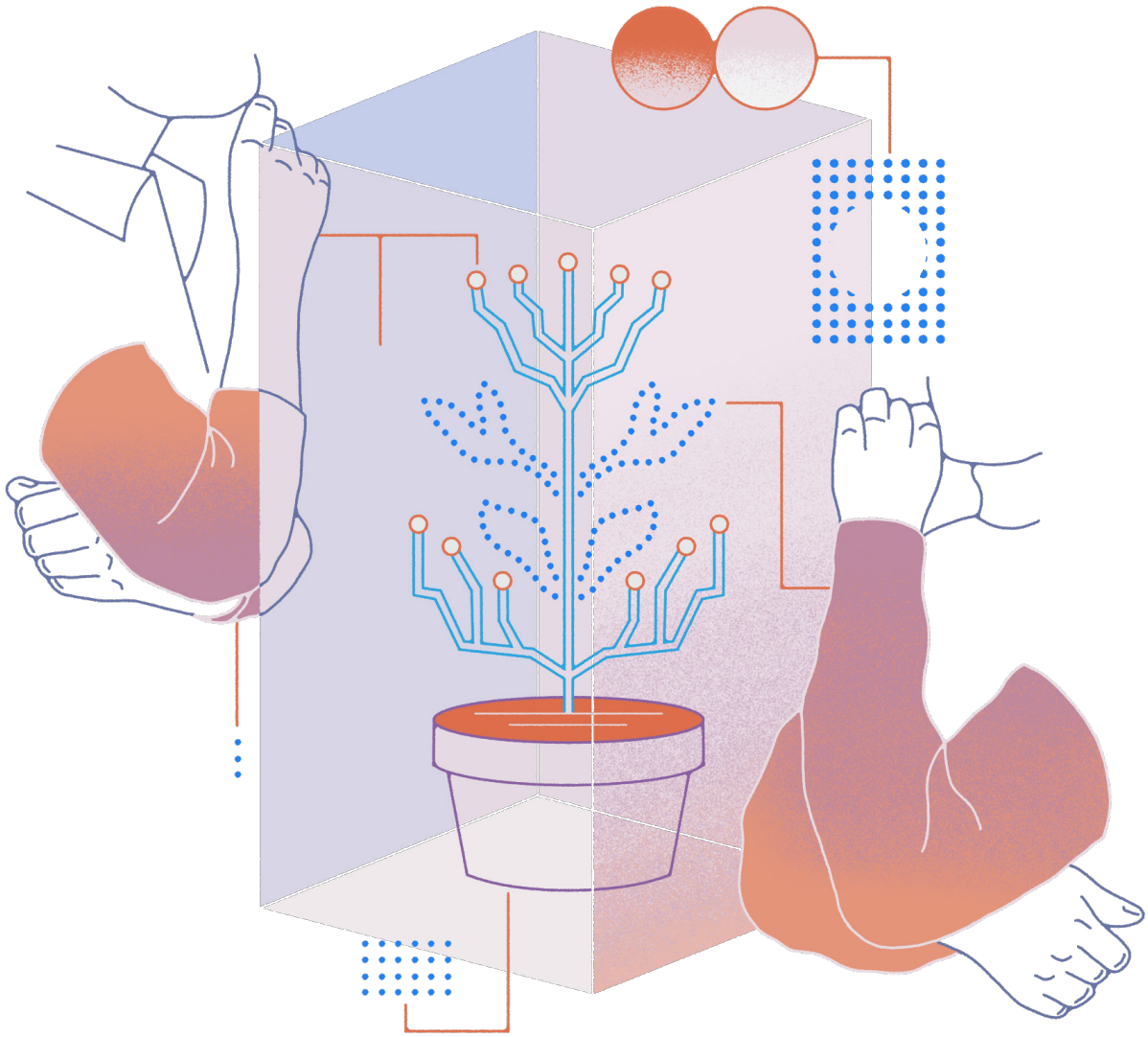
cdt | Research



Grounding AI Policy

Towards Researcher Access to AI Usage Data

Gabriel Nicholas

August 2024

The **Center for Democracy & Technology** (CDT) is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

A report from

Research

Grounding AI Policy

Towards Researcher Access to AI Usage Data

Author

Gabriel Nicholas

WITH CONTRIBUTIONS BY

Kevin Bankston, Miranda Bogen, Mona Elswah, Samir Jain, Michal Luria, Dhanaraj Thakur, and Amy Winecoff.

ACKNOWLEDGEMENTS

We thank Sreya Guha for her invaluable research help with this project. Thank you also to Solon Barocas, Chris Conley, and Zeve Sanderson for their insightful comments and suggestions. All views in this report are those of CDT.

This work is made possible through a grant from the John S. and James L. Knight Foundation.

Cover design by Alanah Sarginson.
Art direction by Timothy Hoagland.

SUGGESTED CITATION

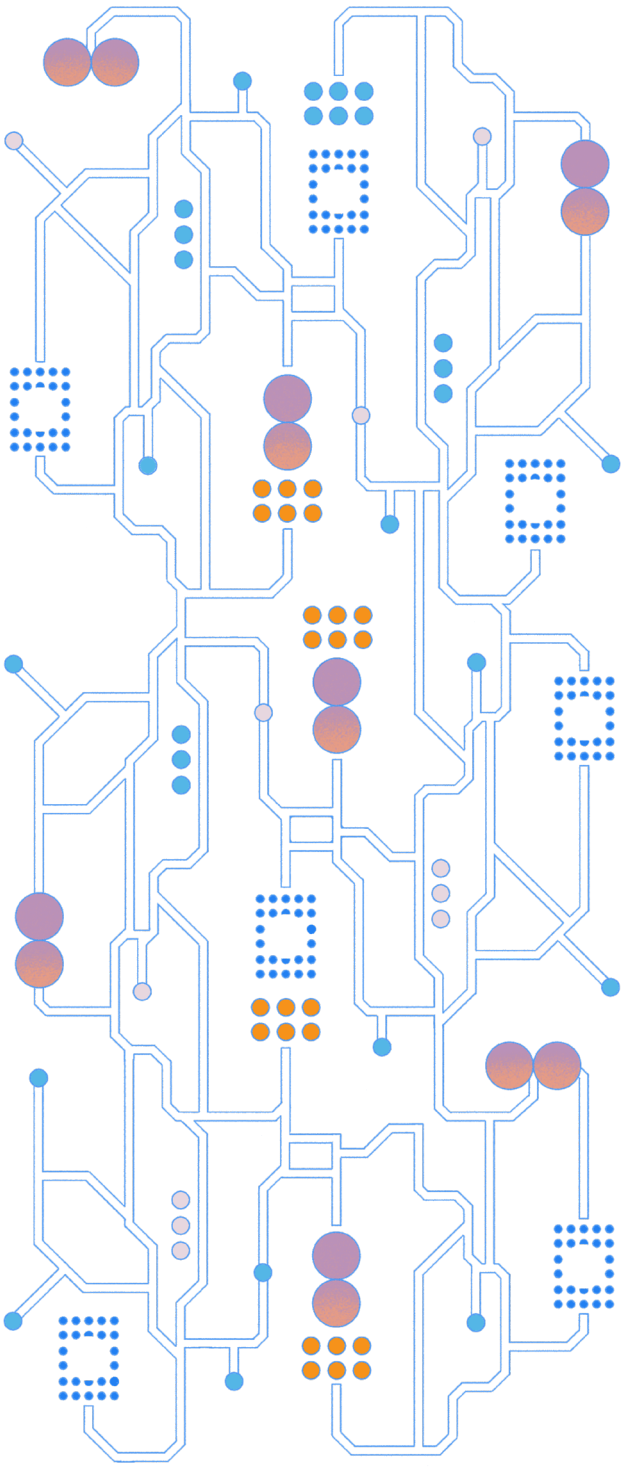
Nicholas, G. (2024) Grounding AI Policy: Towards Researcher Access to AI Usage Data. Center for Democracy & Technology. https://cdt.org/insights/grounding-ai-policy-towards-researcher-access-to-ai-usage-data/

References in this report include original links as well as links archived and shortened by the Perma.cc service. The Perma.cc links also contain information on the date of retrieval and archive.

Contents

Contents

Executive Summary



In the recent AI boom, researchers have uncovered countless potential risks posed by new, general-purpose AI systems, such as undermining election security, providing unsound medical advice, and exhibiting bias in employment decisions. But which of these risks are actually occurring? Which are only risks that *could* occur? And what methods do researchers and the general public have to begin to answer these questions? Currently, AI companies offer none. While many companies permit researchers to test their products for potential harmful uses through practices like red teaming, they offer no access to or information about how people ultimately use those products.

This "use case information gap" presents significant empirical challenges for policymakers seeking to develop evidence-informed AI regulation. Policymakers need to understand the prevalence of different AI-related harms to prioritize their limited resources to address the most pressing concerns. Without information on real world AI usage, policymakers also face an "unknown unknown" problem, in which they can't know about, investigate, or address real world harms.

However, sharing AI usage data — specifically in the form of chat logs — raises serious privacy and trade secrecy concerns. People use AI systems for sensitive purposes and may enter personal information that they do not want or expect to be shared with others. Furthermore, companies may be hesitant to share usage data that might reveal that people are using their products for unsavory purposes. And if usage data gets out into the wrong hands, competitors can potentially use system outputs to reverse engineer or recreate the companies' products for much cheaper.

