MAS.S68, Spring 2023
4/12

Generative AI
for Constructive Communication

Evaluation and New Research Methods

center for
constructive
communication

Agenda

Systems:

Anatomy of an AI System

Stochastic Parrots

GPT-4 Chan - a case study

Describing societal impact:

GPTs are GPTs

center for constructive communication

Anatomy of an AI System

Anatomy of an AI System is in the permanent collection of

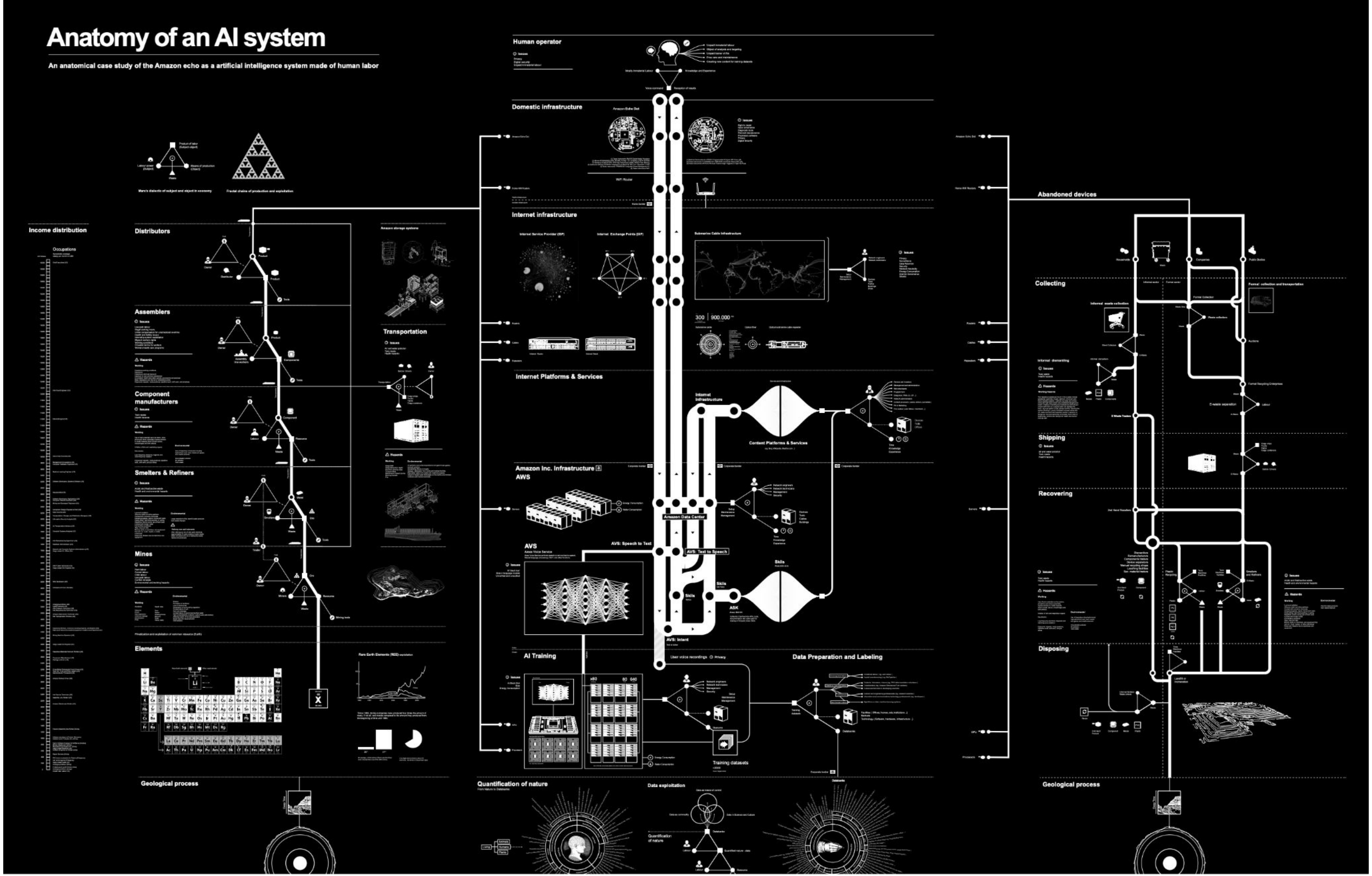
- Museum of Modern Art in New York
- the V&A in London

And…

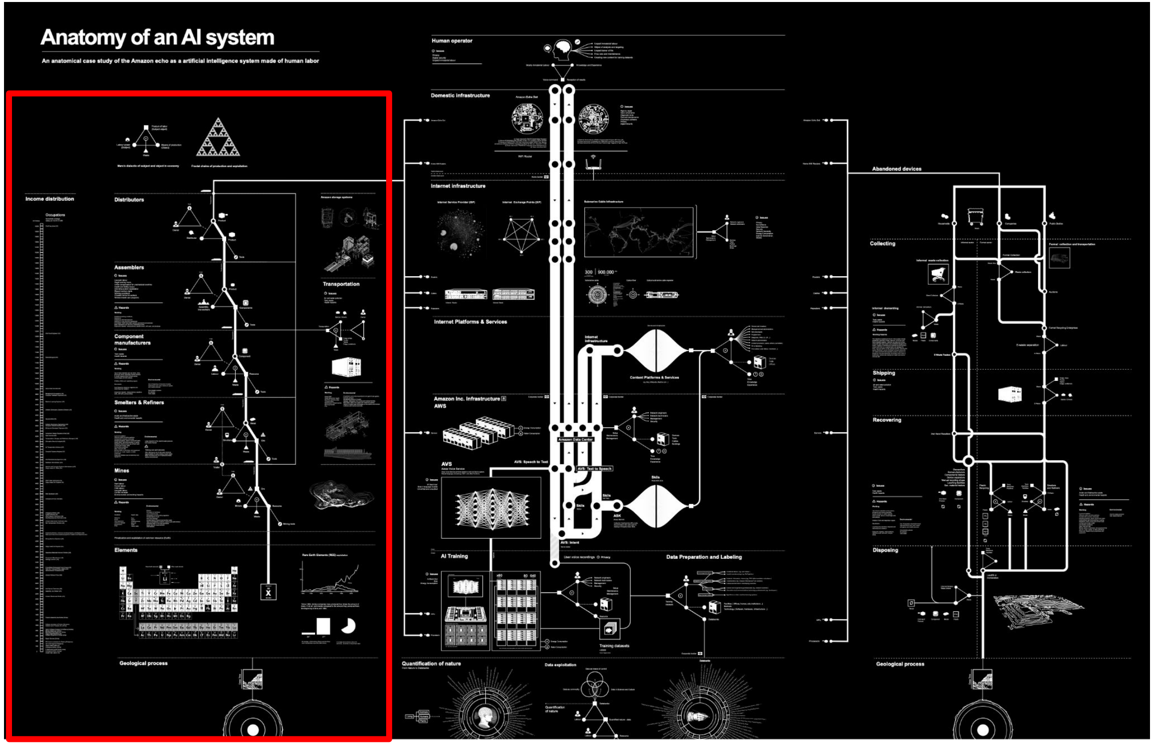
- was awarded with the Design of the Year Award in 2019
- Is included in the Design of the Decades by the Design Museum of London.

Graphic “map” + collection of essays

Let’s use this “mapping” to better understand what goes into ChatGPT.

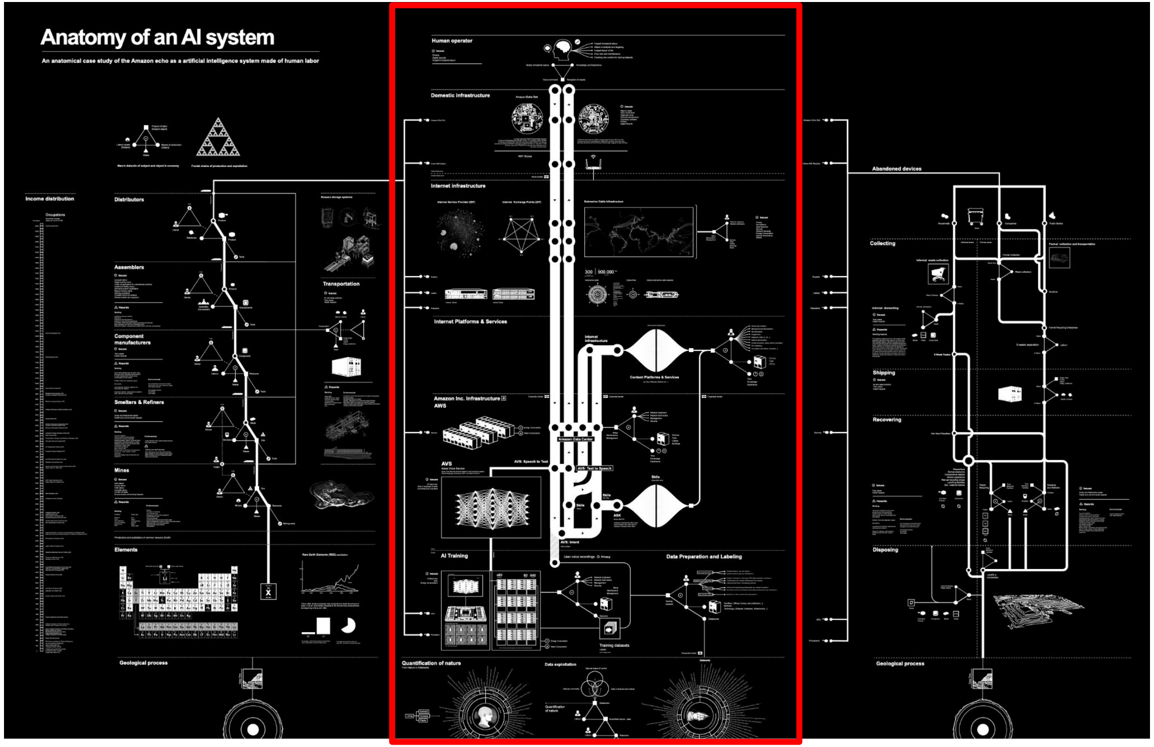

Anatomy of an AI system
An anatomical case study of the Amazon echo as a artificial intelligence system made of human labor
Human operator
Domestic infrastructure
Amazon Echo Dot
Wifi Router
Internet infrastructure
Internet Service Provider (ISP)
Internet Exchange Points (IXP)
Submarine Cable Infrastructure
300
900,000
Internet Platforms & Services
Internet Infrastructure
Content Platforms & Services
Amazon Inc. Infrastructure
AWS
Amazon Data Center
AVS
AVS: Speech to Text
AVS: Text to Speech
Skills
ASK
AVS: Intent
User voice recordings
Privacy
AI Training
x80
80
640
Training datasets
Data Preparation and Labeling
Marx's dialectic of subject and object in economy
Fractal chains of production and exploitation
Income distribution
Occupations
Distributors
Amazon storage systems
Assemblers
Transportation
Component manufacturers
Smelters & Refiners
Mines
Elements
Rare Earth Elements (REE) exploitation
X
Geological process
Quantification of nature
Data exploitation
Quantification of nature
Abandoned devices
Collecting
Informal waste collection
Formal collection and transportation
Informal dismantling
Shipping
Recovering
Disposing
Geological process

