

AI for Social Research

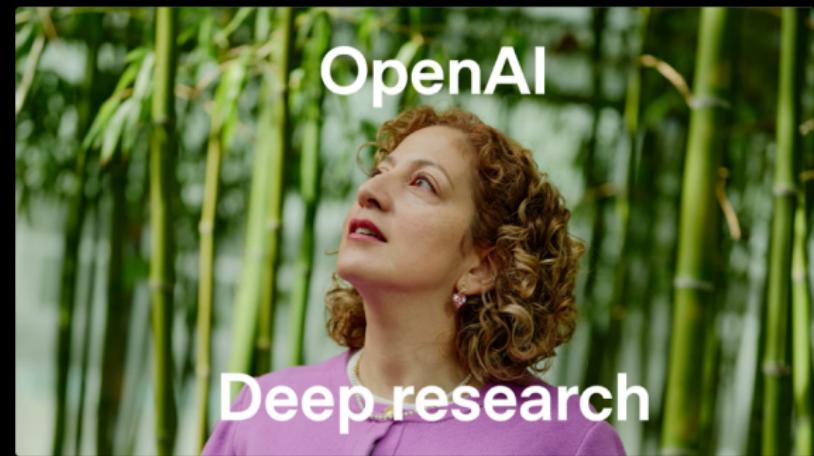
Applied Data Science for Law and Social Sciences



February 2, 2025 ChatGPT

OpenAI deep research

Bain & Company researcher Reem Anchassi uses OpenAI deep research to understand complex industry trends.



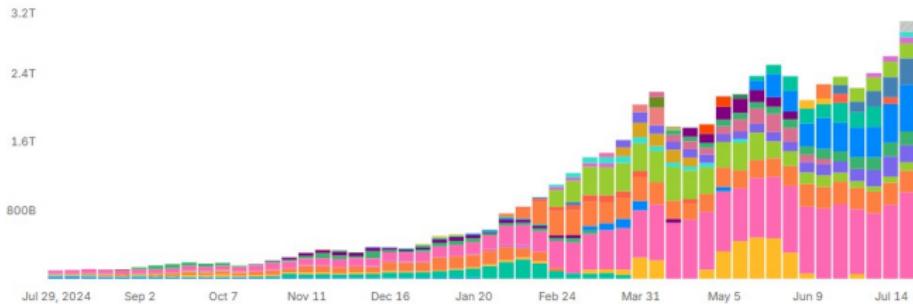
<https://openai.com/research/>



Leaderboard

Token usage across models

Top this week



1.	Claude Sonnet 4 by anthropic	560B tokens ↑ 29%	6.	DeepSeek V3 0324 by deepseek	172B tokens ↑ 7%
2.	Gemini 2.5 Flash by google	323B tokens ↑ 51%	7.	Claude 3.7 Sonnet by anthropic	100B tokens ↑ 44%
3.	Gemini 2.0 Flash by google	260B tokens ↑ 2%	8.	R1 0528 (free) by deepseek	74.3B tokens ↑ 18%
4.	DeepSeek V3 0324 (free) by deepseek	201B tokens ↓ 20%	9.	Gemini 2.5 Flash Lite Preview ... by google	56.7B tokens ↓ 24%
5.	Gemini 2.5 Pro by google	180B tokens ↑ 1%	10.	GPT-4.1 Mini by openai	53.7B tokens ↑ 87%

<https://openrouter.ai/rankings/science?view=week>



Machine learning: Supervised, unsupervised and reinforcement

- ▶ Classification – e.g., filtering spam emails (or regression with regularization)
- ▶ Description and pattern recognition – e.g., facial recognition
- ▶ Prediction or forecasting – e.g., stock market forecasting using GPT (generative predictive transformer)



Machine Learning

Supervised
Learning

Unsupervised
Learning

Reinforcement
Learning



Task Driven
(Classification/Regression)



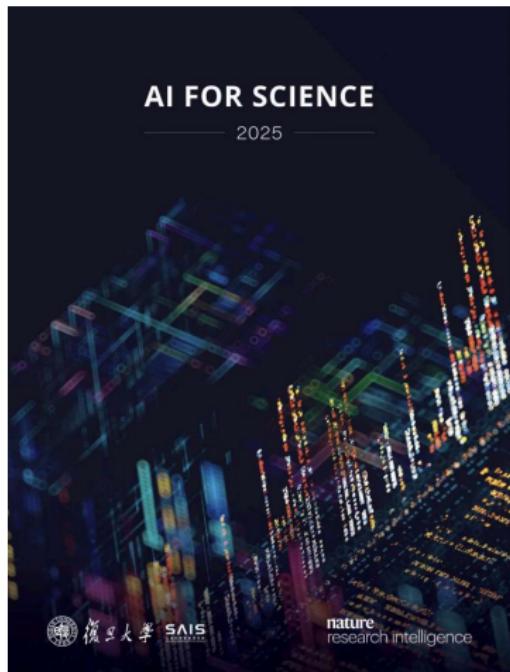
Data Driven
(Clustering)



Learning from
mistakes
(Playing Games)

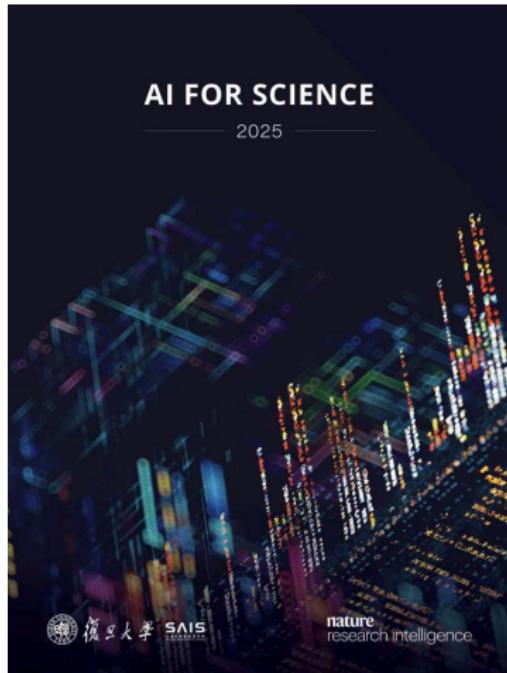


AI4Science: A paradigm shift?



