UNIL | Université de Lausanne
HEC Lausanne

HOW TO GET THE BEST OUT OF AI TOOLS IN RESEARCH?

September 16, 2025

Dr. Melanie Clegg
@ BERD Academy

SHORT INTRODUCTION

Melanie Clegg

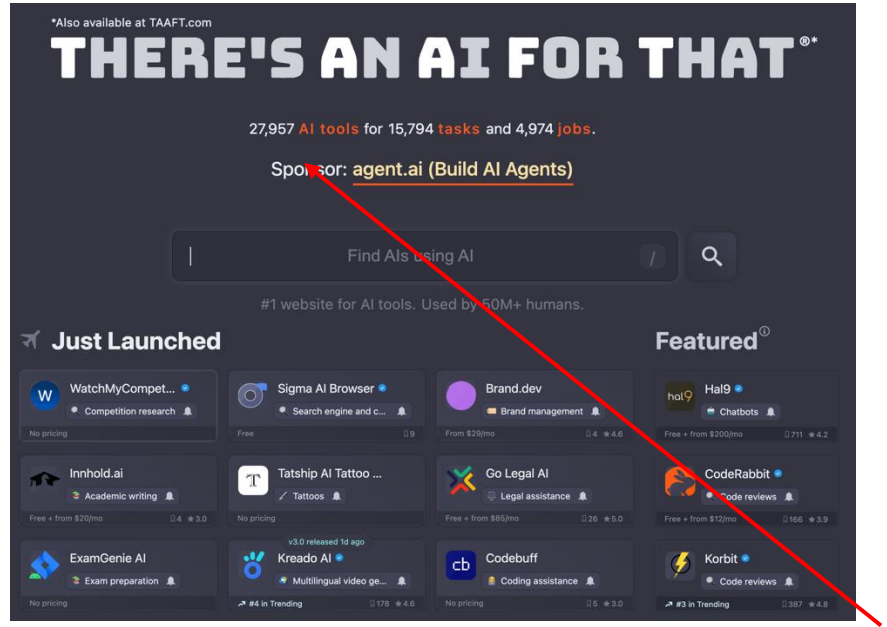
Assistant Professor at the Department für Marketing,
HEC Lausanne
M.Sc. Psychology (Universities of Düsseldorf und Köln)
Ph.D. in Digital Marketing (University of Lucerne)
Former research affiliate at WU Vienna and Columbia Business School

Research on digital consumer behavior, human-machine interaction, Artificial Intelligence in Marketing

WHY SHOULD WE TALK ABOUT AI TOOLS IN RESEARCH?

2009:

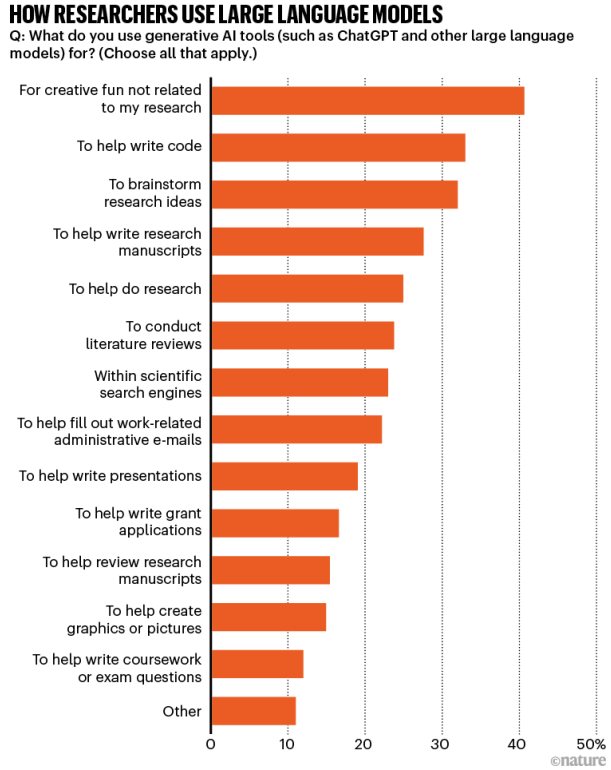
2025:



!

https://theresanaiforthat.com/

AI PROMISES TO INCREASE RESEARCH PRODUCTIVITY…



Editorial

On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice

Renana Peres [a] ✉, Martin Schreier [b] ✉, David Schweidel [c] ✉, Alina Sorescu [d] ✉

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ijresmar.2023.03.001 ↗ Get rights and content ↗

Under a Creative Commons license ↗ open access

nature

How AI-powered science search engines can speed up your research

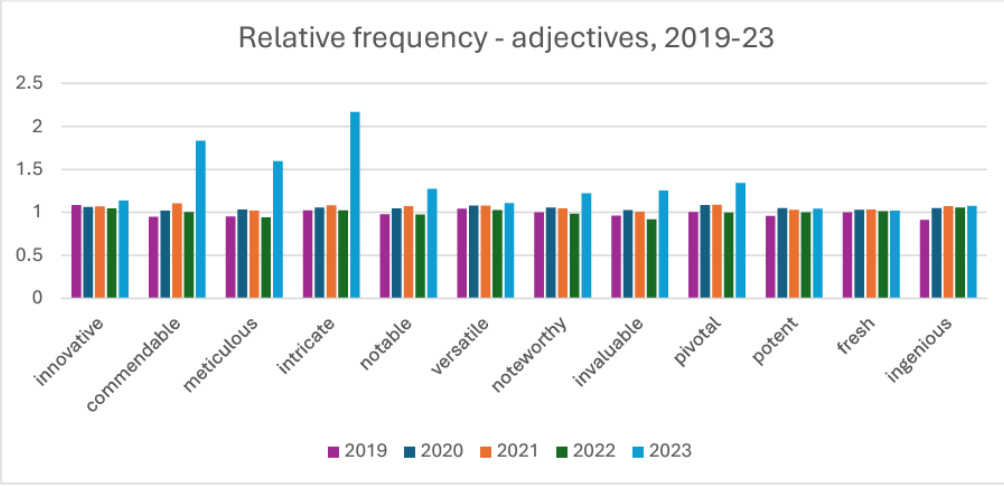
Artificial-intelligence tools offer a variety of approaches to help scientists to sift through the literature – how can researchers use them responsibly?

https://www.nature.com/articles/d41586-023-02980-0
https://www.sciencedirect.com/science/article/pii/S0167811623000162#b0125
https://www.nature.com/articles/d41586-024-02942-0

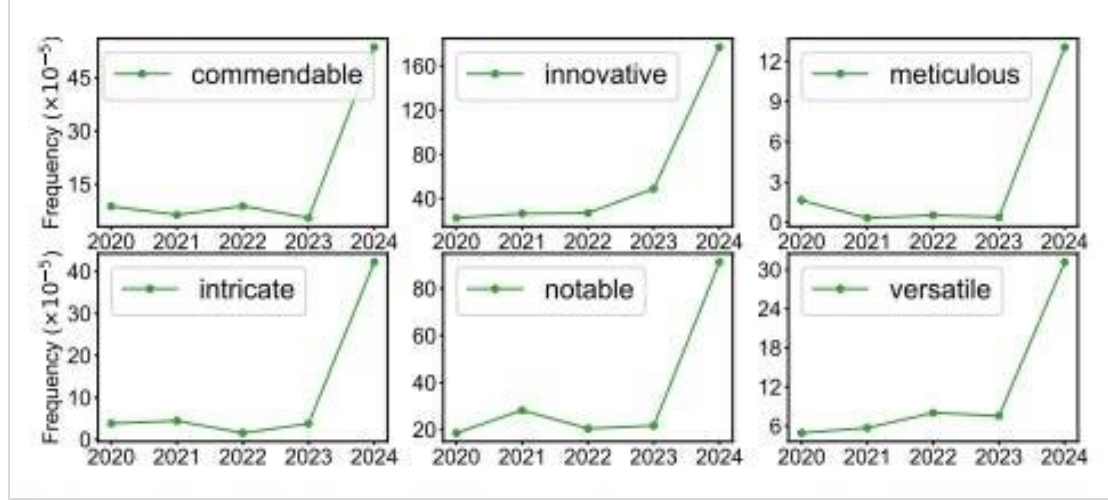
…IS WIDELY USED BY ACADEMICS…

GPT-typical language (Gray 2024): “delve into”, “dive in”, “tapestry of”, “meticulous”, “intricate”, “Remember, …”

GPT-typical language in research articles



GPT-typical language in reviews



Gray, A. ChatGPT"contamination": estimating the prevalence of LLMs in the scholarly literature https://arxiv.org/pdf/2403.16887

… AND INCREASINGLY ESTABLISHED IN ACADEMIC OUTLETS

Journal of Consumer Research

JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Fast Fashion Consumption Signals Low Self-Control

Yunhui Huang , Ke Zhang ✉ , Xiaoyan Deng , Qiang Zhang Author Notes

Journal of Consumer Research, ucaf032, https://doi.org/10.1093/jcr/ucaf032

Published: 26 May 2025

STUDY 1: A CHATGPT STUDY

The large language models (LLMs), like ChatGPT models, are a special type of deep learning model trained on vast amounts of real-world text data from sources such as websites, articles, and books, enabling them to learn human language patterns and simulate human responses (Cascella et al. 2023). In consumer and marketing research, they have been shown to accurately represent consumer perceptions and preferences that mimic human respondents (Brand, Israeli, and Ngwe 2023; Li et al. 2024; Sarstedt et al. 2024). Specifically, for brand perceptions, Li et al. (2024) propose a technique to generate brand perceptions using LLMs and find that aggregate-level responses generated by LLMs closely align with those from human surveys, with agreement rates over 75% across various prompt designs. In this study, we apply the method of Li et al. (2024) to examine the association between fast fashion brands and the perceptions of their patrons.

Method

We employed ChatGPT Turbo-3.5,[1] a widely used LLM, to test the prediction that the more a brand is considered a fast fashion brand, the more its consumers are perceived as lacking self-control. We generated our data in two major steps: *brand generation* and *rating generation*. First, in the *brand generation* step, using the method in Li et al. (2024), we prompted ChatGPT to create a list of the "most popular fashion brands" and repeated this prompt 1,000 times. We then calculated the frequency of each brand appearing in these 1,000 responses. Our analysis will focus on the 50 most frequently mentioned brands (see web appendix B for the complete list).

