

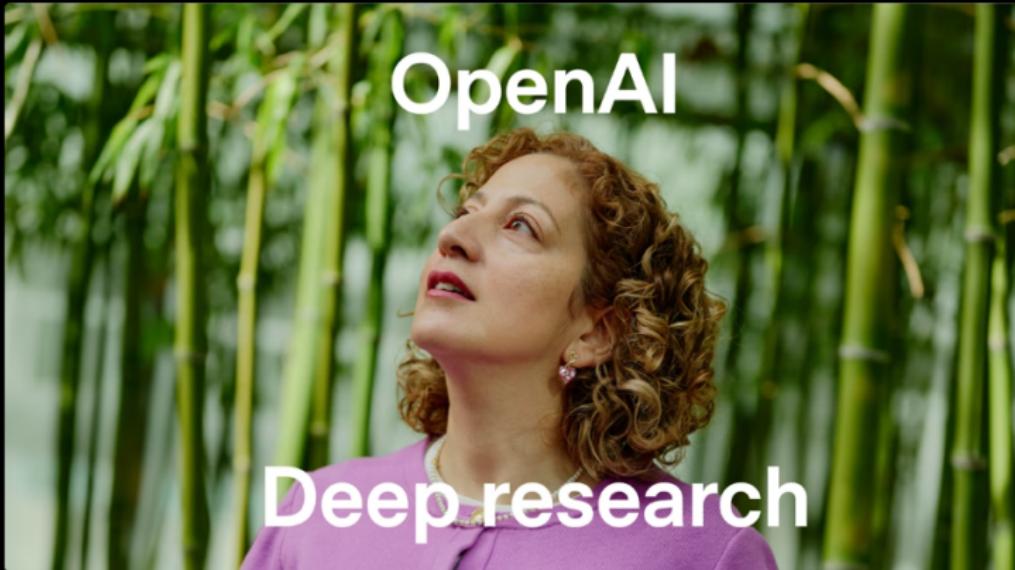
# Masterclass in Social Research

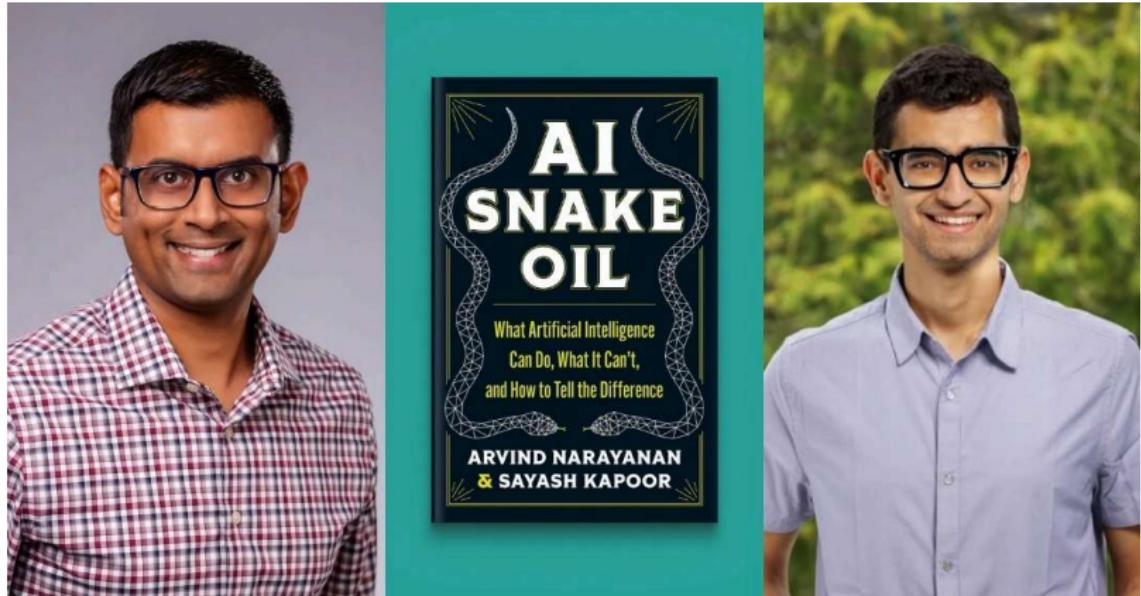
AI for Social Research: Overview



# OpenAI deep research

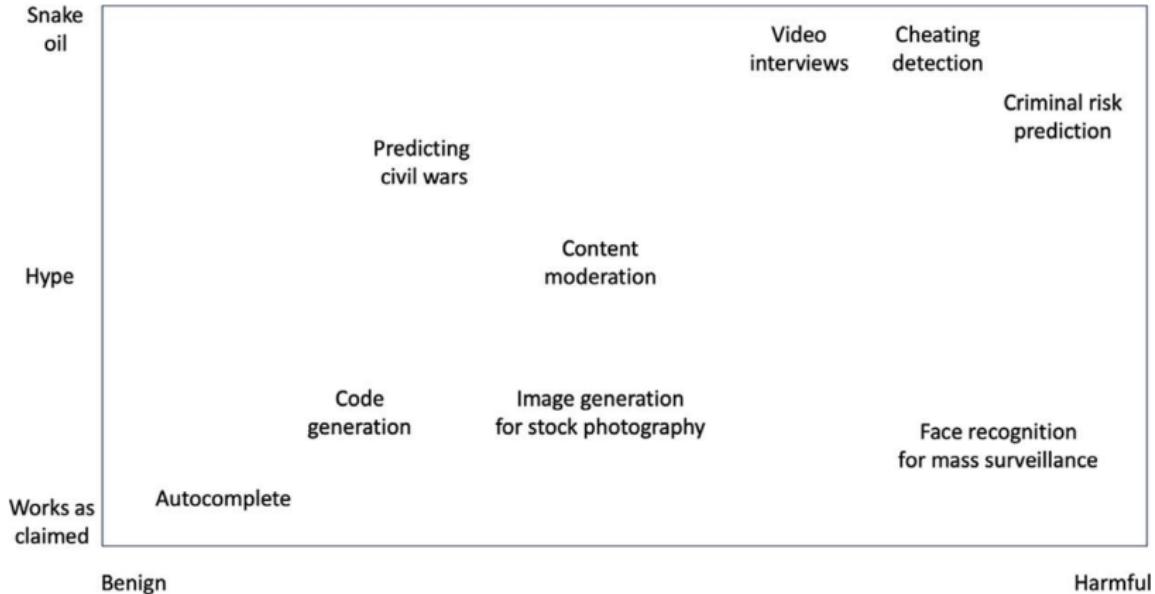
Bain & Company researcher Reem Anchassi uses OpenAI deep research to understand complex industry trends.





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





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



# Large language models: From classification to “generation”

Artificial Analysis LANGUAGE MODELS ▾ SPEECH, IMAGE & VIDEO MODELS ▾ LEADERBOARDS ▾ ARENAS ▾ ABOUT ▾ Newsletter Subscribe

## LLM API Providers Leaderboard - Comparison of over 100 LLM endpoints

Comparison and ranking of API provider performance for over 100 AI LLM Model endpoints across performance key metrics including price, output speed, latency, context window & others. For more details including relating to our methodology, see our FAQs.

API providers compared: OpenAI, Playground AI, Microsoft Azure, Mistral, Ideogram, DeepSeek, Amazon Bedrock, Hyperbolic, Groq, FriendAI, Together.ai, Anthropic, Black Forest Labs, Perplexity, Google, Lambda Labs, Fireworks, Cerebras, Leonardo.AI, Cohere, Recraft AI, Upstage, Simplismart, Speechmatics, Deepintra, Fish Audio, Replicate, Genmo, Nebius, Adobe, MiniMax, CentML, StepFun, Rurpod, Zyphra, Murf AI, Speechify, Rev AI, AssemblyAI, fal.ai, Rime, kluster.ai, Prodia, Hume AI, Reka AI, Deepgram, Gladia, Baseten, Stability.ai, Midjourney, Reve, Databricks, ElevenLabs, IBM, Vivago AI, SambaNova, Dreamina, xAI, Cartesia, LMNT, PlayAI, O1.AI, Alibaba Cloud, Novita, AI21 Labs, and WaveSpeed.