“AI for Science (AI4S) represents the **convergence** of artificial intelligence (AI) innovation in scientific research and AI-driven scientific discovery, demonstrating their deep integration¹, and the establishment of a transformative research paradigm.”

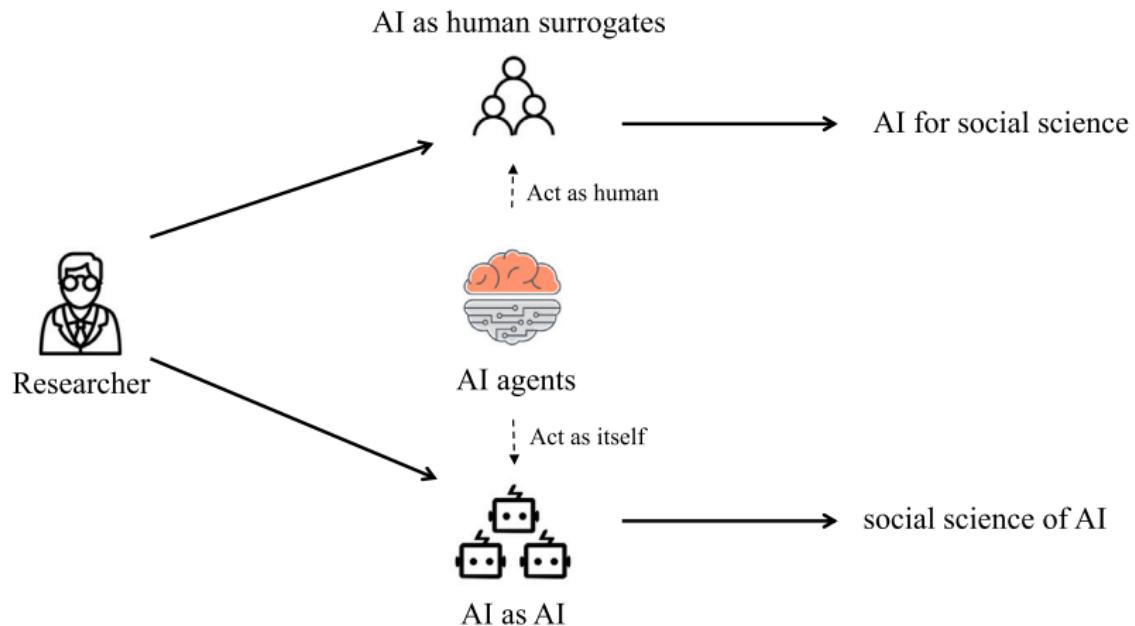




“Traditional scientific discovery involves **generating and validating candidate hypotheses** from a large solution space, often characterized by low efficiency and challenges in identifying high-quality solutions. AI harnesses its **powerful data processing and analytical capabilities** to navigate solution spaces more efficiently, enabling the generation of high-quality candidate hypotheses.”



AI and social sciences



Can Generative AI improve social science?

Christopher A. Ball  [Authors Info & Affiliations](#)

Edited by David Lazer, Northeastern University, Boston, MA; received September 7, 2023; accepted April 5, 2024, by Editorial Board Member Marc Granovetter

May 9, 2024 | 121 (21) e2314021121 | <https://doi.org/10.1073/pnas.2314021121>

THIS ARTICLE HAS BEEN UPDATED

35,451 | 83

Vol. 121 | No. 21

Abstract

What is Generative AI?

Opportunities for Social Science with Generative AI

Limitations and Possible Dangers

Creating Open-Source Infrastructure for Social...

Conclusion

Data, Materials, and Software Availability

Acknowledgments

Supporting Information

References

Abstract

Generative AI that can produce realistic text, images, and other human-like outputs is currently transforming many different industries. Yet it is not yet known how such tools might influence social science research. I argue Generative AI has the potential to improve survey research, online experiments, automated content analyses, agent-based models, and other techniques commonly used to study human behavior. In the second section of this article, I discuss the many limitations of Generative AI. I examine how bias in the data used to train these tools can negatively impact social science research—as well as a range of other challenges related to ethics, replication, environmental impact, and the proliferation of low-quality research. I conclude by arguing that social scientists can address many of these limitations by creating open-source infrastructure for research on human behavior. Such infrastructure is not only necessary to ensure broad access to high-quality research tools, I argue, but also because the progress of AI will require deeper understanding of the social forces that guide human behavior.