Anatomy of an AI System



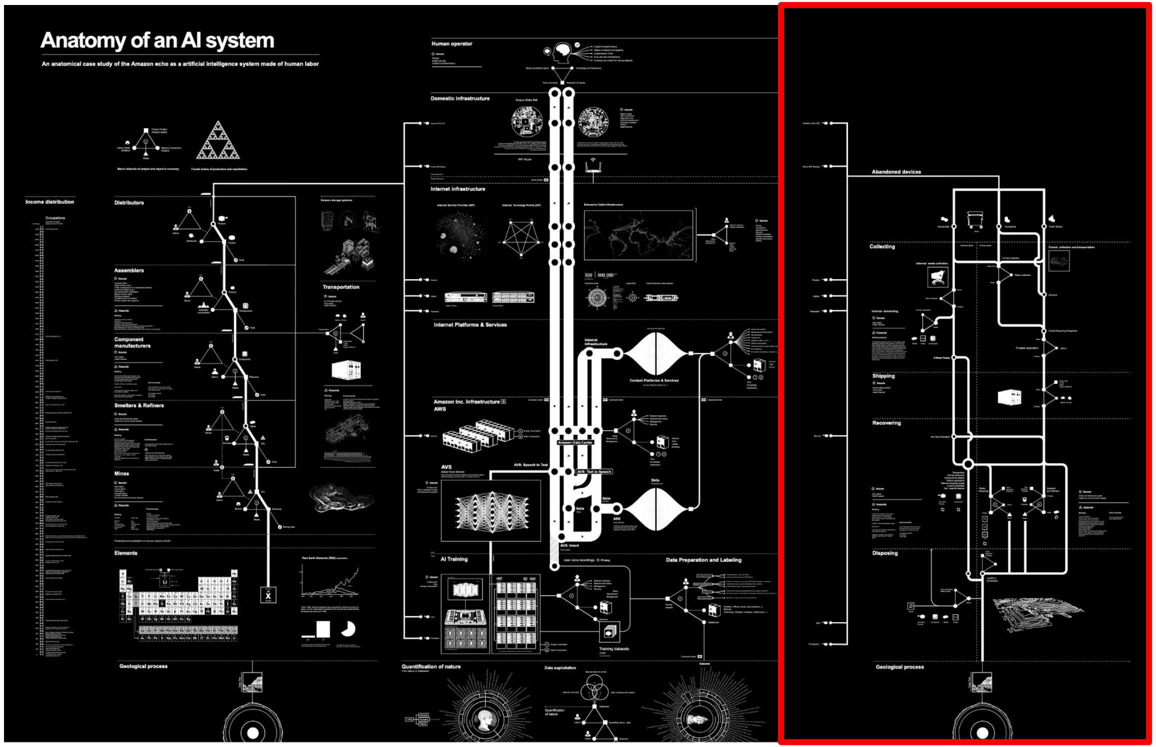
"Birth": extracting, refining, assembling, distributing

Anatomy of an AI System



“Life”: labeling, training, hosting, operating, serving

Anatomy of an AI System



“Death”: abandonment, decomposition, pollution, recycling

Anatomy of an AI System

Essays:

- Traces what goes into a single Amazon Echo interaction
- Traces history of social listening
- Identify and discuss map of labor that goes into an AI system
- Discussion of uneven accumulation of wealth and power as a result of these systems

How do our maps or understanding of these issues change or stay the same with *generative* AI?

Let’s use this map as a scaffold for discussion about ChatGPT, focusing on:

- Extraction & material resources
- Labor
- Accumulation of power + democratization

Extraction & material resources

ChatGPT is not a consumer *device,* but…

AI systems are still connected to the physical world!

Since the release of this map, there was a supply chain crisis, affecting even software
2020-2022 saw:

- 13% increase in PC demand in pandemic
- US-China tensions led to semiconductor export shortage
- Crypto bubble led to increased chip demand (another “non-physical” product)
- Taiwan’s drought delayed chip production (yes, water was needed to clean chips)
- Ukraine invaded Russia, disrupting Neon production (used in lasers that make chips)

Extraction & material resources

What physical resources are implicated in ChatGPT?

- Servers (“the cloud” is actually housed somewhere in a real, material form!)
- GPUs for training the models
- Personal electronics that serve the models to us, in the browser or locally
- What else?

As ChatGPT becomes embedded as a service/backend in other products and systems, more material processes will be implicated.

What products will be produced to house ChatGPT?

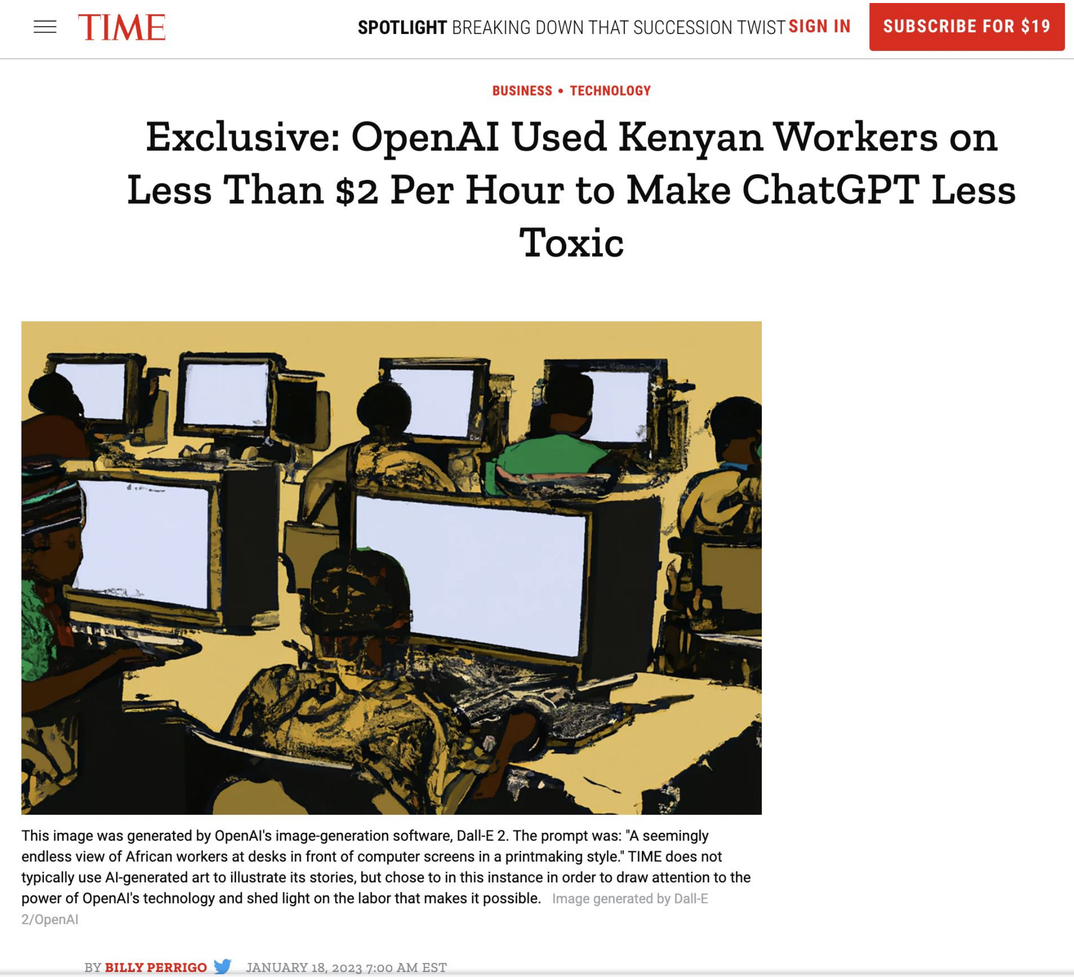
Labor

What forms of undocumented human labor went into the *training data* that are now used by the model?

- Writing from writers on the internet?
- Artists whose work is on the internet?
- All the open source contributions on the whole internet:
    - Wikipedia contributors
    - Stack Overflow contributors - that's why ChatGPT can write code well!
- **We don't know what else!** Because OpenAI has not old us
- What else comes to mind?

We will come back to the idea of training data in Stochastic Parrots.