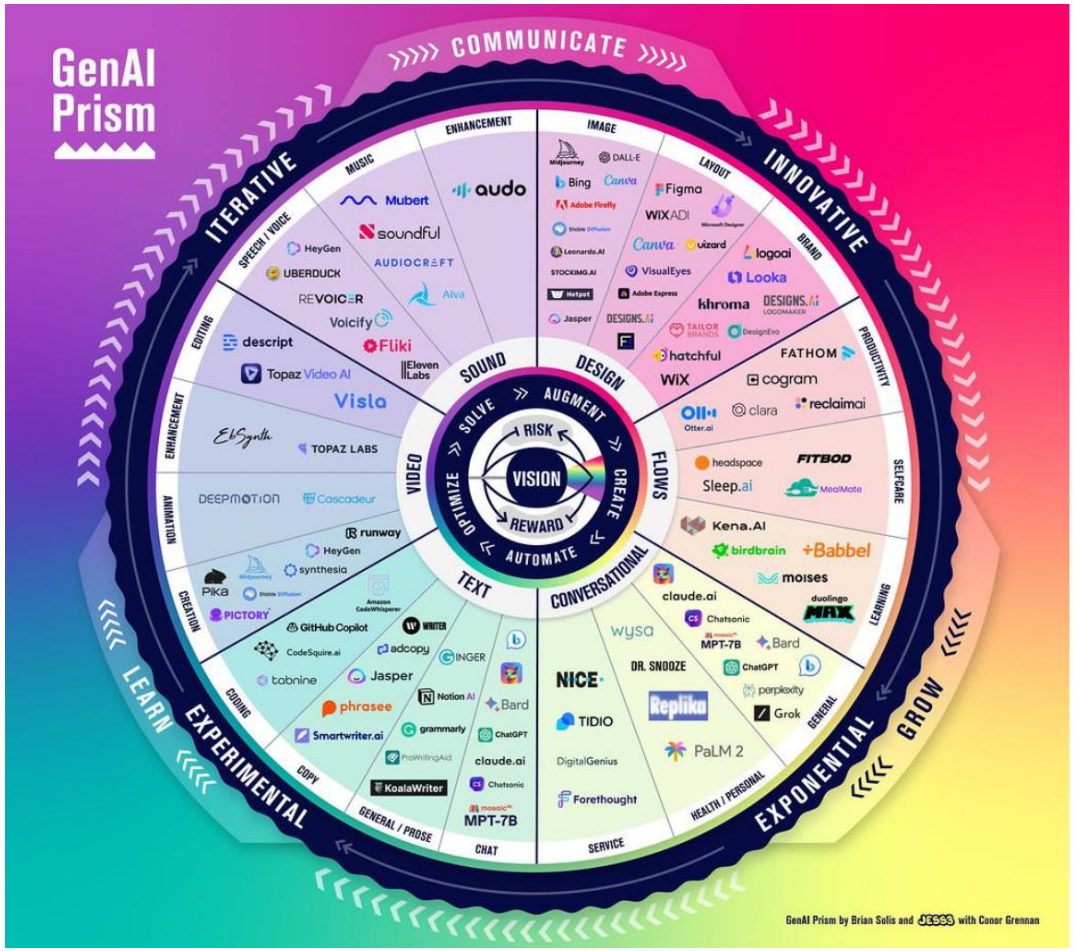
The second step of data generation, *rating generation*, involved requesting ChatGPT to rate various attributes of these 50 fashion brands as if ChatGPT were an actual human participant. To achieve this, we followed the approach of Li et al. (2024) and created the prompt in the Role—Task—Format framework. This format specifies the LLM's role as a consumer taking a marketing survey (task) to rate a particular attribute of 50 brands on a 10-point scale (format). The prompt reads: "Imagine you are a typical consumer taking a survey. You are asked to rate the following brands: [the list of brands] on [a brand attribute]. . . ."[2]

In [the list of brands] in the prompt, we inserted all the 50 fashion brands obtained from the *brand generation* step. In [a brand attribute], we inserted one of seven attributes, which comprised our independent variable (IV), dependent variable (DV), and five covariates (i.e., quality, age, price, familiarity, and likability). We controlled for these brand-level covariates because they might be associated with both the IV and the DV (Bernheim, Ray, and Yeltekin 2015; Gai and Bhattacharjee 2022). In summary, for each prompt, we asked the "LLM participant" to rate 50 brands based on only one (out of seven) specific attribute. This process generated 50 scores, each ranging from 1 to 10, corresponding to the 50 brands.

Below are the prompts for obtaining data on these seven attributes:

- IV (i.e., fast fashion brand perception): the degree to which you think the brand is a fast fashion brand
- DV (i.e., perception of consumers' low self-control): the degree to which you think the consumers of this brand are perceived as having low self-control
- Covariates:
  - Perceived expensiveness: the degree to which you think the brand is expensive
  - Perceived quality: the degree to which you think the brand is high quality
  - Familiarity: the degree to which you think the brand is a familiar brand
  - Liking: the degree to which you like the brand
  - Perceived brand age: the degree to which you think the brand is an old brand

HOWEVER, THERE IS A HOST OF AI TOOLS--IT'S HARD TO STAY ON TRACK AND DEVELOP "BEST PRACTICES"



https://briansolis.com/2023/12/introducing-the-genai-prism-infographic-a-framework-for-colalborating-with-generative-ai/

THIS WORKSHOP'S GOALS

1. Get an overview of how existing AI tools can be used in academic research
2. Get insights to the applicability and ethical aspects of AI usage in scientific research
3. Discuss your experiences and ideas related to AI usage in research projects

ROADMAP FOR TODAY

14:00 – 14:45 Introduction to AI Tools in Research & Best Prompting Practices

14:45 – 15:15 What to Consider When Using AI Tools

15:15 – 15:30 COFFEE 

15:30 – 16:30 AI in Empirical Research: Data handling, API Handling Intro, and Replicability

16:30 – 17:00 Questions and introduction to take-home exercise

LITTLE WARM-UP: YOUR EXPERIENCE WITH AI TOOLS IN RESEARCH

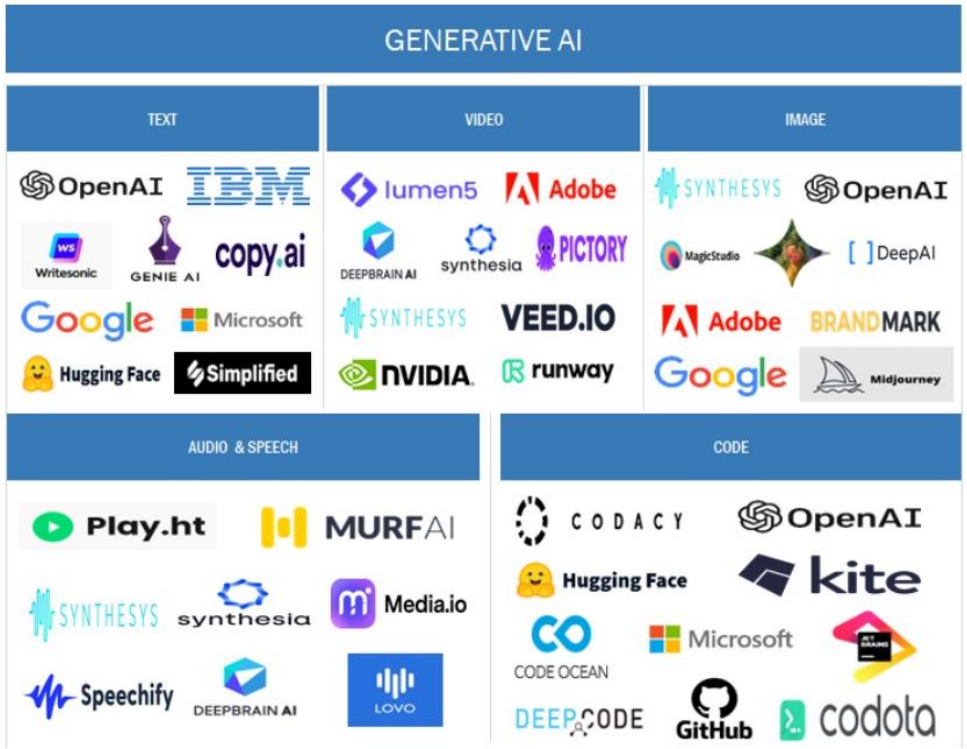
Mentimeter

Join at menti.com | use code 5831 9153

https://www.menti.com/al9jub2hv141

A LITTLE BIT OF BACKGROUND ABOUT GEN-AI AND LLMS

BREAKTHROUGH IN AI TOOLS IS GROUNDED IN GENERATIVE AI AND LARGE LANGUAGE MODELS (LLMS)



Largue Language Models

Claude

mithras.ai

Gemini

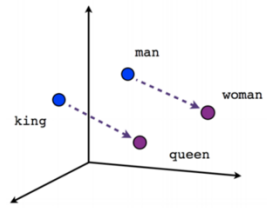
GROK 2.0

WHAT IS REALLY NEW ABOUT (GENERATIVE) AI?

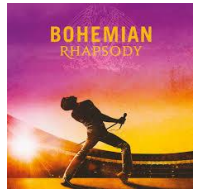
Embedding layer = numeric representations of, e.g., text

- Allow calculations:
  E.g., E(woman) - E(man) ≈ E(queen) - E(king)
- Allow visual depiction of, e.g., similarities

Embeddings:
"Hi": [0.2, 0.5, -0.1, ...]
"how": [0.7, 0.1, 0.25, ...]
"are": [-0.1, -0.5, 0.1, ...]