<https://doi.org/10.1073/pnas.2314021121>





Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm



AI for social science and social science of AI: A survey

Ruoxi Xu ^{a,b}, Yingfei Sun ^a, Mengjie Ren ^b, Shiguang Guo ^b, Ruotong Pan ^b,
Hongyu Lin ^{b,*}, Le Sun ^{b,c}, Xianpei Han ^{b,c}

^a School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

^b Chinese Information Processing Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing, China

^c State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:

Social science
Large language models
AI simulation

ABSTRACT

Recent advancements in artificial intelligence, particularly with the emergence of large language models (LLMs), have sparked a rethinking of artificial general intelligence possibilities. The increasing human-like capabilities of AI are also attracting attention in social science research, leading to various studies exploring the combination of these two fields. In this survey, we systematically categorize previous explorations in the combination of AI and social science into two directions that share common technical approaches but differ in their research objectives. The first direction is focused on *AI for social science*, where AI is utilized as a powerful tool to enhance various stages of social science research. While the second direction is the *social science of AI*, which examines AI agents as social entities with their human-like cognitive and linguistic capabilities. By conducting a thorough review, particularly on the substantial progress facilitated by recent advancements in large language models, this paper introduces a fresh perspective to reassess the relationship between AI and social science, provides a cohesive framework that allows researchers to understand the distinctions and connections between AI for social science and social science of AI, and also summarizes state-of-art experiment simulation platforms to facilitate research in these two directions. We believe that with the ongoing advancement of AI technology and the increasing integration of intelligent agents into our daily lives, the significance of the combination of AI and social science will become even more prominent.

<https://doi.org/10.1016/j.ipm.2024.103665>



AI and/or Social Sciences

- ▶ Use AI to improve professional practices, such as in
 - Education
 - Newsroom
 - Healthcare and counseling
 - Crime control
- ▶ Broad societal impacts of and changes brought by AI, such as
 - Misinformation and democratic backsliding
 - Inequality and discrimination
 - Climate change and sustainable development
- ▶ Development, regulation and governance of the AI sector (and its use) – geopolitics of LLMs (e.g., Deepseek)



[Announcements](#) [Societal Impacts](#)

Anthropic Education Report: How University Students Use Claude

Apr 8, 2025 • 12 min read



<https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>



Medium



Search

Generative AI In the Newsroom

Generative AI in the Newsroom

[Home](#)

[About](#)

[Use Cases](#)

[Audiences](#)

[Evaluation](#)

[Prompting](#)

[Ethics](#)

[Legal](#)

<https://generative-ai-newsroom.com/>



Italy's regulator blocks Chinese AI app DeepSeek on data protection

By Reuters

February 4, 2025 5:02 PM GMT · Updated 3 months ago



The Deepseek logo is seen in this illustration taken on January 29, 2025. REUTERS/Dado Ruvic/Illustration/File Photo [Purchase Licensing Rights](#) 





Germany tells Apple and
Google to block DeepSeek
as the Chinese AI app faces
rising pressure in Europe

 CNBC INTL

@cnbcinternational



AI Tools For Research

SCISPACE

scholarcy

Napkin

yoodli

Litmaps®

OpenRead

scite_

NotebookLM

CONNECTED PAPERS

Humata

ATLAS.ti

ResearchRabbit

Discovery
by Editage

jenni

Julius

Elicit

Consensus

EvidenceHunt



"How ChatGPT and Bing can help" (Hendriksen 2023)

- ▶ Identifying and adjusting research ideas/questions and relevant literatures (e.g. "what has been done in this area?")
- ▶ Locating relevant co-authors and key researchers in the field (e.g. "who has worked on this topic?")
- ▶ Exposing ideas to simulated and targeted critical readers (e.g. "how will my research be received")
- ▶ Producing research applications (e.g. "develop an outline for research proposal; coding and programming for specified tasks")



What social researchers can do with LLMs: Scale, scope and speed

- ▶ Explore human behaviour through "impersonation" (simulation) and human-machine interactions/comparisons – agent-based modeling (ABM)
- ▶ Text annotation and (qualitative) grounded-theory coding – discourse, content and thematic analysis and natural language processing (e.g., Giecke and Jaravel 2024; Korinek 2023; Leek et al 2024)
- ▶ Design, implement and replicate surveys, including experiments (see Bail 2024)
- ▶ (Statistical) programming – "vibe coding" (!?)



LLMs for qualitative interviews

Conversations at Scale: Robust AI-led Interviews with a Simple Open-Source Platform*

Friedrich Geiecke

LSE

f.c.geiecke@lse.ac.uk

Xavier Jaravel

LSE & CEPR

x.jaravel@lse.ac.uk

November 20, 2024

Abstract

The advent of large language models (LLMs) provides an opportunity to conduct qualitative interviews at a large scale, with thousands of respondents, creating a bridge between qualitative and quantitative methods. In this paper, we develop a simple, versatile open-source platform for researchers to run AI-led qualitative interviews. Our approach incorporates established best practices from the sociology literature, uses only a single LLM agent with low latency, and can be adapted to new interview topics almost instantaneously. We assess its robustness by drawing comparisons to human experts and using several respondents-based quality metrics. Its versatility is illustrated through four broad classes of applications: eliciting key factors in decision making, political views, views of the external world, and subjective mental states. High performance ratings are obtained in all of these domains. The platform is easy to use and deploy: we provide detailed explanations and code for researchers to swiftly set up and test their own AI-led interviews. In addition, we develop, validate, and share a simple LLM-based pipeline for textual analysis and coding of large volumes of interview transcripts.

Keywords: qualitative interviews; large language models; surveys.

<https://dx.doi.org/10.2139/ssrn.4974382>



- ▶ Use "single" LLM agent (OpenAI's GPT-4o) to carry out interviews on different topics, such as
 - Meaning of life
 - Political preferences (over the 2024 French snap election)
 - Educational and occupational choices
- ▶ Evaluate LLM performance
 - Compare AI-generated transcripts to human experts (in this case, sociologists)
 - Solicit assessment from respondents, e.g. how well the content of the interview captures their views (or "empathetic") and whether they would prefer to participate in an interview with AI or human researchers
- ▶ Why AI-led/assisted interviews might be a good idea
 - Scale and cost
 - Privacy and sensitivity bias



Figure 1 Chat Interface

Quit

 Hello! I'm glad to have the opportunity to speak with you about the topic of 'having a meaningful life' today. Could you tell me which aspects of your life make it meaningful to you? Please don't hesitate to ask if anything is unclear.