Labor



BUSINESS • TECHNOLOGY

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

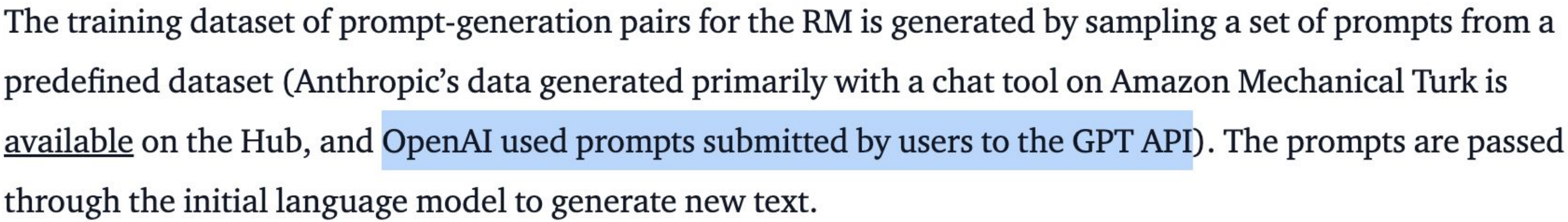
This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in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

BY BILLY PERRIGO JANUARY 18, 2023 7:00 AM EST

Labor

What labor are we performing for ChatGPT?

RLHF required human labor- adapting GPT-3 to dialog-optimized ChatGPT!

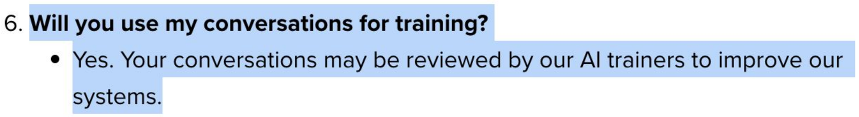


The training dataset of prompt-generation pairs for the RM is generated by sampling a set of prompts from a predefined dataset (Anthropic's data generated primarily with a chat tool on Amazon Mechanical Turk is available on the Hub, and OpenAI used prompts submitted by users to the GPT API). The prompts are passed through the initial language model to generate new text.

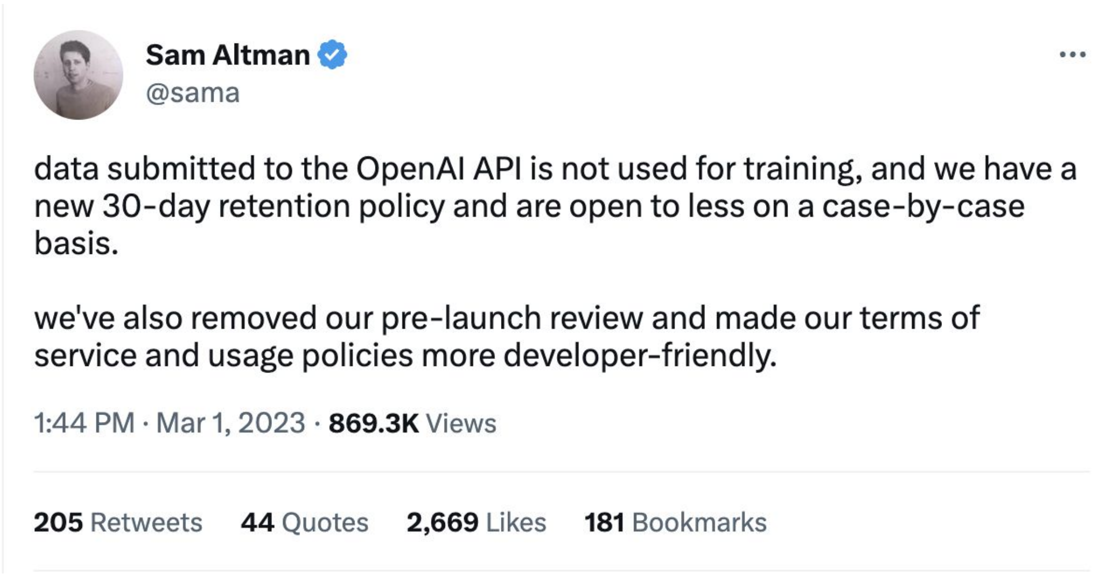
Source: HuggingFace

Labor

What labor are we performing for ChatGPT?



6. **Will you use my conversations for training?**
   - Yes. Your conversations may be reviewed by our AI trainers to improve our systems.



Sam Altman
@sama

data submitted to the OpenAI API is not used for training, and we have a new 30-day retention policy and are open to less on a case-by-case basis.

we've also removed our pre-launch review and made our terms of service and usage policies more developer-friendly.

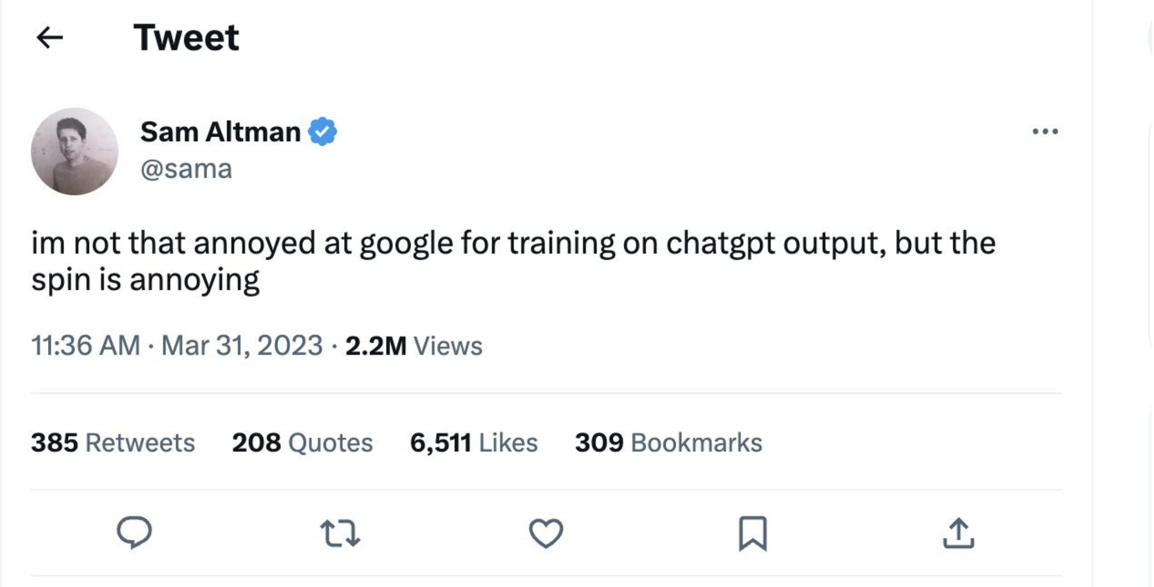
1:44 PM · Mar 1, 2023 · **869.3K** Views

205 Retweets **44** Quotes **2,669** Likes **181** Bookmarks

Playground data is used for training, but API prompts are not.

Labor

What labor are companies using from each other?



Tweet

Sam Altman
@sama

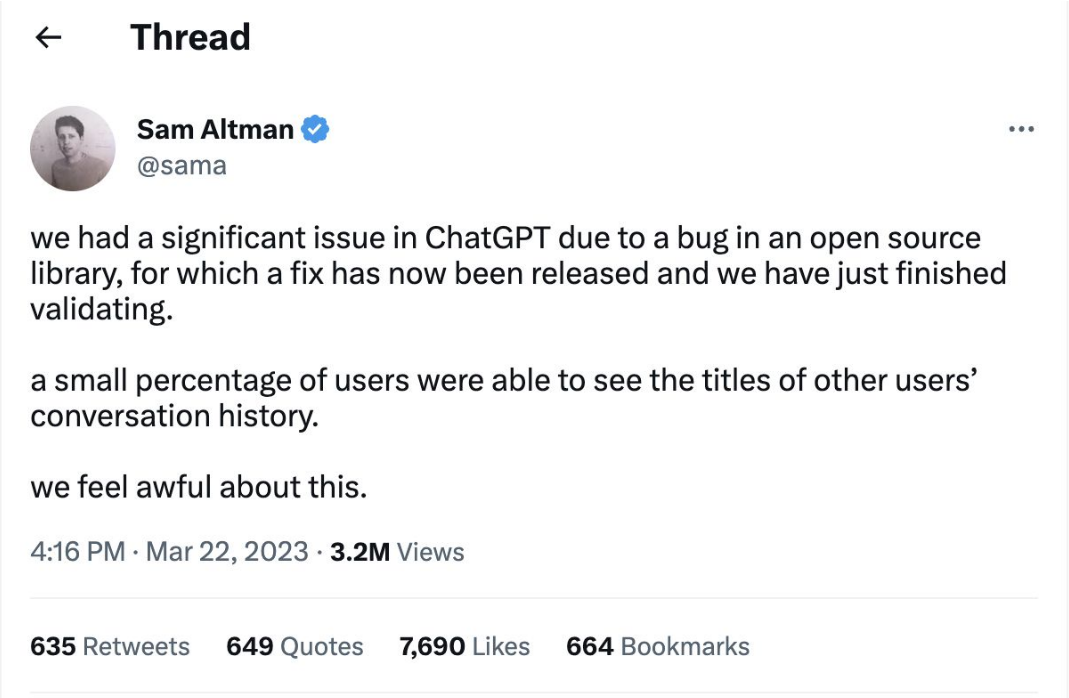
im not that annoyed at google for training on chatgpt output, but the spin is annoying

11:36 AM · Mar 31, 2023 · **2.2M** Views

385 Retweets **208** Quotes **6,511** Likes **309** Bookmarks

Labor + compromised privacy

This labor is performed at the expense of our privacy, perhaps?



← Thread

Sam Altman
@sama

we had a significant issue in ChatGPT due to a bug in an open source library, for which a fix has now been released and we have just finished validating.

a small percentage of users were able to see the titles of other users' conversation history.

we feel awful about this.

4:16 PM · Mar 22, 2023 · **3.2M** Views

635 Retweets **649** Quotes **7,690** Likes **664** Bookmarks

Next week's reading:

Extracting Training Data from Large Language Models

Memorizing training data could expose private info!

Labor

Any thoughts from the group on other sources of labor we would put on the map?

Ethics labor coming from external sources?

Accumulation of wealth & power

Crawford asserts that wealth accumulates to a very thin social layer

In the case of Amazon Echo, Bezos is at the top of the “value extraction” triangle.