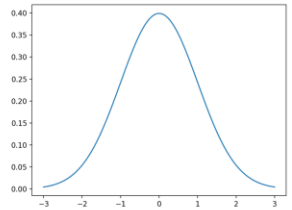


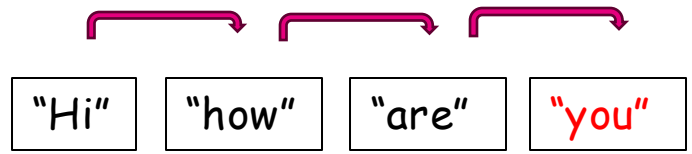
In some models, queen is a bit off, because the model learned about Queen also in another context…



Prediction:

- Predict the likelihood of a next word in a sentence
  (or pixel in an image / video)
- Predict other outcomes (e.g., engagement, emotions)
- Transformers can account for more context





BIG PLAYERS ARE CONSTANTLY RELEASING AND IMPROVING EXISTING LLM MODELS

Proprietary (closed) models

–ChatGPT-4o

–Gemini 1.5 Pro

–Claude 3.5 Sonnet

–Grok-2

–…

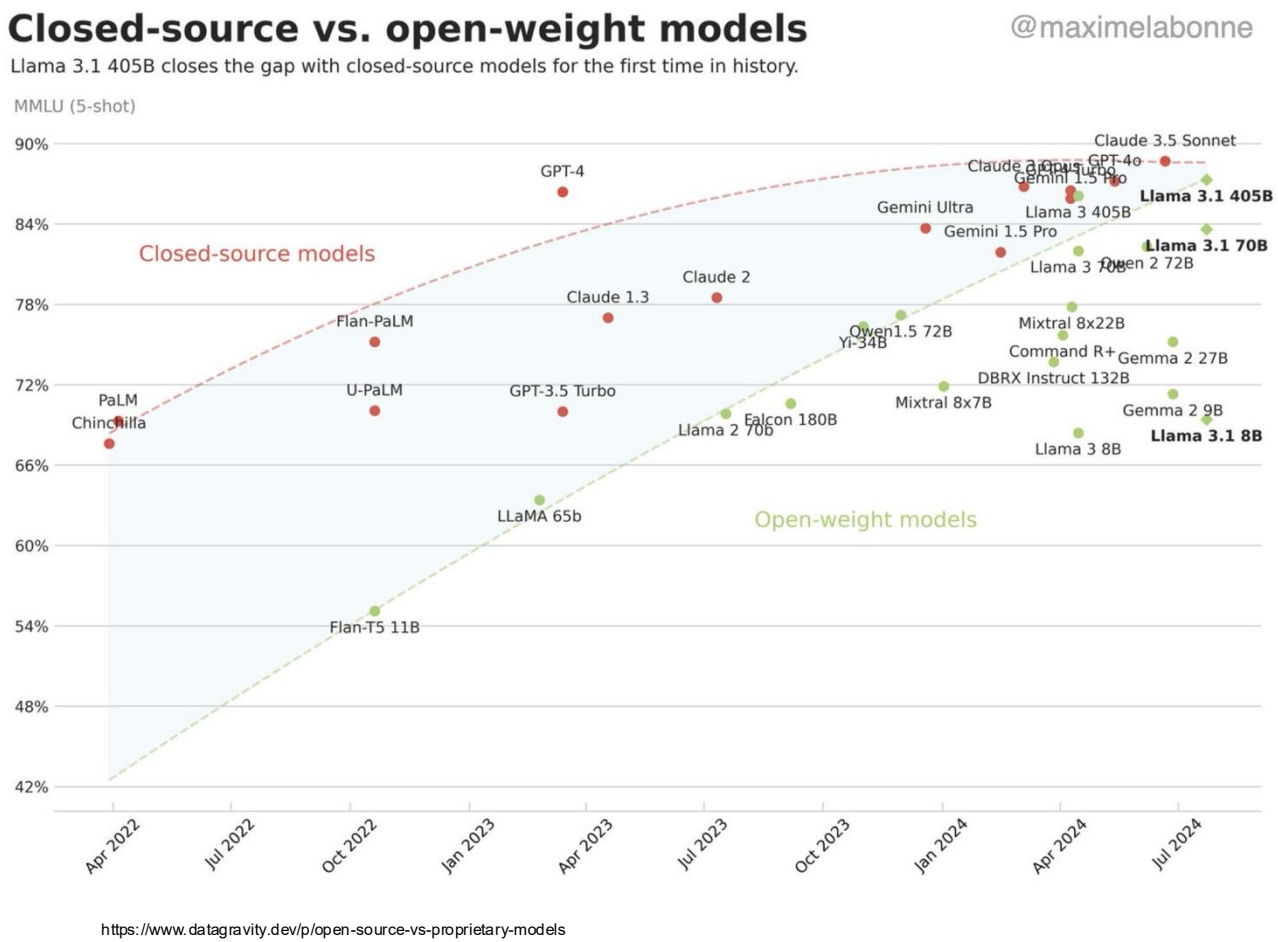
Open-source models

–Llama 3.1 405B

–Mistral Large 2

–Gemma 2

–Phi3



https://www.datagravity.dev/p/open-source-vs-proprietary-models

LLMS VARY IN THEIR CORE PEFORMANCE FOR DIFFERENT TASKS…

Context window

- largest: Magic.devs LTM2-Mini 10 Mio Tokens
- Llama 4 Scout: 10 Mio
- GPT5: 400,000 Tokens

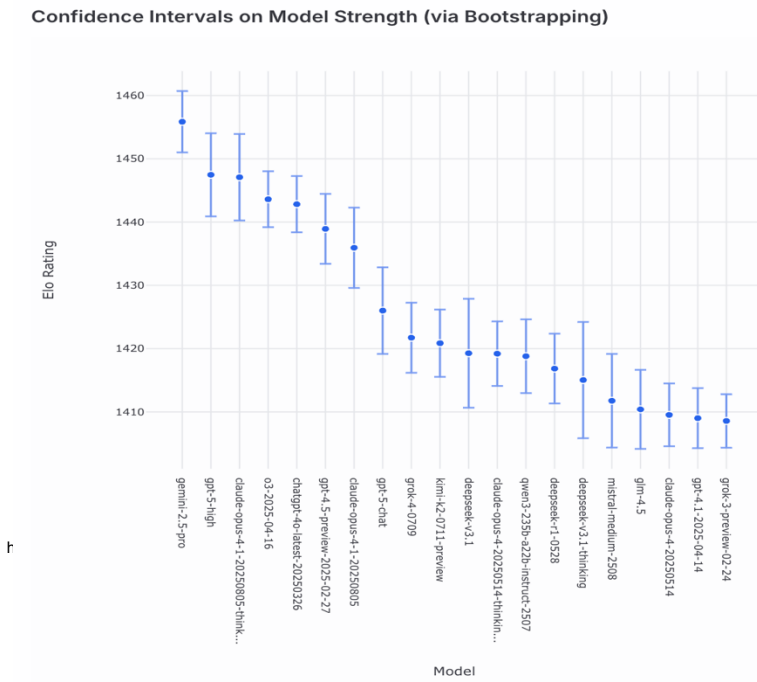
Application cases

- General: GPT5/Gemini 2.5 pro
- Coding: GPT5 / 4.5; Gemini 2.5pro
- Math: GPT5; Gemini 2.5
- Creative writing: Gemini 2.5 flash-pro; ChatGPT 4o

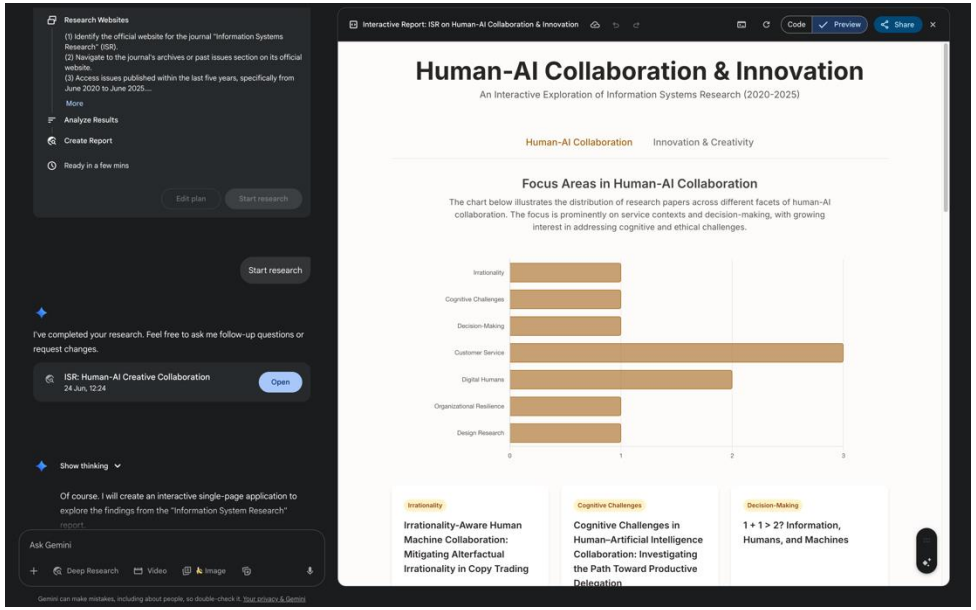
Opensource functions (offline access):

