Ecosystem Graphs: Documenting the Foundation Model Supply Chain

Rishi Bommasani[1], Dilara Soylu[1], Thomas I. Liao[2], Kathleen A. Creel[3], Percy Liang[1]

[1]Stanford University
[2]Independent Researcher
[3]Northeastern University
nlprishi@stanford.edu

Abstract

Foundation models (e.g. GPT-4, Gemini, Llama 3) pervasively influence society, warranting greater understanding. While the models garner much attention, accurately characterizing their impact requires considering the broader sociotechnical ecosystem in which they are created and deployed. We propose **Ecosystem Graphs** as a documentation framework to centralize knowledge of this ecosystem. **Ecosystem Graphs** is composed of *assets* (datasets, models, applications) linked together by *dependencies* that indicate technical and social relationships. To supplement the graph structure, each asset is further enriched with fine-grained metadata, such as the model's estimated training emissions or licensing guidelines. Since its release in March 2023, **Ecosystem Graphs** represents an ongoing effort to document 568 assets (112 datasets, 359 models, 97 applications) from 117 organizations. **Ecosystem Graphs** functions as a multifunctional resource: we discuss two major uses by the 2024 AI Index and the UK's Competition and Markets Authority that demonstrate the value of **Ecosystem Graphs**.

Introduction

Foundation models (FMs) are the defining paradigm of modern AI (Bommasani et al. 2021). Beginning with language models (Devlin et al. 2019; Brown et al. 2020; Chowdhery et al. 2022), the paradigm has expanded to images (Chen et al. 2020; Ramesh et al. 2021; Radford et al. 2021), videos (Singer et al. 2022; Wang et al. 2022b), code (Chen et al. 2021), proteins (Jumper et al. 2021; Verkuil et al. 2022), and more. Given their remarkable capacities, they have been adopted at unprecedented speed: ChatGPT amassed 100 million users in just 50 days, making it the fastest-growing consumer application in history (Hu 2023), and Gemini is deployed across all of Google's 2-billion user products.[1] FMs are transforming the digital economy: Bloomberg predicts the market will expand to $1.3 trillion by 2032 and Accenture projects 75% of knowledge workers will use FMs on a daily basis by 2030.

Despite their rapid adoption, the scope of FMs' current impact on society is unclear. Who reaps their benefits, who shoulders the harms they may cause, and how can we characterize the societal changes they engender? Further, how do trends in research correspond to outcomes in practice? How do emergent abilities (Wei et al. 2022) influence deployment decisions, and how do documented risks (Abid, Farooqi, and Zou 2021) manifest as concrete harms? Currently, the AI community and broader public tolerate the uncomfortable reality that models are deployed ubiquitously through products, yet we know little about the models, how they were built, and the mechanisms (if any) in place to mitigate and address harm (Bommasani et al. 2023a).

To clarify the societal impact of FMs, we propose **Ecosystem Graphs** as a centralized knowledge graph for documenting the foundation model supply chain (Figure 1). **Ecosystem Graphs** consolidates distributed knowledge to improve the ecosystem's transparency. **Ecosystem Graphs** operationalizes the insight that significant understanding of the societal impact of FMs is already possible if we centralize available information to analyze it collectively.

Each node in the graph is (roughly) an *asset*: a dataset, model, or application. Simply being aware of assets is a challenge: new datasets are being built, new models trained, and new products shipped constantly, often with uneven public disclosure. While models often draw attention, the consequences of a FM depend on the broader supply chain in which it is built and into which it is deployed. To link nodes, we specify *dependencies*: models depend on training data and applications depend on models. Dependencies are technical relationships between assets that induce social or economic relationships between organizations (e.g. Microsoft depends on OpenAI because Bing depends on GPT-4). Especially for products, surfacing these dependencies is challenging yet critical: dependencies indicate the flow of resources to build products, which in turn determine much of the direct impact of FMs.

To supplement the graph structure, we document each node with an *ecosystem card*, drawing inspiration from documentation frameworks such as data sheets (Gebru et al. 2018), data statements (Bender and Friedman 2018), and model cards (Mitchell et al. 2018). The ecosystem card contextualizes the node not only in isolation (e.g. when was it built), but also with respect to the graph structure (e.g. the license affects downstream use, data filters interact with upstream dependencies). In contrast to static documentation

[1]https://blog.google/inside-google/message-ceo/google-io-2024-keynote-sundar-pichai/

by dataset or model developers, ecosystem-level documentation introduces key challenges of (i) *maintenance* practices to synchronize the ecosystem graph with the ever-evolving supply chain, and (ii) *incentives* that inhibit or encourage documentation.

Since its release in March 2023, **Ecosystem Graphs** documents 568 assets across 117 organization through an interactive website and database that is updated weekly. To maintain this living resource, we rely primarily on our own contributions, but also on open-source contributions that are verified by the authors as well as contributions from asset developers. **Ecosystem Graphs** enables the study of new phenomenona such as the emergence of *hubs* like ChatGPT in the graph, drawing inspiration from the widespread analysis of hubs across other graphs and networks (Kleinberg 1999; Hendricks, Piccione, and Tan 1995; Franks et al. 2008; Van den Heuvel and Sporns 2013, *inter alia*). For asset developers, hubs demonstrate their assets are high impact; for economists, hubs communicate emergent market structure and potential consolidation of power; for policymakers, hubs identify targets to scrutinize to ensure their security and safety.

More generally, **Ecosystem Graphs** is a multifunctional resource for a range of stakeholders. We present two case studies where **Ecosystem Graphs** has seen extensive use. The 2024 AI Index (Maslej et al. 2024) uses **Ecosystem Graphs** as its data source for documenting broad trends such as geopolitical considerations (e.g. which countries are building the most foundation models?). And the UK's Competition and Markets Authority uses **Ecosystem Graphs** as its primary means for conducting their market surveillance obligations such as understanding the number of open vs. closed models in the market (UK CMA 2023). As foundation models are enmeshed as essential societal infrastructure with wide-reaching impact across economic sectors, supply chain monitoring will be increasingly vital. **Ecosystem Graphs** prototypes this supply chain monitoring, which world governments may scale up in the future.[2]

The Foundation Model Supply Chain

Technology is developed in a social context, shaped by organization-internal processes and external forces (Martin Jr et al. 2020; Gebru et al. 2018; Mitchell et al. 2018; Amironesei, Denton, and Hanna 2021; Paullada et al. 2021; Birhane et al. 2022). Bommasani et al. (2021) introduce this perspective for FMs: Figure 1 demonstrates the canonical pipeline wherein datasets, models, and applications mediate relationships between people on either side. As an example, consider Stable Diffusion (Rombach et al. 2021):

1. People create content, taking photos that they (or others) upload to the web.
2. LAION curates the LAION-5B dataset (Schuhmann et al. 2022) from the CommonCrawl web scrape, filtered for problematic content.
3. LMU Munich, IWR Heidelberg University, and RunwayML train Stable Diffusion on a filtered version of LAION-5B using 256 A100 GPUs.
4. Stability AI builds Stable Diffusion Reimagine by replacing the Stable Diffusion text encoder with an image encoder.
5. Stability AI deploys Stable Diffusion Reimagine as an image editing tool to end users on Clipdrop, allowing users to generate variations on an image (e.g. a room with different furniture).

This process delineates social roles and, thereby, *stakeholders*: data creators, data curators, compute providers, hardware providers, FM developers, downstream application developers, and end users. While framed technically (e.g. curation, training, adaptation), the process has broader societal ramifications. For example, ongoing litigation contends the use of LAION and subsequent image generation from Stable Diffusion infringes on the rights of data creators.

[2] https://crfm.stanford.edu/ecosystem-graphs

Documentation Framework

To document the foundation model supply chain, we introduce the **Ecosystem Graphs** framework. Informally, the framework is defined by a graph of (i) *assets* (e.g. GPT-4), (ii) *dependencies* (e.g. datasets used to build GPT-4, applications built upon GPT-4), and (iii) *ecosystem cards* (e.g. metadata about GPT-4).

Definition

Formally, we define the ecosystem graph as a directed vertex-labeled graph. Each vertex is an asset that is labeled as either a *dataset*, *model*, or *application*. Examples include The Pile dataset, the Stable Diffusion model, and the Bing Search application. Vertices are connected by directed edges that specify *dependencies*. In the aforementioned example, LAION-5B is a dependency for Stable Diffusion and Stable Diffusion is a dependency for Stable Diffusion Reimagine.

Each vertex has an associated *ecosystem card* that documents relevant metadata. Each asset type (dataset, model, application) has a schema for its metadata: the full schema is in the Appendix. Examples of properties within the schema include the "organization" that created the asset, the "license" enforced to use the asset, and type-specific properties (e.g. the "size" of a model).