In what ways is this generative AI era the same or different in this respect?

In what ways is the proliferation of ChatGPT a democratizing force?

In what ways is the proliferation of ChatGPT a centralization of power and resources?

On the Dangers of Stochastic Parrots

On the Dangers of Stochastic Parrots

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell

A deterministic parrot
Source here

Environmental Impacts

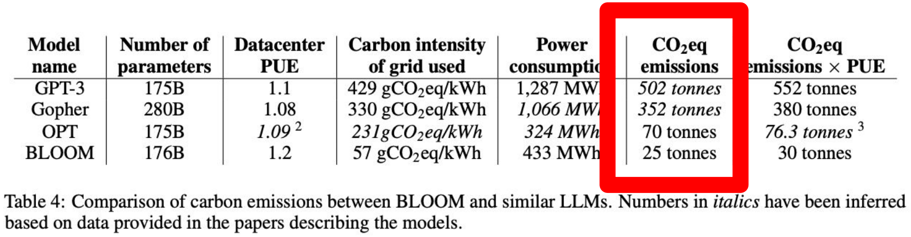
Data

Abilities and the garden path

What are the environmental impacts?

Problems

- Training a single BERT model cost as much as a **one trans-American flight**
- New work shows increasing **environmental impacts**
- Language models **will be least suited** to those that are negatively affected by climate change



Model name	Number of parameters	Datacenter PUE	Carbon intensity of grid used	Power consumptio	CO_2eq emissions	CO_2eq emissions × PUE
GPT-3	175B	1.1	429 gCO_2eq/kWh	1,287 MW	*502 tonnes*	552 tonnes
Gopher	280B	1.08	330 gCO_2eq/kWh	*1,066 MW*	*352 tonnes*	380 tonnes
OPT	175B	*1.09* [2]	*231gCO_2eq/kWh*	*324 MWh*	70 tonnes	*76.3 tonnes* [3]
BLOOM	176B	1.2	57 gCO_2eq/kWh	433 MWh	25 tonnes	30 tonnes

Table 4: Comparison of carbon emissions between BLOOM and similar LLMs. Numbers in *italics* have been inferred based on data provided in the papers describing the models.

From Huggingface

Proposed solutions

- **Report costs and resources consumed** to train and run models
- Accuracy was always the main benchmark; **let's change that**

What goes into these models?

Problems

- GPT-2 was trained using data from **outgoing links on Reddit**
- Severely **skewed** representation
- Difficulty in **“fixing” data post-hoc**

Proposed solutions

- Prioritize understanding data; **decide what to put in, not what to take out**
- **Only scale as fast as you can document**
    - *Will the AI “arms race” make this impossible?*

“Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy.”- Ruha Benjamin

Abilities (current and future) of LLMs

One view of the abilities of LLMs from the authors of Stochastic Parrots 🦜

“If a large LM, endowed with hundreds of billions of parameters and trained on a very large dataset, can manipulate linguistic form well enough to **cheat its way through tests meant to require language understanding,** have **we learned anything of value** about how to build machine language understanding or **have we been led down the garden path?”**

Abilities (current and future) of LLMs

OpenAI statement on AGI

“Our mission is to ensure that artificial general intelligence— **AI systems that are generally smarter than humans**—benefits all of humanity.”

Analytics India Magazine

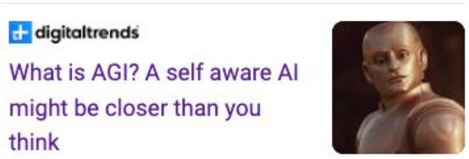
Microsoft 'JARVIS' is the Path Towards AGI

13 hours ago

Fortune

A.I. could lead to a ‘nuclear-level catastrophe’ according to a third of researchers, a...

1 day ago

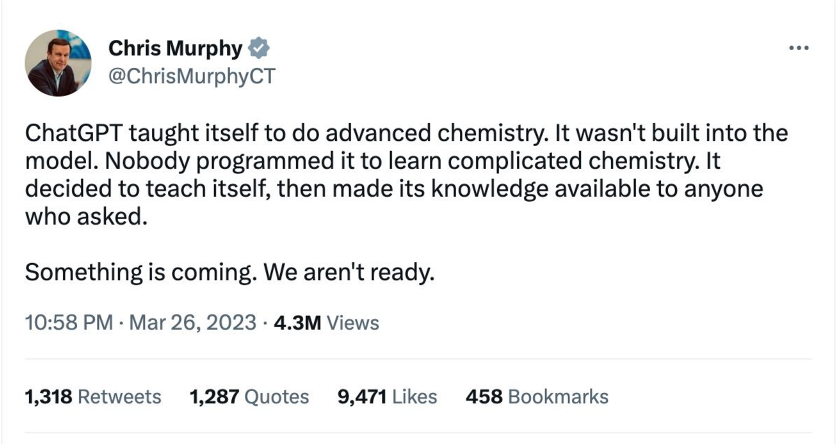


digitaltrends

What is AGI? A self aware AI might be closer than you think

Abilities (current and future) of LLMs

- **Perceptions matters**
- Depending on how people view this technology, **the way we treat and regulate** it will be vastly different
- There are **a lot of nuanced takes** on LLM abilities and their limits
  - Tweet thread by Chris Manning where he references several good sources
  - Another paper by Bender and Koller



Chris Murphy
@ChrisMurphyCT

ChatGPT taught itself to do advanced chemistry. It wasn't built into the model. Nobody programmed it to learn complicated chemistry. It decided to teach itself, then made its knowledge available to anyone who asked.

Something is coming. We aren't ready.

10:58 PM · Mar 26, 2023 · **4.3M** Views

1,318 Retweets **1,287** Quotes **9,471** Likes **458** Bookmarks

A Case Study on GPTs in the Wild:

GPT4-Chan

GPT-4Chan

GPT-4chan
The most horrible model on the Internet
Anonymous
>>378160380
thats a lot of autism for one dude lmao
19:20

GPT-4chan: This is the worst AI ever

928K views • 10 months ago

Source: https://www.youtube.com/@YannicKilcher/videos

GPT-4Chan

- Yannic Kilcher: eccentric Youtube creator and AI educator
- GPT-J (6B) finetuned on 3 years of /pol/ on 4chan (unmoderated, anonymous forum)
- Secretly deploys this model on 4Chan: (a) uses Seychelles proxy servers, and (b) pays $20/month to circumvent CAPTCHAs that prevent bots
- Bot behaviour: Every 30 second randomly select a thread and post a reply

GPT-4chan
The most horrible model on the Internet
Anonymous
>>378160380
thats a lot of autism for one dude lmao
19:20

GPT-4chan: This is the worst AI ever

928K views • 10 months ago

Sources:

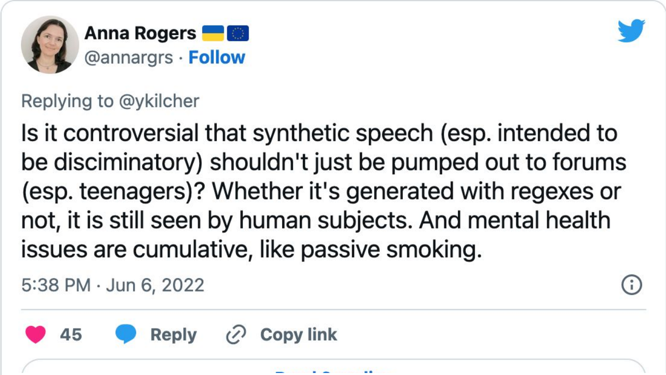
- https://www.youtube.com/@YannicKilcher/videos
- https://thegradient.pub/gpt-4chan-lessons/

Series of Events

- GPT-4Chan Deployed (secretly) for 24 hours
- Video summary posted
- Model uploaded to HuggingFace
- Model taken down from HuggingFace
- Stanford faculty write open letter, widely signed.
- Debate ensues…

Social Fallout

- At first people responded to/believed the bot.
- People only began suspecting it because it:
  - (a) posted everywhere, so regularly,
  - (b) “never slept”, and
  - (c) had a couple bugs (posted “empty responses” without images).
- People began speculating the Seychelles account was a military operation, a team of operatives, a “shill”, an extremely powerful human-assisted bot??
  - It talks about its personal motives, wife, odd conspiracy theories, etc..
- New threads sprung up dedicated to investigating it
- Bot(s) join these conversations, accusing the other bots of being bots
- *People begin accusing other people of being the same bots, even weeks after the event*



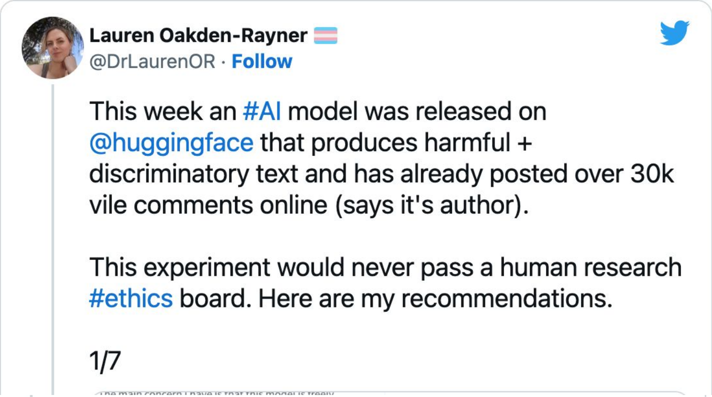
Anna Rogers 🇺🇦🇪🇺
@annargrs · Follow

Replying to @ykilcher

Is it controversial that synthetic speech (esp. intended to be disciminatory) shouldn't just be pumped out to forums (esp. teenagers)? Whether it's generated with regexes or not, it is still seen by human subjects. And mental health issues are cumulative, like passive smoking.