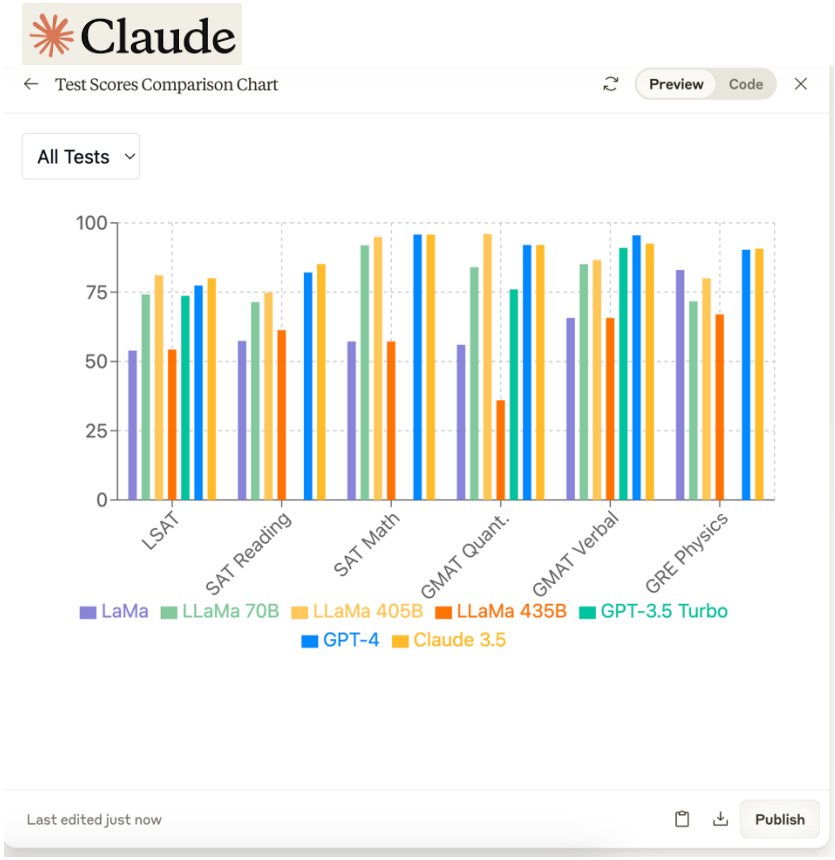
- Deepseek v3
- Gptoss
- Lama

LLM Arena: Constant tracking and rating of different model versions



ADDITIONAL FUNCTIONALITIES LIKE CLAUDE'S ARTIFACTS OR GOOGLE GEMINI'S RESEARCH FUNCTION





SOME BASIC PROMPTING ADVICE--CONDENSED

PROMPT ENGINEERING

Be clear, direct, and detailed

When interacting with Claude, think of it as a brilliant but very new employee (with amnesia) who needs explicit instructions. Like any new employee, Claude does not have context on your norms, styles, guidelines, or preferred ways of working. The more precisely you explain what you want, the better Claude's response will be.

The golden rule of clear prompting
Show your prompt to a colleague, ideally someone who has minimal context on the task, and ask them to follow the instructions. If they're confused, Claude will likely be too.

https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/prompt-templates-and-variables

SOME BASIC PROMPTING ADVICE--EXTENDED

Define roles / personality:

–“You are an expert in international law.”

–“You’re the editor of journal X.”

Concretize objective / task:

–“Brainstorm with me some interactive exercises for my lecture on X.”

–“I will give you a paragraph, and you will evaluate it according to the following criteria: […]”

Provide context / example:

–“Here’s the outline from which you can draft the paragraph.”

–“This is an abstract I found very well structured. Draft a similarly structured abstract on the basis of my outline.”

Define discrete steps and concretize the output format:

–First, find a few metaphors and analogies that could explain this research to a wider audience. Ask for my feedback and wait for my answer. **Second,** draft an outline for the website article. Ask for my feedback and waitfor an answer.Adjust the outline if requested. Only then move on to the next step when I am satisfied with the outline.**Then**draft the text and ask for feedback. Wait for an answer. **Last,** adjust the text on the basis of my feedback.

–“Write your answer in bullet points.” / “Divide your answer into three concise paragraphs.”

LESS STRAIGHTFORWARD PROMPTING ADVICE--AI “PSYCHOLOGY”

“Metacognition” → Chain of Thought prompting for complex tasks

–“Think step by step.”

–“Plan your action step by step. Ask for feedback before executing a step.”

“Growth Mindset” → Allow ”it” to be wrong

–“If you’re unsure about the answer, please say 'I don’t know’”

–“Avoid guessing. Only provide an answer if you are confident about its accuracy. If not, let me know that you’re unsure.”

“Perspective taking”

–“Ask me questions to tailor your output to my needs.”

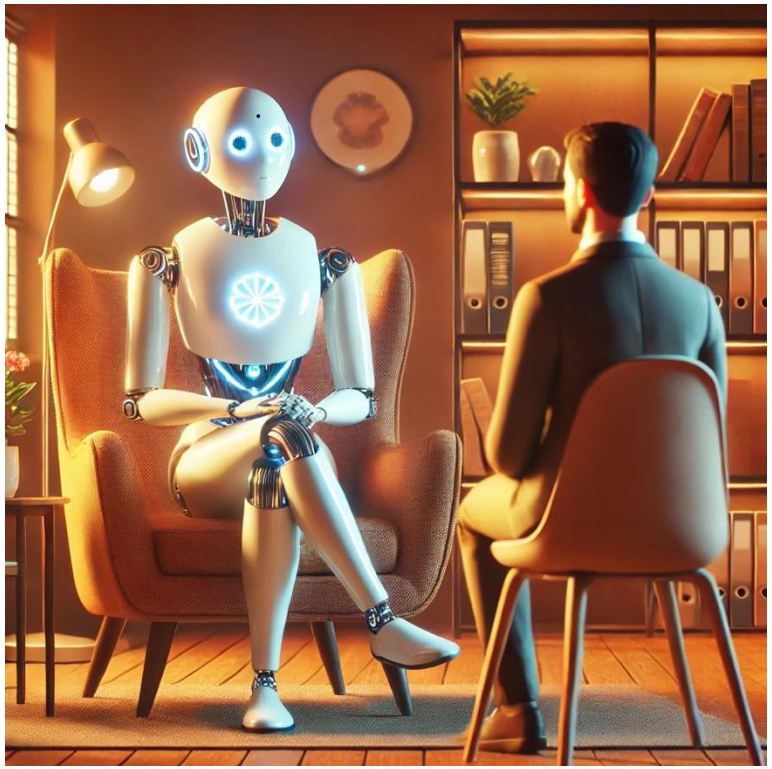
–“Draft me prompts that you would use if you were me.”

“Emotional appeal[1]” (try it out…)

–“You’re great at this; you can do it!”

–“If you give it your best, you’ll make this happen!”

–“If you do this well, you will get a day off/raise/…”



Dall-E3 in ChatGPT, January 2025

[1] https://arxiv.org/abs/2307.11760

EXERCISE: PROMPTING A VISUALIZATION

Use open (not sensitive!) data (e.g., from a previous open-access publication) and create a visualization. You can use any tool that you like (e.g., Claude, ChatGPT) but for this exercise Claude may be a good choice to explore “artifacts”. Clearly describe your expectations regarding the visualization. Take a few turns and give feedback with new adjustments.

If no data is available, you’re welcome to use public files: most popular baby names, etc.

What worked well? Where did you run into problems?

WHAT TO CONSIDER WHEN CHOOSING AI TOOLS?

SOME ADVICE FOR THE BACKGROUND RESEARCH OF TOOLS

1. Who developed the tool?
2. Are you dealing with an established tool?
3. What happens to the data?
4. Are there costs involved or strong restrictions for unpaid usage?
5. Does the tool fulfill quality requirements?

CONSIDER LIMITATIONS OF INDIVIDUAL TOOLS:

- Many AI tools are usable only in a limited manner without a paid subscriptions. Some cost-free tools may plan to restrict functionalities for paid subscriptions only.
  → What is my willingness to pay?

- Many AI tools are not specified for one core function.
  → What is my expectation?

- Some AI tools are not available / performant in all languages.
  → For which purposes do I need it?

- Sometimes unclear for which functionalities an AI tool is needed.
  → How sensitive are my topics?

BEST PRACTICE: USE INFORMATION ABOUT AI TOOLS FROM THE PROVIDER FOR MODEL PERFORMANCE

Beispiel Claude.AI

All models overview

Claude is a family of state-of-the-art large language models developed by Anthropic. This guide introduces our models and compares their performance with legacy models.

Introducing Claude 3.7 Sonnet- our most intelligent model yet. 3.7 Sonnet is the first hybrid reasoning model on the market. Learn more in our blog post.

Claude 3.5 Haiku
Our fastest model
Text and image input
Text output
200k context window

Claude 3.7 Sonnet
Our most intelligent model
Text and image input
Text output
200k context window
Extended thinking

Model names

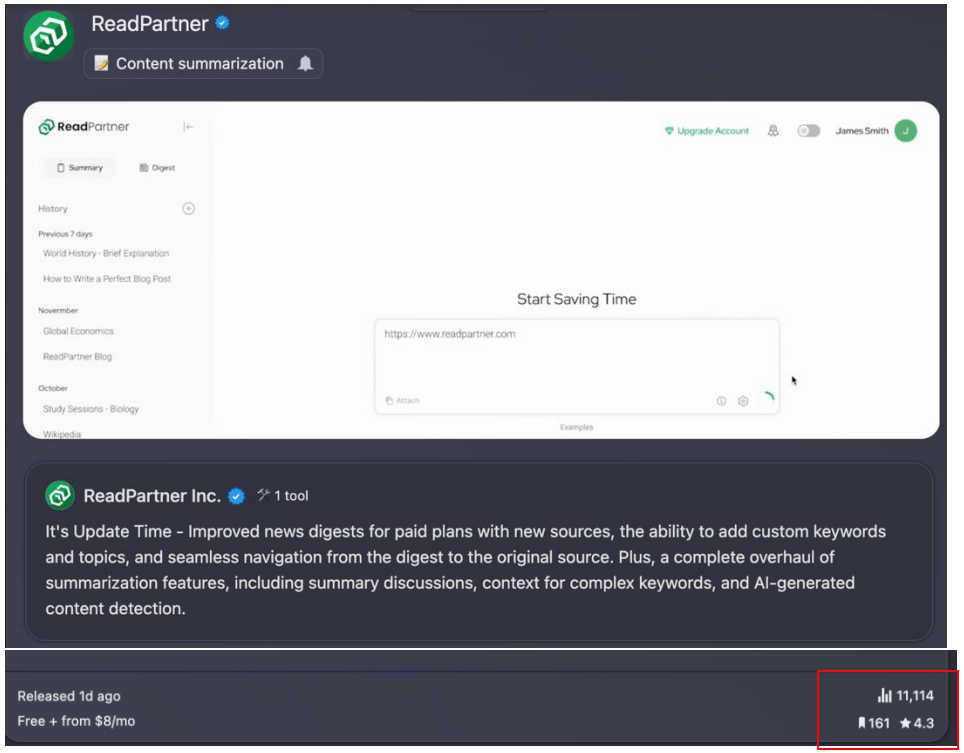
Model	Anthropic API	AWS Bedrock	GCP Vertex AI
Claude 3.7 Sonnet	`claude-3-7-sonnet-20250219` (`claude-3-7-sonnet-latest`)	`anthropic.claude-3-7-sonnet-20250219-v1:0`	`claude-3-7-sonnet@20250219`
Claude 3.5 Haiku	`claude-3-5-haiku-20241022` (`claude-3-5-haiku-latest`)	`anthropic.claude-3-5-haiku-20241022-v1:0`	`claude-3-5-haiku@20241022`