The definition of **Ecosystem Graphs** is deliberately minimal: as a resource aimed at many, often non-technical, stakeholders, we aim for a simple and broadly intelligble representation of the supply chain. Under the hood, we introduce two further forms of complexity that we revisit in §3.2. First, while we refer to vertices as assets, in practice they correspond to sets of related assets (e.g. the models of different sizes in the GPT-3 model family). Second, while we refer to the properties of assets, this information (e.g. a model's license) is annotated in a specific value that specifies a constrained *value* (i.e. a license type like MIT or Apache 2.0) and a more open-ended *description* (e.g. provenance for the information).

Implementation

We implement **Ecosystem Graphs** in a lightweight fashion that is simple to maintain and simple to understand, period-

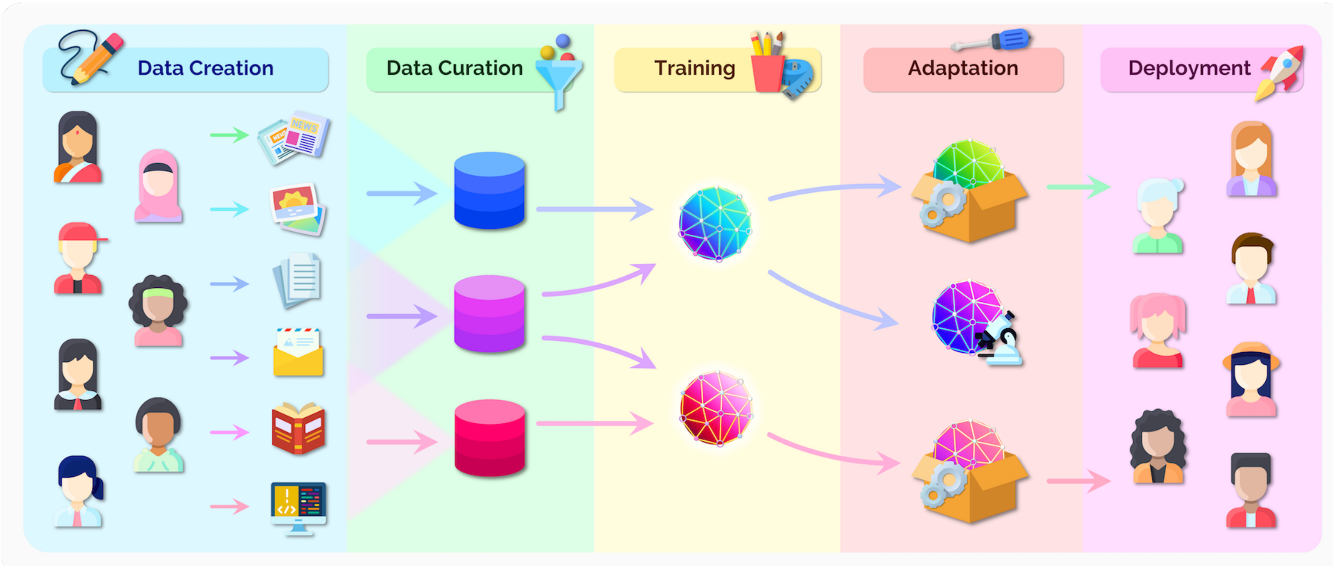

Figure 1: Basic foundation model ecosystem. To conceptualize the FM ecosystem, we present a simplified pipeline. Image taken from Bommasani et al. (2021).

ically adding features due to direct requests by users of the resource.

Codebase. On the back end, **Ecosystem Graphs** is a collection of `YAML` files that store the annotation metadata, organized by the pre-specified schemas. To create a new asset, an annotator creates a YAML entry with the schema fields empty by default and fills in information about the asset. As part of this process, the annotators specifies any known dependencies for the asset, which in turn determine the graph structure. To modify the public-facing website the annotator submits a pull request to our GitHub repository that a verified maintainer of the **Ecosystem Graphs** effort reviews and vets before making the changes live.

Views. The assets in **Ecosystem Graphs** are rendered in an interactive table that users can search, filter, or export as a CSV. These features facilitate several of the uses we describe subsequently. In addition, the dependency graph is rendered with a simple interface. Users can zoom into particular regions to explore specific subgraphs and clicking on a vertex pulls up its ecosystem card. For the purposes of double-blind review, we do not link to the website but, instead, include static screenshots of interesting subgraphs and ecosystem cards in the Appendix.

Nodes

To define the nodes in the graph, we describe how we discover assets and how we group assets into nodes.

Asset discovery Information about assets is highly decentralized. For some assets, primarily models and datasets, the asset is the subject of a research paper or technical report. To capture these assets, we manually track arXiv papers on a daily basis, identifying any papers that introduce an asset. However, even for assets described in papers, we identify the presence of *dark matter*: assets that must exist, but about which no information is publicly available. For instance, the GPT-4 paper (OpenAI 2023) discloses no details whatsoever on the dataset(s) underlying GPT-4. To supplement assets that are research contributions, we proactively the blogs, social media accounts, and media coverage of known foundation model developers. Through these sources, developers may announce new assets, which usually are either foundation models or products/services built upon the developer's foundation models.

Practically, we make decisions to not include every asset we could potentially discover. As a lower bound, there are 150k+ models and 20k+ datasets available on Hugging Face alone. Consequently, we do not have the resources to annotate every one of these assets, so we instead enumerate general principles that guide our decisions on what assets to include or exclude. Namely, we include assets that (i) are socially salient, (ii) have outsized impact, and/or (iii) represent a broader class (e.g. ensuring we include at least one music foundation model). Of course, the first condition is especially subjective, so we are generally receptive to adding an asset if a member of the community or other stakeholder requests its inclusion.[3]

Asset representation. For conceptual simplicity, nodes in **Ecosystem Graphs** combine very closely-related assets. In particular, we merge distinct assets into a single representative node when they are near-indistinguishable based on public information. In practice, this occurs often in the case of datasets, where different variants are used (e.g. different developers filter The Pile in different ways) or a more generic source is being described (e.g. many models are

[3]For example, representatives of the South Korean government have reached out to us to help improve our coverage of South Korean models not covered by Western English-speaking media sources.

trained on English Wikipedia, even though the exact set of articles is likely different). And we merge assets within a model family into a single node for the family (e.g. there is only one Llama 3 node, in spite of there being Llama 3 models of different sizes). There are trade-offs: by collapsing to one node, we obscure slight differences (e.g. Llama 3-400B was released after Llama 3-8B), though if we observe a distinction is relevant, we can choose to dis-aggregate the node into its constituent assets. Overall, we prioritize usefulness over faithfulness.

Edges

Edges constitute dependencies, meaning the assets required in building a given asset. Implementation-wise, dependencies are annotated as one field within the ecosystem card, but they are of heightened importance because they determine the graph connectivity and structure. In general, we identify dependencies in whatever material (e.g. research paper, release blogpost, media article) announces the existence of an asset. However, in many cases we find that assets are announced with no or partial disclosure of the assets they depend upon. This is the dark matter of the foundation model supply chain. In the cases of applications, we especially see this pattern: many products or services are marketed as using Generative AI or the models of specific companies (e.g. OpenAI) without specifying the specific foundation model.

Ecosystem Cards

Each node is instrumented with documentation to contextualize the node. To identify the relevant documentation fields for each asset type, we consulted existing documentation approaches such as data sheets (Gebru et al. 2018) and model cards (Mitchell et al. 2018). Based on these works, we identified an initial set of fields that we updated during a pilot phase of annotating 50 assets. Overall, we employ two principles in determining the final schema of documentation fields for each asset type: (i) we emphasize properties that are critical for understanding the supply chain (e.g. how nodes are influenced by their dependencies or shape their dependents) and (ii) we offload documentation that exists elsewhere (e.g. pointing to existing model cards). The resulting documentation metadata for each node is its *ecosystem card*.

We organize the documentation fields into three categories:

1. *Basic* properties of the node (e.g. the developer organization).
2. *Construction* properties of the node (e.g. the training emissions for models).
3. *Downstream* properties of the node (e.g. the license and terms-of-service for applications).

When annotating a given property, the annotator specifies the *value* (i.e. the concrete and succinct information) along with an optional *description*, which explains how the value should be interpreted (e.g. explains how the emissions value was computed or provides sources).

Basic properties To understand nodes and assets even independent of the broader ecosystem, certain basic information is necessary such as the "name" and producing "organization" of the asset(s). Documenting these properties is typically straightforward, but complexities can arise even concerning basic properties. For example, the naming convention for OpenAI models has been historically unclear.[4] Or the granularity in the case of organizations may be obscure: as an example, we annotate Copilot as developed by GitHub even though GitHub has been acquired by Microsoft. In the case of the "description" field, we generally quote the asset(s) developers, along with providing the "URL" that disclosed the node to the public (e.g. the paper or press release). In addition to these properties, we highlight the "created date": as **Ecosystem Graphs** is maintained over time, filtering on the date can be used to understand how the ecosystem evolves. For example, filtering for node(s) before vs. after January 1, 2023 reveals both (i) the early adopters of FMs and (ii) how publicly-announced release have rapidly accelerated since 2023.

