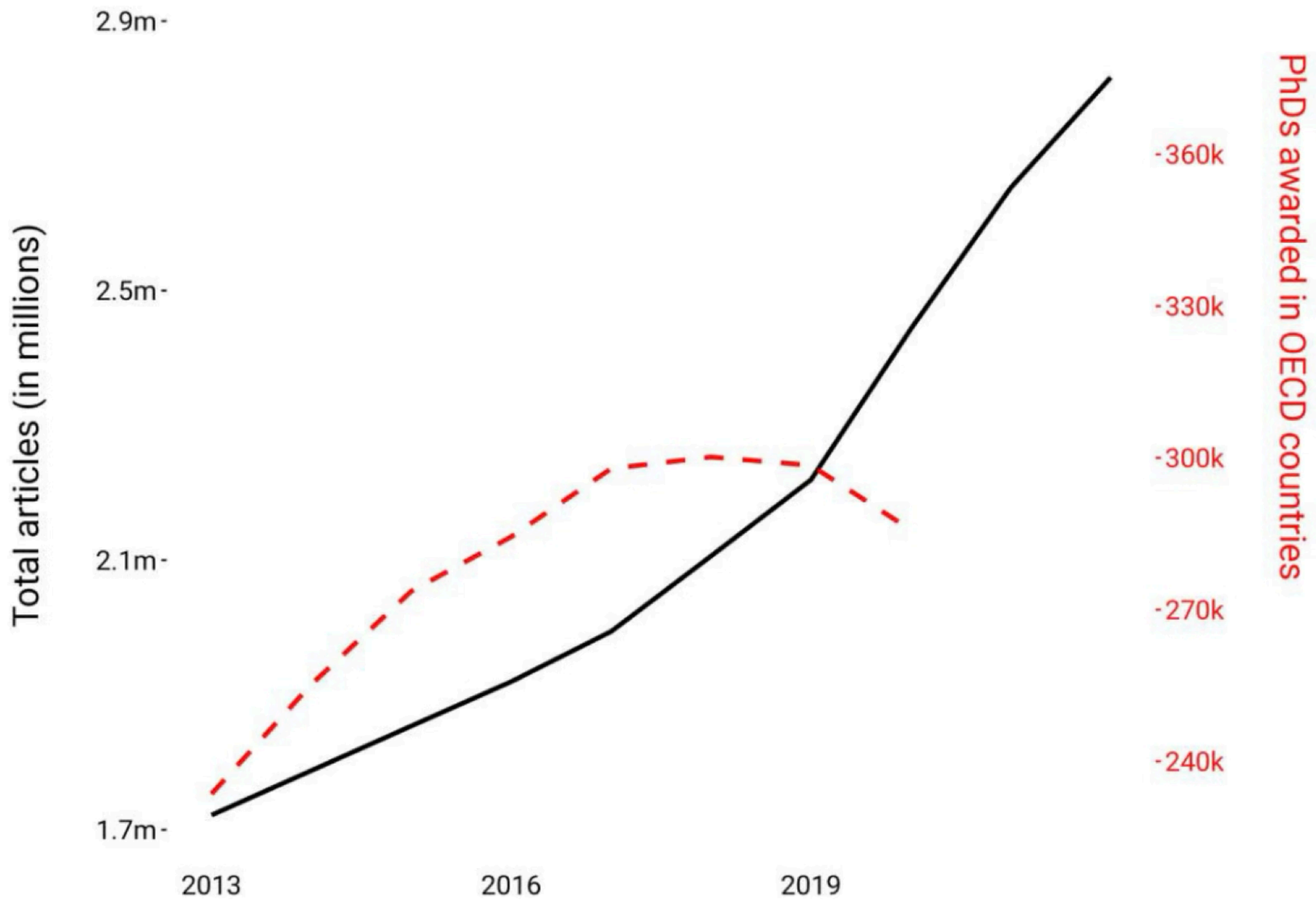
```r
# Install and load required packages
install.packages(c("ggridges", "ggplot2", "dplyr"))
library(ggridges)
library(ggplot2)
library(dplyr)

# Example 1: Using built-in dataset
# Create a sample dataset
set.seed(123)
data <- data.frame(
  value = c(rnorm(100, 20, 5), rnorm(100, 30, 7), rnorm(100, 40, 6)),
  group = rep(c("Group A", "Group B", "Group C"), each = 100),
  time = rep(1:100, 3)
)

# Basic ridgeline plot
ggplot(data, aes(x = value, y = group, fill = group)) +
  geom_density_ridges(alpha = 0.7) +
  theme_minimal() +
  labs(title = "Ridgeline Plot Example",
       x = "Value",
       y = "Group")

# Example 2: Using mtcars dataset
# Prepare data for ridgeline plot
mtcars_grouped <- mtcars %>%
  mutate(cyl = factor(cyl)) %>%
  select(mpg, cyl)
```