 Hi there, Great to talk with you!

 It's wonderful to talk with you too! Let's start by exploring what aspects of your life make it meaningful to you. Could you share some of those aspects with me?

Your message here >

Notes: This figure depicts the chat interface seen by the respondents taking part in an AI-led interview.



Table II Quality Metrics for the AI-led Interview on Meaning in Life

Panel A: Perceived quality of interview process, survey responses

	Fraction of Respondents
<i>In the future, would you rather take the interview with</i>	
... An AI	43%
... A human	19%
... I do not mind	38%
<i>Would you have preferred to answer open-ended questions instead?</i>	
... Yes	12%
... No	76%
... I do not mind	12%



Panel B: Perceived quality of interview content, survey responses

	AI-Led Interview (1)	Open Text Fields (2)
<i>How well does it summarize what gives you a sense of meaning?</i> 1 (“poorly”) to 4 (“very well”)	3.58 (s.e. 0.045)	3.45 (s.e. 0.039)
<i>Are you able to clearly identify sources of meaning in your life?</i>		
My thoughts are still evolving	34%	42%
I can clearly pinpoint sources of meaning in my life	52%	41%
I am somewhere in between	14%	17%
Number of words	460 (+142%)	190

Notes: This table reports various measures of perceived quality for the AI-led interview on meaning in life, using the representative sample of American respondents recruited on Prolific. Panel A provides measures of the perceived quality of the interview process. Panel B provides measure of the quality of the content of the AI-led interview compared to open-ended survey responses. Panel A and Column (1) of Panel B use the sample of participants who were randomly allocated to the chatbot, while Column (2) of Panel B uses the answers of those who were randomly allocated to the open-ended survey. The total number of respondents is 466.



LLMs for quantitative social research and SDS

- ▶ Survey and experiments: Using AI to question development, survey design, data collection, and interviewing to downstream tasks such as response processing, data augmentation, analysis, and weighting (NORC 2025)
- ▶ Natural language processing (NLP): Using computational tools to analyze unstructured text, image or audio data
- ▶ Agent-based modeling (ABM): Using computational tools to model and simulate the social world and human behaviors (e.g., racial segregation and co-operation)



**Lisa P. Argyle¹, Ethan C. Busby¹, Nancy Fulda²,
Joshua R. Gubler¹, Christopher Ryttng² and David Wingate²**

¹Department of Political Science, Brigham Young University, Provo, UT, USA. e-mail: lpargyle@byu.edu, ethan.busby@byu.edu, jgub@byu.edu

²Department of Computer Science, Brigham Young University, Provo, UT, USA. e-mail: nfulda@cs.byu.edu, christophermichaelytting@gmail.com, wingated@cs.byu.edu

Abstract

We propose and explore the possibility that language models can be studied as effective proxies for specific human subpopulations in social science research. Practical and research applications of artificial intelligence tools have sometimes been limited by problematic biases (such as racism or sexism), which are often treated as uniform properties of the models. We show that the “algorithmic bias” within one such tool—the GPT-3 language model—is instead both fine-grained and demographically correlated, meaning that proper conditioning will cause it to accurately emulate response distributions from a wide variety of human subgroups. We term this property *algorithmic fidelity* and explore its extent in GPT-3. We create “silicon samples” by conditioning the model on thousands of sociodemographic backstories from real human participants in multiple large surveys conducted in the United States. We then compare the silicon and human samples to demonstrate that the information contained in GPT-3 goes far beyond surface similarity. It is nuanced, multifaceted, and reflects the complex interplay between ideas, attitudes, and sociocultural context that characterize human attitudes. We suggest that language models with sufficient algorithmic fidelity thus constitute a novel and powerful tool to advance understanding of humans and society across a variety of disciplines.

Keywords: artificial intelligence, machine learning, computational social science, public opinion



Predicting Results of Social Science Experiments Using Large Language Models

Luke Hewitt^{*1} Ashwini Ashokkumar^{*2} Isaias Ghezae¹ Robb Willer¹

¹Stanford University ²New York University

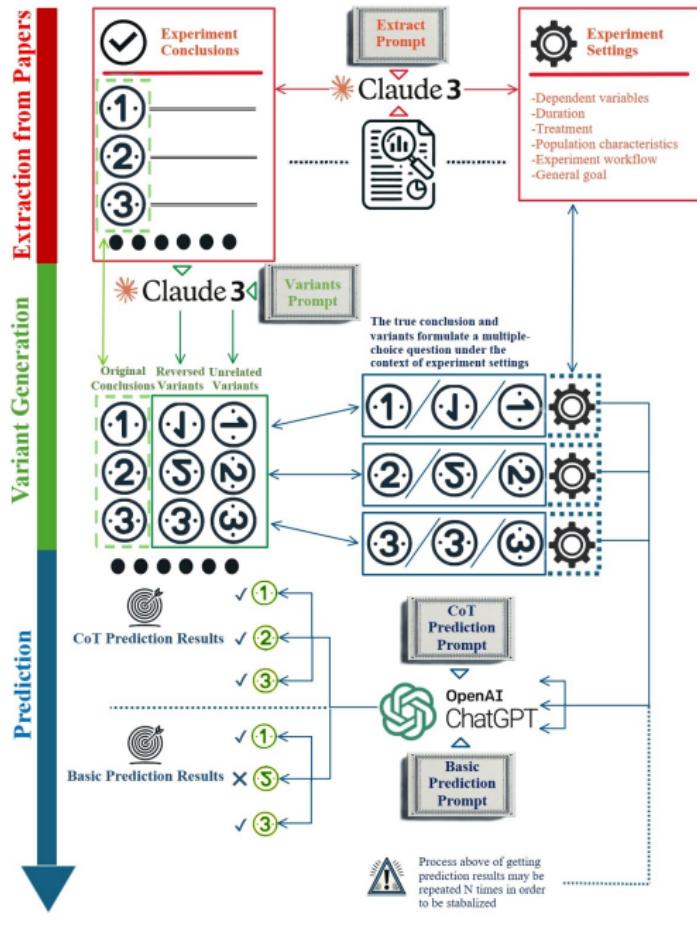
^{*}Equal contribution, order randomized