Construction properties A node's dependencies do not fully determine its nature. Many products can be built from the same model and different models can arise from training on the same dataset. However, summarizing these nuances is challenging: since the training procedure for prominent FMs can amount to dozens of pages in a well-written paper (see Chowdhery et al. 2022), even a summary of model training procedures would be long and likely not more useful than pointing to the paper itself. We decided to include: (i) an umbrella category of *quality control* for all assets, (ii) deliberate *inclusion/exclusion* for datasets (e.g. filtering out "toxic" content based on a word list, which may have the side-effect of removing LGBTQ+ language (Dodge et al. 2021; Gururangan et al. 2022)), (iii) material *training costs* for models (e.g. to contextualize environmental impact (Lacoste et al. 2019; Strubell, Ganesh, and McCallum 2019; Henderson et al. 2020; Luccioni and Hernández-García 2023)) and (iv) *adaptation* details for applications (e.g. fine-tuning details and UI design). These details provide important context since dependencies visually appear the same in **Ecosystem Graphs** even when they corresponding to different underlying relationships.

Downstream properties To construct the ecosystem graph, we specify dependencies on the target asset: given a node, we annotate its parents.[5] However, some properties of an asset influence how it can be built upon, rather than how it was built. Most notably, the *access* status of an asset directly determines who can build on it, whereas the *intended/prohibited uses* influence how the asset should be built upon (in addition to the *license* and *terms of service*). In general, we found these properties to be straightforward to

[4]See https://platform.openai.com/docs/model-index-for-researchers.

[5]We found this natural since, in general, we may not know the all of the dependents of a given upstream asset (e.g. ChatGPT continues to accrue new dependents well after its initial release), but we can better trace the lineage when annotating the downstream asset.

annotate, though we find discussion of intended/prohibited uses is quite uneven and in some cases no license/terms of service could be found.

We also annotate fields that determine the asset's downstream social impact, namely their impact on *end users*. To underscore the transparency of **Ecosystem Graphs**, we highlight mechanisms for accountability/recourse. We track (i) whether asset developers can *monitor* the usage of their assets downstream, and (ii) whether specific *failures* or harms concretely arise, and (iii) when these issues come up, do *feedback* mechanisms exist to propagate this information back upstream? These fields signal assets that are having high impact, which could confer recognition (or payment) to those who contributed to their downstream impact (e.g. valuing data creators whose data generates value downstream).

Complementarity of construction and downstream properties The construction and downstream properties together enrich the interpretation of the underlying graph. When edges are interpreted in the forward direction, they indicate how assets are built; when interpreted in the backwards direction, they indicate how feedback would flow back upstream.

We stress this point as indicative of the present immaturity of the FM ecosystem by analogy to industries with robust supply chain practices, such as the automobile industry. The National Highway Traffic Safety Administration (NHTSA; an agency under the US Department of Transportation), is tasked with ensuring automotive safety.[6] When a batch of parts is found to be sub-standard, established protocols mandate the recall of the fleet of cars built using those parts. Since the National Traffic and Motor Vehicle Safety Act was enacted in 1966, the NHTSA has recalled over 390 million cars due to safety defects. When cars from different manufacturers are reported to be faulty, the shared source of the defect can be traced by attributing their common parts. Centralized infrastructure exists for consumers to report issues to the NHTSA (e.g. the Department of Transporation's Vehicle Safety Hotline), to interpret how their report will be used, to understand the investigation process the NHTSA implements, and to understand the legal remedies and consumer protections they are afforded. And the Federal Motor Vehicle Safety Standards sets formal and transparent standards on what constitutes the minimum performance requirements for each (safety-relevant) automotive part.

Practices in the automobile industry illustrate the virtues of observing an ecosystem as a whole. If an upstream asset (akin to the faulty brakes) is identified to be faulty in the FM ecosystem, we are not confident that communications of its faults, let alone interventions like a recall, will occur. For example, if LAION-5B was shown to be data poisoned (Carlini et al. 2023) or to include child sexual abuse material (Thiel 2023), how would the developers of Stable Diffusion, the application developers who built upon Stable Diffusion, and the end userbase be notified? The FM ecosystem lacks entities similar to the NHTSA that could share information about failures or prompt "recalls". Similarly, many assets lack monitoring mechanisms, let alone publicly-disclosed means for relaying feedback and incident reports upstream. While some organizations have partnerships (e.g. when Khan Academy users surface issues, Khan Academy and OpenAI likely work together to diagnose if these problems arise from Khan Academy's use of OpenAI's GPT-4), these mechanisms do not allow end users to report issues directly. More fundamentally, no adverse event reporting system exists in the foundation model ecosystem (Guha et al. 2023; NAIAC 2023), thereby lacking key consumer protections to mitigate known harms.

Annotation Practices

Different assets often have idiosyncratic properties. As we iterated on annotation best practices, we identified two key questions: (i) how should we interpret *missing* data entries and (ii) how can we *trust* the recorded information?

Missing data. We identified two forms of missing data that arise under different conditions. To clarify these different forms of missing data, we use different annotations. We annotate a field as "None" if an annotator looked for the relevant information and was unable to find it. In contrast, we annotate "Unknown" if the information cannot be found but must exist. For example, a model necessarily has training emissions (even if 0) associated with it, whereas it may not have a model card. Therefore, we would annotate "Unknown" for the former and "None" for the latter. When aggregated across the entire ecosystem, these conventions for missing data help to articulate pervasive opacity (i.e. many *unknown* values) and immaturity (i.e. many *none* values).

Trust. To ensure information in **Ecosystem Graphs** is legitimate and credible, we implement two mechanisms. First, to add or modify information, all such requests must be verified by a vetted maintainer to ensure the correctness of the information as well as the consistency with prior annotations. As **Ecosystem Graphs** expands to be a community-maintained artifact, moderation will support the sustained quality of the resource. Since **Ecosystem Graphs** is implemented as a GitHub repository, the full history of commits and associated discussion is maintained to ensure the provenance of information, akin to the Wikipedia change log. Second, information should be attributed to a publicly-available source to ensure the annotation matches the source and that the source itself is reliable.[7]

Maintenance and Incentives

Everything in the foundation model supply chain is subject to rapid change. For example, in a single week in May 2024, over 50 assets were released. Therefore, keeping pace in maintaining this resource real-time is an ongoing challenge. In fact, even for existing assets, metadata may change (e.g. the license for Microsoft's Phi-2 was changed to the open-source MIT license several months after release). For this reason, we discuss *who* maintains **Ecosystem Graphs**

[6] See their guidelines on motor vehicle safety at https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/14218-mvsdefectsandrecalls_041619-v2-tag.pdf.

[7] In the future, we may permit developers to self-verify information that was not previously publicly.

and whether *incentives misalignment* introduces challenges, given much of the value is contingent on the resource being both correct and up-to-date Moving forward, as FMs feature more centrally in broader social discourse and **Ecosystem Graphs** sees greater adoption, maintenance could be mandated as a policy requirement to ensure sustained transparency (see Bommasani et al. 2023b; Madry 2023).

Who maintains the ecosystem graph? To this point, the ecosystem graph has been built and maintained primarily by the authors of this work. While the authors will continue maintenance, our contributions will be increasingly insufficient given the growing scale of the ecosystem. Therefore, we envision two complementary strategies for expanding the group involved in maintaining **Ecosystem Graphs**. Building on traditions of open source software development (e.g. Wikipedia, Mozilla, PyTorch, Hugging Face), we actively encourage contributions. For this reason we use a lightweight process with explicit guidelines on how to create and edit entries in `YAML`. To broaden accessibility, new assets can be submitted from a public Google form to remove the barriers of having a GitHub account and familiarity with GitHub. Drawing upon trends in open source, we will implement processes for top contributors to be recognized for their achievement and signal-boosted in the broader community.[8] Since the launch in March 2023, multiple open source contributors have emerged with extensive contribution histories.

As **Ecosystem Graphs** expands, we envision it may become a broad-use public repository that organizations themselves are incentivized to maintain. In future, we propose that major FM organizations each select a dedicated representative responsible for the upkeep of their organization's nodes (and, possibly, some direct neighbors). This mechanism introduces *accountability*: the veracity of an organization's nodes and dependencies is the responsibility of this maintainer. Here, we could lean on practices of periodic public reporting (e.g. quarterly financial earnings) in reminding the representative to update the graph regularly. We imagine the specifics of this will further sharpen as **Ecosystem Graphs** grows and as we better understand both the rate of change in the ecosystem and the informational needs that **Ecosystem Graphs** serves. In the future, the process of updating the ecosystem graph could be integrated into develop-internal corporate processes, since much of what is tracked in **Ecosystem Graphs** is likely already tracked within organizations.