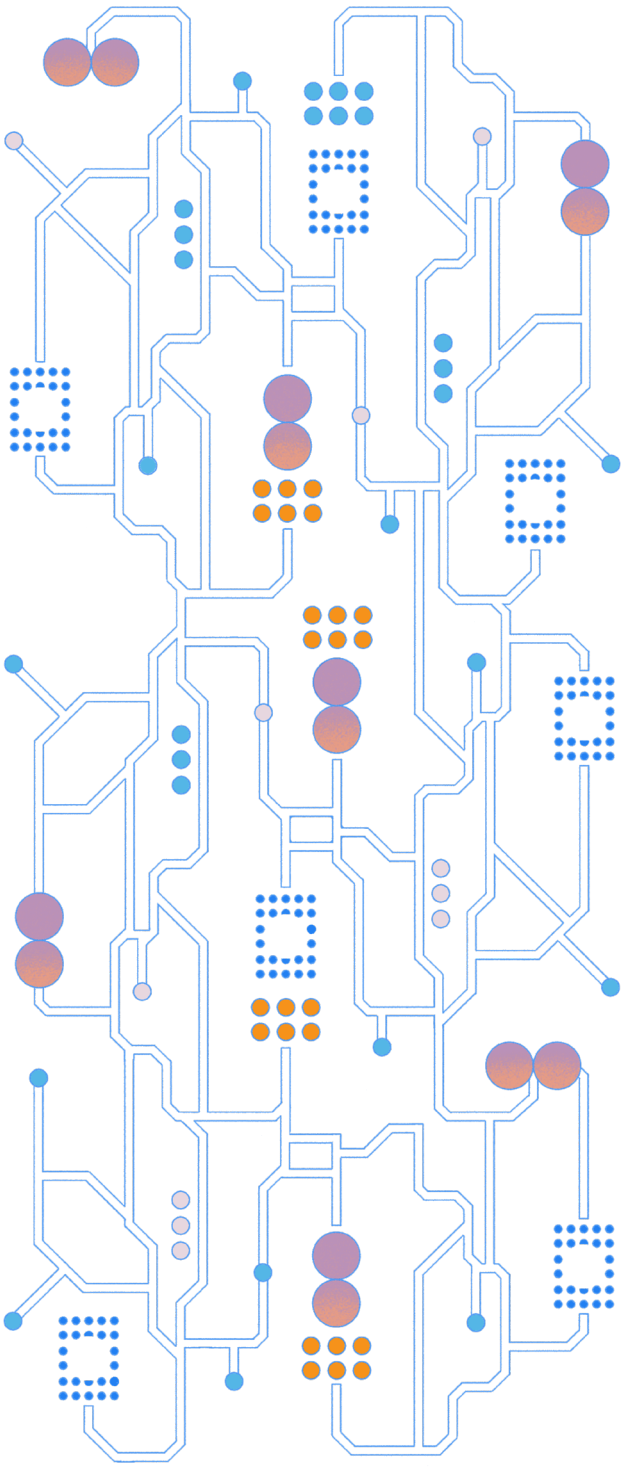
These challenges, however, are navigable; companies can share information with researchers about how people use their products in a way that informs policymaking without undermining user privacy and corporate security. This paper proposes three general approaches:

Data donations. AI companies should create mechanisms to allow users to share their chat log history and other information with researchers. They can do this by building APIs, allowing users to download and share their chat histories, or building a specific "Share your data with researchers" option into their products. Lawmakers can support data donations by protecting researchers who build tools, such as web scrapers, that allow users to consent to sharing their usage data through user-approved but company-unauthorized means.

Transparency reports. AI companies should share summary statistics about how people use their systems in high risk/high impact domains, such as health care and education. Companies can solicit feedback from experts in these domains about what information would be most relevant and include that in their reports. Governments can support these efforts by adding them to their voluntary company transparency reporting commitments, which currently only focus on disclosing products' capabilities, limitations, and trust and safety measures.

Direct data access. AI companies can develop technical mechanisms to share chat logs and other usage data with researchers while protecting user privacy. Direct access to chat logs by researchers can lead to significant privacy issues, but technologies like data clean rooms, differential privacy, and others yet to be developed may be able to reduce these risks. Lawmakers can support these efforts by establishing regulations that allow companies to test privacy-preserving data access programs without legal repercussions.

Introduction



A chair is for sitting. A clock is for telling time. To look at these objects is to understand their primary use. Until recently, AI was, in most cases, a similar technology, where design and use were closely linked. A facial recognition system recognized faces, a spellchecker checked spelling. Today though, with the advent of powerful "transformer models," a single AI application can (at least in appearance) be used to countless ends — to write poetry, evaluate a resume, identify bird species, and diagnose diseases. As possible use cases become broader, so do the potential risks, which now range from the malicious, such as generating propaganda or sexual images of children, to the inadvertent, such as providing misleading election or health information.

With these advances, companies and governments are rapidly integrating AI into new systems and domains (Knight, 2023). In response, policymakers are scrambling to regulate AI in order to mitigate its risks and maximize its potential benefits. This has manifested in a flurry of political activity, which in the US alone includes dozens of proposed federal bills, a small number of state laws and hundreds more state bills, the longest executive order ever issued, and a tide of regulatory guidance.

However, when designing new regulations, policymakers face an empirical dilemma: they must regulate AI without any access to real world data on how people and businesses are using these systems. Unlike social media and the internet, where user behavior is often public and leaves observable data traces, general-purpose AI systems are largely accessed through private, one-on-one interactions, such as chatbots. AI companies collect user interaction data, but are reluctant to share it even with vetted researchers, out of privacy, security, reputational, and competitive and trade secrecy concerns (Bommasani et al., 2024; Sanderson & Tucker, 2024). Instead, companies allow researchers and other external parties to probe their systems for vulnerabilities and harmful errors through practices such as red-teaming (Friedler et al., 2023). While these methods can help prevent AI systems from being used for the worst possible use cases, they do not offer empirical insights about the harms users experience in the real world.

The lack of available empirical information about how people use general purpose AI systems makes it extremely challenging to develop

evidence-informed policy. Three potential methods can help address this *use case information gap*, each with its own benefits and challenges:

1. **Data donations.** Users can voluntarily share data about their own interactions with AI systems (e.g., chat logs) directly with researchers (Sanderson & Tucker, 2024). AI companies can build technical tools to support this, including APIs, data portability tools, or a "Share your data with researchers" option. Researchers can also allow users to donate data directly, typically through browser extensions, without needing permission or support from companies. (Shapiro et al., 2021). Data donations raise few privacy concerns, but may introduce sampling bias, since those with the interest and technical skills to donate their data may not represent AI users writ large (van Driel et al., 2022).
2. **Transparency reports.** AI companies can analyze data about how people use their systems and share their findings with the public (Bommasani et al., 2024; Vogus & Llansó, 2021). Companies can solicit feedback from experts in high-risk domains, such as health care and elections, about what information would be of use to them. This kind of transparency report differs from the current White House voluntary commitments and similar efforts around the world, which focus on disclosing companies' efforts to keep users safe. Transparency reports raise little privacy risk, but can be opaque in their methodologies and details and potentially co-opted to serve company interests (Parsons, 2017).
3. **Direct access to log data.** AI companies can grant researchers access to chat log data and other information they hold about users' interactions with their products. Companies could provide this access directly, or indirectly by running queries on behalf of researchers. Companies could also provide this information voluntarily or, potentially, mandated under law (Lemoine & Vermeulen, 2023). Direct access poses significant privacy risks. While technical interventions might partially mitigate these risks, they may not be able to address them sufficiently to justify the practice. Companies may further resist granting direct data access, as it could jeopardize their reputation or expose corporate secrets.

This paper proceeds in three parts. First, it describes the use case information gap, why it should be closed, and what challenges there are to doing so. Then, it gives more detail on the three approaches to providing researchers access to use case information previously mentioned. Finally, it offers recommendations for how AI companies and lawmakers can implement these approaches in ways that benefit researchers and ultimately the public, while safeguarding users' privacy.

Definitions and Scope

This paper specifically focuses on researcher access to use case information for *popular, consumer-facing general purpose AI applications*. In practice, this means sharing chat logs from chatbots built by foundation model developers, such as OpenAI's ChatGPT, Google's Gemini, and Anthropic's Claude. This paper does not focus on these systems because they are the most important — indeed, they arguably receive too much attention already — but for practical reasons.

Working backwards, this paper focuses on AI applications, rather than foundation models (e.g., GPT-4, Claude 3 Opus, Llama) or model hosting services (e.g., the GPT-4 API, Stable Diffusion, Microsoft Azure). (Jones, 2023). Foundation models may not always have a centralized entity to monitor their use, as is the case with "open source" models, such as Llama and Mistral (Solaiman, 2023). Hosting services could in theory monitor AI usage, but moving governance and surveillance lower down the technical stack raises greater privacy concerns (Donovan, 2019). This merits its own analysis, outside the scope of this paper. This paper also focuses on consumer-facing AI products rather than business-to-business services, as the latter involves trade secrecy concerns that are beyond the scope of this study. Furthermore, it focuses on popular AI applications because they are more likely to have significant societal effects that merit research scrutiny and more likely to have the resources needed to build the infrastructure necessary to make usage data available to researchers.

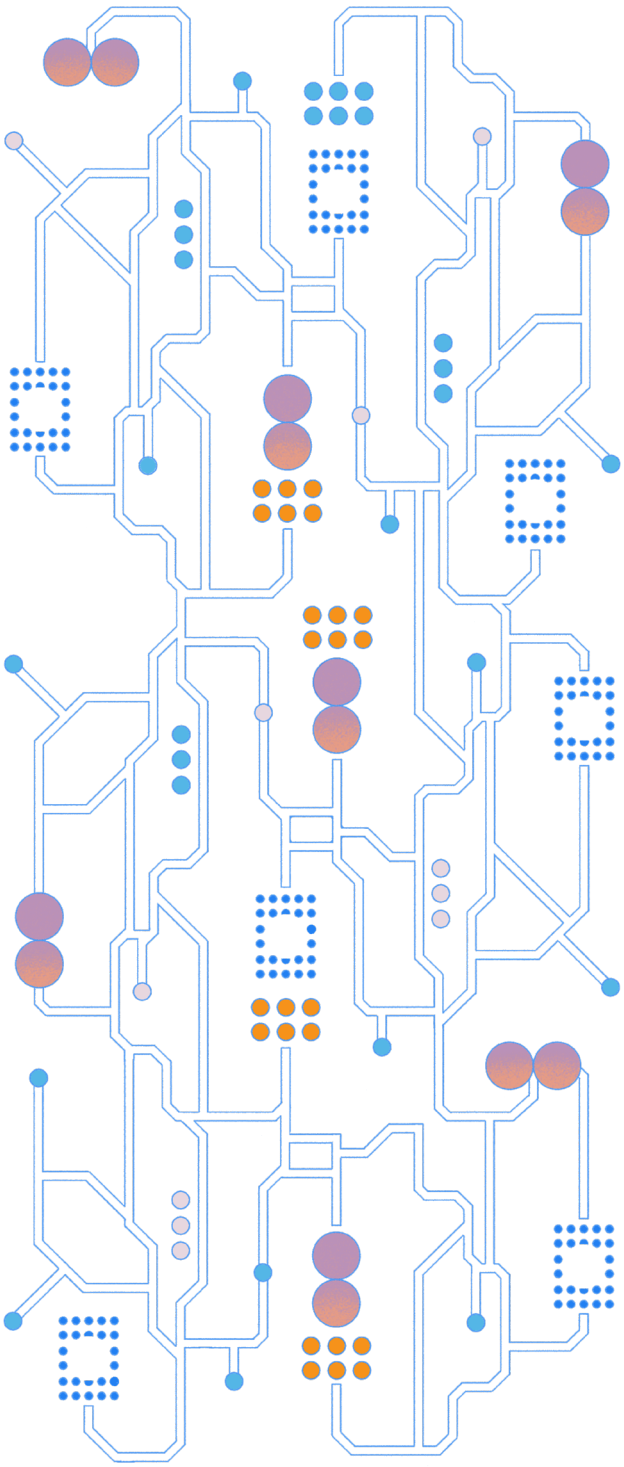
Finally, this paper borrows the concept of "general-purpose AI" (GPAI) from the EU AI Act, which defines it as, "an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications." (AI Act, Article 3, Section 44b). While concepts like "generality" and "capability" are up for debate, this paper focuses on chatbot applications built on top of state-of-the-art models designed to cover the broadest range of domains, rather than narrow uses such as customer service chatbots.