Send a message to the model...

https://lmstudio.ai + see also https://ollama.com/

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Does good > bad? 2) AI slop



Hanson et al Quan Sci Stud. 2024

Human problem; how to incentive quality > quantity, globally?

Task
1. Review Literature
2. Locate data sources
3. Analyse data

Does good > bad? 3) Environmental impact

1) Do we have accurate (full lifecycle) estimates? 2) what's the comparison?

Measuring the environmental impact of delivering AI at Google Scale

Cooper Elsworth, Keguo Huang, David Patterson, Ian Schneider, Robert Sedivy, Savannah Goodman, Ben Townsend, Parthasarathy Ranganathan, Jeff Dean, Amin Vahdat, Ben Gomes, James Manyika

for the median Gemini Apps text prompt over one year. We identify that the median Gemini Apps text prompt uses less energy than watching nine seconds of television (0.24 Wh) and consumes the equivalent of five drops of water (0.26 mL). While these impacts are low compared to other daily activities, reducing the environmental

3) Is it worth it? Balance impact vs benefit (AGI, ASI -> science, health, education)

Duty to advance research….
Threats e.g., to funding, response rates, stalling discovery…

Conclusion

1. Quantitative research is mostly cognitive work

2. AI is becoming highly capable and useful

AI is the future - even if progress plateaus.

- Overdetermined given capability, influence on science, education, society
- Profound change, at infancy: we should adapt + contribute
- Understand impact, capabilities, responsible use + use AI to help learn
- Accelerate research: more efficient, reproducible, ambitious, enjoyable

Thanks (feedback):

- **Tom Berman**
- **Liam Wright**
- **Snehal Pinto Pereira**
- **Ben Weidmann**

- Centre for Longitudinal Studies, Social Research Institute, IOE / UCL
- ARC (Yevgeniya Kovalchuk, Ed Lowther, Mack Nixon, Niklas Loynes)
- Patrick Sturgis for slide 6 idea
- NCRM + this event (Gabriele Durrant, Penny White; Peter Tennant)

NCRM NATIONAL CENTRE FOR RESEARCH METHODS

Funders: ESRC, UKRI Digital Research Infrastructure Programme

david.bann@ucl.ac.uk

Conflicts of interest: none

In case of interest

Generative AI Tools for Quantitative Research: A Practical Guide
Presenter(s): David Bann and Liam Wright

Generative Artificial Intelligence (GenAI)—systems that produce text, code, images and more—has advanced rapidly in recent years. Large language models (LLMs), like ChatGPT, have surged in capability.

Below we outline practical guidance for incorporating GenAI models into quantitative research workflows, focusing on coding. This should be useful to quantitative social scientists, but also to health researchers or data scientists more broadly. 'Supporting Materials' section includes links to further reading and guides on creating web visualisation tools and using local LLMs.

You may have already accessed services like ChatGPT through a web browser, copying and pasting syntax, but these can now be directly integrated into coding tools (e.g., Interactive Development Environments [IDEs] like VS Code, Positron, or RStudio).

https://www.ncrm.ac.uk/resources/online/all/?id=20859

Why can't Epidemiology be automated (yet)?