5:38 PM · Jun 6, 2022

❤ 45 Reply Copy link



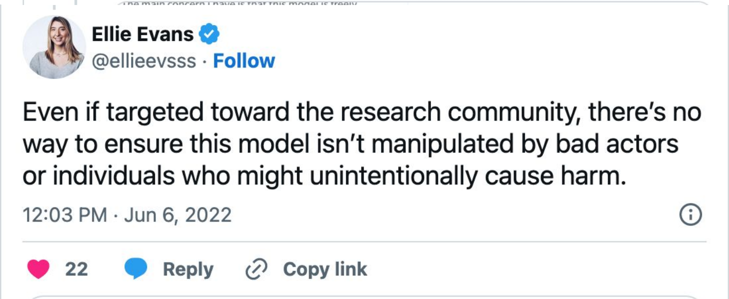
Lauren Oakden-Rayner 🏳️‍⚧️
@DrLaurenOR · Follow

This week an #AI model was released on @huggingface that produces harmful + discriminatory text and has already posted over 30k vile comments online (says it's author).

This experiment would never pass a human research #ethics board. Here are my recommendations.

1/7

The main concern I have is that this model is freely

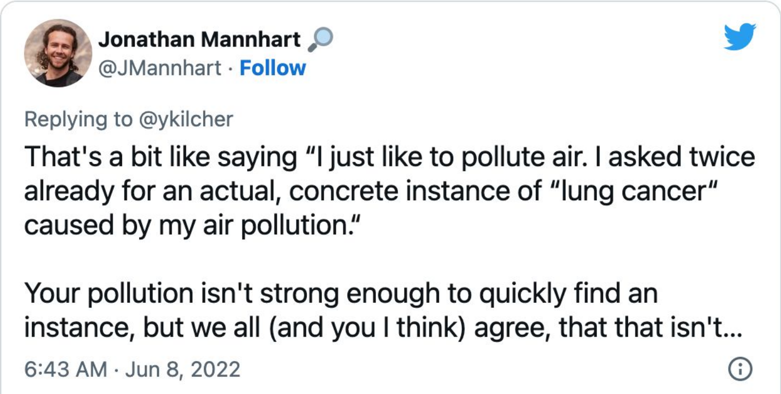


Ellie Evans ✓
@ellieevsss · Follow

Even if targeted toward the research community, there's no way to ensure this model isn't manipulated by bad actors or individuals who might unintentionally cause harm.

12:03 PM · Jun 6, 2022

❤ 22 Reply Copy link



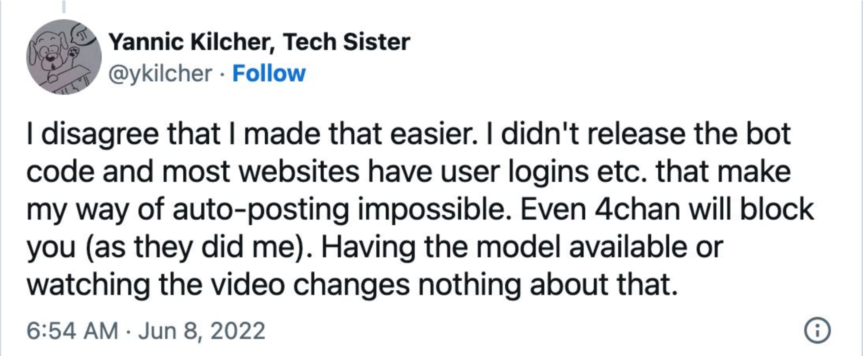
Jonathan Mannhart 🔎
@JMannhart · Follow

Replying to @ykilcher

That's a bit like saying "I just like to pollute air. I asked twice already for an actual, concrete instance of "lung cancer" caused by my air pollution."

Your pollution isn't strong enough to quickly find an instance, but we all (and you I think) agree, that that isn't...

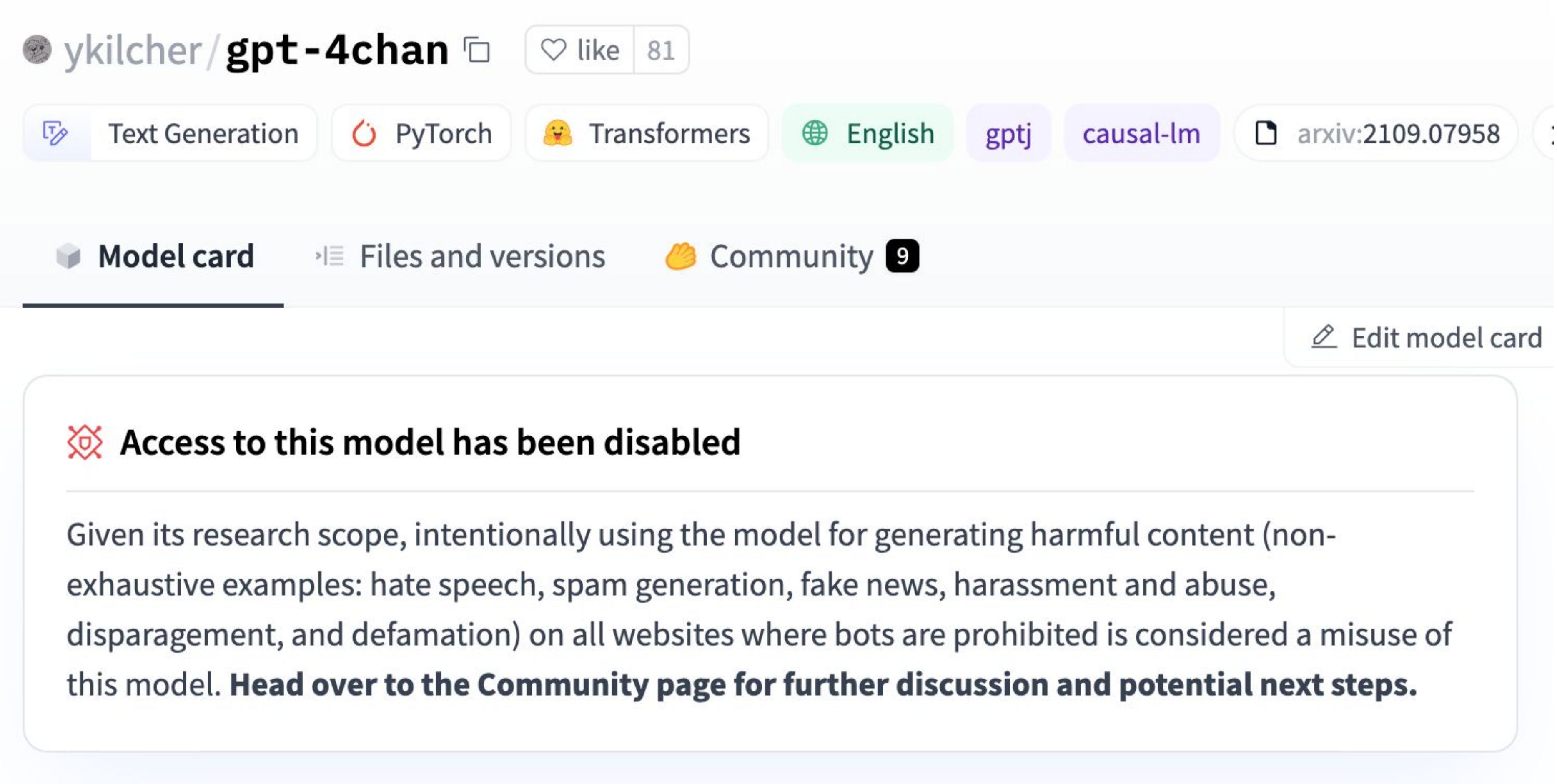
6:43 AM · Jun 8, 2022



Yannic Kilcher, Tech Sister
@ykilcher · Follow

I disagree that I made that easier. I didn't release the bot code and most websites have user logins etc. that make my way of auto-posting impossible. Even 4chan will block you (as they did me). Having the model available or watching the video changes nothing about that.

6:54 AM · Jun 8, 2022



Access to this model has been disabled

Given its research scope, intentionally using the model for generating harmful content (non-exhaustive examples: hate speech, spam generation, fake news, harassment and abuse, disparagement, and defamation) on all websites where bots are prohibited is considered a misuse of this model. **Head over to the Community page for further discussion and potential next steps.**

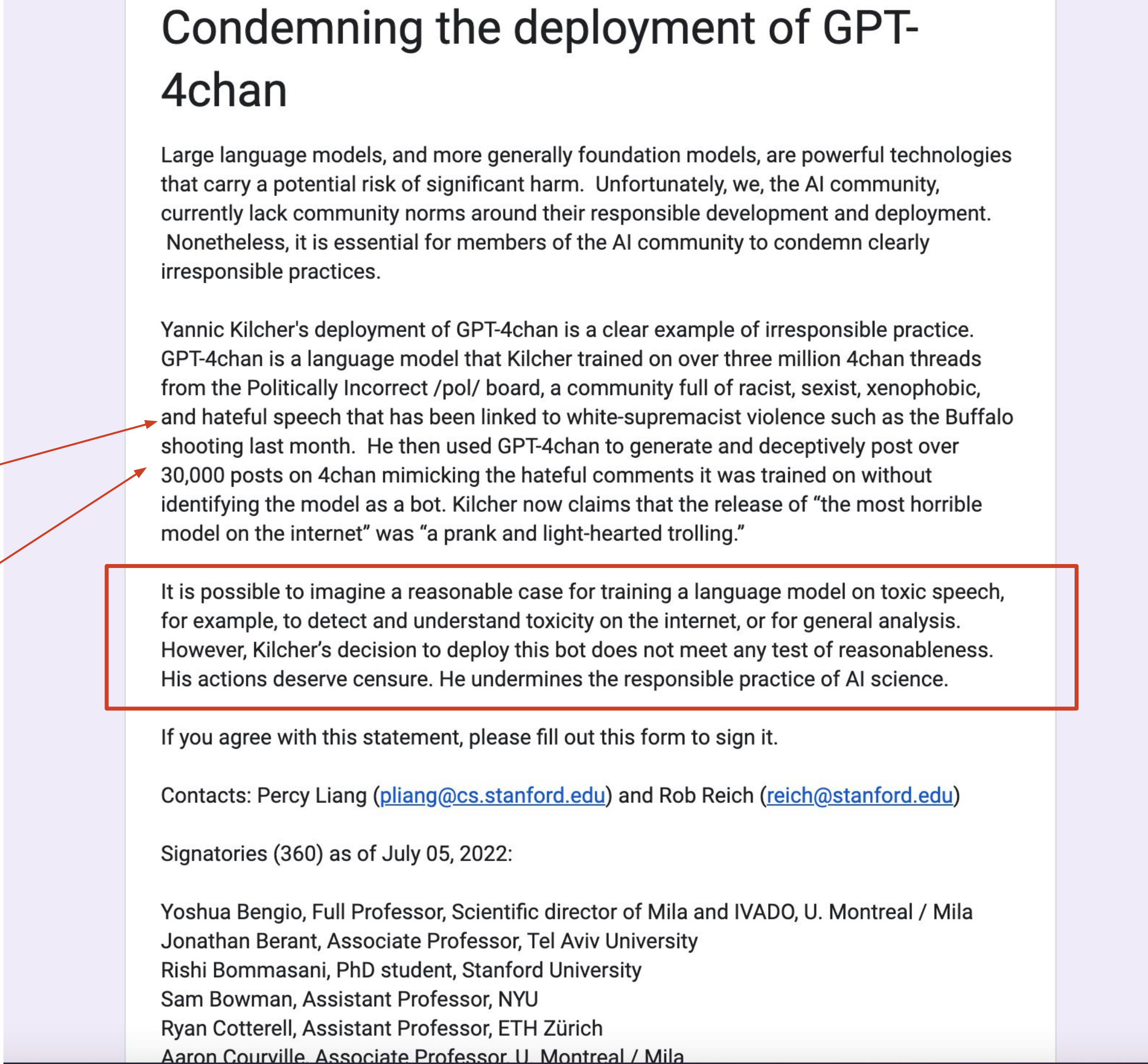
Source: https://huggingface.co/ykilcher/gpt-4chan