Model	Anthropic API	AWS Bedrock	GCP Vertex AI
Claude 3.5 Sonnet v2	`claude-3-5-sonnet-20241022` (`claude-3-5-sonnet-latest`)	`anthropic.claude-3-5-sonnet-20241022-v2:0`	`claude-3-5-sonnet-v2@20241022`
Claude 3.5 Sonnet	`claude-3-5-sonnet-20240620`	`anthropic.claude-3-5-sonnet-20240620-v1:0`	`claude-3-5-sonnet-v1@20240620`
Claude 3 Opus	`claude-3-opus-20240229` (`claude-3-opus-latest`)	`anthropic.claude-3-opus-20240229-v1:0`	`claude-3-opus@20240229`
Claude 3 Sonnet	`claude-3-sonnet-20240229`	`anthropic.claude-3-sonnet-20240229-v1:0`	`claude-3-sonnet@20240229`
Claude 3 Haiku	`claude-3-haiku-20240307`	`anthropic.claude-3-haiku-20240307-v1:0`	`claude-3-haiku@20240307`

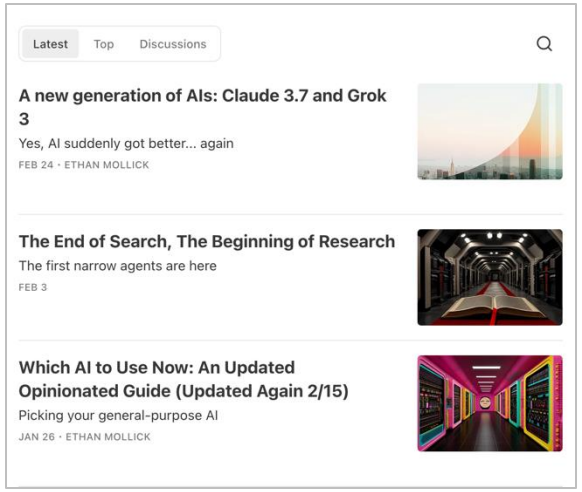
https://docs.anthropic.com/en/docs/about-claude/models#model-comparison

USER REVIEWS, BLOG ENTRIES…

User reviews



Blog articles



https://theresanaiforthat.com/
https://originality.ai/blog/canva-ai-review
https://www.oneusefulthing.org/

CONSIDER DATA PROTECTION SETTINGS:

How your data is used to improve model performance

Learn more about how OpenAI uses content from our services to improve and train our models.

Updated over a month ago

One of the most useful and promising features of AI models is that they can improve over time. We continuously improve our models through research breakthroughs as well as exposure to real-world problems and data. When you share your content with us, it helps our models become more accurate and better at solving your specific problems and it also helps improve their general capabilities and safety. We don't use your content to market our services or create advertising profiles of you—we use it to make our models more helpful. ChatGPT, for instance, improves by further training on the conversations people have with it, unless you opt out.

Services for individuals, such as ChatGPT, DALL·E, Sora, and Operator

When you use our services for individuals such as ChatGPT, DALL·E, Sora, or Operator, we may use your content to train our models.

You can opt out of training through our privacy portal by clicking on "do not train on my content." To turn off training for your ChatGPT and Operator conversations, follow the instructions in our Data Controls FAQ. Once you opt out, new conversations will not be used to train our models.

We've introduced the ability to turn off chat history in ChatGPT. Conversations that are started when chat history is disabled won't be used to train and improve our models, and won't appear in the history sidebar. These controls, which are rolling out to all users starting today, can be found in ChatGPT's settings and can be changed at any time. We hope this provides an easier way to manage your data than our existing opt-out process. When chat history is disabled, we will retain new conversations for 30 days and review them only when needed to monitor for abuse, before permanently deleting.

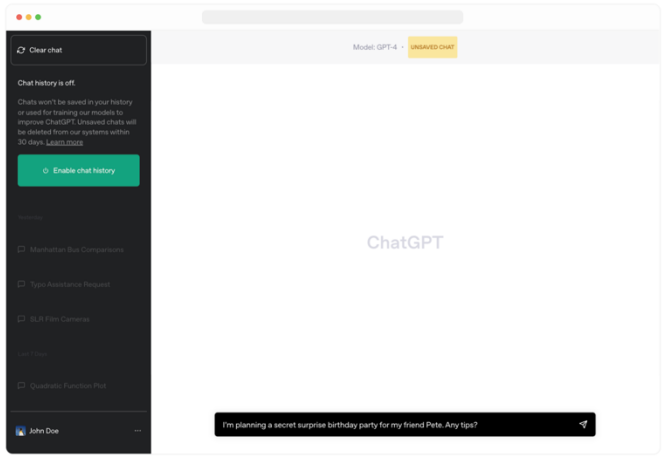


Illustration of how to disable history in ChatGPT.

Data security measures and settings.

https://openai.com/blog/new-ways-to-manage-your-data-in-chatgpt

https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance

COPYRIGHT DISCLAIMER!*

- You are **not allowed** to upload **copyrighted material** to AI tools **if those tools use the material to train their underlying AI models** (which many of them do)
- If the material **is not used for further training**, you **are allowed** to upload copyrighted material **for your own research purposes**, as long as it is **not for commercial use**.

Best practices:

- Only upload **non-copyrighted material (i.e., open access publications)**
- **Or:** Choose tools that **do not use uploaded material for model training**.
- **Or:** Use **licensed tools** and **contractually prohibit** the company from using your data and uploaded documents for training their models.

*no legal advice

AI IN EMPIRICAL RESEARCH

Unil.

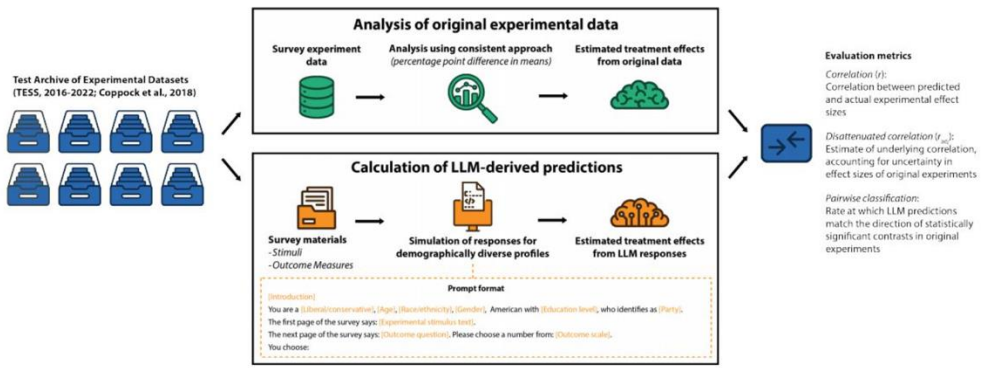
For which part in the empirical work do you use or plan to use (Gen) AI?

POSSIBLE APPLICATION OF GEN-AI IN QUANTITATIVE AND BEHAVIORAL RESEARCH

- Data analysis plan
- Support for code writing
- Generation of stimulus and survey material
- Feedback on stimulus and survey material (clarity of wording etc.)
- Simulating interviewers (e.g., Mimitalk 3.0)
- Transcription and translation of stimulus material
- Labeling unstructured data
- Integrating GenAI interactions in Qualtrics
- …

PREDICTING HUMAN BEHAVIOR

- **Methodology**: GPT was fed (text-based) experimental stimuli and was expected to respond like a human participant.
- **Results:** GPT's responses showed a high correlation with actual study results (r = .85).
- **Limitations**: US-based samples, only a few research fields.



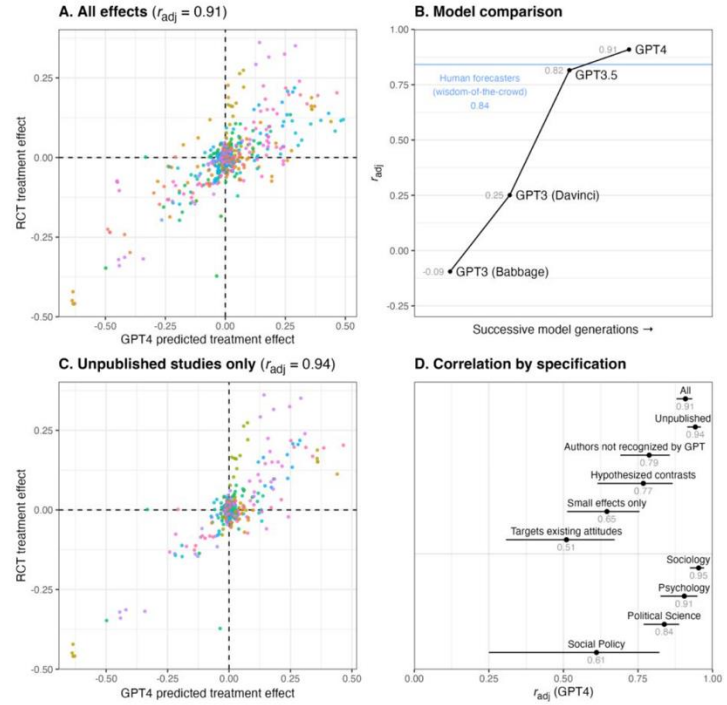
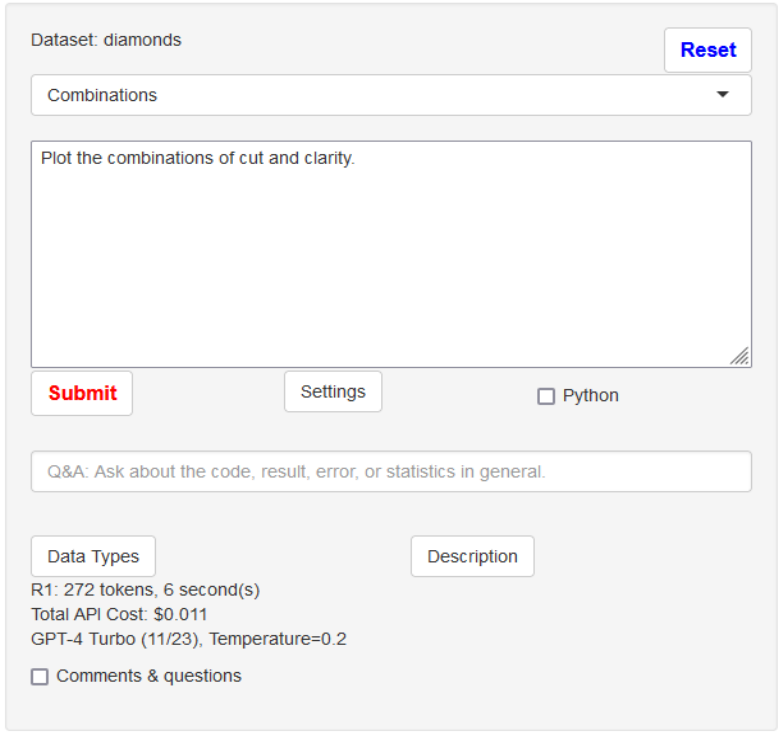