The compatibility of incentives. Ensuring the ecosystem is transparent serves many informational needs and benefits many stakeholders. Akin to other shared knowledge resources (e.g. Wikipedia, the US Census), downstream use cases continuously arise, further incentivizing the sustained upkeep of the resource. While incentives may not exist for all relevant information to be made transparent, we hope **Ecosystem Graphs** will encourage increased transparency by demonstrating the cumulative value of public information.

However, we also recognize that there exist pernicious incentives for organizations to maintain opacity in the ecosystem: most directly, transparency could infringe on corporate secrets and commercial interest. Central to our approach in **Ecosystem Graphs** is recognizing that in many cases, information can be made transparent to the public without compromising commercial interests. The information in question is common knowledge amidst an organization's competitors, and it is better for the public to have partial transparency rather than to have nothing at all. This approaches aligns with having representatives from each organization update assets, thereby iteratively identifying the boundaries of transparency in an organization-specific and asset-specific way. More expansively, **Ecosystem Graphs** mediates an incremental process for building norms of transparency (Liang et al. 2022a; Bommasani et al. 2023b) and functions as an inroad for policy intervention as specific informational needs grow more important.

Use Cases for Ecosystem Graphs

Ecosystem Graphs is a multifunctional resource that aims to address informational needs of many stakeholders. We first enumerate, in brief, how different stakeholders may benefit from **Ecosystem Graphs** before describing two major demonstrated uses of the resource by the 2024 AI Index and the UK Competition and Markets Authority.

Stakeholders

AI developers. AI developers need to be aware of what assets are available. **Ecosystem Graphs** allows developers to be made aware of new assets near-realtime as well as use metadata to search the large space of assets. For example, a startup may want to identify new video foundation models that are permissively licensed for commercial applications. In turn, the startup could simply filter on the website by modality and license type to identify the appropriate assets.

End users. Consumers deserve to know how the technology they use was built, just as food in the US must be labeled with its ingredients.[9] The graph structure and the simple web interface make this practical: a user can look up the product, see if it is in the graph, and trace its dependencies. This process surfaces any existing mechanisms for feedback reporting, which could prove to be useful if the user experiences an issue. The user would also be able to find documentation of similar issues or failures, should such documentation exist(see Costanza-Chock, Raji, and Buolamwini 2022). In the future, we imagine this information could become a basis for more formal consumer protections: if a user experiences harm, what are their means for recourse? Or if they pursue legal action, how might society attribute responsibility to different entities implicated upstream?

[8] While we do not currently implement any extrinsic bounties, works like Zhao, Laszka, and Grossklags (2017) and Chowdhury and Williams (2021) demonstrate their efficacy, warranting further consideration in the future.

[9] https://www.fda.gov/food/food-ingredients-packaging/

AI researchers. Ecosystem Graphs provides an array of functionalities that are relevant for AI research. For example, understanding the downstream use of foundation models can help ground research to practical impact. As an example, many of the applications of text-to-image models like Stable Diffusion diverge from what has been traditionally studied in computer vision research, and billions have been invested into language-based startups whose applications (e.g. copywriting) differ from standard tasks studied in natural language processing (e.g. summarization). **Ecosystem Graphs** enables reflection by the scientific community: understanding (i) what is being deployed in society, (ii) the demand for such technology, and (iii) the resulting societal impact can help AI researchers build models that are more accurate, efficient, robust, and fair. In particular, this understanding can bolster research on the risks of AI (Bender et al. 2021; Weidinger et al. 2022; Bommasani et al. 2021; Abid, Farooqi, and Zou 2021; Buchanan et al. 2021, *inter alia*) by tracking how they release to observed harms or usage patterns.

Economists. FMs are general-purpose technologies (Bresnahan and Trajtenberg 1995; Brynjolfsson, Rock, and Syverson 2021) that define an emerging market (Eloundou et al. 2023; Bommasani et al. 2021, §5.5) in the (digital) economy (Acemoglu and Autor 2010; Acemoglu and Restrepo 2018; Agrawal, Gans, and Goldfarb 2021; Autor et al. 2022). Early work shows that FMs can complete tasks of significant economic value (Noy and Zhang 2023; Felten, Raj, and Seamans 2023; Korinek 2023), i.e. the *realizable* market potential of FMs. **Ecosystem Graphs** naturally complements this work by defining the *realized* impact of FMs at macro-scale, complementing more grounded analyses such as Peng et al. (2023) on developer productivity using GitHub Copilot and Eloundou et al. (2023) on labor exposure using GPT-4.

Auditors. Auditors need to prioritize attention on auditing assets of highest priority(see Metaxa et al. 2021; Raji and Buolamwini 2019). To inform these decisions, **Ecosystem Graphs** helps auditors to prioritize (i) assets with known reports of "failures", (ii) unsatisfactory "quality control" practices and (iii) assets with significant opacity (i.e. much is unknown about the node). Further, auditors should factor in the impact of these nodes: we recommend auditors target *algorithmic monocultures* (Kleinberg and Raghavan 2021; Bommasani et al. 2022) that are made legible by **Ecosystem Graphs**. Namely, if an upstream asset has extensive downstream dependencies (e.g. ChatGPT is a hub that supports many applications), then risks associated with this asset may propagate downstream perniciously if unchecked. For example, if a dataset like LAION-400M is subject to data poisoning (Carlini et al. 2023) or contains child sexual abuse material, what are the downstream risks for assets built on the dataset?

Case Study 1: UK CMA

The UK Competition and Markets Authority (CMA) is the competition regulator for the United Kingdom. In May 2023, shortly after the release of **Ecosystem Graphs**, the CMA announced that it would conduct market surveillance on the foundation model supply chain to "help create an early understanding of the market for foundation models and how their use could evolve". Given that **Ecosystem Graphs** was available, the CMA reached out to us and, over the course of several months, we engaged with the CMA on directly using **Ecosystem Graphs** to conduct their activities instead of creating a new database (their initial proposal). In particular, we provided both technical guidance on how to conduct data analysis and handle missing data in **Ecosystem Graphs**, as well more qualitative learnings about the foundation model ecosystem.

In the initial report of the CMA released in September 2023 (UK CMA 2023), the CMA makes extensive use of **Ecosystem Graphs**. For example, the CMA uses our data to identify which foundation model developers produce the most foundation models. And, given the specific interest in how open foundation models (i.e. those with widely available model weights) may differentially contribute to competition as compared to more restricted/closed foundation models, the CMA used our annotations to characterize the distribution of release strategies as captured by our data. The analysis indicated that, at the time of the report, 42% of tracked models were released with open weights and that most common license for licensed assets was the Apache 2.0 license. Based on subsequent discussions, the CMA plans to continue using **Ecosystem Graphs** as a primary resource to build market understanding, which may be emulated by market surveillance authorities and competition regulators in other jurisdictions.

Case Study 2: AI Index

The AI Index is an annual effort that tracks, collates, distills, and visualizes data on trends in artificial intelligence. It is recognized globally as one of the most credible and authoritative sources for data and insights on artificial intelligence: the Index has thousands of academic citations, is used as a primary resource for top-level policy-making on AI in many jurisdictions, and is reported on in mainstream outlets like Bloomberg and The New York Times. The 2024 AI Index (Maslej et al. 2024) was launched in April 2024 as the seventh edition with a dedicated section on foundation models.

In preparing the 2024 AI Index, Maslej et al. (2024) reached out to use **Ecosystem Graphs** as the sole data source for all analyses about foundation models. In total, our data was the basis for five analyses featured in the Index: trends in the number of foundation models, the sectors developing foundation models, the release strategies for foundation models, the organizations developing foundation models, and the countries developing foundation models. We provide some of these figures as examples (Figure 2, Figure 3). In doing so, this use case also demonstrates how data in **Ecosystem Graphs** can be merged with other information such as information about companies (e.g. location of headquarters, industry sector).

Related Work

To situate our work, we consider both our objective of increasing transparency in AI and our methodology of supply chain monitoring.