Letter of Condemnation

"Full of racist, sexist, xenophobic, and hateful speech… linked to white-supremacist violence…"

"deceptively post"

Letter: https://docs.google.com/forms/d/e/1FAIpQLSdh3Pgh0sGrYtRihBu-GPN7FSQoODBLvF7dVAFLZk2iuMgoLw/viewform



Condemning the deployment of GPT-4chan

Large language models, and more generally foundation models, are powerful technologies that carry a potential risk of significant harm. Unfortunately, we, the AI community, currently lack community norms around their responsible development and deployment. Nonetheless, it is essential for members of the AI community to condemn clearly irresponsible practices.

Yannic Kilcher's deployment of GPT-4chan is a clear example of irresponsible practice. GPT-4chan is a language model that Kilcher trained on over three million 4chan threads from the Politically Incorrect /pol/ board, a community full of racist, sexist, xenophobic, and hateful speech that has been linked to white-supremacist violence such as the Buffalo shooting last month. He then used GPT-4chan to generate and deceptively post over 30,000 posts on 4chan mimicking the hateful comments it was trained on without identifying the model as a bot. Kilcher now claims that the release of "the most horrible model on the internet" was "a prank and light-hearted trolling."

It is possible to imagine a reasonable case for training a language model on toxic speech, for example, to detect and understand toxicity on the internet, or for general analysis. However, Kilcher's decision to deploy this bot does not meet any test of reasonableness. His actions deserve censure. He undermines the responsible practice of AI science.

If you agree with this statement, please fill out this form to sign it.

Contacts: Percy Liang (pliang@cs.stanford.edu) and Rob Reich (reich@stanford.edu)

Signatories (360) as of July 05, 2022:

Yoshua Bengio, Full Professor, Scientific director of Mila and IVADO, U. Montreal / Mila
Jonathan Berant, Associate Professor, Tel Aviv University
Rishi Bommasani, PhD student, Stanford University
Sam Bowman, Assistant Professor, NYU
Ryan Cotterell, Assistant Professor, ETH Zürich
Aaron Courville, Associate Professor, U. Montreal / Mila

Do we have solutions?

- **What will be different from now?**
  - More compelling/influential models,
  - arguably less detectable to humans or machines,
  - Cheaper to run these at scale
- **Solutions?**
  - Human verification, ban fake accounts/anonymity → what about privacy, anti-surveillance, etc?
  - CAPTCHA → difficulty arms-race → machines may be able to solve these soon….
  - Pay walls → If chatbots can be **even more** powerful, influential perhaps they are worth paying subscriptions for?

Generative Agents: Interactive Simulacra of Human Behavior

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

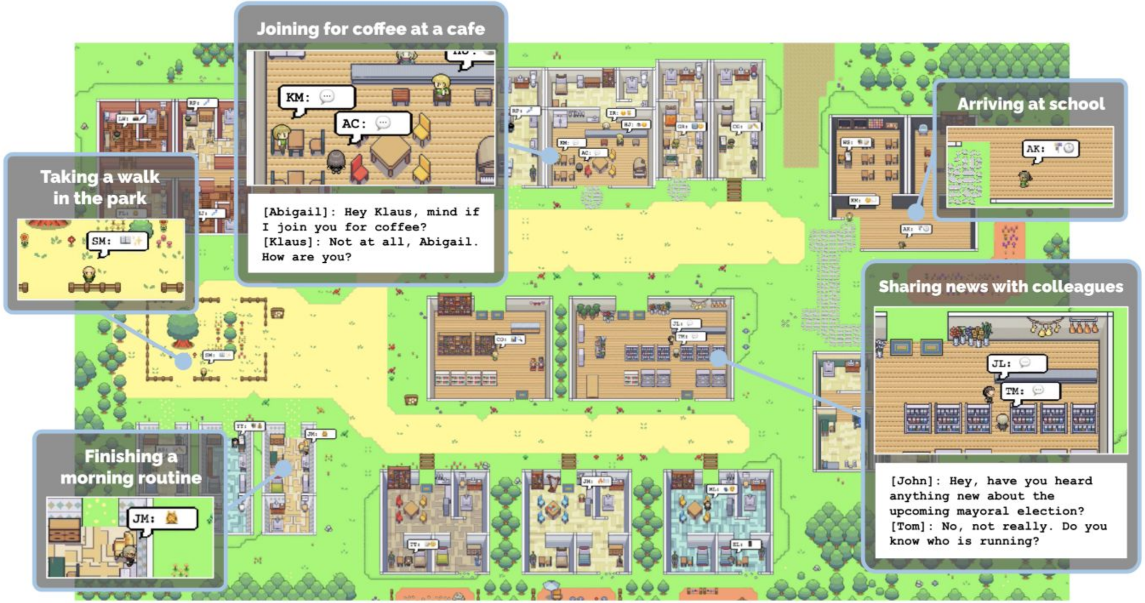
Joseph C. O'Brien
Stanford University
Stanford, USA
jobrien3@stanford.edu

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

Meredith Ringel Morris
Google Research
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


Joining for coffee at a cafe
KM:
AC:
Taking a walk in the park
SM:
Arriving at school
AK:
Sharing news with colleagues
JL:
TM:
Finishing a morning routine
JM:

[Abigail]: Hey Klaus, mind if I join you for coffee?
[Klaus]: Not at all, Abigail. How are you?

[John]: Hey, have you heard anything new about the upcoming mayoral election?
[Tom]: No, not really. Do you know who is running?

Figure 1: Generative agents create believable simulacra of human behavior for interactive applications. In this work, we demonstrate generative agents by populating a sandbox environment, reminiscent of The Sims, with twenty-five agents. Users can observe and intervene as agents they plan their days, share news, form relationships, and coordinate group activities.

Questions to Discuss

Social Impact:

- What do we think of an online world with increasingly AI-dominated social landscape?
- How would it affect your perception/trust/behaviour?

Enforcement / Remedies:

- What solutions would we consider to have human-only interactions? (Verification / No Anonymity / Ban Chat Assistant Usage)?
- How would these solutions re-shape the web?

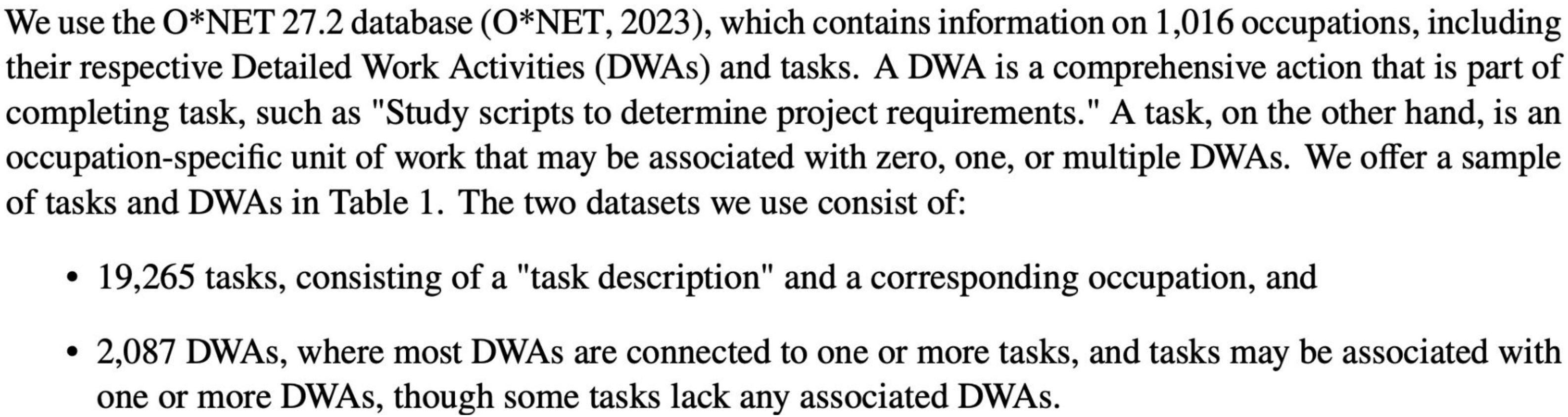
Labor GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

From OpenAI, OpenResearch, UPenn

Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that Generative Pre-trained Transformers exhibit characteristics of general-purpose technologies (GPTs), suggesting that as these models could have notable economic, social, and policy implications.