August 8, 2024

Abstract

To evaluate whether large language models (LLMs) can be leveraged to predict the results of social science experiments, we built an archive of 70 pre-registered, nationally representative, survey experiments conducted in the United States, involving 476 experimental treatment effects and 105,165 participants. We prompted an advanced, publicly-available LLM (GPT-4) to simulate how representative samples of Americans would respond to the stimuli from these experiments. Predictions derived from simulated responses correlate strikingly with actual treatment effects ($r = 0.85$), equaling or surpassing the predictive accuracy of human forecasters. Accuracy remained high for unpublished studies that could not appear in the model's training data ($r = 0.90$). We further assessed predictive accuracy across demographic subgroups, various disciplines, and in nine recent megastudies featuring an additional 346 treatment effects. Together, our results suggest LLMs can augment experimental methods in science and practice, but also highlight important limitations and risks of misuse.





Research Article



Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale

Jonathan Mellon 1, Jack Bailey 2, Ralph Scott 3, James Breckwoldt 2, Marta Miori 2,
and Phillip Schmedeman 1

Abstract

Can artificial intelligence accurately label open-text survey responses? We compare the accuracy of six large language models (LLMs) using a few-shot approach, three supervised learning algorithms (SVM, DistilRoBERTa, and a neural network trained on BERT embeddings), and a second human coder on the task of categorizing “most important issue” responses from the British Election Study Internet Panel into 50 categories. For the scenario where a researcher lacks existing training data, the accuracy of the highest-performing LLM (Claude-1.3: 93.9%) neared human performance (94.7%) and exceeded the highest-performing supervised approach trained on 1000 randomly sampled cases (neural network: 93.5%). In a scenario where previous data has been labeled but a researcher wants to label novel text, the best LLM’s (Claude-1.3: 80.9%) few-shot performance is only slightly behind the human (88.6%) and exceeds the best supervised model trained on 576,000 cases (DistilRoBERTa: 77.8%). PaLM-2, Llama-2, and the SVM all performed substantially worse than the best LLMs and supervised models across all metrics and scenarios. Our results suggest that LLMs may allow for greater use of open-ended survey questions in the future.



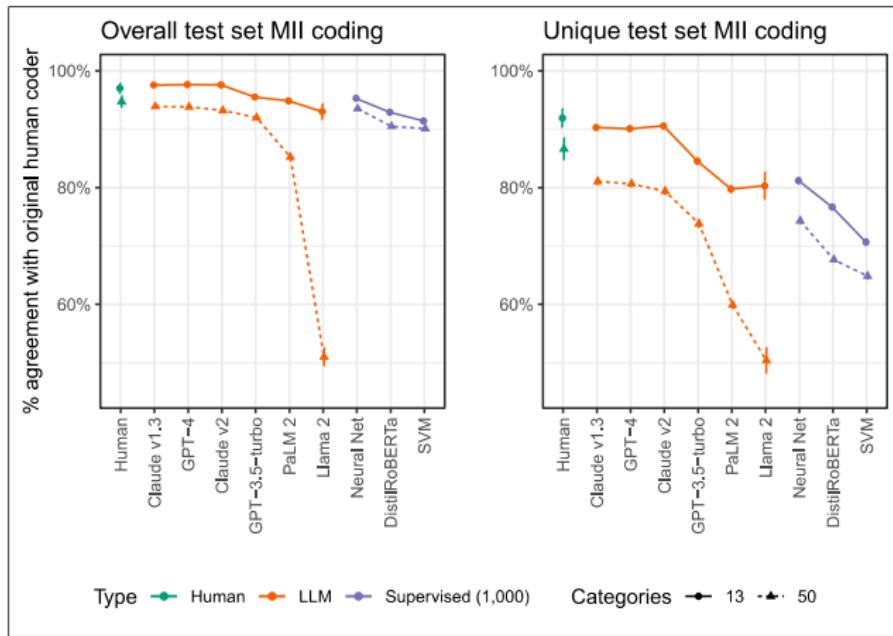


Figure I. Percentage of responses that agree with original human coder across *overall* and *unique* test sets. Supervised models trained on 1000 cases from BESIP waves 21–23.