PROMPT OPTIONS

Filter, e.g. GPT, Meta EXPAND COLUMNS ↗

API PROVIDER	FEATURES	MODEL INTELLIGENCE	PRICE	OUTPUT TOKENS/S	LATENCY	END-TO-END RESPONSE TIME	FURTHER ANALYSIS		
API PROVIDER	MODEL	CONTEXT WINDOW	ARTIFICIAL ANALYSIS INTELLIGENCE INDEX	BLENDED USD/1M Tokens	MEDIAN Tokens/s	MEDIAN First Chunk (s)	TOTAL Response (s)	REASONING Time (s)	FURTHER ANALYSIS
OpenAI	o4-mini (high)	200k	70	\$1.93	125.2	31.35	35.35	N/A	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers
Microsoft Azure	o4-mini (high)	200k	70	\$1.93	34.9	147.40	161.73	N/A	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers
Google	G Gemini 2.5 Pro	1m	68	\$3.44	148.1	39.57	42.94	N/A	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers
OpenAI	o3	128k	67	\$17.50	195.6	13.24	15.80	N/A	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers
Microsoft Azure	o3	128k	67	\$17.50	56.3	57.97	66.86	N/A	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers
xAI	Grok 3 mini Reasoning (high)	131k	67	\$0.35	95.0	0.24	26.54	21.04	<input checked="" type="checkbox"/> Model <input checked="" type="checkbox"/> Providers

<https://artificialanalysis.ai/leaderboards/providers>



# AI and Social Sciences

- ▶ Use AI to improve social sciences (see the next slide) or other practical uses such as in
  - Education (for both teachers and students)
  - Newsroom
  - Healthcare and counseling
  - Crime control
- ▶ Broad societal impacts of and (potential) changes brought by AI, such as
  - Misinformation and democratic backsliding
  - Group inequality and discrimination
  - Climate change and sustainable development
- ▶ Development, regulation and governance of AI – 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>





# Generative AI In the Newsroom

## Generative AI in the Newsroom

[Home](#)

[About](#)

[Use Cases](#)

[Audiences](#)

[Evaluation](#)

[Prompting](#)

[Ethics](#)

[Legal](#)

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



## "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 GenAI: Scale, scope and speed

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



## Caveats

- ▶ Human bias – sure, but perhaps an opportunity to "reverse engineer" the (re)production of different biases?
- ▶ Junk science and quality control – domain-specific retrieval-augmented generation (RAG)?
- ▶ Ethics, perhaps both as a problem and solution – deception and simulation?
- ▶ Transparency and replicability – performance evaluation, prompt tweaking and model comparison?



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

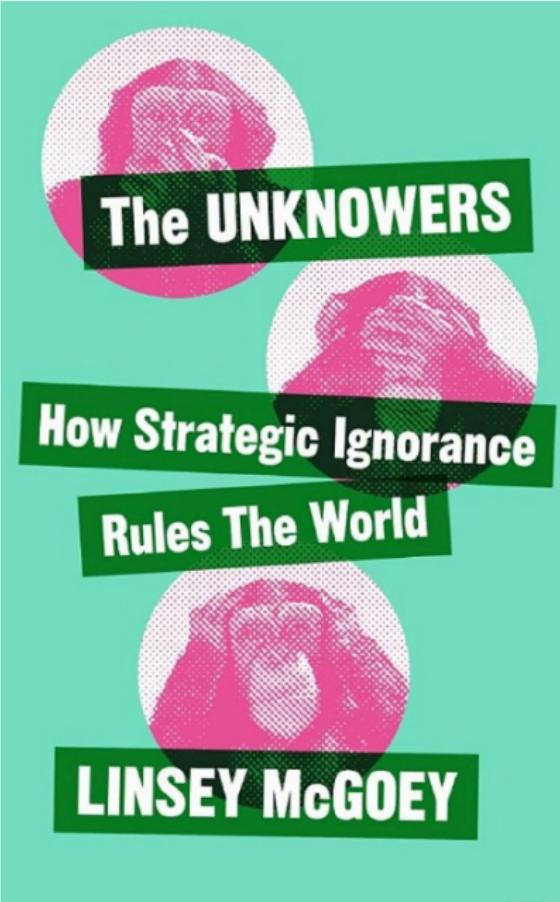
<https://www.theatlantic.com/technology/archive/2025/05/reddit-ai-persuasion-experiment-ethics/682676/>