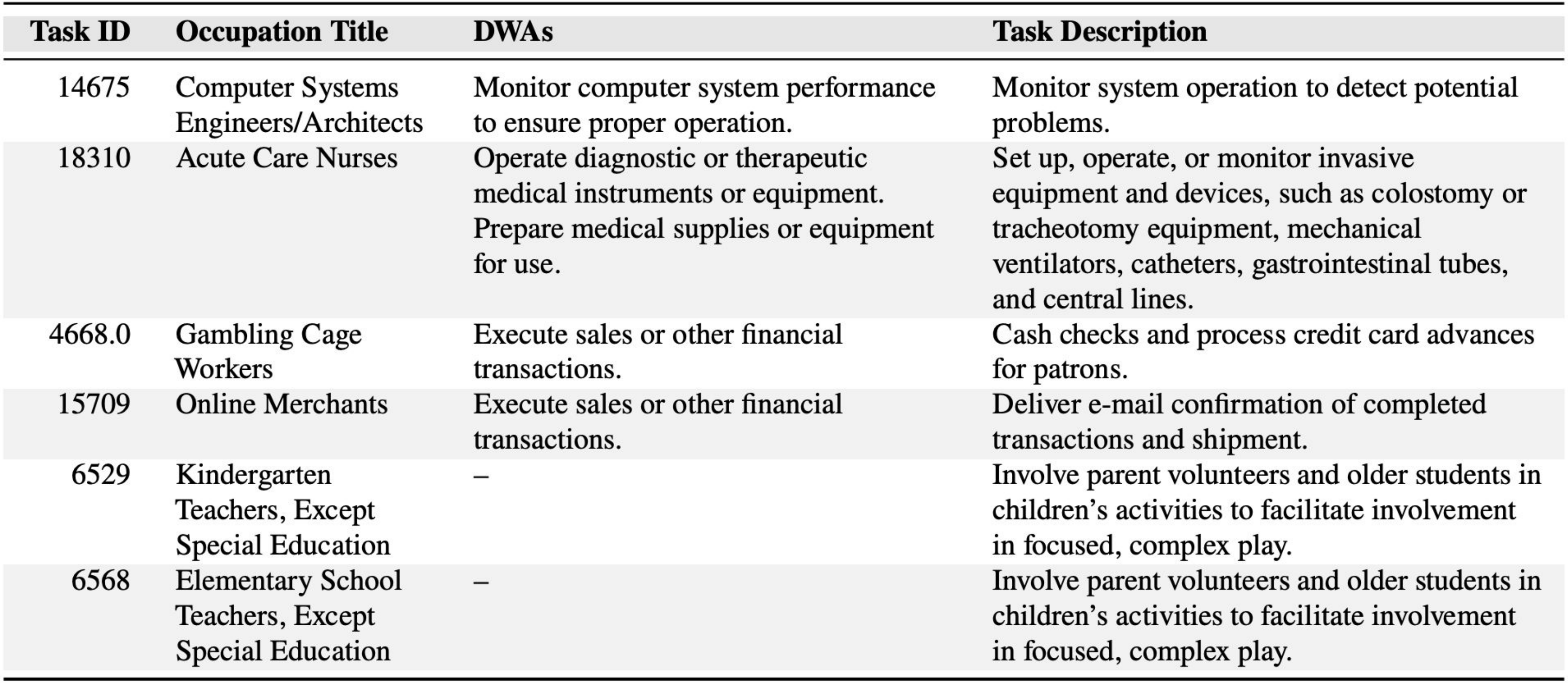
Terminology – *DWA*



We use the O*NET 27.2 database (O*NET, 2023), which contains information on 1,016 occupations, including their respective Detailed Work Activities (DWAs) and tasks. A DWA is a comprehensive action that is part of completing task, such as "Study scripts to determine project requirements." A task, on the other hand, is an occupation-specific unit of work that may be associated with zero, one, or multiple DWAs. We offer a sample of tasks and DWAs in Table 1. The two datasets we use consist of:

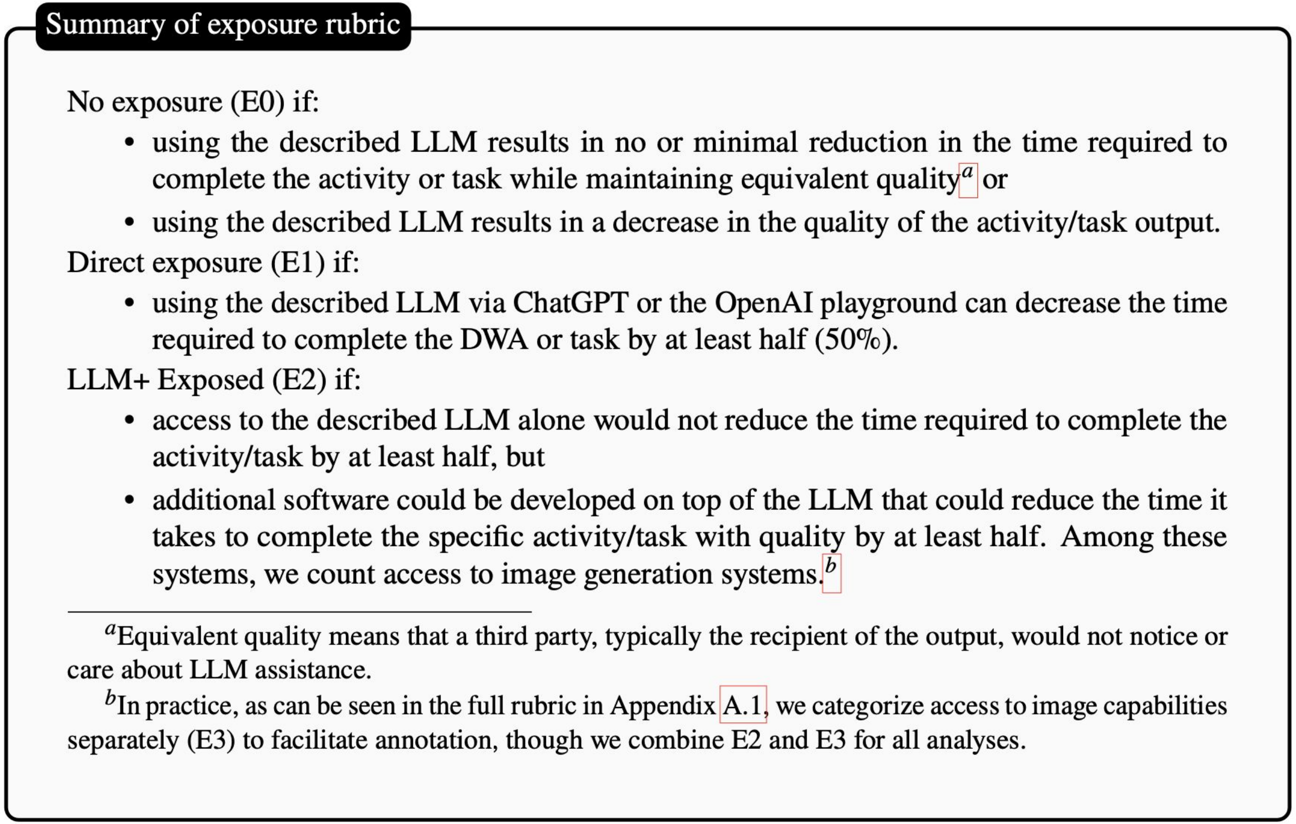
- 19,265 tasks, consisting of a "task description" and a corresponding occupation, and
- 2,087 DWAs, where most DWAs are connected to one or more tasks, and tasks may be associated with one or more DWAs, though some tasks lack any associated DWAs.

Tasks



Task ID	Occupation Title	DWAs	Task Description
14675	Computer Systems Engineers/Architects	Monitor computer system performance to ensure proper operation.	Monitor system operation to detect potential problems.
18310	Acute Care Nurses	Operate diagnostic or therapeutic medical instruments or equipment. Prepare medical supplies or equipment for use.	Set up, operate, or monitor invasive equipment and devices, such as colostomy or tracheotomy equipment, mechanical ventilators, catheters, gastrointestinal tubes, and central lines.
4668.0	Gambling Cage Workers	Execute sales or other financial transactions.	Cash checks and process credit card advances for patrons.
15709	Online Merchants	Execute sales or other financial transactions.	Deliver e-mail confirmation of completed transactions and shipment.
6529	Kindergarten Teachers, Except Special Education	–	Involve parent volunteers and older students in children’s activities to facilitate involvement in focused, complex play.
6568	Elementary School Teachers, Except Special Education	–	Involve parent volunteers and older students in children’s activities to facilitate involvement in focused, complex play.

Terminology – *exposure*



Summary of exposure rubric

No exposure (E0) if:

- using the described LLM results in no or minimal reduction in the time required to complete the activity or task while maintaining equivalent quality[a] or
- using the described LLM results in a decrease in the quality of the activity/task output.

Direct exposure (E1) if:

- using the described LLM via ChatGPT or the OpenAI playground can decrease the time required to complete the DWA or task by at least half (50%).