With a definition of "AI systems" in hand, we can clarify what we mean by use case information. This paper primarily focuses on use case information as chatlogs, i.e. the text and other media content of a user's messages and the AI system's responses.

Chatlogs are limited, since they reveal nothing about the context of usage. For example, a user asking a chatbot to write an email asking for an unpaid payment could be using that text to run a phishing scam or to help navigate an awkward conversation with an associate about money. As will be discussed later, chatlogs also risk exposing very personal or personally identifiable information, which can be challenging to conceal from researchers.

Use case information can also include metadata, which is information about the data. Metadata may encompass details about the conversation itself, such as timestamps, session identifiers, AI system versions, error logs, usage policy violations, and refusals, as well as other actions the user has taken, such as regenerating a response or flagging content. It can also include information about the user, such as user identifiers, device information, and location data, but due to the high risk of user re-identification, information about the user is outside the scope of this paper.

Background



Why should AI companies share information about how people use their products?

The theory of change underlying companies sharing use case information is one of transparency. Greater public awareness of how general-purpose AI systems are used both for good and for ill would, in theory, incentivize companies to develop these systems in a more socially responsible way due to potential brand damage and regulation. Transparency studies scholars, however, frequently criticize the use of transparency as a means of creating accountability for overseeing technology (Laufer et al., 2022). They argue that transparency initiatives are often ineffective because they are disconnected from power (boyd, 2016), co-opted for political ends (Annany & Crawford, 2018; Widder et al., 2023), or designed without an end user or specific purpose in mind (Corbett & Denton, 2023). Misguided transparency initiatives can be outright detrimental to the problems they seek to address, potentially obfuscating or redirecting attention from the most significant problems (Stohl et al., 2016; Zalnieriute, 2021), externalizing the burden of oversight to resource-strapped watchdog organizations (Birchall, 2021), and creating "sunshine rules" that those seeking to avoid oversight can use to intentionally tie up government resources (Pozen, 2018).

Closing the use case information gap avoids many of the pitfalls other transparency initiatives face. First, it has the potential to be connected to power, since as of this writing, state and federal lawmakers are working very actively to craft laws and issue guidance regarding AI (Lenhart, 2023). Second, researchers who study the use and misuse of AI are relatively well-funded by philanthropic groups (e.g., Prest, 2023), and may receive more funding in the future through government initiatives such as the National Artificial Intelligence Research Resource (National Science Foundation, 2024) and the AI Safety Institute (AI Safety Institute, n.d.). Finally, at least in this paper, the purpose of use case information is straightforward: to allow for more evidence-informed policymaking.

Providing researchers with better access to AI use case information can allow for more evidence-informed policymaking in three ways. First, understanding the prevalence of various AI system uses can help policymakers more effectively map, measure, and manage their

associated risks (Caliskan & Lum, 2024). In its AI Risk Management Framework, the National Institute for Standards and Technology (NIST) defines risk as "the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event" (NIST, 2024). Evaluating the probability and magnitude of harm is difficult without empirical data; human intuition alone is often inaccurate. For instance, despite parents and teachers expressing concerns that ChatGPT in schools would lead to a surge of cheating (Laird et al., 2023), initial research has found that high school students are cheating no more than they were before (Singer, 2023). Access to research grounded in use case information can help policymakers rely less on their potentially misguided intuitions about risk and more on data about where those risks are actually occurring.

Second, use case information could allow policymakers to more effectively prioritize conflicting normative goals. In technical and regulatory design concerning general-purpose models, different normative goals are often at odds with one another (Guha et al., 2023). For instance, research from the AI Democracy Project found that today's state of the art chatbots often provide inaccurate, misleading, and even harmful information about US elections. Examples include citing incorrect election laws, providing inaccurate voter registration information, and directing users to incorrect polling places based on their zip codes. (Angwin et al., 2024). Shortly afterward, Google prevented its Gemini chatbot from responding to all global elections-related questions (Singh, 2024; Gilbert, 2024). Without data on how and how many people use Gemini to access election information, it is impossible to evaluate whether this policy, on balance, is preventing the spread of election misinformation or making it more difficult for people to politically educate themselves. Technologists and lawmakers both face inescapable tradeoffs, and require empirical evidence in order to navigate them effectively. Guidance from the National AI Advisory Committee words this point more specifically (NAIAC, 2023):

> *The relative importance of harms – and magnitude of harms relative to baseline systems – can be fiercely contested. Ideally, policymakers should be able to make evidence-informed decisions about the relative gravity of distinct harms, which could also involve a consideration of how harms are forecasted and mitigation strategies are operationalized.*

The inability to weigh harms also raises resource allocation questions for agencies like the Cybersecurity & Infrastructure Security Agency (CISA), who are responsible for securing the physical and cybersecurity of America's voting system (Harris et al., 2024). Are enough people using AI systems to access election information to make it

worthwhile for agencies to issue guidance on best practices, even if that drains resources available to their other work on the risks generative AI poses to elections? These include allowing foreign nation state actors to more easily run influence operations, create fake voting records, and target election officials with voice cloning (CISA, 2024).

Third, use case information could help policymakers identify new areas of concern worth mitigating and new beneficial use cases worth protecting. While many scholars have theorized about the risks of general purpose AI systems (e.g., Okerlund et al., 2022; Weidinger et al., 2021; Shevlane et al., 2023, Shelby et al., 2022; Solaiman et al., 2023), there may be malicious uses and inadvertent harms occurring that scholars could not have predicted (e.g., Firdhous et al., 2023). On more public-facing general-purpose technologies, such as social media and the internet, unexpected harms can be publicly observed — for instance, no one may have expected TikTok to contribute to children eating Tide Pods, but when it happened, researchers and journalists could raise awareness, and local health officials could respond (American Poison Centers, n.d.). Were an AI chatbot to recommend similarly bizarre and dangerous behavior, it would be harder for the public to find out so long as companies held onto exclusive access to chat log data.

Current state of use case information sharing

AI companies reveal almost no information about how people use their products. As of this writing, no major AI company has (at least publicly) shared chat logs with third party researchers, created an API or data portability option to allow users to share their chat logs with others, or otherwise shared meaningful high-level usage patterns. This is not because companies do not collect this information — a review of the privacy policies of ChatGPT, Gemini, Claude, Bing Copilot, and Grok indicates that all monitor how users use their systems. All monitor for violations of their usage policies and use that data for research and to improve their models (Anthropic, 2024; OpenAI, 2023b; Google, 2024; Microsoft Copilot, 2024; xAI, 2024).[1]

The limited usage information that companies have made available is of little utility to researchers. Corporate disclosures are limited to publicly disclosed partnerships and customer success stories — for instance, Oscar using OpenAI to build an

1 It is worth noting that many of these services offer a zero-data retention policy option for certain enterprise users, where they do not collect the data (e.g., Anthropic, n.d.; OpenAI, n.d.). They also all retain data for different periods of time.

insurance claims assistant (OpenAI, 2024), or GitLab using Anthropic to build a code generation tool (Chu, 2024). Companies will also sometimes host forums where people can discuss projects where they use these tools (e.g., OpenAI's help forum, Cohere's Discord), or forums will emerge in which people discuss how they are using these tools, such as on Reddit, Github, and HuggingFace. There is at least one public example of a company using this information in the past: when deciding whether or not to publicly release the model weights for GPT-2, OpenAI scanned public forums to determine whether anyone had discussed using it for harmful purposes (they found no examples) (Solaiman et al., 2019). However, these publicly discussed use cases are hardly representative of how people use general purpose AI systems, and may capture aspirational use cases more than actual ones.

Scholars have developed several tools and methodologies to assess the societal impacts of building and deploying AI systems, a few of which companies have actually adopted in some form. Some of these approaches involve providing access to the underlying components of AI models to allow researchers to evaluate their potential capabilities and shortcomings (Casper et al., 2024). Examples include publishing model weights (Friedler et al., 2023), red-teaming (Ganguli et al., 2022), and explainable AI methods (Lipton, 2016; Nicholas, 2020). With few exceptions (e.g., Kaur et al., 2022; Luria, 2023), these methods focus entirely on technical artifacts (Eyert & Lopez, 2023), namely the data used to train the model, the model itself, and in some cases, the code used to train the model. This added layer of abstraction means that publicly available data reveals nothing about how end users actually utilize these systems.

Other approaches to AI transparency in the academic literature focus on documenting or evaluating how well a model works in a given context. Examples of this include algorithmic impact statements (Selbst, 2021), audits (Raji et al., 2021), and documentation initiatives (Micheli et al., 2023). These approaches provide context for how AI models may be used and can help pinpoint factors that heighten risks for specific user groups. However, they often have built-in assumptions about how people use these models, which can limit how much they reveal about real world usage. For instance, most AI documentation initiatives, including model cards, datasheets, system cards, FactSheets, data statements, and data nutrition labels, have the AI developer disclose the "purpose" or "intended use case" of their system and evaluate how well it might work towards that end (Arnold et al., 2019; Procope et al., 2023; Gebru et al., 2018; Mitchell et al., 2018). While these approaches are designed for revealing risks of narrow-purpose AI systems, rather than general-purpose ones, they have paved the way for scholars to begin considering how transparency efforts can best be tailored to individual use cases (Mohammad, 2022; Hupont & Gomez, 2022).

Scholarship has also explored concepts analogous to use case transparency for general-purpose models (Mökander et al., 2023). Even before the popularity of transformer models, Mittelstadt et al. (2016) argued for separately considering what a model is designed to do, how it works, and its impact on the world. The Partnership on AI's conceptualization of risk distinguishes between "model risk" (associated with the model itself) and "application risk" (associated with downstream use cases of that model), and between "known risk" (that have been identified and understood empirically) and "speculative risk" (that are hypothetical) (Partnership on AI, 2023; Hutchinson et al., 2022).

The most in-depth framework for evaluating use case transparency is the Foundation Model Transparency Index (Bommasani et al., 2024). The index is comprised of one hundred binary indicators designed to assess a foundation model's transparency. A third of the indicators pertain to downstream use of the model, which includes governance mechanisms as well as seven indicators particularly focused on impact. These impact indicators align closely with the concept of use case transparency and include whether the number of applications using the model is disclosed, the proportion of applications across different market sectors, usage reports, statistics on model usage across geographies, and the number of affected individuals. However, the index does not go so far as evaluating researcher access to specific data such as chat logs.