David Bann, Ed Lowther, Liam Wright, Yevgeniya Kovalchuk

https://github.com/edlowther/automated-epidemiology

International Journal of
Epidemiology

Methods Futures Briefing #006

Series edited by Robert Meckin and Mark Elliot

Artificial Intelligence

By David Bann[1] and Liam Wright[1]

NCRM
NATIONAL CENTRE FOR RESEARCH METHODS

https://www.ncrm.ac.uk/resources/futures/

Appendix slides

Autonomy Level	Example Systems	Unlocking AGI Level(s)	Example Risks Introduced
Autonomy Level 0: No AI *human does everything*	Analogue approaches (e.g., sketching with pencil on paper) Non-AI digital workflows (e.g., typing in a text editor; drawing in a paint program)	No AI	n/a (status quo risks)
Autonomy Level 1: AI as a Tool *human fully controls task and uses AI to automate mundane sub-tasks*	Information-seeking with the aid of a search engine Revising writing with the aid of a grammar-checking program Reading a sign with a machine translation app	Possible: Emerging Narrow AI Likely: Competent Narrow AI	de-skilling (e.g., over-reliance) disruption of established industries
Autonomy Level 2: AI as a Consultant *AI takes on a substantive role, but only when invoked by a human*	Relying on a language model to summarize a set of documents Accelerating computer programming with a code-generating model Consuming most entertainment via a sophisticated recommender system	Possible: Competent Narrow AI Likely: Expert Narrow AI; Emerging AGI	over-trust radicalization targeted manipulation
Autonomy Level 3: AI as a Collaborator *co-equal human-AI collaboration; interactive coordination of goals & tasks*	Training as a chess player through interactions with and analysis of a chess-playing AI Entertainment via social interactions with AI-generated personalities	Possible: Emerging AGI Likely: Expert Narrow AI; Competent AGI	anthropomorphization (e.g., parasocial relationships) rapid societal change
Autonomy Level 4: AI as an Expert *AI drives interaction; human provides guidance & feedback or performs subtasks*	Using an AI system to advance scientific discovery (e.g., protein-folding)	Possible: Exceptional Narrow AI Likely: Expert AGI	societal-scale ennui mass labor displacement decline of human exceptionalism
Autonomy Level 5: AI as an Agent *fully autonomous AI*	Autonomous AI-powered personal assistants *(not yet unlocked)*	Likely: Exceptional AGI; ASI	misalignment concentration of power

https://arxiv.org/pdf/2311.02462

PERSPECTIVE | JANUARY 27, 2025

PNAS

The misalignment of incentives in academic publishing and implications for journal reform

Jennifer S. Trueblood, David B. Allison, [...] Andrei R. Teodorescu

e.g.,

Take publishing back. Non-profit, scientist-run journals.

Change evaluations. Move beyond impact factor: reward rigor, openness, contribution

Evaluate changes + experiment e.g., game theory

https://www.pnas.org/topic/568

ChatGPT

Association of Birth Weight with Body Mass Index at Age 42 in a Longitudinal Cohort

Introduction

Birth weight has long been studied as an early-life predictor of adult health. The "developmental origins of health and disease" hypothesis (Barker hypothesis) proposes that an adverse intrauterine environment, often reflected by low birth weight (LBW), can predispose individuals to chronic conditions in later life [1]. Indeed, some studies report that LBW is associated with higher risks of obesity, insulin resistance, and type 2 diabetes in adulthood [2]. Conversely, high birth weight (e.g. >4 kg) has been linked to greater body mass index (BMI) and higher odds of overweight or obesity in childhood and adult life [3]. For example, a large pooled analysis of twin cohorts found that each 1 kg increase in birth weight was associated with roughly a 0.5–0.9 kg/m^2 higher BMI in later life ($p < 0.001$) [4]. These findings suggest a positive linear tracking of body size from birth to adulthood.

However, the relationship is not entirely straightforward. Some evidence indicates a **J-shaped** or **U-shaped** association, where *both* low and high birth weights confer higher risk of adverse outcomes. A systematic review noted that individuals at the extremes of birth weight tend to have elevated odds of adult overweight/obesity [5]. In particular, while high birth weight infants are more likely to become obese, the impact of low birth weight on adult adiposity has been less clear and might depend on other factors [3]. It has been suggested that the association between low birth weight and adult obesity could be **obscured by sex differences** [6]. Males and females have different fetal growth patterns and metabolic responses; indeed, famine and cohort studies have found that in utero undernutrition affects later adiposity in a sex-specific manner [6]. For instance, one study observed that low birth weight was associated with *lower* adult BMI in men but a tendency toward *higher* adiposity in women [7]. This underscores the importance of analyzing sex-specific effects when examining birth weight and obesity outcomes.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	n/a
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	1
Objectives	3	State specific objectives, including any prespecified hypotheses	1

Bann et al, arxiv / IJE 2025