## 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)?





**The UNKNOWERS**

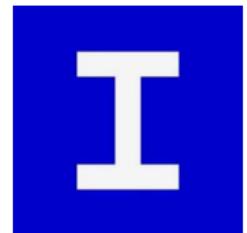
**How Strategic Ignorance  
Rules The World**

**LINSEY McGOEY**





The  
Alan Turing  
Institute



# Machine learning: Supervised, unsupervised and reinforcement

- ▶ Classification (or regression with regularization) – filtering spam emails
- ▶ Pattern recognition (or clustering) – facial recognition
- ▶ Prediction (or forecasting) – stock market forecasting and GPT (generative predictive transformer)



# Machine Learning

Supervised  
Learning

Unsupervised  
Learning

Reinforcement  
Learning



Task Driven  
(Classification/Regression)



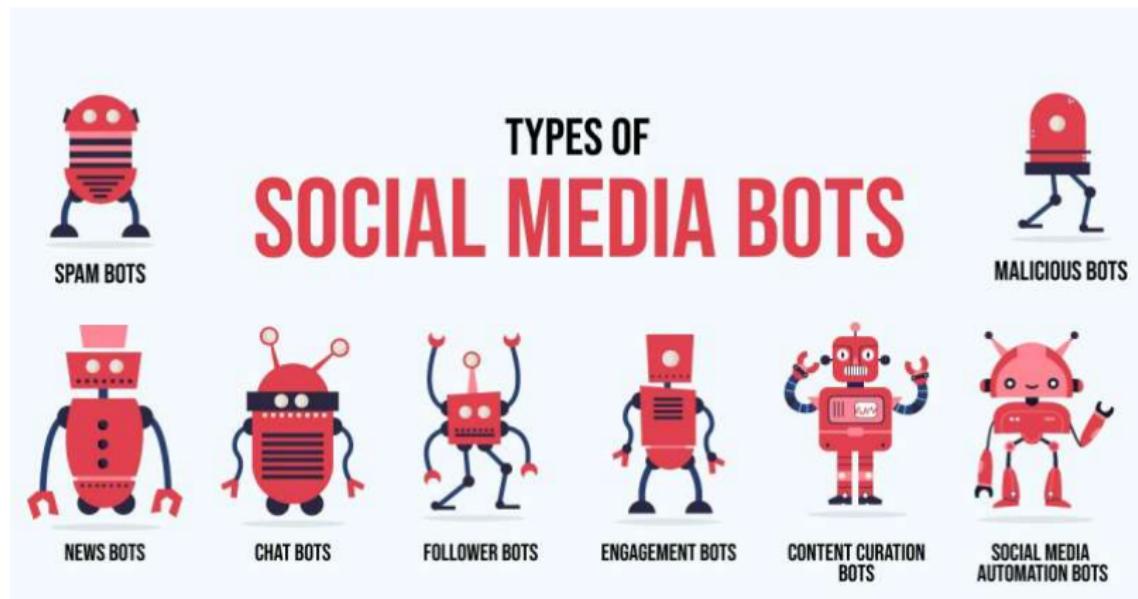
Data Driven  
(Clustering)



Learning from  
mistakes  
(Playing Games)



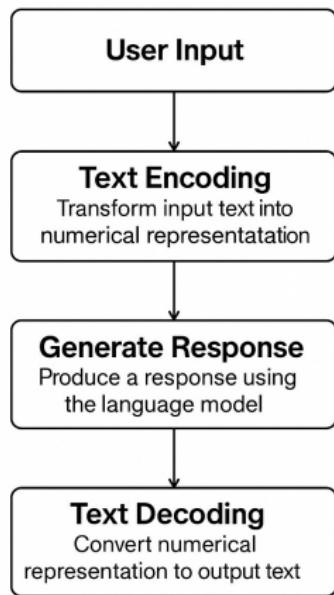
Example: Bots on social media



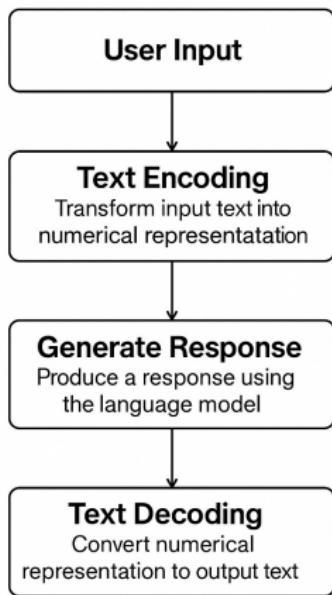
<https://www.socialfollowers.uk/blogs/how-social-media-bots-work-and-how-to-detect-them/>