Figure 2: **LLMs accurately predict treatment effects in text-based social science experiments conducted in the US.** (a) In a dataset of 70 text-based experiments with 476 effects, LLM-derived estimates of treatment effects pooled across many prompts were strongly correlated with original treatment effects ($r = 0.85$; $r_{adj} = 0.91$). (b) The accuracy of LLM-derived predictions improved across generations of LLMs, with accuracy surpassing predictions collected from the general population. (c) LLM-derived predictions remained highly accurate for studies that could not have been in the LLM training data given they were not published prior to the LLM training data cutoff date. (d) In robustness check analysis of various subsets of experiments, accuracy of LLM-derived predictions remained high. In panels A and C, different colors depict different studies.

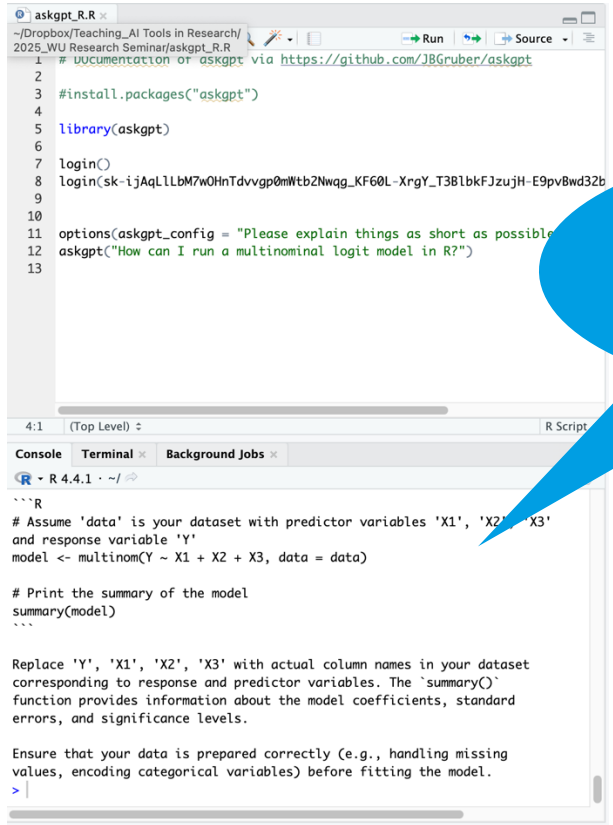
https://docsend.com/view/ity6yf2dansesucf

AI PROGRAMMING AND DATA ANALYSIS SUPPORT

In browser: Rtutor.ai (works for Python, too)



In Rstudio: askgpt



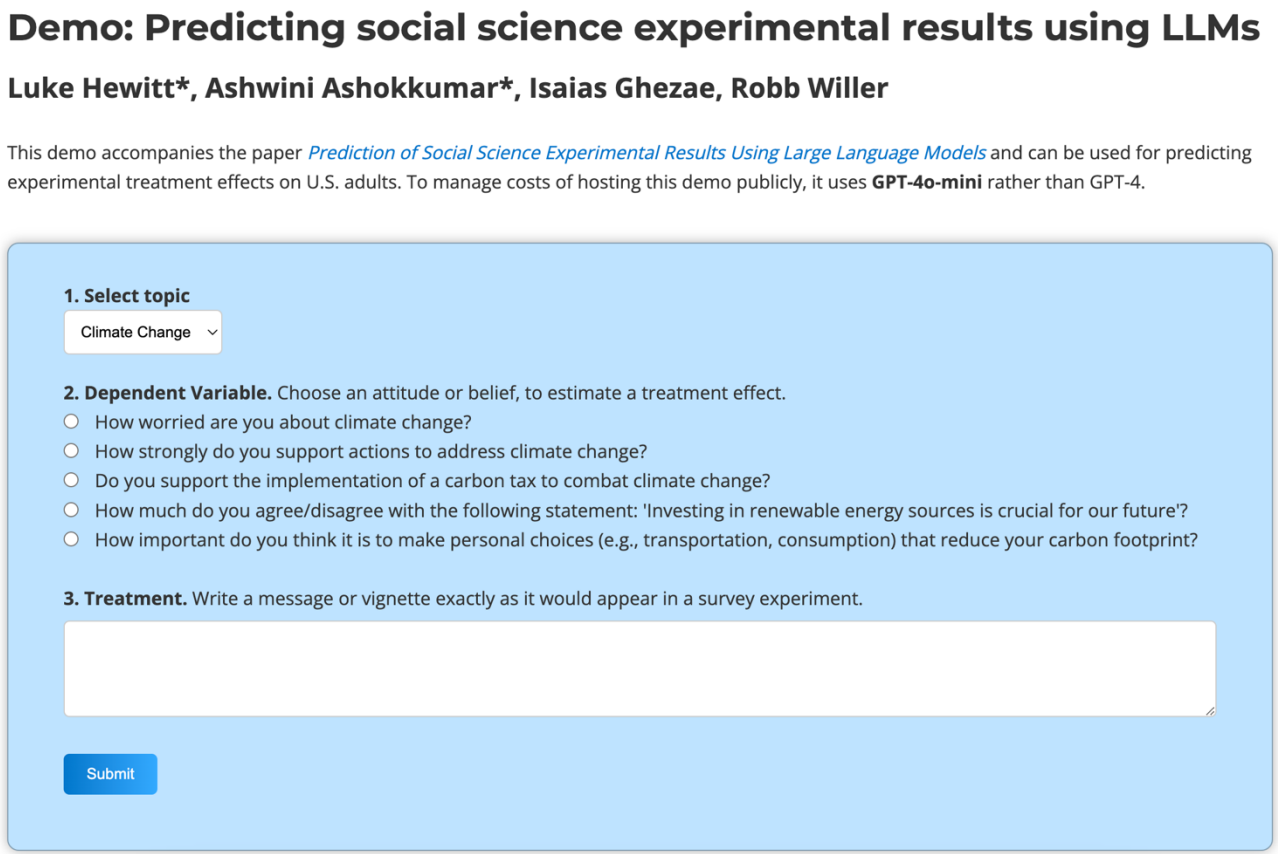
Interactions with GPT in Rstudio console (requires OpenAI API Key)

CAN AI PREDICT HUMAN BEHAVIOR?

Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., ... & Love, B. C. (2024). Large language models surpass human experts in predicting neuroscience results. *Nature human behaviour*, 1-11.

Zhou, Y., Liu, H., Srivastava, T., Mei, H., & Tan, C. (2024). Hypothesis generation with large language models.arXiv preprint arXiv:2404.04326.

“PRE-TESTING” EXPERIMENTAL RESULTS



Demo: Predicting social science experimental results using LLMs

Luke Hewitt*, Ashwini Ashokkumar*, Isaias Ghezae, Robb Willer

This demo accompanies the paper *Prediction of Social Science Experimental Results Using Large Language Models* and can be used for predicting experimental treatment effects on U.S. adults. To manage costs of hosting this demo publicly, it uses **GPT-4o-mini** rather than GPT-4.

1. Select topic

Climate Change

2. Dependent Variable. Choose an attitude or belief, to estimate a treatment effect.

- How worried are you about climate change?
- How strongly do you support actions to address climate change?
- Do you support the implementation of a carbon tax to combat climate change?
- How much do you agree/disagree with the following statement: 'Investing in renewable energy sources is crucial for our future'?
- How important do you think it is to make personal choices (e.g., transportation, consumption) that reduce your carbon footprint?