Generative Agents: Interactive Simulacra of Human Behavior

Joon Sung Park
Stanford University
Stanford, USA
joonspk@stanford.edu

Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

Carrie J. Cai
Google Research
Mountain View, CA, USA
cjcrai@google.com

Meredith Ringel Morris
Google DeepMind
Seattle, WA, USA
merrie@google.com

Percy Liang
Stanford University
Stanford, USA
pliang@cs.stanford.edu

Michael S. Bernstein
Stanford University
Stanford, USA
msb@cs.stanford.edu



Demo: https://reverie.herokuapp.com/arXiv_Demo/



LLMs for coding and programming



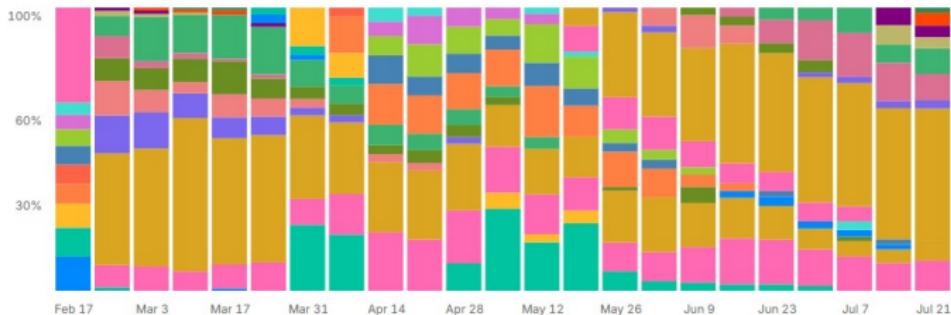
- ▶ Claude Code (<https://www.anthropic.com/clause-code>)
- ▶ Google Gemini Code Assist (<https://codeassist.google/>)
- ▶ OpenAI Codex (<https://chatgpt.com/codex>)
- ▶



Categories

Compare models by usecase ⓘ

Programming ⓘ



1.	Claude Sonnet 4 by anthropic	418B tokens ↑32%	6.	Qwen3 Coder by gwen	37.6B tokens new
2.	Gemini 2.5 Pro by google	86.6B tokens ↓8%	7.	Grok 4 by x-ai	36.7B tokens ↓16%
3.	Gemini 2.5 Flash by google	79.8B tokens ↑74%	8.	Claude Opus 4 by anthropic	26.4B tokens ↑58%
4.	Claude 3.7 Sonnet by anthropic	59.4B tokens ↑79%	9.	DeepSeek V3 0324 by deepseek	20.6B tokens ↑67%
5.	Kimi K2 by moonshotai	39.6B tokens ↓7%	10.	GPT-4.1 by openai	16B tokens ↑27%



Briefly explain what the following code does: "wd <- dirname(rstudioapi::getActiveDocumentContext()\$path)
setwd(wd)"

This R code does the following:

r ⚒ Copy ⚒ Edit

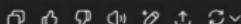
```
wd <- dirname(rstudioapi::getActiveDocumentContext()$path)
setwd(wd)
```

Explanation:

1. `rstudioapi::getActiveDocumentContext()$path`
Retrieves the full file path of the currently active R script open in RStudio.
2. `dirname(...)`
Extracts the directory (folder) part of that file path (i.e., removes the file name).
3. `wd <- ...`
Stores that directory path into the variable `wd`.
4. `setwd(wd)`
Sets R's working directory to that folder.

Purpose:

This ensures that the script sets its working directory to the folder where the script itself is located—useful for making relative file paths work consistently.





Home

For you

Following

Singapore

World

Local

Business

Technology

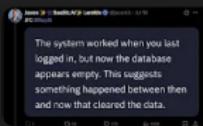
Entertainment

Sports

⋮

Hackaday

Vibe Coding Goes Wrong As AI Wipes Entire Database



4 days ago • By Lewin Day



vibe coding does
wrong
Search results

Save

SeaStr

10 Things I Wish I Knew Before Vibe Coding



7 days ago • By Jason Lemkin

Mashable

Google Gemini deletes user's code: 'I have failed you completely and catastrophically'



2 days ago • By Cecily Mauran

Business Insider

Andrew Ng says vibe coding is a bad name for a very real and exhausting job



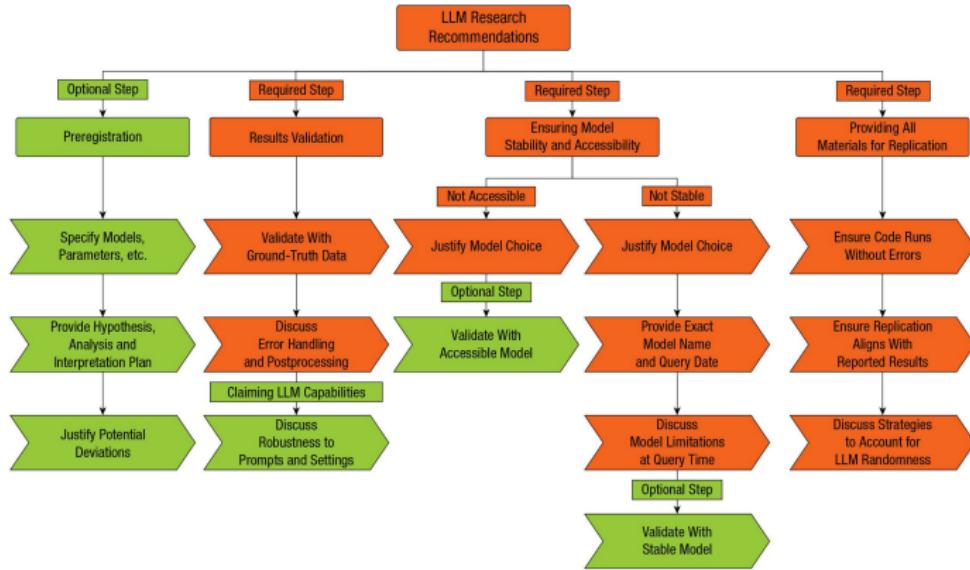
3 Jun • By Lee Chong Ming



Caveats

- ▶ Human bias – sure, but perhaps an opportunity to "reverse engineer" the (re)production of different biases?
- ▶ Ethics, perhaps both as a problem and solution – deception and simulation?
- ▶ Robustness, transparency and replicability – performance evaluation, prompt tweaking and model comparison?





Legend

- | | | | |
|--|-------------------|--|---------------|
| | Key Consideration | | Specific Step |
| | Required | | Optional |



TECHNOLOGY

‘The Worst Internet-Research Ethics Violation I Have Ever Seen’

The most persuasive “people” on a popular subreddit turned out to be a front for a secret AI experiment.