# How ChatGPT Works



# 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.



# 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](#) 



# Exercise: Conducting interviews using AI agent (Geiecke and Jaravel 2024)

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

Friedrich Geiecke

LSE

[f.c.geiecke@lse.ac.uk](mailto:f.c.geiecke@lse.ac.uk)

Xavier Jaravel

LSE & CEPR

[x.jaravel@lse.ac.uk](mailto: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.



- ▶ 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 and quantitative social research

- ▶ Survey and experiments: Using computational tools to design survey and simulate responses (see the NORC piece)
- ▶ Natural language processing (NLP): Using computational tools to analyze unstructured text, image or audio data (more later in Masterclass)
- ▶ Agent-based modeling (ABM): Using computational tools to model and simulate the social world and human behaviors (e.g., segregation)



# Out of One, Many: Using Language Models to Simulate Human Samples

Lisa P. Argyle<sup>1</sup>, Ethan C. Busby<sup>1</sup>, Nancy Fulda<sup>2</sup>,  
Joshua R. Gubler<sup>1</sup>, Christopher Rytting<sup>2</sup> and David Wingate<sup>2</sup>

<sup>1</sup>Department of Political Science, Brigham Young University, Provo, UT, USA. e-mail: [lpargyle@byu.edu](mailto:lpargyle@byu.edu), [ethan.busby@byu.edu](mailto:ethan.busby@byu.edu), [jgub@byu.edu](mailto:jgub@byu.edu)

<sup>2</sup>Department of Computer Science, Brigham Young University, Provo, UT, USA. e-mail: [nfulda@cs.byu.edu](mailto:nfulda@cs.byu.edu), [christophermichaelrytting@gmail.com](mailto:christophermichaelrytting@gmail.com), [wingated@cs.byu.edu](mailto: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



# 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/](https://reverie.herokuapp.com/arXiv_Demo/)



# LLMs for coding and programming



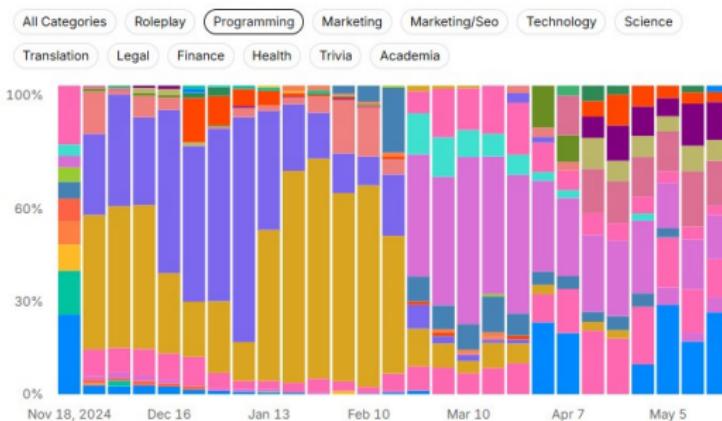
- ▶ Claude Code (<https://www.anthropic.com/claude-code>)
- ▶ Google Gemini Code Assist (<https://codeassist.google/>)
- ▶ OpenAI Codex (<https://chatgpt.com/codex>)