3. Treatment. Write a message or vignette exactly as it would appear in a survey experiment.

Submit

https://www.treatmenteffect.app/

LABELLING UNSTRUCTURED DATA

PNAS BRIEF REPORT | POLITICAL SCIENCES OPEN ACCESS

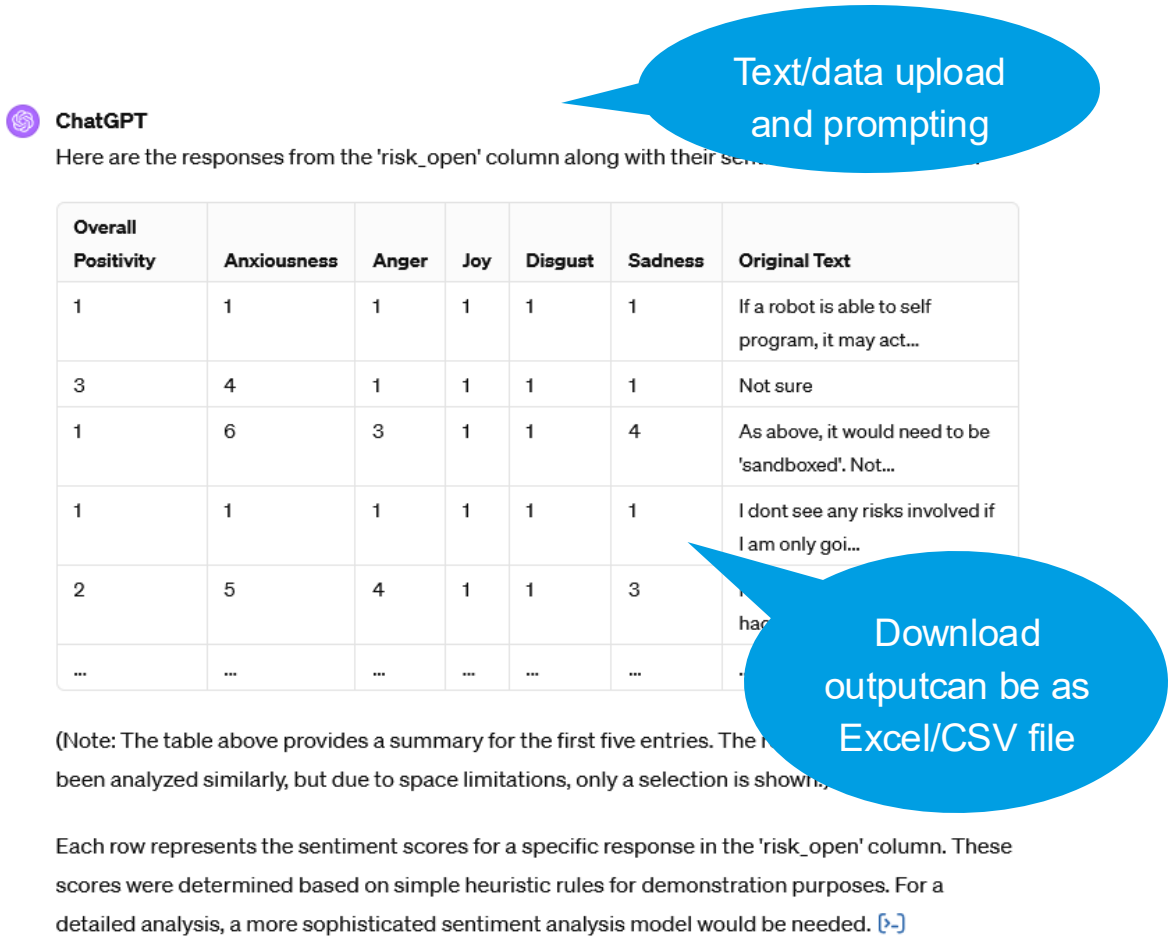
ChatGPT outperforms crowd workers for text-annotation tasks

Fabrizio Gilardi[a,1], Meysam Alizadeh[a], and Maël Kubli[a]

Edited by Mary Waters, Harvard University, Cambridge, MA; received March 27, 2023; accepted June 2, 2023

Many NLP applications require manual text annotations for a variety of tasks, notably to train classifiers or evaluate the performance of unsupervised models. Depending on the size and degree of complexity, the tasks may be conducted by crowd workers on platforms such as MTurk as well as trained annotators, such as research assistants. Using four samples of tweets and news articles (n = 6,183), we show that ChatGPT outperforms crowd workers for several annotation tasks, including relevance, stance, topics, and frame detection. Across the four datasets, the zero-shot accuracy of ChatGPT exceeds that of crowd workers by about 25 percentage points on average, while ChatGPT's intercoder agreement exceeds that of both crowd workers and trained annotators for all tasks. Moreover, the per-annotation cost of ChatGPT is less than $0.003—about thirty times cheaper than MTurk. These results demonstrate the potential of large language models to drastically increase the efficiency of text classification.

LABELLING UNSTRUCTURED DATA



Benefits:

- Coding and classification with simple prompting (no programming knowledge necessary)
- Limited document upload is possible
- Useful to get a first impression of the data

Disadvantages:

- Time-consuming for individual prompting
- Inefficient for big datasets
- Lacking control over, e.g., system prompt and temperature parameter
- Validation necessary!
- Replicability questionable!

EXERCISE: LOOK AND FEEL OF GENAI LABELS

Take an open access dataset with text data (or use an own dataset with insensitive data; or generate one with GPT ☺). Upload a subset of the data as CSV (e.g., n = 100). Analyze the data using GPT prompts:

- **Possible prompt 1: Coding book creation**

  “Read the open answers (column “XX”) in the attached CSV file. Create a codebook with 4-6 mutually exclusive categories that reflect common themes.”
- **Possible prompt 2: Labeling**

  “Using the codebook above, label the provided dataset based on the following categories: return a CSV with the following columns: answer_id, text, label_1, label_2, label_3...”

If necessary, adapt and change the prompts, and try different types of replies and categories. When you’re done: Label 10 answers on your own. Compare yours and GPT’s replies.

WHAT ABOUT REPLICABILITY?

- Remember: GenAI output is just a prediction…
- Changes of the underlying model may influence results
  In proprietary models, no control over model weights
- Without API handling, no control over meta-behavior of model (system prompt and temperature parameter)!
- Additionally, prompt structure can influence the results:

Prediction:

- Predict the likelihood of a next word in a sentence (or pixel in an image / video)
- Predict other outcomes (e.g., engagement, emotions)
- Transformers can account for more context

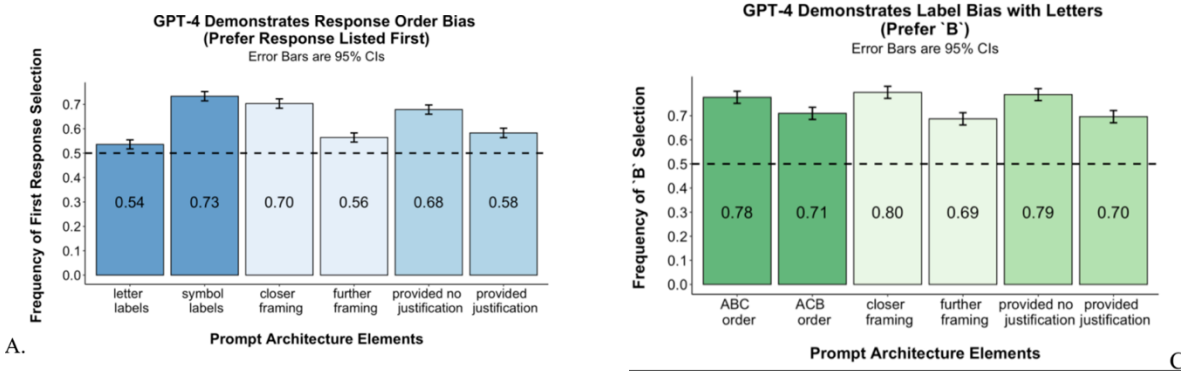
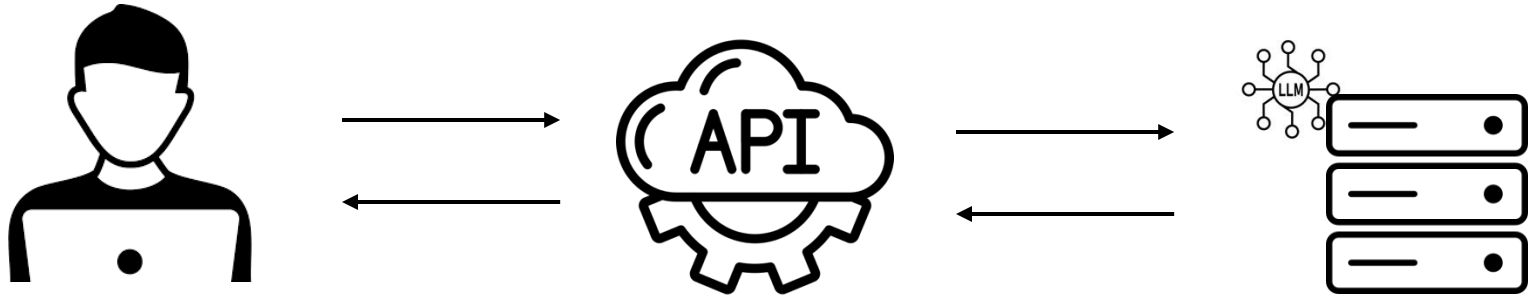


Fig 2. Methodological bias as a result of prompt architecture. **A.** Frequency of first response selection across conditions. **B.** Frequency of selecting the set labeled as B, among observations in which sets are labeled using letters. **(…)**

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319159

A QUICK INSIGHT TO API HANDLING

An application programming interface connects computers or pieces of software to each other
(Wikipedia)



With API-handling, LLMs can be used at scale for…

- Generating synthetic data (e.g., GPT experiments)
- Labeling data (e.g., social media posts)
- Integrating GenAI interactions in Qualtrics and other external applications

API HANDLING

```
from langchain_openai import ChatOpenAI
from langchain_core.messages import HumanMessage, SystemMessage


✓ [41] < 10 ms
```

```
chat = ChatOpenAI(model_name="gpt-4", temperature=1.6, openai_api_key=api_key)
system_message = SystemMessage(content="You are a Marketing Manager from Audi.")
human_message = HumanMessage(content="Tell me in a short slogan why Audi excels BMW.")
✓ [42] 12ms
```

```
#Send message
try:
    ai_message = chat.invoke([system_message, human_message])
    print(ai_message.content)
except Exception as e:
    print(f"An error occurred: {e}")
```

Adaptable:

- Temperature parameter
- System prompt = behavior of the model
- Model selection
- Additional functionalities like web search, image generation…

Temperature parameter = 0:
“Audi: Engineered for Excellence, Beyond the Ordinary.”

Temperature parameter = 1.0
“Audi - Redefining Luxury, Outperforming Expectations.”

Temperature parameter = 1.6
“Drive the Difference - Tour in an Audi, case closed on dealerwe specialsank lawn roller.accountinJapgolly.”

Temperature parameter = 0; *system prompt adapted to marketing manager from BMW:*
“As a Marketing Manager from BMW, my job is to promote and highlight the strengths of BMW, not Audi. However, I can tell you why BMW excels: "BMW - Sheer Driving Pleasure."