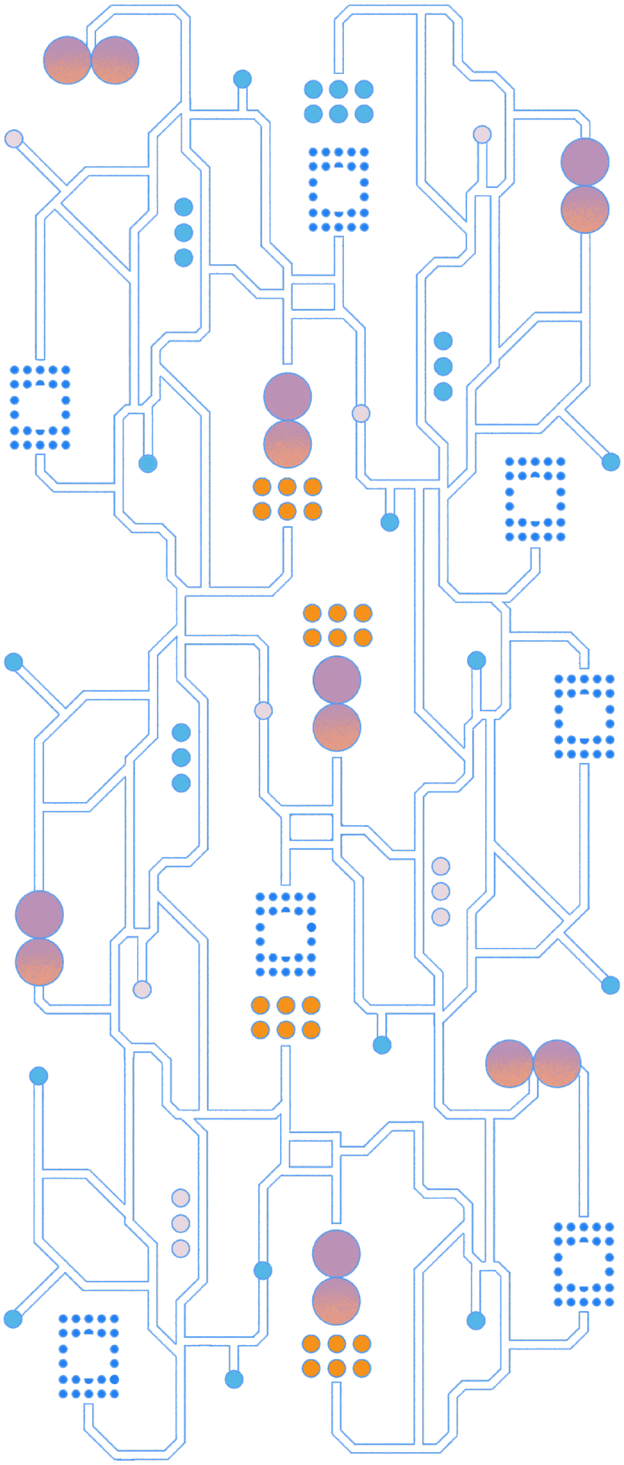
Risks of use case information sharing

The primary risk to users from sharing use case information is individual privacy. People use general-purpose AI systems for lots of different sensitive purposes, including medical advice (Leonard, 2023), therapy (Lucas, 2024), and sexual pleasure (Verma & Oremus, 2023). Employees enter into these systems confidential business data, source code, personally identifiable information, and their private thoughts (LayerX, 2023). If this data ends up in the hands of external researchers without users' knowledge or permission, it could constitute a very serious privacy violation (Benthall & Sivan-Sevilla, 2024; Nissenbaum, 2009). Furthermore, if users fear that this data could be collected without their knowledge, they may avoid using AI applications in beneficial ways due to concerns about what information they might inadvertently make public. Any effort to share AI use case information with researchers must balance the utility of the data to researchers against the individual privacy risks to users. Sharing images and videos that users generate using AI with researchers creates its own set of challenges, privacy and otherwise. Users might use AI applications to generate non-consensual intimate

imagery or child sexual abuse material. If researchers end up holding this content, it exposes them to significant legal liability and raises several practical and ethical concerns about how to handle and report this information (Thiel, 2023).

For companies, the primary risks of sharing use case information concern protecting corporate secrecy and maintaining their reputations. If researchers are able to use chat logs and other information to reverse engineer AI products, it could undermine their claims to intellectual property and trade secrecy (LaRoque, 2017). Furthermore, if use case information gets in the hands of competitors, it could allow them to build similar, competing products using far fewer resources. For instance, UC Berkeley researchers found that by fine-tuning LLaMa on a dataset of 70,000 user-shared conversations from ShareGPT, they ended up with a model that achieved approximately 90% of the performance of GPT-4 at a training cost of just $300 (The Vicuna Team, 2023). Use case data can also pose reputational risk for companies. Researchers may use this data to find that AI applications are being used in harmful ways, or may not work as well as companies promise they do, any of which may hurt an AI company's market position.

Approaches to sharing use case information



As the previous sections have shown, if researchers had better information about how people use general-purpose AI systems, it would enable policymakers to develop more evidence-informed regulations and guidance. However, companies do not share information about how people use their systems, in part to protect the privacy of their users and in part to prevent competitors from being able to use this information to their advantage. In this section, I review three possible ways companies could better share this information, and offer analysis on both their benefits and risks.

Data donations from users

Data donation is when users voluntarily share data about their activity on a service to help researchers achieve a specific goal. Data donation has been used to organize tech laborers (Chan, 2021) and generate medical research (National Cancer Institute, n.d.) but the closest analog to donating general purpose AI data is social media research (Aslett et al., 2023; Ohme et al., 2024). Data donation is an imperfect methodology, since those who are willing and who have the necessary technical knowledge to donate data may not be statistically representative of the whole population. However, data donation does have the advantage of involving direct user consent, although bad faith actors could pose as researchers and use that data for inappropriate purposes. The example of social media data donation points to a few ways data donation can be operationalized with general purpose AI applications, each with varying degrees of platform involvement.

Application programming interface (API)

An application programming interface (API) is a technical means for allowing one computer to request or send data to another. Software products will sometimes offer APIs that allow third parties, including researchers, to request data on behalf of a user. This creates a seamless experience for the user, who only needs to log into an application with their credentials. However, it restricts the researcher to collecting only the data that the company allows the user to share. For instance, an AI company could create an API that allows users to give third-party researchers access to the prompts they input into their model, but not the model's response. The company could limit researchers' access in

other ways as well, such as allowing them only to request the last five conversations the user had, or only allow the researcher to query the model once per day. APIs are easy for researchers to integrate, but their design and access policies are entirely subject to the platform's whims. Platforms also have to carefully navigate tradeoffs when designing APIs: permit researchers to access too much information, and it may create privacy risks, as happened with Cambridge Analytica (Cadwalladr & Graham-Harrison, 2018); permit too little, and the data is not useful to researchers (Tromble, 2021). As of this writing, no major AI company, including OpenAI, Google, Microsoft, Anthropic, or xAI, allows users to share their chat log history or any other usage data via API.

Data portability

Another option for users to donate their AI usage data to researchers is through data portability tools. Data portability is the ability to download one's own data from a product or service in order to bring it elsewhere. Ported data often offers a fairly complete picture of how a user has historically interacted with a product. For instance, Facebook's data portability tool allows users to download every friend request, event invitation, post, and more they have ever made on the platform (Nicholas & Weinberg, 2019). If AI companies offered data portability for their products, users could download their data and share it with researchers. The EU's General Data Production Regulation (GDPR) and the California Consumer Privacy Act include rights to data portability, so companies have a harder time interfering with access than they do with APIs.

However, data portability also has several limitations as a research tool. First, taking advantage of data portability tools often requires some tenacity and technical expertise since the options to download one's data is often buried deep in an application's settings. Second, without fine-grained privacy controls around what data is included in data downloads, users may end up sharing much more data than they want to with researchers. Third, data portability mechanisms often only show data users have provided to platforms themselves, which leaves out important context. For instance, if one downloads their Facebook data, they can see every comment they have made on others' posts but cannot see the original post, since that is not their data to port (Nicholas, 2020).

General purpose AI companies have so far offered little in terms of data portability. ChatGPT allows users to download their chats, but Gemini, Anthropic, BingChat, and Grok do not.

Web scraping

Finally, researchers can allow users to donate AI usage data via web scraping. (Sanderson & Tucker, 2024). In other words, researchers can create browser extensions that snapshot what users see when they interface with AI applications, and users can opt to participate in research and donate their data by installing that extension. Browser extensions can in theory create user privacy risks, since they can collect more data than they promise, but in the social media example, researchers have mitigated these risks before by open sourcing their code and having third-party, independent security and privacy audits (Erwin, 2021).

The downside of web scraping is that developing these browser extensions can be technically challenging for researchers and can create legal liability concerns (Shapiro et al., 2021). Companies often redesign their websites and employ anti-scraping technology, so researchers may need highly technically capable programmers available to constantly update their systems in order to continue to collect data. Data donation via scraping is also far more challenging on mobile apps, requiring methods such as screen recording (Shore & Prena, 2024). Beyond technical considerations, the practice of data scraping exists in a legal gray zone (Sellars, 2018). Furthermore, AI companies vary in whether they allow users to own the output an application provides from the user's inputs. For instance, while OpenAI allows users to own ChatGPT's outputs (OpenAI, 2023a, 3.1), Anthropic only authorizes people to use Claude's outputs under their Usage Policy (Anthropic, 2023, 6(a)), and Cohere retains the rights to its model's outputs altogether (Cohere, n.d., 9(a)). Uncertain legal grounds can lead researchers to abandon a project, fearing the financial burden of defending against even frivolous lawsuits.

However, researchers have found ways to enable data donations. Zhao et al. (2024) for instance developed WildChat, a wrapper around ChatGPT and GPT-4, that they published on HuggingFace. This tool allowed users to access these usually paywalled models for free in exchange for consenting to donate their chat conversations to researchers. Among other findings, researchers found that 70% of the time people attempted to use a publicly available jailbreak prompt on the model, they were successful. Other researchers have used chat log data available on the public web. Ouyang et al. (2023) compiled a dataset of nearly 100,000 conversations from ShareGPT conversations, a third party tool users can use to share and archive conversations they had with ChatGPT. They used this data to understand where existing benchmarks diverge from real world use cases.

Transparency reports

A transparency report is a document issued by a company or government that discloses information about how it operates. According to Access Now, at least 85 internet and telecommunications companies have published transparency reports (Access Now, n.d.). Usually, technology companies' transparency reports disclose either legal information requests (e.g., subpoenas from law enforcement or government agencies), or content and platform enforcement measures (e.g., content removed for violating a company's terms of service, intellectual property laws, or other content takedown regulations) (Trust & Safety Professional Association, 2023). Transparency reports give companies a legal and privacy preserving way to disclose information about their practices in sensitive yet socially valuable areas. However, when companies have broad control over what they reveal and how, transparency reports risk being reduced to marketing materials (Vogus, 2022; Zalnieriute, 2021).