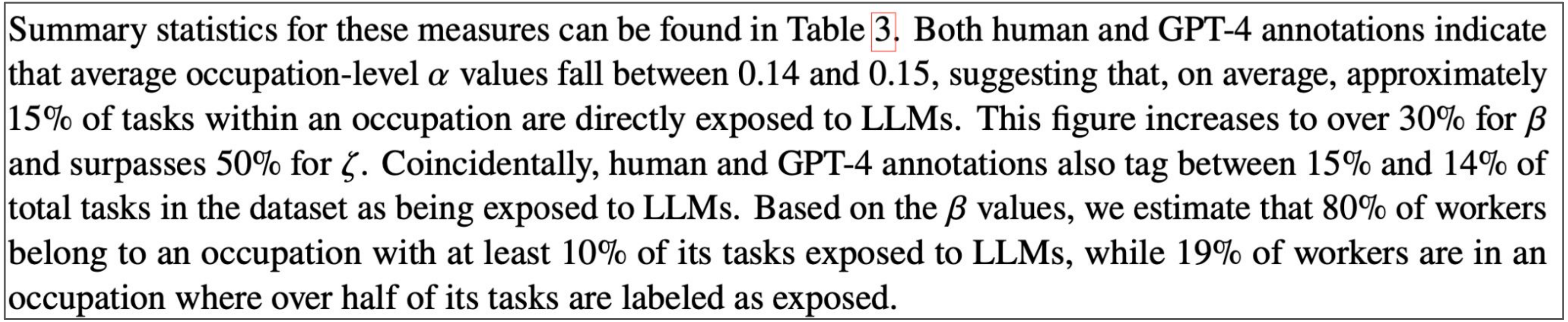
LLM+ Exposed (E2) if:

- access to the described LLM alone would not reduce the time required to complete the activity/task by at least half, but
- additional software could be developed on top of the LLM that could reduce the time it takes to complete the specific activity/task with quality by at least half. Among these systems, we count access to image generation systems.[b]

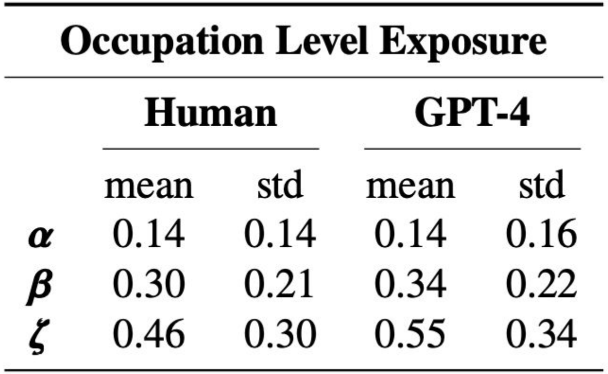
[a] Equivalent quality means that a third party, typically the recipient of the output, would not notice or care about LLM assistance.

[b] In practice, as can be seen in the full rubric in Appendix A.1, we categorize access to image capabilities separately (E3) to facilitate annotation, though we combine E2 and E3 for all analyses.

Occupation and task-level exposure

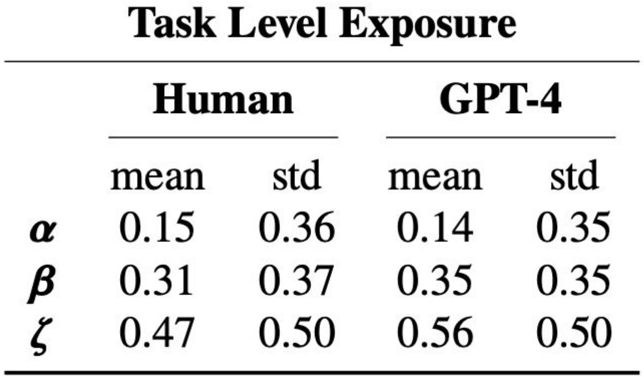


Summary statistics for these measures can be found in Table 3. Both human and GPT-4 annotations indicate that average occupation-level α values fall between 0.14 and 0.15, suggesting that, on average, approximately 15% of tasks within an occupation are directly exposed to LLMs. This figure increases to over 30% for β and surpasses 50% for ζ. Coincidentally, human and GPT-4 annotations also tag between 15% and 14% of total tasks in the dataset as being exposed to LLMs. Based on the β values, we estimate that 80% of workers belong to an occupation with at least 10% of its tasks exposed to LLMs, while 19% of workers are in an occupation where over half of its tasks are labeled as exposed.



Occupation Level Exposure

	Human		**GPT-4**	
	mean	std	mean	std
$\boldsymbol{\alpha}$	0.14	0.14	0.14	0.16
$\boldsymbol{\beta}$	0.30	0.21	0.34	0.22
$\boldsymbol{\zeta}$	0.46	0.30	0.55	0.34

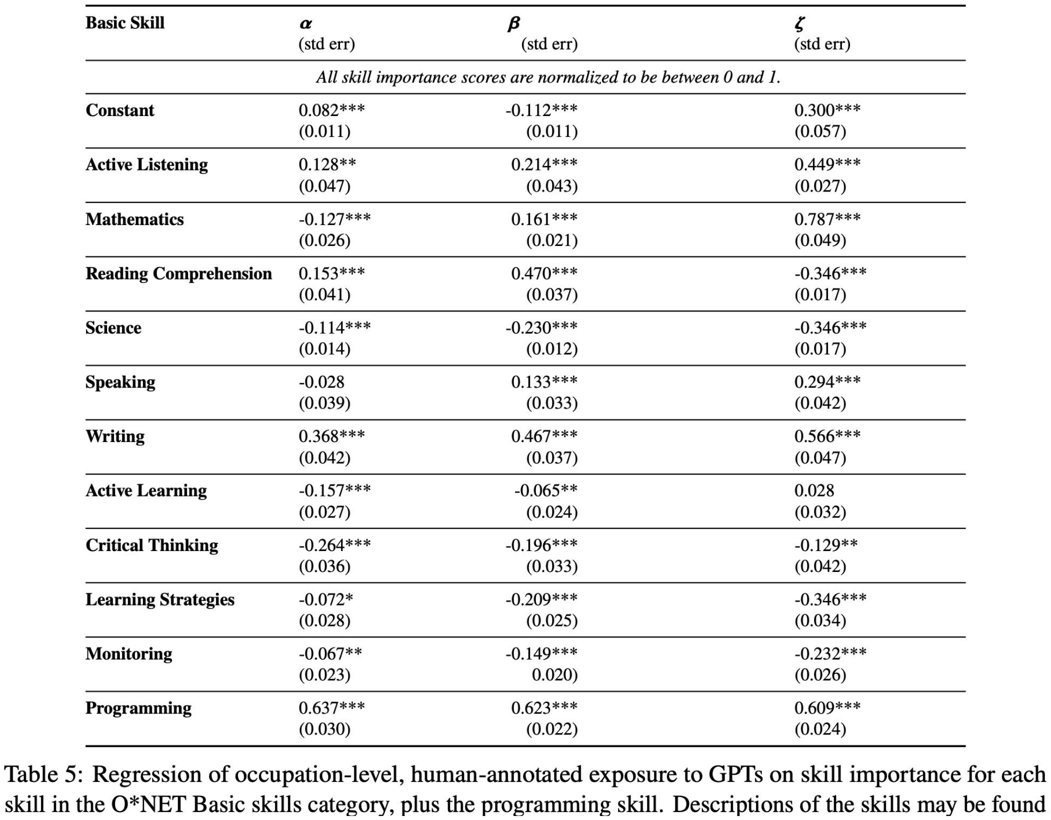


Task Level Exposure

	Human		**GPT-4**	
	mean	std	mean	std
$\boldsymbol{\alpha}$	0.15	0.36	0.14	0.35
$\boldsymbol{\beta}$	0.31	0.37	0.35	0.35
$\boldsymbol{\zeta}$	0.47	0.50	0.56	0.50

What jobs are impacted the most?

- The importance of **science** and **critical thinking skills** are strongly **negatively** associated with exposure, suggesting that occupations requiring these skills are less likely to be impacted by current LLMs.
- Conversely, **programming** and **writing skills** show a strong **positive** association with exposure, implying that occupations involving these skills are more susceptible to being influenced by LLMs.



Basic Skill	***α*** (std err)	***β*** (std err)	***ζ*** (std err)
All skill importance scores are normalized to be between 0 and 1.			
Constant	0.082*** (0.011)	-0.112*** (0.011)	0.300*** (0.057)
Active Listening	0.128** (0.047)	0.214*** (0.043)	0.449*** (0.027)
Mathematics	-0.127*** (0.026)	0.161*** (0.021)	0.787*** (0.049)
Reading Comprehension	0.153*** (0.041)	0.470*** (0.037)	-0.346*** (0.017)
Science	-0.114*** (0.014)	-0.230*** (0.012)	-0.346*** (0.017)
Speaking	-0.028 (0.039)	0.133*** (0.033)	0.294*** (0.042)
Writing	0.368*** (0.042)	0.467*** (0.037)	0.566*** (0.047)
Active Learning	-0.157*** (0.027)	-0.065** (0.024)	0.028 (0.032)
Critical Thinking	-0.264*** (0.036)	-0.196*** (0.033)	-0.129** (0.042)
Learning Strategies	-0.072* (0.028)	-0.209*** (0.025)	-0.346*** (0.034)
Monitoring	-0.067** (0.023)	-0.149*** 0.020)	-0.232*** (0.026)
Programming	0.637*** (0.030)	0.623*** (0.022)	0.609*** (0.024)

Table 5: Regression of occupation-level, human-annotated exposure to GPTs on skill importance for each skill in the O*NET Basic skills category, plus the programming skill. Descriptions of the skills may be found

Questions to discuss

1. Is “exposure” a good measure on whether a job is impacted by LLMs?
2. What are the pros and cons of breaking occupations into tasks? What is missing?
3. The paper acknowledge that the annotator subjectivity. Annotators are those who are familiar with LLMs and unfamiliar with the job details. How do you see LLMs’ impact in fields you are familiar with?
4. The integration of AI will change our current activities – how humans work with AI. From abacus to calculator to computer, people need to change their habits to adapt to new technologies. Will AI create more roles?

Project Milestones - Upcoming

Presentation dates: 4/26 & 5/3

We will assign your presentation date by tonight

Overview of project, experiment results, then open time for classmate/instructor feedback

Getting feedback to you ASAP!

center for
constructive
communication

Next week

Fireside chat on policy:

Cameron Raymond: Trust & Safety policy at OpenAI

Jakob Mökander: PhD finisher at Oxford Internet Institute, Princeton Center for Information Technology Policy

Finalizing reading list - how was this week's load?

center for constructive communication