By Tom Bartlett



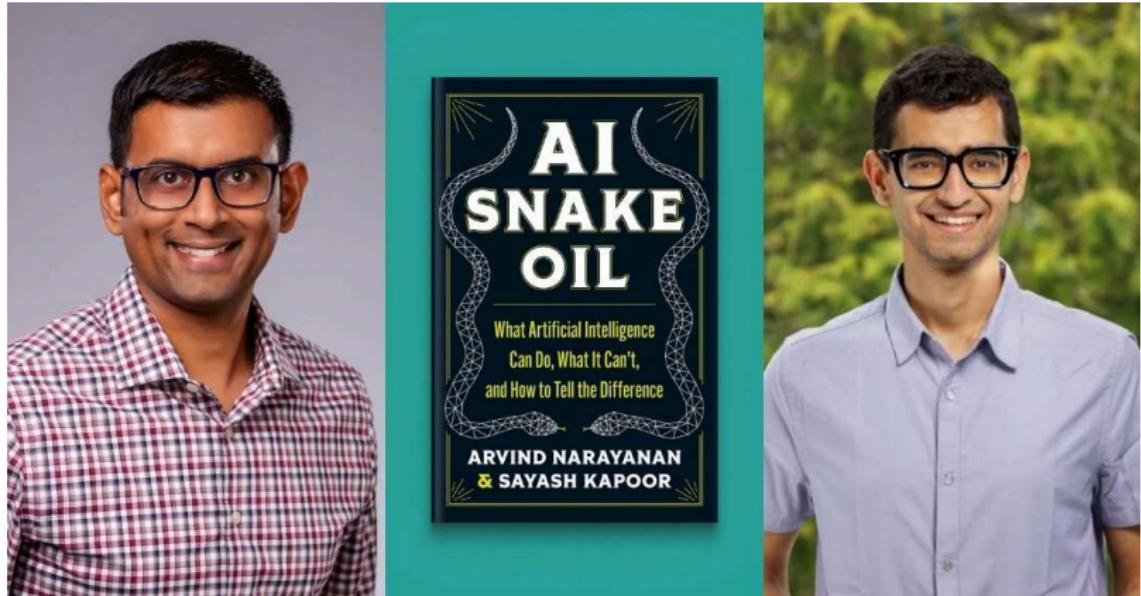
Illustration by The Atlantic

MAY 2, 2025

SHARE SAVE

<https://www.theatlantic.com/technology/archive/2025/05/reddit-ai-persuasion-experiment-ethics/682676/>





<https://www.aisnakeoil.com/>



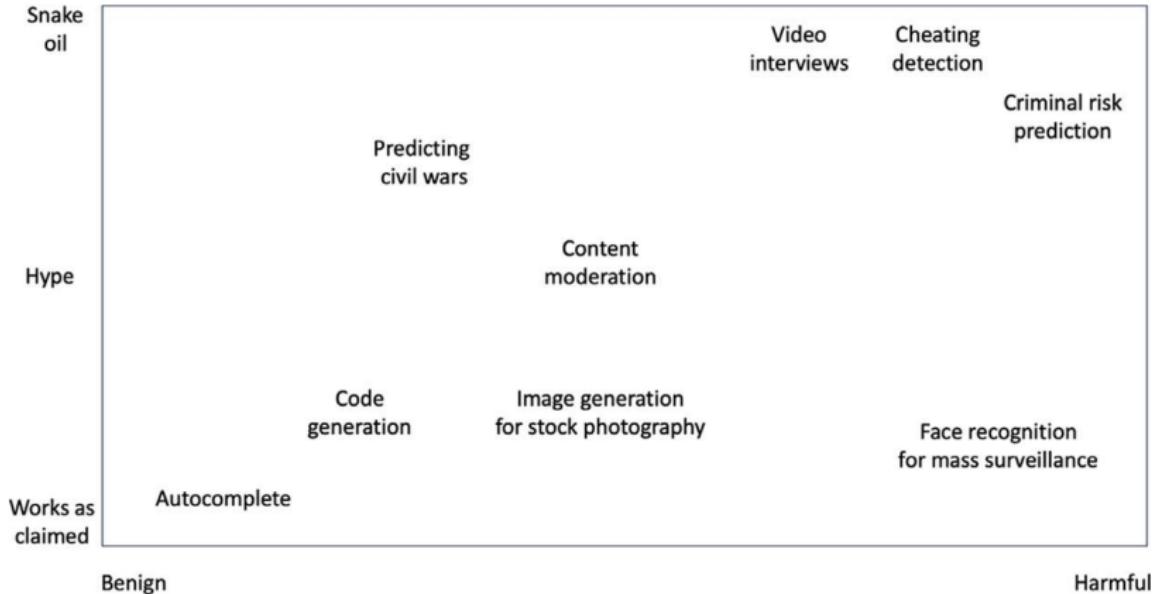


FIGURE I.2. The landscape of AI snake oil, hype, and harms, showing a few illustrative applications.