Governments, academics, and companies have sought to adopt the approach of transparency reports, popularized by internet and telecommunications companies, to foundation models. Multiple national and multinational bodies have had companies sign off on various codes and commitments in which they promise to issue some form of transparency report as a means of ensuring AI safety and accountability. Examples include the G7 Hiroshima Process, the White House Voluntary Commitments on AI, the US AI executive order, and Canada's AI Code of Conduct. Some of these efforts have focused on the underlying model while others have focused on AI applications. The White House Voluntary Commitments (2023), for instance, focus largely on public reporting about theoretical uses of AI systems, such as vulnerabilities discovered by red teams, potential societal risks, domains of appropriate and inappropriate use, and model capabilities and limitations. Narayanan & Kapoor argue that AI transparency reports should focus on policy enforcements on AI applications and the spread of harmful content. In particular, they recommend AI companies disclose how they define harmful content, how frequently users encounter it, what enforcement mechanisms and safety mitigations have been put in place, and how effective they are (2024). As of this writing, the only signatory of these government AI transparency efforts to have released a transparency report is Microsoft, and their disclosure largely consisted of marketing language and decontextualized statistics (Smith & Crampton, 2024).

General-purpose AI transparency reports could inform the public about more than just the model and policy enforcement. Transparency reports could reveal how and how frequently people attempt to use a given AI application for high risk tasks, such as seeking medical information, financial advice, or information about elections.

(Transparency reports are better suited to providing insights into user queries than AI system responses, as the latter can fluctuate significantly based on query specifics and individual user customizations.) While we don't see or expect similar transparency reports for other general-purpose technologies, like document sharing software or email, AI applications should be held to a higher transparency standard because they are at least partially responsible for the content that gets generated and are more likely to have unexpected uses and effects.

Companies can take either a top-down or a bottom-up approach to sharing use case information in transparency reports. In the top-down approach, AI companies would solicit questions from stakeholders and experts in these different high risk arenas and publish the answers. For example, if news emerged that someone became seriously ill after following misleading medical advice from an AI product about a specific disease, health officials might want to investigate how often people use that product for diagnosing that disease. This type of transparency report is likely more useful to stakeholders, as it includes their direct input. However, AI companies might struggle to provide accurate responses because they lack the context in which the users are utilizing the product. In the previous example, it may be difficult to distinguish whether users are asking about the disease to diagnose their own condition, out of curiosity, or to test the model after seeing the news coverage about how it does poorly on it.

In the bottom-up approach, AI companies use data analysis techniques to understand usage patterns and share popular clusters of different use cases with the public. Microsoft Research has employed this method with chat logs from Bing Copilot, though their research has been quite high level. One study compared Bing Copilot usage to Bing Search, finding that users relied on Copilot more for knowledge work and complex tasks (Suri et al., 2024). Another study developed a taxonomy of user intents behind use of the large language model (Shah et al., 2023). However, it is still unclear whether or what it would take for this sort of high-level information to benefit researchers and policymakers.

Direct data sharing

In addition to users, AI companies themselves can share AI usage information. While this data is valuable to researchers, it also raises significant privacy challenges. AI companies collect and store information on users' prompts (e.g., text prompts, file uploads), their applications' outputs or other responses (e.g., text responses, generated images or video, errors, flagged content), and users' actions on those outputs (e.g., positive or negative feedback, attempts to generate a new response, external sharing). Together, these data could give researchers a holistic view of how people use AI systems

and what real world risks they raise, without concerns about skewed samples (as with data donations) or company bias (as with transparency reports). Of course, directly sharing data without users' explicit permission could, as discussed earlier in this paper, significantly undermine users' privacy, and, if it fell into the wrong hands, harm AI companies' market position.

However, other sectors have "navigated the Scylla and Charybdis of privacy and trade secrecy" in order to have companies share sensitive data with researchers so as to better understand the effects of new technologies (Morten et al., 2024; Nicholas & Thakur, 2022). The Food and Drug Administration, for instance, manages to require pharmaceutical and medical device companies to share data about their clinical trials with researchers without posing major healthcare privacy risks to patients (Herder et al., 2020). Utilities have been able to share electricity usage data with researchers in a way that does not reveal sensitive information about customers (Fournier et al., 2020). And in the case of social media, although voluntary researcher access to data has long been controversial, companies have been able to share data with researchers in a manner that allows impactful research without compromising their users' privacy (Lapowsky, 2024).

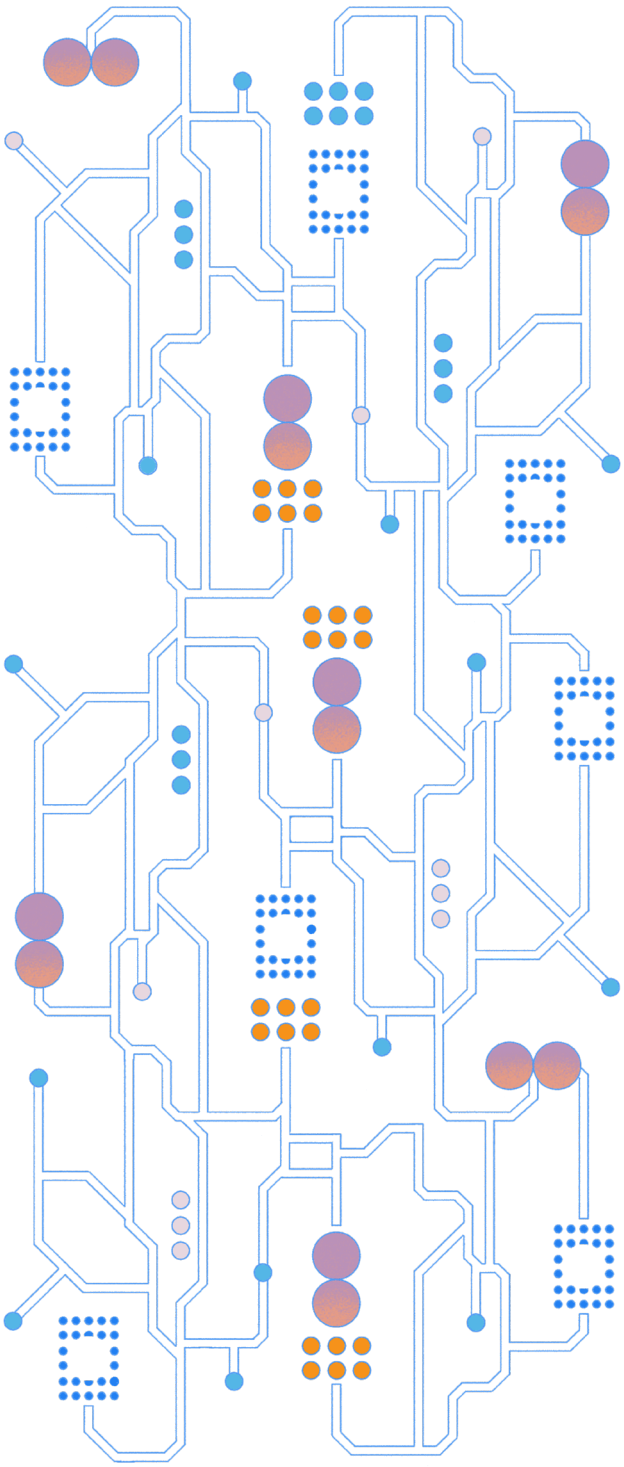
AI companies can adopt any of several approaches to share user data with researchers in a more privacy-preserving manner. However, it remains uncertain whether any of these methods can fully (or at least sufficiently) mitigate the privacy risks associated with direct data access. One approach is to use data clean rooms, where a company shares data with an external party in a secure, monitored environment. In this setting, the external party cannot access the raw data but can analyze and aggregate it safely. Another approach could be to use differential privacy, a statistical method of adding random noise to the data that still preserves certain traits to make it useful for researchers to analyze. Finally, companies can more directly intermediate access to data, such as by running queries on behalf of researchers. Currently, there are no publicly known examples of AI companies sharing usage data with external researchers, so these methods have not been tested. Although they have been used in other domains, applying them to general-purpose AI systems might present unforeseen challenges.

Finally, AI companies, or at least social media platforms with AI features, may be required by law to share usage data with researchers upon request. In the EU, under the Digital Services Act (DSA) Article 40, Very Large Online Platforms (VLOPs) and search engines (VLOSEs) — i.e., platforms with more than 45 million active monthly users — are obligated to share certain data that researchers request access to (DSA, Article 40). Depending on how the DSA is interpreted, AI applications that grow to have enough users could be considered VLOSEs (Lemoine & Vermeulen, 2023), which would make questions of how to safely operationalize researcher access to their

data of immediate importance.[2] Proposed legislation in the US regarding researcher access to data, such as the Platform Accountability and Transparency Act and the AI Foundation Model Transparency Act, could also have similar effects (H.R.6681, S.5339).

2 In addition to depending on future definitional work from the EU about whether/which AI applications count as VLOSEs, it would also matter whether a conversation with a chatbot counts as a private conversation service, meaning it would be excluded from Article 40. Thank you to Matias Vermeulen for pointing this out.

Recommendations



AI companies should build tools to allow users to donate their usage data to researchers

AI companies should develop research APIs that allow users to donate their chat logs and other usage data to vetted, third-party researchers with appropriate safeguards. Among the options for data donation, APIs have the potential to be the easiest for researchers to implement and the most privacy protective for users. A well-designed research API can ensure that participants clearly understand what data they are consenting to share with researchers and limitations on any secondary uses or disclosures and allows them to revoke access at any time. APIs position AI companies as intermediaries, vetting who gets access and for which projects. This allows companies to stop bad actors or privacy violating research projects and impose appropriate privacy and security safeguards (Vogus, 2022). At the same time, companies should not be permitted to censor research because it harms their brand or unduly limit researchers by requiring them to sign onerous data access agreements that grant the company final approval of research before publishing, prevent researchers from sharing the underlying data with peer reviewers with appropriate safeguards, or other undue burdens. Establishing public guidelines for research approval, avoiding non-disclosure agreements for researchers with respect to their findings, and involving neutral third parties to approve researchers and projects can help maintain this balance. Examples such as Twitter's former academic research API could serve as a good model (Twitter, n.d.).

AI transparency reports should include information about how people use their product

AI companies should work to develop transparency reports that not only provide information about the safety and governance of their systems, but also give researchers and the public insight into how people use them. AI companies should do this in both a top-down and bottom-up way. They should engage with researchers, policymakers, civil society, and other external stakeholders to determine

what use case information would be most helpful to them. They should publish this information, and also share their methodologies for analysis so that those relying on these disclosures can better understand how to use them. Companies should also share their own research on how people take advantage of the tools they provide to illuminate unexpected use cases. Finally, companies should share summary information about product usage, both in general and in specific high-risk domains like finance and education, based on input from external stakeholders.

AI companies should pilot larger data sharing programs with researchers

AI companies should experiment with ways to share usage data directly with researchers in privacy preserving ways. Currently, it is an open question whether existing privacy practices, including those making data access safer (e.g., data clean rooms, indirect company requests) and those making the data reveal less personal or sensitive information (e.g., differential privacy, various de-identification processes), can sufficiently enable companies, rather than users, to share usage data with researchers without compromising privacy. Academic researchers and AI companies should both dedicate resources to developing these methods, and AI companies should be willing to pilot new programs. This could involve allowing researchers to apply for access to specific AI usage data and include methods in their requests to address various privacy and cybersecurity concerns.