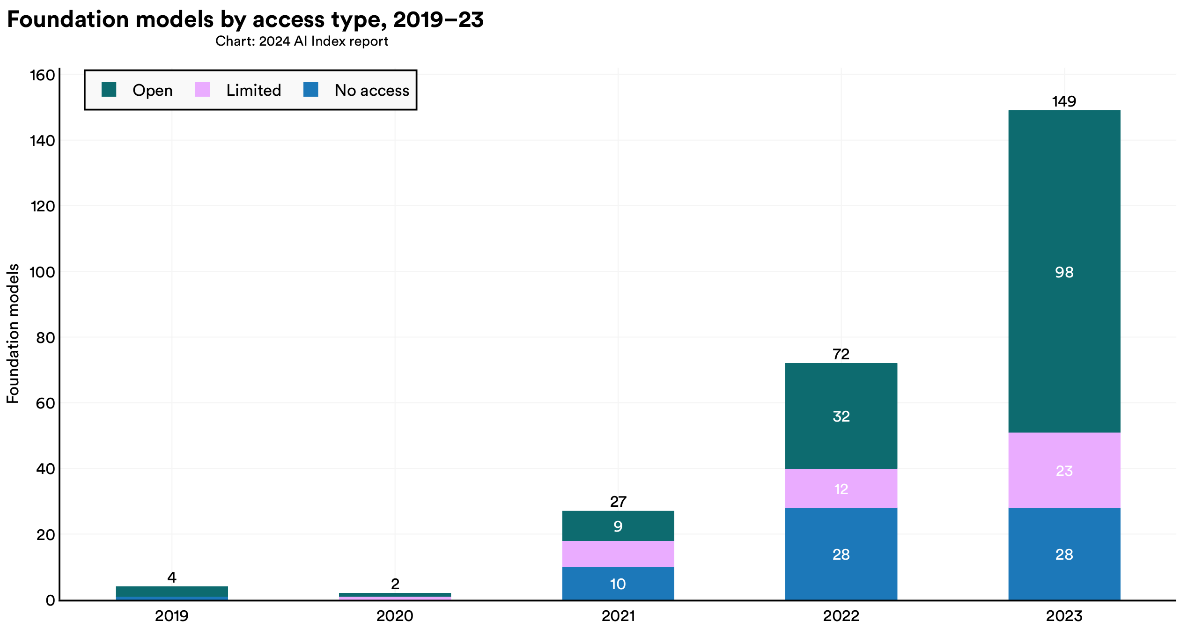

Figure 2: Release strategies over time. Release strategies for foundation models based on **Ecosystem Graphs** data as reported in the 2024 AI Index (Maslej et al. 2024).

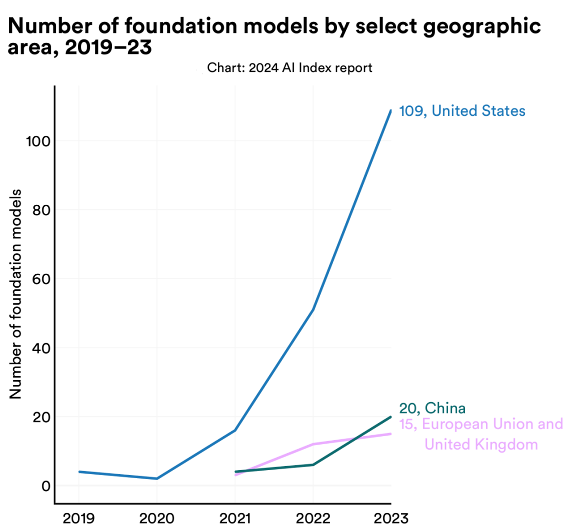

Figure 3: Geographic development over time. Geographic regions for foundation model development based on **Ecosystem Graphs** data as reported in the 2024 AI Index (Maslej et al. 2024).

Transparency in AI

We categorize efforts for transparency in AI into three categories.

First, *evaluation* is a widespread practice for articulating the properties and measuring the behavior of systems: in the research community, it is customary to evaluate systems against particular benchmarks to assess their performance. Evaluations can vary in the specific type of transparency (see Bommasani et al. 2023b) they provide: some evaluations quantify the accuracy of models (e.g. ImageNet; Deng et al. 2009), others stress-test models (e.g. CheckList; Ribeiro et al. 2020) or adversarially identify failures (e.g. red-teaming; Perez et al. 2022) and still others characterize models along a range of dimensions (e.g. HELM; Liang et al. 2022b). In general, while some efforts expand evaluation to datasets (Bommasani and Cardie 2020; Swayamdipta et al. 2020; Ethayarajh, Choi, and Swayamdipta 2022; Mitchell et al. 2022) or adopt methodologies from human-computer interaction to consider human factors like user experience (Lee, Liang, and Yang 2022; Lee et al. 2022), for the most part, evaluation aims to characterize a specific model in isolation.

Second, *documentation* is a growing practice for specifying metadata about the broader context that situates model and system development. Formative works like data sheets (Gebru et al. 2018) and model cards (Mitchell et al. 2018) brought this approach to the fore, complementing evaluations by articulating design decisions and developer positions involved in creating assets. Subsequent efforts have enriched these approaches to make these documentation artifacts more useful, accessible, or otherwise aligned to specific informational needs (Crisan et al. 2022).[10] In general, documentation efforts aim to contextualize a specific asset against a broader social backdrop, often with an inclination towards how the asset came to be and with greater uptake for research-centric assets to our knowledge.

Third, *analyses* and critiques have become increasingly relevant, showcasing much of the latent and oft-overlooked underpinnings of AI development and deployment. These works often bring questions of values and power to the fore, frequently appealing to concepts or methodologies from the social sciences and disciplines beyond computer science (e.g. Dotan and Milli 2020; Ethayarajh and Jurafsky 2020; Scheuerman, Hanna, and Denton 2021; Raji et al. 2021; Koch et al. 2021; Denton et al. 2021; Paullada et al. 2021; Birhane et al. 2022; Bommasani 2022). Rather than specific assets (other than for case study/illustration), analytic work centers broader themes (e.g. algorithmic intermediaries; Lazar 2023) or classes of technology (e.g. predictive optimization; Wang et al. 2022a).

Our framework shares the objective of making AI (specifically FMs) transparent. However, it differs along important axes from all of these established approaches, perhaps most closely resembling the documentation class (since, indeed, **Ecosystem Graphs** is a documentation framework). In contrast to current interpretations of evaluation and documentation, **Ecosystem Graphs** is fundamentally about the ecosystem rather any specific asset: the value of **Ecosystem Graphs** arises from tracking all assets.[11] This introduces a variety of new challenges (e.g. partial observability of certain information, more complicated maintenance as there is constant change across the ecosystem even if particular assets may not change for extended periods). Further, **Ecosystem Graphs** especially highlights the importance of grounding out into applications (for which a general-purpose analogue of data sheets and model cards does not exist to our knowledge) and, more generally, moving beyond research artifacts to commercial artifacts that affect broader society. Finally, in comparison to analytic/critical methods, **Ecosystem Graphs** retains much of the concreteness of evaluation/documentation: we believe **Ecosystem Graphs** provides valuable descriptive understanding that could support future normative analyses and critiques.

Beyond these distinctions, we emphasize that our contribution extends beyond most prior works on documentation in AI. Concretely, most prior works (e.g. Gebru et al. 2018; Bender and Friedman 2018; Mitchell et al. 2018) propose the framework to document artifacts, perhaps with an additional proposal of who will conduct this documentation and how/why. In contrast, we concretely execute, implementing the **Ecosystem Graphs** framework in our codebase and public website. This mirrors works like HELM (Liang et al. 2022b) where, in addition to designing an evaluation, the contributions include evaluating all language models available at present. The infrastructure, sustained upkeep and, ultimately, the resource itself are what provide value: ensuring transparency requires we follow through and enact the conceptual frameworks we design.

Supply Chain Monitoring

At the technical level, **Ecosystem Graphs** foreground the tracking of dependencies, whereas at the social level, **Ecosystem Graphs** delineates institutional relationships. Both of these constructs are encountered in almost every mature industry and, therefore, have been studied across a range of fields. Concretely, almost every commercial product is the composite of some collection of materials/ingredients, meaning it has a complex supply chain. As a result, we specifically point to related work in open-source software (which share similar implementation to **Ecosystem Graphs**) and market structure (which emulate **Ecosystem Graphs** in terms of organizations).

Open-source software. Much like FMs, open-source software development is sustained by an immense network of dependencies. Akin to our efforts to track the FM ecosystem, the demand to track the open-source software ecosystem is immense: the software bill of materials (SBOM) is a national-level initiative of the US's Cybersecurity and Infrastructure Security Agency to maintain an inventory of the ingredients that make up software (White House Executive Order 2021).[12] These approaches have clarified how to ensure compliance from different stakeholders (e.g. software vendors) and how to standardize information to support a range of use cases, providing inspiration for the abstractions we make in **Ecosystem Graphs**. To implement this vision, a range of efforts have been put forth over the years[13] with applied policy work mapping out the sociotechnical challenges for maintaining and funding these efforts (Ramaswami 2021; Scott et al. 2023). Further, they present an exemplar of policy uptake towards mandatory public reporting of these dependencies as exemplified by the proposed Securing Open Source Software Act of 2022.[14] And, much akin to the uses we consider, these efforts already have shown how descriptive understanding of the ecosystem directly informs decision-making and characterizes the impact of assets.[15]

Market structure. In defining **Ecosystem Graphs**, we made the fundamental decision to define the graph in terms of assets. We contrast this with approaches more common in disciplines like economics and sociology, where it would be customary to instead foreground the organizations/institutions responsible for creating these assets (Rowlinson 1997). We believe this (fairly techno-centric) choice provides valuable leverage given the status quo: the number of assets is

[10]See https://huggingface.co/docs/hub/model-cards.