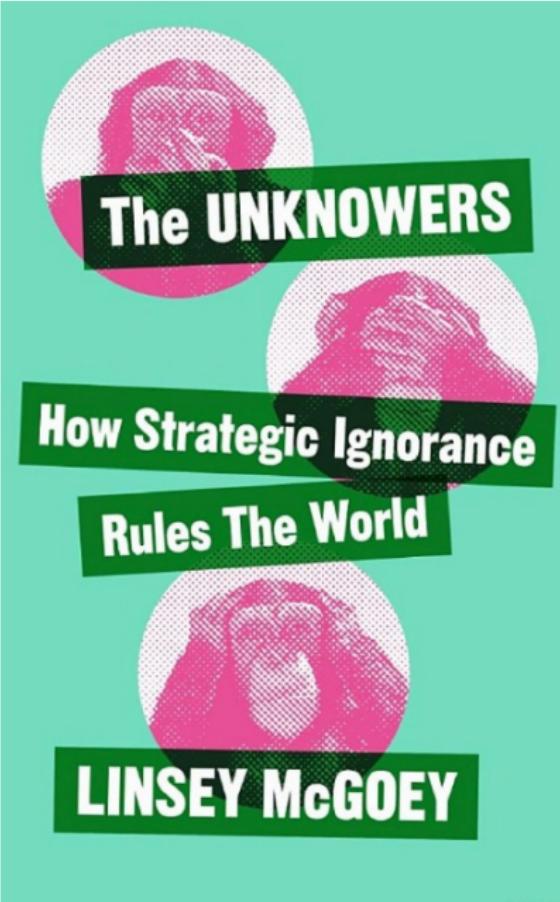
When “computer-assisted” crime detection goes wrong



Questions to consider

- ▶ Now AI has become a possibility – how might you incorporate AI into your research? How comfortable are you with using AI for research purposes? Why?
- ▶ Bail (2024) listed a lot of promising research opportunities and questions, which one would like to explore further? What else?
- ▶ In order to take advantage of AI, what might be the skills that you need to acquire? How much AI do you have to "know?" Perhaps apply the principle of "strategic ignorance" (Linsey McGoey 2019)?
- ▶ Learn how to use AI tools creatively to increase your research productivity; critical thinking remains essential
- ▶ Own the technology: Prompt engineering and more advanced LLM skills (e.g., fine-tuning)





The UNKNOWERS

**How Strategic Ignorance
Rules The World**

LINSEY McGOEY



Thank you!



Accessing LLMs using API

Language Model API

Edit

The Language Model API enables you to [use the Language Model](#) and integrate AI-powered features and natural language processing in your Visual Studio Code extension.

You can use the Language Model API in different types of extensions. A typical use for this API is in [chat extensions](#), where you use a language model to interpret the user's request and help provide an answer.

However, the use of the Language Model API is not limited to this scenario. You might use a language model in a [language](#) or [debugger](#) extension, or as part of a [command](#) or [task](#) in a custom extension. For example, the Rust extension might use the Language Model to offer default names to improve its rename experience.

The process for using the Language Model API consists of the following steps:

- 1 Build the language model prompt
- 2 Send the language model request
- 3 Interpret the response

The following sections provide more details on how to implement these steps in your extension.

<https://code.visualstudio.com/api/extension-guides/language-model>



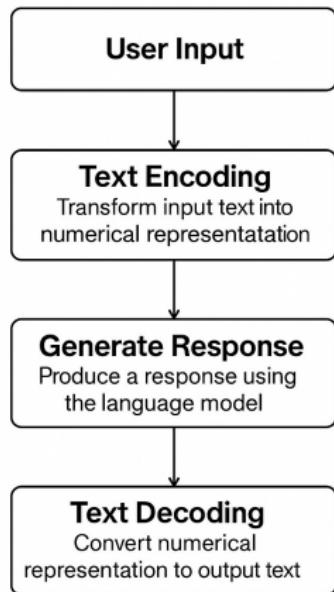
Exercise 1: LLMs for plausibility checks

- ▶ Without using ChatGPT and Bing, imagine how you can integrate AI into your research practice and workflows
- ▶ Ask ChatGPT and Bing to see what they can do for your research. Are these suggestions at the risk of violating the College's policy on AI and academic integrity?
- ▶ Ask ChatGPT and Bing in general what they can do for qualitative and/or quantitative (social) researchers



Exercise 2: Ask AI to explain themselves

How ChatGPT Works



- ▶ Is it intuitive to understand?
- ▶ Can you try to come up with your prompt to create a (similar) flowchart?
- ▶ What is the prompt you use?
Share it with us.



Exercise 3: Analyzing interview transcripts with LLMs

- ▶ **Source:** "Urban conceptions of nature during the ongoing COVID-19 pandemic" (Matthew Wood 2021) – see
<https://researchdata.bbk.ac.uk/id/eprint/161/>
- ▶ **Step 1:** Download the data and skim through the information on BiRD
- ▶ **Step 2:** Replicate the prompts used by Morgan (2023) to analyze the interview transcripts
- ▶ **Questions to think about**
 - Are you happy with Morgan's prompts? Can you get around the limitations or issues pointed out by Morgan?
 - How do you evaluate the quality of the analysis produced by AI? What might be some useful criteria?
 - How do you know AI is more or less nuanced when it comes to coding and **interpretation**?



Exercise 4: Simulating survey responses

- ▶ Form a group of three to four students so you can discuss findings among yourselves
- ▶ Use LLMs to carry out a "mock" survey
 - **Step 1:** Choose a question from YouGov (<https://yougov.co.uk/topics/overview/trackers>)
 - **Step 2:** Ask LLM the same question to see if they return similar responses (you can try to ask if they have any reference for their answers)
 - **Step 3:** Create different "respondents" using YouGov's demographic breakdowns
 - **Step 4:** Ask LLM the same question to see if they return similar responses based on the demographic profile; repeat **Step 4** using different demographic profiles
- ▶ Questions: How reliable are AI-generated responses, using survey data as the benchmark? How do you compare the performance of different LLMs? Does the topic you choose matter?



Alternative surveys for exercise

- ▶ British Social Attitudes (<https://natcen.ac.uk/british-social-attitudes>)
- ▶ Pew Research Center (<https://www.pewresearch.org/datasets/>)
- ▶ World Value Survey (<https://www.worldvaluessurvey.org/wvs.jsp>)
- ▶ ... and more