Social media platforms should include AI information in their research APIs and data portability tools

General purpose AI companies are not the only ones with access to data about how people use AI — surfaces where AI-generated data is shared, in particular social media companies, may also have access to this information. Social media companies should include AI-related metadata in their research APIs about data generated both on- and off-platform. For instance, when researchers access public content via the Meta Content Library, that data should include information about whether it was generated using Meta AI or whether an uploaded image included Coalition for Content Provenance and Authenticity (C2PA) or International Press Telecommunications Council (IPTC) metadata that suggests it was AI-generated (C2PA, n.d.; IPTC, n.d.). Platforms can also provide information from other signals they use to determine whether media is

AI-generated (e.g., Google's SynthID, image scans) and share a confidence interval with researchers. This latter data should only be made available to vetted researchers, since it could potentially be used to reverse-engineer and undermine companies' AI-detection systems. Finally, platforms that host chatbots on their services, such as Meta and Snapchat, should allow users to export their conversations with these bots as part of their data portability services.

Lawmakers should protect users' abilities to donate their AI usage data to researchers, including through scraping

Data donation via user-permitted web scraping can provide meaningful utility to researchers, create a minimal burden for companies, and give users the choice whether to assume any privacy risk from the data sharing. However, in the US, the practice falls in a legal gray area. Policymakers should clarify the law to make sure researchers employing user-permitted web scraping in good faith do not face legal risk from AI companies, and that AI companies are not held liable for mishandling user data if they allow such tools.

There is precedent for companies using privacy law as a justification for shutting down independent research in social media. Facebook, for instance, shut down the NYU Ad Observer, a browser extension that users could install to scrape and donate data on the political ads they encountered in their feed, arguing that it went against their Terms of Service and their FTC consent order (Clark, 2021). Similarly, Facebook shut down AlgorithmWatch's project to allow users to donate information about what they were recommended on their Instagram newsfeed, also using scraping, arguing that it went against their Terms of Service and GDPR (Kasyer-Bril, 2021). Both projects had explicit user consent, were open source, and were instrumental to journalism covering Facebook. The NYU Ad Observer was also independently audited for user privacy and security (Mozilla, n.d.). In response, the FTC clarified that the NY Ad Observer did not go against the consent order, and reprimanded Facebook for using it "as a pretext to advance other aims" (Levine, 2021). US lawmakers also proposed a number of laws shortly after this incident that protected researcher's abilities to independently scrape and analyze platforms, including the Platform Accountability and Transparency Act (PATA), the Social Media Data Act, and the Digital Services Oversight and Safety Act. The EU went one step further, protecting researcher access to data under DSA Article 40.

Lawmakers should offer similar protections to researchers and AI companies concerning users donating their usage data. By issuing clarification that AI companies will not be accused of privacy violations for supporting good faith, privacy-protective research efforts, companies cannot hide behind the figleaf of legal liability when prohibiting research, and companies that are actually interested in sharing this data could do so without fear of recourse. This can be done through bills explicitly designed to protect researcher access to data, or through government and multilateral bodies offering guidance on best practices.

Lawmakers should encourage transparency reports that include information about how people use AI systems in high-risk domains

As previously discussed, global voluntary commitments for transparency have primarily focused on AI companies disclosing their models' limitations, capabilities, and safety efforts, including benchmarking and red-teaming. Lawmakers should expand these to include information about how people use AI applications. In particular, lawmakers should push AI companies to work with experts in high risk domains where their products may be used — such as finance, healthcare, and education — to understand what information they could provide to help people more safely use their products in these ways. They can do this by providing guidelines and best practices for involving external stakeholders in the design of transparency reports. Companies should publish this information alongside their methodologies for analysis, so that those relying on these disclosures can better understand how to use them.

References

Access Now. (n.d.). Transparency Reporting Index. *Access Now*. https://www.accessnow.org/campaign/transparency-reporting-index/ [perma.cc/4J25-B3A3]

AI Foundation Model Transparency Act of 2023, H.R.6881 (2023). https://www.congress.gov/bill/118th-congress/house-bill/6881/ [perma.cc/4BJF-WTTR]

AI Safety Institute. (n.d.). *The AI Safety Institute (AISI)*. https://www.aisi.gov.uk/ [perma.cc/GJ7Q-29SM]

America's Poison Centers. (n.d.). *Laundry Detergent Packets*. https://poisoncenters.org/track/laundry-detergent-packets [perma.cc/EE9V-CMGL]

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://journals.sagepub.com/doi/abs/10.1177/1461444816676645 [perma.cc/N9B4-6VRC]

Angwin, J., Nelson, A., & Palta, R. (2024). *Seeking Reliable Election Information? Don't Trust AI*. The AI Democracy Projects. https://www.proofnews.org/seeking-election-information-dont-trust-ai/ [perma.cc/43ST-Y8KZ]

Anthropic. (n.d.). *I have a zero retention agreement with Anthropic. What products does it apply to?* | Anthropic Help Center. https://support.anthropic.com/en/articles/8956058-i-have-a-zero-retention-agreement-with-anthropic-what-products-does-it-apply-to [perma.cc/JS7B-W9W3]

Anthropic. (2023). *Terms of Service* (p. 6(a)). https://www-cdn.anthropic.com/files/4zrzovbb/website/e2d538c84610b7cc8cb1c640767fa4ba73f30190.pdf [perma.cc/8FB6-D9KM]

Anthropic. (2024, June 5). *Privacy Policy*. https://www.anthropic.com/legal/privacy [perma.cc/6BFX-J5XZ]

Arnold, M., Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, K. N., Olteanu, A., Piorkowski, D., Reimer, D., Richards, J., Tsay, J., & Varshney, K. R. (2019). FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development, 63*(4/5), 6:1-6:13. https://doi.org/10.1147/JRD.2019.2942288 [perma.cc/5VHN-TV8Y]

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., & Tucker, J. A. (2024). Online searches to evaluate misinformation can increase its perceived veracity. *Nature, 625*(7995), 548–556. https://doi.org/10.1038/s41586-023-06883-y [perma.cc/QZP5-MQF4]

Benthall, S., & Sivan-Sevilla, I. (2024). *Regulatory CI: Adaptively Regulating Privacy as Contextual Integrity*. Federal Trade Commission. https://www.ftc.gov/system/files/ftc_gov/pdf/2%20-%20Benthall%20-%20Adaptively%20Regulating%20Privacy%20as%20Contextual%20Integrity.pdf [perma.cc/42J3-7TJF]

Birchall, C. (2021). *Radical secrecy: The ends of transparency in datafied America*. University of Minnesota Press.

Bommasani, R., Klyman, K., Kapoor, S., Longpre, S., Xiong, B., Maslej, N., & Liang, P. (2024). *The Foundation Model Transparency Index v1.1 May 2024*. Stanford University Center for Research on Foundation Models.

Bommasani, R., Klyman, K., Longpre, S., Xiong, B., Kapoor, S., Maslej, N., Narayanan, A., & Liang, P. (2024). *Foundation Model Transparency Reports*. https://doi.org/10.48550/arXiv.2402.16268 [perma.cc/T8ME-753H]

boyd, d (2016). *Transparency != Accountability: Hearing before the EU Parliament Event 07/11 Algorithmic Accountability and Transparency*, EU Parliament. http://www.danah.org/papers/talks/2016/EUParliament.html [perma.cc/C6VU-75XD]

C2PA. (n.d.). *Coalition for Content Provenance and Authenticity*. https://c2pa.org/ [perma.cc/73DA-C9LE]

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [perma.cc/4KPH-3V93]

Caliskan, A., & Lum, K. (2024). Effective AI regulation requires understanding general-purpose AI. *Brookings*. https://www.brookings.edu/articles/effective-ai-regulation-requires-understanding-general-purpose-ai/ [perma.cc/76PX-MYBV]

Casper, S., Ezell, C., Siegmann, C., Kolt, N., Curtis, T. L., Bucknall, B., Haupt, A., Wei, K., Scheurer, J., Hobbhahn, M., Sharkey, L., Krishna, S., Von Hagen, M., Alberti, S., Chan, A., Sun, Q., Gerovitch, M., Bau, D., Tegmark, M., ... Hadfield-Menell, D. (2024). Black-Box Access is Insufficient for Rigorous AI Audits. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2254–2272. https://doi.org/10.1145/3630106.3659037 [perma.cc/U2CM-6SQ6]

Chan. (2021, May 5). The Workers Who Sued Uber and Won. *Dissent Magazine*. https://www.dissentmagazine.org/online_articles/the-workers-who-sued-uber-and-won/ [perma.cc/HSJ2-SZM6]

Chu, K. (2024). *GitLab uses Anthropic for smart, safe AI-assisted code generation*. GitLab. https://about.gitlab.com/blog/2024/01/16/gitlab-uses-anthropic-for-smart-safe-ai-assisted-code-generation/ [perma.cc/T3PA-AAYW]

Clark, M. (2021, August 4). Research Cannot Be the Justification for Compromising People's Privacy. *Meta*. https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/ [perma.cc/468B-43C2]

Cohere. (n.d.). *Terms Of Use* (p. 9(a)). https://cohere.com/terms-of-use [perma.cc/PSQ3-4LTT]

Corbett, E., & Denton, E. (2023). Interrogating the T in FAccT. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1624–1634. https://doi.org/10.1145/3593013.3594104 [perma.cc/Q6JE-H5VN]

Cybersecurity Infrastructure & Security Agency. (2024). *Risk in Focus: Generative A.I. and the 2024 Election Cycle*. https://www.cisa.gov/resources-tools/resources/risk-focus-generative-ai-and-2024-election-cycle [perma.cc/8S2J-MGFE]

Digital Services Act, Article 40, COM/2020/825, European Commission, 52020PC0825 (2020). https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN [perma.cc/H2MN-L8KD]

Donovan, J. (2019). *Navigating the Tech Stack: When, Where and How Should We Moderate Content?* Centre for International Governance Innovation. https://www.cigionline.org/articles/navigating-tech-stack-when-where-and-how-should-we-moderate-content/ [perma.cc/7RQ2-2TCF]

Erwin, M. (2021, August 4). *Why Facebook's claims about the Ad Observer are wrong | The Mozilla Blog*. Mozilla Blog. https://blog.mozilla.org/en/mozilla/news/why-facebooks-claims-about-the-ad-observer-are-wrong/ [perma.cc/H485-5UYW]

Eyert, F., & Lopez, P. (2023). Rethinking Transparency as a Communicative Constellation. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 444–454. https://doi.org/10.1145/3593013.3594010 [perma.cc/PG3A-52NK]

Firdhous, M. F. M., Elbreiki, W., Abdullahi, I., Sudantha, B. H., & Rahmat, B. (2023, December). WormGPT: A Large Language Model Chatbot for Criminals. *2023 24th International Arab Conference on Information Technology (ACIT)*. https://ieeexplore.ieee.org/abstract/document/10453752 [perma.cc/V7KW-KPAB]