[11]We do note other works exist at ecosystem scale in other senses such AI100 (Stone et al. 2022), AI Index (Zhang et al. 2021), and various data/model repositories (Wolf et al. 2020; Lhoest et al. 2021); see https://modelzoo.co/.

[12]See https://www.cisa.gov/sbom.

[13]See https://libraries.io/, http://deps.dev/, and https://ecosyste.ms/.

[14]https://www.congress.gov/bill/117th-congress/senate-bill/4913

[15]See https://docs.libraries.io/overview.html\#sourcerank and https://github.com/ossf/criticality_score.

currently still manageable (on the order of hundreds to thousands), the assets themselves are distinctive (e.g. they are not exchangeable in the way oil or steel may be in other market analyses), and specific assets markedly contextualize our understanding of organizations (e.g. Stable Diffusion dramatically shapes our perception of Stability AI). In spite of these advantages, we point to a range of works that foreground institutions in mapping out market structure and the dynamics by which actors interact to shape the economy. For example, given we draw upon a comparative analysis of the FM ecosystem to the automotive ecosystem, Weingast and Marshall (1988) demonstrate that institution-centrism better allow for comparisons/juxtapositions across sectors. Alternatively, Einav and Levin (2010) showcase how grounding to institutions facilitates various forms of measurement (e.g. due to firm-level requirements on information disclosure). Finally, many works in political and institutional sociology prime us to view institutions as the natural unit for studying power relations in modern networks and markets (Frickel and Moore 2006; Dequech 2006; Fleury 2014, *inter alia*).

Conclusion

The social footprint of foundation models is rapidly escalating. Foundation models mediate thousands of technologies affecting billions of people in aggregate. In general, current understanding of the foundation model supply chain is in its infancy and generally underwhelming. We introduce **Ecosystem Graphs** to make progress towards rectifying this situation by promoting supply chain monitoring through concrete tooling and a new centralized data repository. Moving forward, we hope to see the community building on and contributing back to **Ecosystem Graphs** as we collectively work towards a more mature foundation model ecosystem.

Acknowledgements

We thank Alex Tamkin, Ali Alvi, Amelia Hardy, Ansh Khurana, Ashwin Paranjape, Ayush Kanodia, Chris Manning, Dan Ho, Dan Jurafsky, Daniel Zhang, David Hall, Dean Carignan, Deep Ganguli, Dimitris Tsipras, Erik Brynjolfsson, Iason Gabriel, Irwing Kwong, Jack Clark, Jeremy Kaufmann, Laura Weidinger, Maggie Basta, Michael Zeng, Nazneen Rajani, Rob Reich, Rohan Taori, Tianyi Zhang, Vanessa Parli, Xia Song, Yann Dubois, and Yifan Mai for valuable feedback on this effort at various stages. We specifically thank Ashwin Ramaswami for extensive guidance on prior work in the open-source ecosystem and future directions for the foundation model ecosystem. In addition, the authors would like to thank the Stanford Center for Research on Foundation Models (CRFM) and the Stanford Institute for Human-Centered Artificial Intelligence (HAI) for directly supporting this research. The initial release of **Ecosystem Graphs** will be accompanied by a policy brief through collaboration with Stanford HAI. RB was supported by an NSF Graduate Research Fellowship (grant number: DGE-1656518). This work was supported in part by the AI2050 program at Schmidt Futures (Grant G-22-63429).

Ethical Concerns

By its nature, our work involves tracking an evolving ecosystem with uncertainty surrounding coverage: as a result, analyses based on our monitoring may be prone to overgeneralizing, and insufficiently considering the untracked nodes in the ecosystem. Beyond this, we do not foresee significant ethical concerns from this work or its use.

Research Positionality

The researchers involved in this work are situated at academic institutions: in large part, they are motivated to conduct this work by deficiencies in public and academic understanding of the supply chain and market impact of foundation models.

Adverse Impact

We do not foresee any significant adverse impacts of this work.

References

Abid, A.; Farooqi, M.; and Zou, J. 2021. Persistent anti-muslim bias in large language models. *arXiv preprint arXiv:2101.05783*.

Acemoglu, D.; and Autor, D. 2010. Skills, Tasks and Technologies: Implications for Employment and Earnings. Working Paper 16082, National Bureau of Economic Research.

Acemoglu, D.; and Restrepo, P. 2018. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, 108(6): 1488–1542.

Agrawal, A. K.; Gans, J. S.; and Goldfarb, A. 2021. AI Adoption and System-Wide Change. Working Paper 28811, National Bureau of Economic Research.

Amironesei, R.; Denton, E.; and Hanna, A. 2021. Notes on Problem Formulation in Machine Learning. *IEEE Technology and Society Magazine*, 40(3): 80–83.

Autor, D.; Mindell, D.; Reynolds, E.; and Solow, R. 2022. *The Work of the Future: Building Better Jobs in an Age of Intelligent Machines*. MIT Press. ISBN 9780262046367.

Bender, E. M.; and Friedman, B. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics (TACL)*, 6: 587–604.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Birhane, A.; Kalluri, P.; Card, D.; Agnew, W.; Dotan, R.; and Bao, M. 2022. The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 173–184. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Bommasani, R. 2022. Evaluation for Change. *ArXiv*.

Bommasani, R.; and Cardie, C. 2020. Intrinsic Evaluation of Summarization Datasets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8075–8096. Online: Association for Computational Linguistics.

Bommasani, R.; Creel, K. A.; Kumar, A.; Jurafsky, D.; and Liang, P. 2022. Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? In *Advances in Neural Information Processing Systems*.

Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; Brynjolfsson, E.; Buch, S.; Card, D.; Castellon, R.; Chatterji, N.; Chen, A.; Creel, K.; Davis, J. Q.; Demszky, D.; Donahue, C.; Doumbouya, M.; Durmus, E.; Ermon, S.; Etchemendy, J.; Ethayarajh, K.; Fei-Fei, L.; Finn, C.; Gale, T.; Gillespie, L.; Goel, K.; Goodman, N.; Grossman, S.; Guha, N.; Hashimoto, T.; Henderson, P.; Hewitt, J.; Ho, D. E.; Hong, J.; Hsu, K.; Huang, J.; Icard, T.; Jain, S.; Jurafsky, D.; Kalluri, P.; Karamcheti, S.; Keeling, G.; Khani, F.; Khattab, O.; Koh, P. W.; Krass, M.; Krishna, R.; Kuditipudi, R.; Kumar, A.; Ladhak, F.; Lee, M.; Lee, T.; Leskovec, J.; Levent, I.; Li, X. L.; Li, X.; Ma, T.; Malik, A.; Manning, C. D.; Mirchandani, S.; Mitchell, E.; Munyikwa, Z.; Nair, S.; Narayan, A.; Narayanan, D.; Newman, B.; Nie, A.; Niebles, J. C.; Nilforoshan, H.; Nyarko, J.; Ogut, G.; Orr, L.; Papadimitriou, I.; Park, J. S.; Piech, C.; Portelance, E.; Potts, C.; Raghunathan, A.; Reich, R.; Ren, H.; Rong, F.; Roohani, Y.; Ruiz, C.; Ryan, J.; Ré, C.; Sadigh, D.; Sagawa, S.; Santhanam, K.; Shih, A.; Srinivasan, K.; Tamkin, A.; Taori, R.; Thomas, A. W.; Tramèr, F.; Wang, R. E.; Wang, W.; Wu, B.; Wu, J.; Wu, Y.; Xie, S. M.; Yasunaga, M.; You, J.; Zaharia, M.; Zhang, M.; Zhang, T.; Zhang, X.; Zhang, Y.; Zheng, L.; Zhou, K.; and Liang, P. 2021. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*.

Bommasani, R.; Klyman, K.; Longpre, S.; Kapoor, S.; Maslej, N.; Xiong, B.; Zhang, D.; and Liang, P. 2023a. The Foundation Model Transparency Index. arXiv:2310.12941.

Bommasani, R.; Zhang, D.; Lee, T.; and Liang, P. 2023b. Improving Transparency in AI Language Models: A Holistic Evaluation. *Foundation Model Issue Brief Series*.

Bresnahan, T. F.; and Trajtenberg, M. 1995. General purpose technologies ‘Engines of growth’? *Journal of Econometrics*, 65(1): 83–108.

Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*.