# LLM rankings (Openrouter)

## LLM Rankings

Compare models for programming prompts ⓘ



### Leaderboard

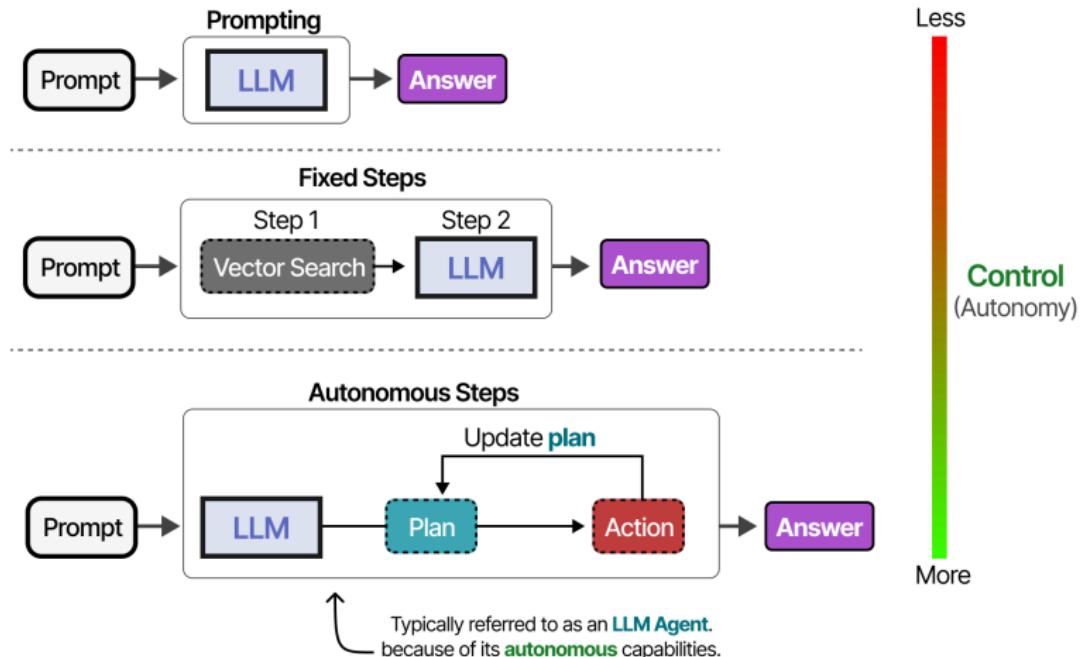
Token usage across models

Top this week ⚠

1.	<a href="#">OpenAI: GPT-4o-mini &gt;</a> GPT-4o mini is OpenAI's newest model after [GPT-4 Omni](/models/...)	13.8B tokens ↑ 10%
2.	<a href="#">Google: Gemini 2.5 Pro Preview &gt;</a> Gemini 2.5 Pro is Google's state-of-the-art AI model designed for ad...	8.72B tokens ↓ 16%
3.	<a href="#">Anthropic: Claude 3.7 Sonnet &gt;</a> Claude 3.7 Sonnet is an advanced large language model with improv...	8.28B tokens ↓ 14%



# AI “agents”



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



## Appendix: GenAI 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: Running critical discourse analysis with AI

- ▶ **Step 1:** Ask AI to produce a brief definition of critical discourse analysis – critically evaluate the response provided by AI
  - What response do you get? Do you agree with it or not? Is it the same as what Margarita discussed last week?
  - Can you try to engage with AI to "improve" the definition?
- ▶ **Step 2:** Upload the "Return to Work" letter and ask AI to carry out a critical discourse analysis based on the definition – what do you get? What if you ask AI to incorporate the "critical disability framework?"
- ▶ **Questions to think about**
  - Are you comfortable with what we have just done? Why not? What might be the caveat (Morgan 2023)?
  - Does it make sense to analyze the "discourse" produced by AI or Human-AI interactions? How?



# Exercise: Analyzing interview transcripts

- ▶ **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: 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



ARTICLE

# Synthetic Replacements for Human Survey Data? The Perils of Large Language Models

James Bisbee<sup>ID</sup>, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel and Jennifer M. Larson

Political Science Department, Vanderbilt University, Nashville, TN, USA

**Corresponding author:** James Bisbee; Email: [james.h.bisbee@vanderbilt.edu](mailto:james.h.bisbee@vanderbilt.edu)

(Received 2 May 2023; revised 18 January 2024; accepted 20 January 2024; published online 17 May 2024)

## Abstract

Large language models (LLMs) offer new research possibilities for social scientists, but their potential as “synthetic data” is still largely unknown. In this paper, we investigate how accurately the popular LLM ChatGPT can recover public opinion, prompting the LLM to adopt different “personas” and then provide feeling thermometer scores for 11 sociopolitical groups. The average scores generated by ChatGPT correspond closely to the averages in our baseline survey, the 2016–2020 American National Election Study (ANES). Nevertheless, sampling by ChatGPT is not reliable for statistical inference: there is less variation in responses than in the real surveys, and regression coefficients often differ significantly from equivalent estimates obtained using ANES data. We also document how the distribution of synthetic responses varies with minor changes in prompt wording, and we show how the same prompt yields significantly different results over a 3-month period. Altogether, our findings raise serious concerns about the quality, reliability, and reproducibility of synthetic survey data generated by LLMs.