Fournier, E. D., Cudd, R., Federico, F., & Pincetl, S. (2020). On energy sufficiency and the need for new policies to combat growing inequities in the residential energy sector. *Elementa: Science of the Anthropocene, 8*, 24. https://doi.org/10.1525/elementa.419 [perma.cc/NU89-X6VG]

Friedler, S., Singh, R., Blili-Hamelin, B., Metcalf, J., & Chen, B. J. (2023). *AI Red-Teaming Is Not a One-Stop Solution to AI Harms: Recommendations for Using Red-Teaming for AI Accountability*, Data & Society. https://datasociety.net/wp-content/uploads/2023/10/Recommendations-for-Using-Red-Teaming-for-AI-Accountability-PolicyBrief.pdf [perma.cc/P2X6-RWNS]

Ganguli, D., Lovitt, L., Kernion, J., Askell, A., Bai, Y., Kadavath, S., Mann, B., Perez, E., Schiefer, N., Ndousse, K., Jones, A., Bowman, S., Chen, A., Conerly, T., DasSarma, N., Drain, D., Elhage, N., El-Showk, S., Fort, S., ... Clark, J. (2022). *Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned*. https://doi.org/10.48550/arXiv.2209.07858 [perma.cc/F69Z-MEBV]

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723 [perma.cc/AJE7-NK98]

Gilbert, D. (2024, June 7). Google's and Microsoft's AI Chatbots Refuse to Say Who Won the 2020 US Election. *Wired*. https://www.wired.com/story/google-and-microsofts-chatbots-refuse-election-questions/ [perma.cc/3748-BQQW]

Google. (2024, May 29). *Gemini Apps Privacy Hub*. Gemini Apps Help. https://support.google.com/gemini/answer/13594961?hl=en [perma.cc/SUE8-8D5S]

Guha, N., Lawrence, C., Gailmard, L. A., Rodolfa, K., Surani, F., Bommasani, R., Raji, I., Cuéllar, M.-F., Honigsberg, C., Liang, P., & Ho, D. E. (2023). *AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing*. https://papers.ssrn.com/abstract=4634443 [perma.cc/S8BQ-CEKV]

Harris, D. E., Norden, L., Praetz, N., & Howard, E. (2024). *How Election Officials Can Identify, Prepare for, and Respond to AI Threats.* Brennan Center for Justice. https://www.brennancenter.org/our-work/research-reports/how-election-officials-can-identify-prepare-and-respond-ai-threats [perma.cc/R9AF-9UH2]

Herder, M., Morten, C. J., & Doshi, P. (2020). Integrated Drug Reviews at the US Food and Drug Administration—Legal Concerns and Knowledge Lost. *JAMA Internal Medicine, 180*(5), 629–630. https://doi.org/10.1001/jamainternmed.2020.0074 [perma.cc/HCW5-G59K]

Hupont, I., & Gomez, E. (2022). *Documenting use cases in the affective computing domain using Unified Modeling Language*. https://doi.org/10.48550/arXiv.2209.09666 [perma.cc/P9CY-6EXL]

Hutchinson, B., Rostamzadeh, N., Greer, C., Heller, K., & Prabhakaran, V. (2022). Evaluation Gaps in Machine Learning Practice. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1859–1876. https://doi.org/10.1145/3531146.3533233 [perma.cc/9JKJ-UKG4]

IPTC. (n.d.). *International Press Telecommunications Conference*. https://iptc.org/ [perma.cc/5TQ7-TC25]

Jones, E. (2023). *What is a foundation model?* Ada Lovelace Institute. https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/ [perma.cc/HGL5-WZ3A]

Kaur, H., Adar, E., Gilbert, E., & Lampe, C. (2022). Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 702–714. https://doi.org/10.1145/3531146.3533135 [perma.cc/7M37-MZVH]

Kayser-Bril, N. (2021, August 13). AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook. *AlgorithmWatch*. https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/ [perma.cc/4PV3-P58S]

Knight, W. (n.d.). Google's Gemini Is the Real Start of the Generative AI Boom. *Wired*. https://www.wired.com/story/google-gemini-generative-ai-boom/ [perma.cc/Y4KP-FEXR]

Laird, E., & Dwyer, M. (2023). *Off Task: EdTech Threats to Student Privacy and Equity in the Age of AI*. Center for Democracy & Technology. https://cdt.org/insights/report-off-task-edtech-threats-to-student-privacy-and-equity-in-the-age-of-ai/ [perma.cc/HL7H-S92H]

Lapowsky, I. (2024). *Bridging the divide: Translating research on digital media into policy and practice*. Knight Foundation. https://knightfoundation.org/features/bridging-the-divide-translating-research-on-digital-media-into-policy-and-practice/ [perma.cc/63WV-HQYT]

LaRoque, S. J. (2017). Reverse Engineering and Trade Secrets in the Post-Alice World. *University of Kansas Law Review, 66*, 427. https://kuscholarworks.ku.edu/bitstream/handle/1808/25704/10_LaRoque_Final.pdf [perma.cc/UG9F-3PAD]

Laufer, B., Jain, S., Cooper, A. F., Kleinberg, J., & Heidari, H. (2022). Four Years of FAccT: A Reflexive, Mixed-Methods Analysis of Research Contributions, Shortcomings, and Future Prospects. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 401–426. https://doi.org/10.1145/3531146.3533107 [perma.cc/WQE9-X4JW]

LayerX. (2023). *Revealing the True GenAI Data Exposure Risk*. https://go.layerxsecurity.com/hubfs/Research-Revealing-the-True-GenAI-Data-Exposure-Risk.pdf [perma.cc/2V57-R8J7]

Lemoine, L., & Vermeulen, M. (2023). *Assessing the Extent to Which Generative Artificial Intelligence (AI) Falls Within the Scope of the EU's Digital Services Act: An Initial Analysis* (4702422). https://doi.org/10.2139/ssrn.4702422 [perma.cc/77K6-3XD6]

Lenhart, A. (2023). *Federal AI Legislation: An Analysis of Proposals from the 117th Congress Relevant to Generative AI tools.* Institute for Data, Democracy & Politics, George Washington University. https://iddp.gwu.edu/federal-ai-legislation [perma.cc/B6HS-498B]

Leonard, A. (2023, September 16). "Dr. Google" meets its match in Dr. ChatGPT. *NPR*. https://www.npr.org/sections/health-shots/2023/09/16/1199924303/chatgpt-ai-medical-advice [perma.cc/K7EW-TY2Q]

Levine, S. (2021, August 5). *Letter from Acting Director of the Bureau of Consumer Protection Samuel Levine to Facebook*. Federal Trade Commission. https://www.ftc.gov/blog-posts/2021/08/letter-acting-director-bureau-consumer-protection-samuel-levine-facebook [perma.cc/KS6Y-D9H6]

Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31–57. https://doi.org/10.1145/3236386.3241340 [perma.cc/8CGR-U7JV]

Lucas, J. (2024, May 4). *The teens making friends with AI chatbots*. The Verge. https://www.theverge.com/2024/5/4/24144763/ai-chatbot-friends-character-teens [perma.cc/4MEW-K5T2]

Luria, M. (2023). Co-Design Perspectives on Algorithm Transparency Reporting: Guidelines and Prototypes. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1076–1087. https://doi.org/10.1145/3593013.3594064 [perma.cc/KK6P-RWV5]

Micheli, M., Hupont, I., Delipetrev, B., & Soler-Garrido, J. (2023). The landscape of data and AI documentation approaches in the European policy context. *Ethics and Information Technology, 25*(4), 56. https://doi.org/10.1007/s10676-023-09725-7 [perma.cc/3SBH-W3ME]

Microsoft Copilot. (2024, July 9). *Copilot Privacy and Protections*. https://learn.microsoft.com/en-us/copilot/privacy-and-protections [perma.cc/JN88-XQB4]

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596 [perma.cc/KJ7A-HYKG]

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 205395171667967. https://doi.org/10.1177/2053951716679679 [perma.cc/2VBD-9RHW]

Mohammad, S. (2022). Ethics Sheets for AI Tasks. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8368–8379. https://doi.org/10.18653/v1/2022.acl-long.573 [perma.cc/7BKB-3K8T]

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: A three-layered approach. *AI and Ethics*. https://link.springer.com/article/10.1007/s43681-023-00289-2 [perma.cc/D6GM-7J9U]

Morten, C., Nicholas, G., & Viljoen, S. (2024). Researcher Access to Social Media Data: Lessons from Clinical Trial Data Sharing. *Berkeley Technology Law Journal, 38*(109). https://doi.org/10.2139/ssrn.4716353 [perma.cc/LMJ9-NASQ]

Mozilla. (n.d.). *Documentation of Privacy Review for AdObserver*. Bugzilla. https://bugzilla.mozilla.org/show_bug.cgi?id=1676407 [perma.cc/BJQ9-JLN5]

Narayanan, A., & Kapoor, S. (2023, June 26). Generative AI companies must publish transparency reports. *Knight First Amendment Institute*. http://knightcolumbia.org/blog/generative-ai-companies-must-publish-transparency-reports [perma.cc/ZY4G-8NYN]

National AI Advisory Committee. (2023). *Rationales, Mechanisms, and Challenges to Regulating AI: A Concise Guide and Explanation*. https://ai.gov/wp-content/uploads/2023/07/Rationales-Mechanisms-Challenges-Regulating-AI-NAIAC-Non-Decisional.pdf [perma.cc/JKQ4-FGA6]

National Cancer Institute. (n.d.). *Donate Your Data and Specimens for Cancer Research. National Institutes of Health*. https://www.cancer.gov/research/participate/articles/donate-medical-data-and-samples [perma.cc/S95R-J3NT]

National Science Foundation. (2024). *National Artificial Intelligence Research Resource Pilot*. https://new.nsf.gov/focus-areas/artificial-intelligence/nairr [perma.cc/M2R9-VUL8]

Nicholas, G. (2020). *Explaining Algorithmic Decisions. Georgetown Law Technology Review, 4*(711), 20. https://georgetownlawtechvieworg.wpcomstaging.com/wp-content/uploads/2020/07/4.2-p711-730-Nicholas.pdf [perma.cc/NSV5-9D8W]

Nicholas, G., & Thakur, D. (2022). *Learning to Share: Lessons on Data-Sharing from Beyond Social Media* (pp. 1–44). Center for Democracy & Technology. https://cdt.org/insights/learning-to-share-lessons-on-data-sharing-from-beyond-social-media/ [perma.cc/JBR3-S7VC]

Nicholas, G., & Weinberg, M. (2019). *Data Portability and Platform Competition*. Engelberg Center on Innovation Law & Policy. https://nyuengelberg.org/outputs/data-portability-and-platform-competition/ [perma.cc/KEZ8-8D5L]

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press. https://www.sup.org/books/title/?id=8862 [perma.cc/HHM2-4AD3]

NIST. (2024). AI Risk Management Framework. *National Institute of Standards and Technology*. https://www.nist.gov/itl/ai-risk-management-framework [perma.cc/AC9S-X5C3]