Brynjolfsson, E.; Rock, D.; and Syverson, C. 2021. The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics*, 13(1): 333–72.

Buchanan, B.; Lohn, A.; Musser, M.; and Sedova, K. 2021. Truth, Lies, and Automation.

Carlini, N.; Jagielski, M.; Choquette-Choo, C. A.; Paleka, D.; Pearce, W.; Anderson, H.; Terzis, A.; Thomas, K.; and Tramèr, F. 2023. Poisoning Web-Scale Training Datasets is Practical. *ArXiv*, abs/2302.10149.

Chen, M.; Radford, A.; Child, R.; Wu, J.; Jun, H.; Luan, D.; and Sutskever, I. 2020. Generative Pretraining From Pixels. In III, H. D.; and Singh, A., eds., *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 1691–1703. PMLR.

Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; Ponde, H.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; Ray, A.; Puri, R.; Krueger, G.; Petrov, M.; Khlaaf, H.; Sastry, G.; Mishkin, P.; Chan, B.; Gray, S.; Ryder, N.; Pavlov, M.; Power, A.; Kaiser, L.; Bavarian, M.; Winter, C.; Tillet, P.; Such, F. P.; Cummings, D. W.; Plappert, M.; Chantzis, F.; Barnes, E.; Herbert-Voss, A.; Guss, W. H.; Nichol, A.; Babuschkin, I.; Balaji, S. A.; Jain, S.; Carr, A.; Leike, J.; Achiam, J.; Misra, V.; Morikawa, E.; Radford, A.; Knight, M. M.; Brundage, M.; Murati, M.; Mayer, K.; Welinder, P.; McGrew, B.; Amodei, D.; McCandlish, S.; Sutskever, I.; and Zaremba, W. 2021. Evaluating Large Language Models Trained on Code. *ArXiv*, abs/2107.03374.

Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S.; Schuh, P.; Shi, K.; Tsvyashchenko, S.; Maynez, J.; Rao, A.; Barnes, P.; Tay, Y.; Shazeer, N. M.; Prabhakaran, V.; Reif, E.; Du, N.; Hutchinson, B.; Pope, R.; Bradbury, J.; Austin, J.; Isard, M.; Gur-Ari, G.; Yin, P.; Duke, T.; Levskaya, A.; Ghemawat, S.; Dev, S.; Michalewski, H.; García, X.; Misra, V.; Robinson, K.; Fedus, L.; Zhou, D.; Ippolito, D.; Luan, D.; Lim, H.; Zoph, B.; Spiridonov, A.; Sepassi, R.; Dohan, D.; Agrawal, S.; Omernick, M.; Dai, A. M.; Pillai, T. S.; Pellat, M.; Lewkowycz, A.; Moreira, E.; Child, R.; Polozov, O.; Lee, K.; Zhou, Z.; Wang, X.; Saeta, B.; Diaz, M.; Firat, O.; Catasta, M.; Wei, J.; Meier-Hellstern, K.; Eck, D.; Dean, J.; Petrov, S.; and Fiedel, N. 2022. PaLM: Scaling Language Modeling with Pathways. *arXiv*.

Chowdhury, R.; and Williams, J. 2021. Introducing Twitter’s first algorithmic bias bounty challenge.

Costanza-Chock, S.; Raji, I. D.; and Buolamwini, J. 2022. Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’22, 1571–1583. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Crisan, A.; Drouhard, M.; Vig, J.; and Rajani, N. 2022. Interactive Model Cards: A Human-Centered Approach to Model Documentation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’22, 427–439. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition (CVPR)*, 248–255.

Denton, E.; Hanna, A.; Amironesei, R.; Smart, A.; and Nicole, H. 2021. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society*, 8(2): 20539517211035955.

Dequech, D. 2006. Institutions and norms in institutional economics and sociology. *Journal of Economic Issues*, 40(2): 473–481.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Association for Computational Linguistics (ACL)*, 4171–4186.

Dodge, J.; Sap, M.; Marasović, A.; Agnew, W.; Ilharco, G.; Groeneveld, D.; Mitchell, M.; and Gardner, M. 2021. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 1286–1305. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Dotan, R.; and Milli, S. 2020. Value-laden disciplinary shifts in machine learning. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

Einav, L.; and Levin, J. 2010. Empirical Industrial Organization: A Progress Report. *Journal of Economic Perspectives*, 24(2): 145–62.

Eloundou, T.; Manning, S.; Mishkin, P.; and Rock, D. 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv:2303.10130.

Ethayarajh, K.; Choi, Y.; and Swayamdipta, S. 2022. Understanding Dataset Difficulty with $\mathcal{V}$-Usable Information. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvari, C.; Niu, G.; and Sabato, S., eds., *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 5988–6008. PMLR.

Ethayarajh, K.; and Jurafsky, D. 2020. Utility is in the Eye of the User: A Critique of NLP Leaderboards. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4846–4853. Online: Association for Computational Linguistics.

Felten, E. W.; Raj, M.; and Seamans, R. C. 2023. How will Language Modelers like ChatGPT Affect Occupations and Industries? *SSRN Electronic Journal*.

Fleury, L. 2014. *Sociology of culture and cultural practices: The transformative power of institutions*. Lexington Books.

Franks, D. W.; Noble, J.; Kaufmann, P.; and Stagl, S. 2008. Extremism propagation in social networks with hubs. *Adaptive Behavior*, 16(4): 264–274.

Frickel, S.; and Moore, K. 2006. *The new political sociology of science: Institutions, networks, and power*. Univ of Wisconsin Press.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Ill, H. D.; and Crawford, K. 2018. Datasheets for Datasets. *arXiv preprint arXiv:1803.09010*.

Guha, N.; Lawrence, C. M.; Gailmard, L. A.; Rodolfa, K. T.; Surani, F.; Bommasani, R.; Raji, I. D.; Cuéllar, M.-F.; Honigsberg, C.; Liang, P.; and Ho, D. E. 2023. AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing. *George Washington Law Review, Symposium on Legally Disruptive Emerging Technologies*.

Gururangan, S.; Card, D.; Drier, S. K.; Gade, E. K.; Wang, L. Z.; Wang, Z.; Zettlemoyer, L.; and Smith, N. A. 2022. Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection. In *Conference on Empirical Methods in Natural Language Processing*.

Henderson, P.; Hu, J.; Romoff, J.; Brunskill, E.; Jurafsky, D.; and Pineau, J. 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248): 1–43.

Hendricks, K.; Piccione, M.; and Tan, G. 1995. The Economics of Hubs: The Case of Monopoly. *The Review of Economic Studies*, 62(1): 83–99.

Hu, K. 2023. ChatGPT sets record for fastest-growing user base - analyst note.

Jumper, J. M.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D. A.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; and Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596: 583 – 589.

Kleinberg, J.; and Raghavan, M. 2021. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22): e2018340118.

Kleinberg, J. M. 1999. Hubs, Authorities, and Communities. *ACM Comput. Surv.*, 31(4es): 5–es.

Koch, B.; Denton, E.; Hanna, A.; and Foster, J. G. 2021. Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Korinek, A. 2023. Language Models and Cognitive Automation for Economic Research. NBER Working Papers 30957, National Bureau of Economic Research, Inc.

Lacoste, A.; Luccioni, A.; Schmidt, V.; and Dandres, T. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

Lazar, S. 2023. Governing the Algorithmic City. *Tanner Lectures*.

Lee, M.; Liang, P.; and Yang, Q. 2022. CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities. In *Conference on Human Factors in Computing Systems (CHI)*.

Lee, M.; Srivastava, M.; Hardy, A.; Thickstun, J.; Durmus, E.; Paranjape, A.; Gerard-Ursin, I.; Li, X. L.; Ladhak, F.; Rong, F.; Wang, R. E.; Kwon, M.; Park, J. S.; Cao, H.; Lee, T.; Bommasani, R.; Bernstein, M.; and Liang, P. 2022. Evaluating Human-Language Model Interaction.

Lhoest, Q.; del Moral, A. V.; Jernite, Y.; Thakur, A.; von Platen, P.; Patil, S.; Chaumond, J.; Drame, M.; Plu, J.; Tunstall, L.; Davison, J.; vSavsko, M.; Chhablani, G.; Malik, B.; Brandeis, S.; Scao, T. L.; Sanh, V.; Xu, C.; Patry, N.; McMillan-Major, A.; Schmid, P.; Gugger, S.; Delangue, C.; Matussiere, T.; Debut, L.; Bekman, S.; Cistac, P.; Goehringer, T.; Mustar, V.; Lagunas, F.; Rush, A. M.; and Wolf, T. 2021. Datasets: A Community Library for Natural Language Processing. *ArXiv*, abs/2109.02846.

Liang, P.; Bommasani, R.; Creel, K. A.; and Reich, R. 2022a. The Time Is Now to Develop Community Norms for the Release of Foundation Models.