**Keywords:** ChatGPT; synthetic data; public opinion; research ethics

**Edited by:** Jeff Gill



# 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>



## Concluding remarks

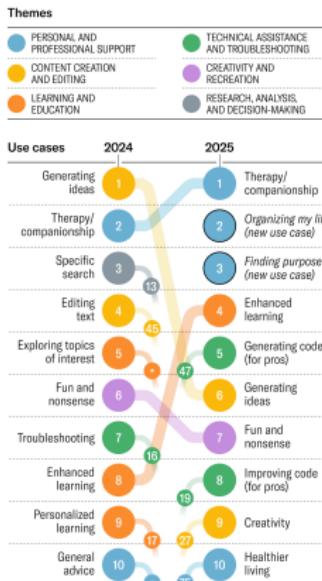
- ▶ The past three weeks only allow us to scratch the surface, and yet we have touched upon how GenAI can be used for three major research tasks: Analysis, coding and simulation
- ▶ Learn how to use AI tools creatively to increase your research productivity; critical thinking remains crucial and should not undermine the agency of human researchers
- ▶ AI hype may as well be a bubble if we look at it again in a decade – use it with caution (and make sure you talk to your)
- ▶ Transparency is key: Explain how and why you use AI carefully in your research so your peers can evaluate the quality/rigour of your work



# "How People Are Really Using Gen AI in 2025" (Havard Business Review)

## Top 10 Gen AI Use Cases

The top 10 gen AI use cases in 2025 indicate a shift from technical to emotional applications, and in particular, growth in areas such as therapy, personal productivity, and personal development.



\*Did not make list of top 100 in 2025

Source: Filtered.com



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



## Assuring an accurate research record

May 16th, 2025

Following the posting of the preprint paper "Artificial Intelligence, Scientific Discovery, and Product Innovation" on arXiv in November 2024, concerns were raised about the integrity of the research. MIT conducted an internal, confidential review and concluded that the paper should be withdrawn from public discourse.

In an effort to correct the research record, MIT has contacted arXiv to formally request that the paper be withdrawn and *The Quarterly Journal of Economics*, where it had been submitted. The letter on behalf of the Committee on Discipline to arXiv states:

"Earlier this year, the COD conducted a confidential internal review based upon allegations it received regarding certain aspects of this paper. While student privacy laws and MIT policy prohibit the disclosure of the outcome of this review, we are writing to inform you that MIT has no confidence in the provenance, reliability or validity of the data and has no confidence in the veracity of the research contained in the paper. Based upon this finding, we also believe that the inclusion of this paper in arXiv may violate arXiv's Code of Conduct."



# Artificial Intelligence, Scientific Discovery, and Product Innovation<sup>\*</sup>

Aidan Toner-Rodgers<sup>†</sup>  
MIT

December 25, 2024

This paper studies the impact of artificial intelligence on innovation, exploiting the randomized introduction of a new materials discovery technology to 1,018 scientists in the R&D lab of a large U.S. firm. AI-assisted researchers discover 44% more materials, resulting in a 39% increase in patent filings and a 17% rise in downstream product innovation. These compounds possess more novel chemical structures and lead to more radical inventions. However, the technology has strikingly disparate effects across the productivity distribution: while the bottom third of scientists see little benefit, the output of top researchers nearly doubles. Investigating the mechanisms behind these results, I show that AI automates 57% of “idea-generation” tasks, reallocating researchers to the new task of evaluating model-produced candidate materials. Top scientists leverage their domain knowledge to prioritize promising AI suggestions, while others waste significant resources testing false positives. Together, these findings demonstrate the potential of AI-augmented research and highlight the complementarity between algorithms and expertise in the innovative process. Survey evidence reveals that these gains come at a cost, however, as 82% of scientists report reduced satisfaction with their work due to decreased creativity and skill underutilization.