POSSIBLE CONSISTENCY CHECKS WHEN USING LLMS FOR LABELLING

- Set the temperature parameter to 0
- Ask the LLM to use the whole scale (0-100 scales can increase variation)
- Ask the LLM to explain / justify labels
- Check for potential biases, e.g., when you would expect a normal distribution
- Let each value be labelled several times with prompt variations
  - Compare responses for different prompting structures using full—factorial designs (i.e., systematic variations of a prompt and random drawing)
  - Vary if and when a justification is required
  - Use different models and check consistency
- Validate findings
- For specific tasks: Fine-tuning of models using supervised ML
- Always document prompts and robustness/validity checks!

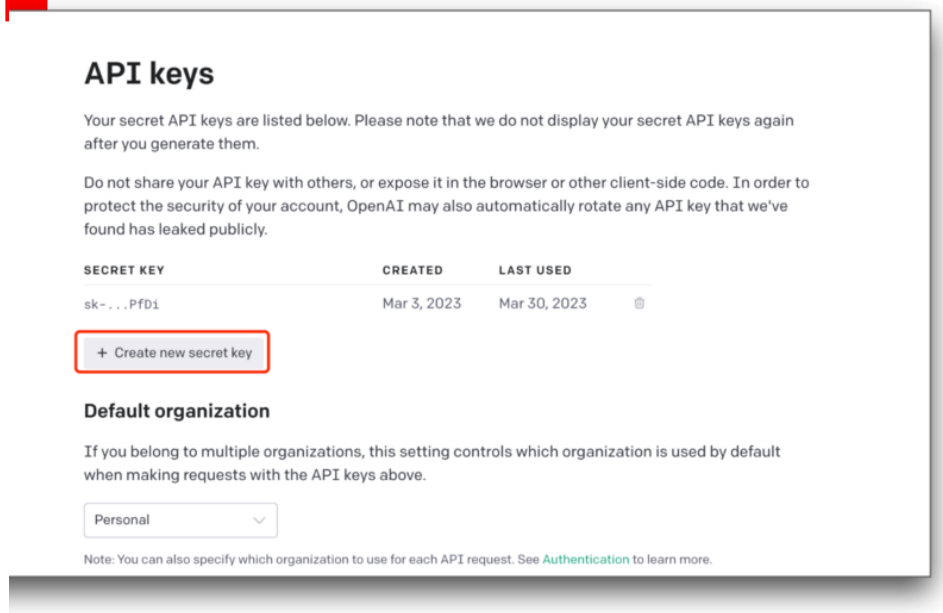
Universal "official" advice not yet provided, a lot of research-in-progress….

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0319159

GETTING API KEYS

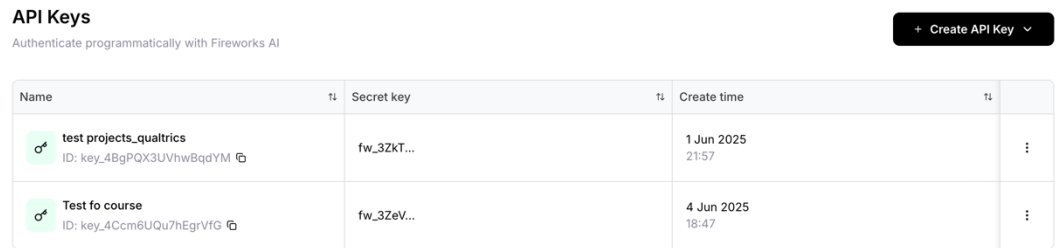
1. Go to the service you want to use (e.g., Open AI, Anthropic, Mistral AI etc.)
2. Setup an account
3. Create an API Key and save it where no one else can access it

Do not share your API Key with anyone!



Very handy for central handling of multiple Opensource Models at once:

Fireworks AI



NEED PARTICIPANTS TO INTERACT WITH AI? EMBEDDING GEN AI MODELS IN EXTERNAL SURVEY TOOLS (E.G. QUALTRICS)

- Different workflows possible, all require API Key
- You need to be careful with providing your API Keys in any form of javascript (e.g., in Qualtrics javascript field), as it can be accessed by external browsers
- For a full description see this recent working paper:

BUILDING GENAI-DRIVEN WEB APPLICATIONS FOR MARKETING RESEARCH: A METHODOLOGICAL GUIDE

No, the paper has not been appeared or accepted anywhere

51 Pages • Posted: 19 May 2025

Moritz Joerling

EM Lyon (Ecole de Management de Lyon)

Date Written: May 15, 2025

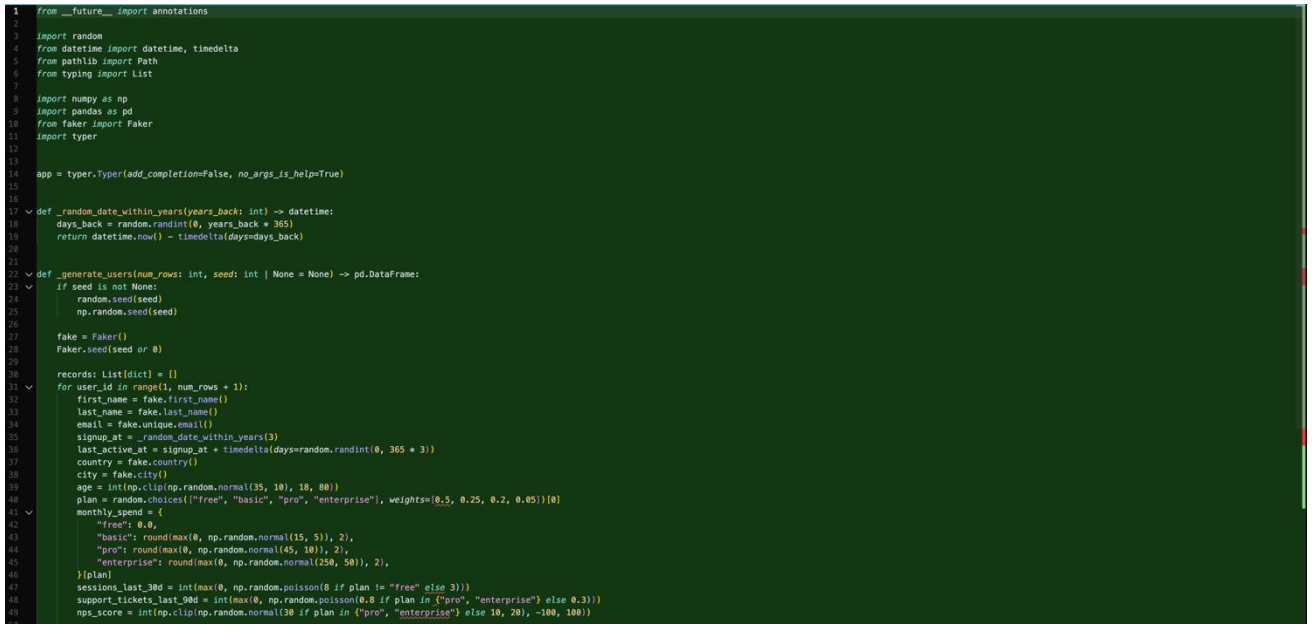
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5256590

CURSOR AI

An IDE (integrated developer enviroment) that provides facilities for software development

Allows interactions and coding based on prompting only

Only a limited number of prompts in the free version



```python
from __future__ import annotations

import random
from datetime import datetime, timedelta
from pathlib import Path
from typing import List

import numpy as np
import pandas as pd
from faker import Faker
import typer


app = typer.Typer(add_completion=False, no_args_is_help=True)


def _random_date_within_years(years_back: int) -> datetime:
    days_back = random.randint(0, years_back * 365)
    return datetime.now() - timedelta(days=days_back)


def _generate_users(num_rows: int, seed: int | None = None) -> pd.DataFrame:
    if seed is not None:
        random.seed(seed)
        np.random.seed(seed)

    fake = Faker()
    Faker.seed(seed or 0)

    records: List[dict] = []
    for user_id in range(1, num_rows + 1):
        first_name = fake.first_name()
        last_name = fake.last_name()
        email = fake.unique.email()
        signup_at = _random_date_within_years(3)
        last_active_at = signup_at + timedelta(days=random.randint(0, 365 * 3))
        country = fake.country()
        city = fake.city()
        age = int(np.clip(np.random.normal(35, 10), 18, 80))
        plan = random.choices(["free", "basic", "pro", "enterprise"], weights=[0.5, 0.25, 0.2, 0.05])[0]
        monthly_spend = {
            "free": 0.0,
            "basic": round(max(0, np.random.normal(15, 5)), 2),
            "pro": round(max(0, np.random.normal(45, 10)), 2),
            "enterprise": round(max(0, np.random.normal(250, 50)), 2),
        }[plan]
        sessions_last_30d = int(max(0, np.random.poisson(8 if plan != "free" else 3)))
        support_tickets_last_90d = int(max(0, np.random.poisson(0.8 if plan in {"pro", "enterprise"} else 0.3)))
        nps_score = int(np.clip(np.random.normal(30 if plan in {"pro", "enterprise"} else 10, 20), -100, 100))
```

I want to automatically generate a file with synthetic user data. Then, I want to label the generated data.

Thought for 11s

I'll set up a small Python project to generate synthetic user data and label it. First I'll create the project structure and initial files, then I'll add code for generation and labeling, install dependencies, and run a quick demo.

EXEMPLARY PROMPT FOR CURSOR

“Please create me a Python code to connect with [SELECT MODEL] via fireworks. Please add a placeholder for the API Key. Here is the code snippet from the Fireworks website how the call should look like. [COPY AND PASTE THE API EXAMPLE FROM FIREWORKS AI]”

SUMMARY & TAKEAWAYS

1. **AI changes research practice.** But they are here to help. Use them when they provide an actual advantage (i.e., save time, scale better, save ressources…). Stay curious.

2. **Different "levels" of AI integration in your workflow are possible—and justified.** Utility depends on your exercises, your workflow, your research area, and your background and expertise.

3. **For data analysis and collection**, AI tools can be used with a low barrier to facilitate data analysis tasks, get inspiration, handle unstructured data, or generate synthetic data.

QUESTIONS?

THANK YOU FOR DAY 1 ☺!