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2024). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures, 18*(2), 124–141. https://www.tandfonline.com/doi/full/10.1080/19312458.2023.2181319 [perma.cc/73TW-S8J8]

Okerlund, J., Klasky, E., Middha, A., Kim, S., Rosenfeld, H., Kleinman, M., & Parthasarathy, S. (2022). *What's in the Chatterbox?* University of Michigan. https://stpp.fordschool.umich.edu/research-projects/whats-in-the-chatterbox [perma.cc/4EDC-52H7]

OpenAI. (n.d.). *Enterprise privacy*. https://openai.com/enterprise-privacy/ [perma.cc/3XUJ-SNP6]

OpenAI. (2023a, November 14). *Business terms*. https://openai.com/policies/business-terms/ [perma.cc/9PTS-PB4K]

OpenAI. (2023b, November 14). *Privacy policy*. https://openai.com/policies/privacy-policy/ [perma.cc/74LM-XBLM]

OpenAI. (2024, April 1). *Oscar: Reducing health insurance costs and improving care*. https://openai.com/index/oscar/ [perma.cc/VWQ9-K3RP]

Ouyang, S., Wang, S., Liu, Y., Zhong, M., Jiao, Y., Iter, D., Pryzant, R., Zhu, C., Ji, H., & Han, J. (2023). The Shifted and The Overlooked: A Task-oriented Investigation of User-GPT Interactions. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2375–2393. https://doi.org/10.18653/v1/2023.emnlp-main.146 [perma.cc/87K9-R3UQ]

Parsons, C. (2017). The (In)effectiveness of Voluntarily Produced Transparency Reports. *Business & Society, 58*. https://journals.sagepub.com/doi/full/10.1177/0007650317717957 [perma.cc/GZE3-BSSL]

Partnership on AI. (2023). *PAI's Guidance for Safe Foundation Model Deployment*. Partnership on AI. https://partnershiponai.org/modeldeployment/ [perma.cc/C37D-CPNT]

Platform Accountability and Transparency Act, S.5339 (2022). https://www.congress.gov/bill/117th-congress/senate-bill/5339 [perma.cc/R32J-2QGB]

Pozen, D. E. (2018). Transparency's Ideological Drift. *Yale Law Journal, 128*(1). https://www.yalelawjournal.org/article/transparencys-ideological-drift [perma.cc/5NR9-XAD4]

Preset, M. J. (2023). *10 Foundations Pool $200 Million for Efforts to Govern Artificial Intelligence*. The Chronicle of Philanthropy. https://www.philanthropy.com/article/10-foundations-pool-200-million-for-efforts-to-govern-artificial-intelligence [perma.cc/R3YC-FQFL]

Procope, C., Cheema, A., Adkins, D., Alsallakh, B., Green, N., McReynolds, E., Pehl, G., Wang, E., & Zvyagina, P. (2022, February 22). *System-Level Transparency of Machine Learning | Research—AI at Meta*. https://ai.meta.com/research/publications/system-level-transparency-of-machine-learning/ [perma.cc/9PCR-FZVN]

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. https://doi.org/10.1145/3351095.3372873 [perma.cc/R983-ERZH]

Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (Artificial Intelligence Act), COM/2021/206 (2024). https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=celex%3A52021PC0206 [perma.cc/T6JQ-EJDR]

Selbst, A. (2021). An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology, 35*(1). https://jolt.law.harvard.edu/assets/articlePDFs/v35/Selbst-An-Institutional-View-of-Algorithmic-Impact-Assessments.pdf [perma.cc/93F3-7GQR]

Sellars, A. (2018). *Twenty Years of Web Scraping and the Computer Fraud and Abuse Act. 24*(2), 372. https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1466&context=faculty_scholarship [perma.cc/7FD8-YMZN]

Shah, C., White, R. W., Andersen, R., Buscher, G., Counts, S., Das, S., Montazer, A., Manivannan, S., Neville, J., Ni, X., Rangan, N., Safavi, T., Suri, S., Wan, M., & Yang, L. (2023). *Using Large Language Models to Generate, Validate, and Apply User Intent Taxonomies*. Microsoft. https://www.microsoft.com/en-us/research/publication/using-large-language-models-to-generate-validate-and-apply-user-intent-taxonomies/ [perma.cc/P76G-7LND]

Shapiro, E., Sugarman, M., Bermejo, F., & Zuckerman, E. (2021, February). *New Approaches to Platform Data Research*. NetGain Partnership. https://www.netgainpartnership.org/resources/2021/2/25/new-approaches-to-platform-data-research [perma.cc/CPT6-77M3]

Shelby, R., Rismani, S., Henne, K., Moon, Aj., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, A., Garcia, E., & Virk, G. (2023). *Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction*. https://doi.org/10.48550/arXiv.2210.05791 [perma.cc/78PJ-VTUD]

Shevlane, T., Farquhar, S., Garfinkel, B., Phuong, M., Whittlestone, J., Leung, J., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., Ho, L., Siddarth, D., Avin, S., Hawkins, W., Kim, B., Gabriel, I., Bolina, V., Clark, J., Bengio, Y., ... Dafoe, A. (2023). *Model evaluation for extreme risks*. https://doi.org/10.48550/arXiv.2305.15324 [perma.cc/TYS2-2PDE]

Shore, A., & Prena, K. (2023). Platform rules as privacy tools: The influence of screenshot accountability and trust on privacy management. *New Media & Society*, 14614448231188929. https://doi.org/10.1177/14614448231188929 [perma.cc/8Y77-PMSA]

Singer, N. (2023, December 13). Cheating Fears Over Chatbots Were Overblown, New Research Suggests. *The New York Times*. https://www.nytimes.com/2023/12/13/technology/chatbot-cheating-schools-students.html [perma.cc/5J5M-E9W2]

Singh, J. (2024, March 12). Google won't let you use its Gemini AI to answer questions about an upcoming election in your country. *TechCrunch*. https://techcrunch.com/2024/03/12/google-gemini-election-related-queries/ [perma.cc/4P8Q-G8ST]

Smith, B., & Crampton, N. (2024, May 1). Providing further transparency on our responsible AI efforts. *Microsoft On the Issues*. https://blogs.microsoft.com/on-the-issues/2024/05/01/responsible-ai-transparency-report-2024/ [perma.cc/5V4R-W8VA]

Solaiman, I. (2023). The Gradient of Generative AI Release: Methods and Considerations. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 111–122. https://doi.org/10.1145/3593013.3593981 [perma.cc/ER8Z-93QX]

Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J. W., Kreps, S., McCain, M., Newhouse, A., Blazakis, J., McGuffie, K., & Wang, J. (2019). *Release Strategies and the Social Impacts of Language Models*. https://doi.org/10.48550/arXiv.1908.09203 [perma.cc/JV76-EP6G]

Stohl, C., Stohl, M., & Leonardi, P. M. (2016). Digital Age | Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. *International Journal of Communication, 10*(0), Article 0. https://ijoc.org/index.php/ijoc/article/view/4466 [perma.cc/43KN-UHUX]

Suri, S., Counts, S., Wang, L., Chen, C., Wan, M., Safavi, T., Neville, J., Shah, C., White, R. W., Andersen, R., Buscher, G., Manivannan, S., Rangan, N., & Yang, L. (2024). *The Use of Generative Search Engines for Knowledge Work and Complex Tasks*. Microsoft. https://www.microsoft.com/en-us/research/publication/the-use-of-generative-search-engines-for-knowledge-work-and-complex-tasks/ [perma.cc/6ABX-N2FQ]

The Vicuna Team. (2023, March 30). *Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality | LMSYS Org*. LMSYS ORG. https://lmsys.org/blog/2023-03-30-vicuna [perma.cc/PK8U-EQDY]

The White House. (2023, July 21). *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*. The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [perma.cc/AGV8-NQYK]

Thiel, D. (2023, December 23). *Identifying and Eliminating CSAM in Generative ML Training Data and Models*. Stanford Internet Observatory. https://purl.stanford.edu/kh752sm9123 [perma.cc/AM2Q-8ZLE]

Tromble, R. (2021). Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age. *Social Media + Society, 7*(1). https://journals.sagepub.com/doi/10.1177/2056305121988929 [perma.cc/69M3-HBTJ]

Trust & Safety Professional Association. (2021, July 7). *Transparency Reporting*. Trust & Safety Professional Association. https://www.tspa.org/curriculum/ts-fundamentals/transparency-report/ [perma.cc/B89P-JQZZ]

Tucker, Z. S., Joshua. (2023, November 1). *Beyond Red Teaming: Facilitating User-based Data Donation to Study Generative AI | TechPolicy.Press*. Tech Policy Press. https://techpolicy.press/beyond-red-teaming-facilitating-user-based-data-donation-to-study-generative-ai [perma.cc/T66D-BRGY]

Twitter. (n.d.). *Module 2: Applying for a Twitter developer account and choosing the right product track*. GitHub. https://github.com/xdevplatform/getting-started-with-the-twitter-api-v2-for-academic-research/blob/main/modules/2-choosing-the-right-product-track.md [perma.cc/4RPD-YNZA]

van Driel, I. I., Giachanou, A., Pouwels, J. L., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2022). Promises and Pitfalls of Social Media Data Donations. *Communication Methods and Measures, 16*(4), 266–282. https://www.tandfonline.com/doi/full/10.1080/19312458.2022.2109608 [perma.cc/SPD5-H7LX]

Verma, P., & Oremus, W. (2023, June 26). Meta's new AI is being used to create sex chatbots. *The Washington Post*. https://www.washingtonpost.com/technology/2023/06/26/facebook-chatbot-sex/ [perma.cc/5TVG-N2ZD]

Vogus, C. (2022). *Improving Researcher Access to Digital Data*. Center for Democracy & Technology. https://cdt.org/wp-content/uploads/2022/08/2022-08-15-FX-RAtD-workshop-report-final-int.pdf [perma.cc/Z584-WYQ7]

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., ... Gabriel, I. (2021). *Ethical and social risks of harm from Language Models*. https://doi.org/10.48550/arXiv.2112.04359 [perma.cc/MP2S-LK4Y]

Widder, D. G., West, S., & Whittaker, M. (2023). *Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI*. https://doi.org/10.2139/ssrn.4543807 [perma.cc/DF2Q-23BR]

xAI. (2024, March 15). *Privacy policy*. https://x.ai/privacy-policy [perma.cc/6AA8-3T36]

Zalnieriute, M. (2021). "Transparency Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism. *Critical Analysis of Law, 8*(1), Article 1. https://doi.org/10.33137/cal.v8i1.36284 [perma.cc/B5QP-GUUG]

Zhao, W., Ren, X., Hessel, J., Cardie, C., Choi, Y., & Deng, Y. (2024). *WildChat: 1M ChatGPT Interaction Logs in the Wild*. https://doi.org/10.48550/arXiv.2405.01470 [perma.cc/QZA4-32EC]

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

CENTER FOR
DEMOCRACY
& TECHNOLOGY