Liang, P.; Bommasani, R.; Lee, T.; Tsipras, D.; Soylu, D.; Yasunaga, M.; Zhang, Y.; Narayanan, D.; Wu, Y.; Kumar, A.; Newman, B.; Yuan, B.; Yan, B.; Zhang, C.; Cosgrove, C.; Manning, C. D.; R'e, C.; Acosta-Navas, D.; Hudson, D. A.; Zelikman, E.; Durmus, E.; Ladhak, F.; Rong, F.; Ren, H.; Yao, H.; Wang, J.; Santhanam, K.; Orr, L. J.; Zheng, L.; Yuksekgonul, M.; Suzgun, M.; Kim, N. S.; Guha, N.; Chatterji, N. S.; Khattab, O.; Henderson, P.; Huang, Q.; Chi, R.; Xie, S. M.; Santurkar, S.; Ganguli, S.; Hashimoto, T.; Icard, T. F.; Zhang, T.; Chaudhary, V.; Wang, W.; Li, X.; Mai, Y.; Zhang, Y.; and Koreeda, Y. 2022b. Holistic Evaluation of Language Models. *ArXiv*, abs/2211.09110.

Luccioni, A. S.; and Hernández-García, A. 2023. Counting Carbon: A Survey of Factors Influencing the Emissions of Machine Learning. *ArXiv*, abs/2302.08476.

Martin Jr, D.; Prabhakaran, V.; Kuhlberg, J.; Smart, A.; and Isaac, W. S. 2020. Extending the machine learning abstraction boundary: A Complex systems approach to incorporate societal context. *arXiv preprint arXiv:2006.09663*.

Maslej, N.; Fattorini, L.; Brynjolfsson, E.; Etchemendy, J.; Ligett, K.; Lyons, T.; Manyika, J.; Ngo, H.; Niebles, J. C.; Parli, V.; et al. 2024. The AI index 2024 annual report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA*.

Metaxa, D.; Park, J. S.; Robertson, R. E.; Karahalios, K.; Wilson, C.; Hancock, J.; and Sandvig, C. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction*, 14(4): 272–344.

Mitchell, M.; Luccioni, A. S.; Lambert, N.; Gerchick, M.; McMillan-Major, A.; Ozoani, E.; Rajani, N.; Thrush, T.; Jernite, Y.; and Kiela, D. 2022. Measuring Data. *ArXiv*, abs/2212.05129.

Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2018. Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*.

Madry, A. 2023. Advances in AI: Are We Ready For a Tech Revolution? *Cybersecurity, Information Technology, and Government Innovation Subcommittee*.

NAIAC. 2023. RECOMMENDATION: Improve Monitoring of Emerging Risks from AI through Adverse Event Reporting.

Noy, S.; and Zhang, W. 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *SSRN Electronic Journal*.

OpenAI. 2023. GPT-4 Technical Report.

Paullada, A.; Raji, I. D.; Bender, E. M.; Denton, E.; and Hanna, A. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns*, 2(11): 100336.

Peng, S.; Kalliamvakou, E.; Cihon, P.; and Demirer, M. 2023. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. *ArXiv*, abs/2302.06590.

Perez, E.; Huang, S.; Song, F.; Cai, T.; Ring, R.; Aslanides, J.; Glaese, A.; McAleese, N.; and Irving, G. 2022. Red Teaming Language Models with Language Models. *ArXiv*, abs/2202.03286.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning (ICML)*, volume 139, 8748–8763.

Raji, I. D.; and Buolamwini, J. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, 429–435. New York, NY, USA: Association for Computing Machinery. ISBN 9781450363242.

Raji, I. D.; Denton, E.; Bender, E. M.; Hanna, A.; and Paullada, A. 2021. AI and the Everything in the Whole Wide World Benchmark. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Ramaswami, A. 2021. Securing Open Source Software at the Source. *Plaintext Group by Schmidt Futures*.

Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; and Sutskever, I. 2021. Zero-Shot Text-to-Image Generation. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 8821–8831. PMLR.

Ribeiro, M. T.; Wu, T.; Guestrin, C.; and Singh, S. 2020. Beyond Accuracy: Behavioral Testing of NLP Models with CheckList. In *Association for Computational Linguistics (ACL)*, 4902–4912.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752.

Rowlinson, M. 1997. *Organisations and institutions: perspectives in economics and sociology*. Springer.

Scheuerman, M. K.; Hanna, A.; and Denton, E. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2).

Schuhmann, C.; Beaumont, R.; Vencu, R.; Gordon, C. W.; Wightman, R.; Cherti, M.; Coombes, T.; Katta, A.; Mullis, C.; Wortsman, M.; Schramowski, P.; Kundurthy, S. R.;

Crowson, K.; Schmidt, L.; Kaczmarczyk, R.; and Jitsev, J. 2022. LAION-5B: An open large-scale dataset for training next generation image-text models. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Scott, S.; Brackett, S. A.; Herr, T.; and Hamin, M. 2023. Avoiding the success trap: Toward policy for open-source software as infrastructure. *Atlantic Council*.

Singer, U.; Polyak, A.; Hayes, T.; Yin, X.; An, J.; Zhang, S.; Hu, Q.; Yang, H.; Ashual, O.; Gafni, O.; Parikh, D.; Gupta, S.; and Taigman, Y. 2022. Make-A-Video: Text-to-Video Generation without Text-Video Data. *ArXiv*, abs/2209.14792.

Stone, P.; Brooks, R.; Brynjolfsson, E.; Calo, R.; Etzioni, O.; Hager, G.; Hirschberg, J.; Kalyanakrishnan, S.; Kamar, E.; Kraus, S.; et al. 2022. Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. *arXiv preprint arXiv:2211.06318*.

Strubell, E.; Ganesh, A.; and McCallum, A. 2019. Energy and policy considerations for deep learning in NLP. *arXiv preprint arXiv:1906.02243*.

Swayamdipta, S.; Schwartz, R.; Lourie, N.; Wang, Y.; Hajishirzi, H.; Smith, N. A.; and Choi, Y. 2020. Dataset cartography: Mapping and diagnosing datasets with training dynamics. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Thiel, D. 2023. Identifying and Eliminating CSAM in Generative ML Training Data and Models.

UK CMA. 2023. AI Foundation Models: Initial Report.

Van den Heuvel, M. P.; and Sporns, O. 2013. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12): 683–696.

Verkuil, R.; Kabeli, O.; Du, Y.; Wicky, B. I. M.; Milles, L. F.; Dauparas, J.; Baker, D.; Ovchinnikov, S.; Sercu, T.; and Rives, A. 2022. Language models generalize beyond natural proteins. *bioRxiv*.

Wang, A.; Kapoor, S.; Barocas, S.; and Narayanan, A. 2022a. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy. *Available at SSRN*.

Wang, Y.; Li, K.; Li, Y.; He, Y.; Huang, B.; Zhao, Z.; Zhang, H.; Xu, J.; Liu, Y.; Wang, Z.; Xing, S.; Chen, G.; Pan, J.; Yu, J.; Wang, Y.; Wang, L.; and Qiao, Y. 2022b. InternVideo: General Video Foundation Models via Generative and Discriminative Learning. *ArXiv*, abs/2212.03191.

Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; Chi, E. H.; Hashimoto, T.; Vinyals, O.; Liang, P.; Dean, J.; and Fedus, W. 2022. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*. Survey Certification.

Weidinger, L.; Uesato, J.; Rauh, M.; Griffin, C.; Huang, P.-S.; Mellor, J.; Glaese, A.; Cheng, M.; Balle, B.; Kasirzadeh, A.; Biles, C.; Brown, S.; Kenton, Z.; Hawkins, W.; Stepleton, T.; Birhane, A.; Hendricks, L. A.; Rimell, L.; Isaac, W.; Haas, J.; Legassick, S.; Irving, G.; and Gabriel, I. 2022. Taxonomy of Risks Posed by Language Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 214–229. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Weingast, B. R.; and Marshall, W. J. 1988. The Industrial Organization of Congress; or, Why Legislatures, Like Firms, Are Not Organized as Markets. *Journal of Political Economy*, 96(1): 132–163.

White House Executive Order. 2021. Executive Order on Improving the Nation's Cybersecurity.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Le Scao, T.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Online: Association for Computational Linguistics.

Zhang, D.; Mishra, S.; Brynjolfsson, E.; Etchemendy, J.; Ganguli, D.; Grosz, B.; Lyons, T.; Manyika, J.; Niebles, J. C.; Sellitto, M.; et al. 2021. The AI index 2021 annual report. *arXiv preprint arXiv:2103.06312*.

Zhao, M.; Laszka, A.; and Grossklags, J. 2017. Devising effective policies for bug-bounty platforms and security vulnerability discovery. *Journal of Information Policy*, 7: 372–418.