Mapping the AI Supply Chain: Tracking Relationships, Partnerships, and Dependencies in the AI Industry

Sarah H. Cen
Carnegie Mellon University

Lindsey A. Gailmard
Stanford University

Joohan Kim
Carnegie Mellon University

Rishi Bommasani
Stanford University

Daniel E. Ho
Stanford University

Percy Liang
Stanford University

Working draft — January 2026.
Updated version forthcoming.

Abstract

The rapid proliferation of artificial intelligence (AI) systems, particularly general-purpose large language models (LLMs), has spawned a complex, interdependent ecosystem of organizations that produce, deploy, and maintain AI systems. This project provides the first real-time, large-scale empirical mapping of the AI supply chain, defined as the network of upstream and downstream relationships that structure how AI resources, products, services, capital, and labor flow. Assembling a novel dataset enables us to analyze how actors interact, how supply chain structures evolve, how economic power concentrates, and where chokepoints and strategic dependencies may emerge. We argue that a detailed understanding of the AI supply chain is essential for AI governance: it highlights bottlenecks in compute, data, talent, and model access; clarifies responsibilities and potential liabilities; enhances market transparency; and enables policymakers to anticipate systemic vulnerabilities. This paper introduces our data, methodology, and preliminary analytical results, situating them within a growing, but nascent literature on AI ecosystems, accountability, and supply-chain mapping. We conclude by outlining priority areas for policy intervention and future research.

1 Introduction

Artificial intelligence has drawn sustained scientific, commercial, and political attention. The recent emergence of large language models (LLMs) and other foundation models, however, marks a discontinuity in how AI systems are developed, accessed, and used. By enabling users to interact with AI systems through natural language rather than specialized interfaces or domain-specific APIs, these systems substantially lower barriers to entry for both downstream developers and end users. This shift has transformed AI into a general-purpose technology with wide-ranging applications.

As adoption has accelerated, the production of AI-enabled technologies has become increasingly distributed across a diverse set of organizations. Rather than building end-to-end systems entirely in-house, firms now assemble AI capabilities from multiple sources, including model developers, data companies, and cloud infrastructure providers. These actors, in turn, rely on contributions from upstream actors, like semiconductor manufacturers and human data workers who annotate and evaluate training inputs. A growing body of work has described this constellation of actors as an "AI supply chain" or "algorithmic supply chain" (Cobbe et al. 2023, Cen et al. 2023, Hopkins et al. 2025a).

Building on this body of work, we define the AI supply chain as a set of upstream-downstream relationships through which AI resources, products, and services flow. Formally, we say that an actor is upstream from another if it provides inputs that materially enable the latter's AI-related activities. An actor that receives them is downstream. Under this definition, the AI supply chain spans contributors as far upstream as human annotators and hardware manufacturers, and as far downstream as the end-users of AI-enabled products. Actors may occupy different roles across relationships (e.g., upstream in some relationships and downstream in others) and feedback loops (e.g., user data to improve upstream models) are common.

Scholars of political economy and industrial organization have long emphasized that supply chains are not neutral conduits of production, but structures where power, dependency, and governance are negotiated [Gereffi et al., 2005, Baldwin, 2012]. In traditional manufacturing, supply-chain structure determines vulnerability to shocks, bargaining power between firms, and the feasibility of regulation. Similar dynamics are emerging in AI, with recent concerns arising around the ability of general-purpose AI technologies to impact a diverse array of sectors and applications and, as a result, risks. Because AI development relies on resources, like compute, data, and technical talent, that are themselves highly concentrated and geopolitically contested [Muldoon et al., 2025], actors at the top of the AI supply chain who maintain control over resources are in a position to set prices and indirectly control downstream risks. Coordination and dealing-making at the top of the supply chain fuel worries that dominant AI actors are creating "kingmakers" [Gurun et al., 2025], fueling an unsustainable "AI bubble" [Caballero, 2026, Floridi, 2024], and neutralizing policymakers through regulatory capture [Wei et al., 2024].

Fundamentally, these issues can be traced to the simultaneous concentration of resources and control at the top of the supply chain, and a rapid expansion of downstream use and risk at the bottom. We argue that understanding and tackling such issues requires a systematic examination of the *relationships* between AI actors and not merely the AI stack. It requires investigating the flow of data, models, compute, labor, and capital through the AI supply chain.

Despite its importance, the AI supply chain remains poorly documented in a systematic way. Recent work has proposed conceptual frameworks and provided case studies highlighting the importance of supply chains for accountability (Cobbe et al. 2023), as well as empirical mappings of particular platforms (Laufer et al. 2025) or model families (Bommasani et al. 2023). Yet we lack an empirical, real-time view of the organizational relationships that span multiple sectors, firm types, and layers of AI production.

This paper addresses that gap. We build a fully automated pipeline for mapping the AI supply chain in real-time. The result is a dynamic, up-to-date network of upstream-downstream relationships derived from documents, such as news articles, press releases, company blogs, and SEC filings. We classify and analyze the structure of these relationships by developing a taxonomy of relationship types and use network analyses to study its dependencies and chokepoints. In the process, we gather a dataset of over 37,000 documents dating from late 2021 to early 2026, resulting in a network of relationships between nearly 3,000 actors. We provide several quantitative and qualitative analyses of the network, concluding with illustrative examples of the relationships found in the data. We conclude with implications of the AI supply chain on decision making, discussing how this methodology and data enable empirical analyses of relationships and trends that can inform governance efforts and support more targeted, robust policy.

Our contribution is both descriptive and normative. Descriptively, we provide a framework and methodology for constructing a real-time, dynamic view of the AI supply chain based on a combination of data sources. Normatively, we argue that AI governance efforts should be informed by explicit empirical models of the supply chain structure, rather than assumptions about the market structure. Ultimately, understanding the ecosystem creates better policy.

2 Background and Related Work

Our work draws on three partially overlapping strands of research: (1) conceptual treatments of AI and algorithmic supply chains; (2) empirical efforts to map AI ecosystems; and (3) governance and accountability frameworks that address distributed responsibility in complex sociotechnical systems. These literatures motivate both the need for and feasibility of empirical supply chain mapping for AI.

2.1 Conceptualizing Algorithmic and AI Supply Chains

Early discussions of AI supply chains emerged from concerns about distributed responsibility and the "many hands" problem in algorithmic systems (Slota et al. 2023). As systems become more complex and modular, harms and failures are often the outcome of the actions of multiple actors rather than a single developer or deployer (Cobbe et al. 2023). This complicates traditional approaches to accountability, which often assume more centralized control.

Cobbe et al. [2023] propose the concept of an *algorithmic supply chain* to describe flows “where several actors contribute towards the production, deployment, use, and functionality of AI technologies.” They emphasize that accountability debates must consider not only technical design choices, but also the political economy in which they are embedded, including general-purpose AI technology and regulatory arbitrage across jurisdictions. Widder and Nafus [2023], drawing on interviews with AI engineers, similarly describe a modular “supply chain” of AI components and services that pass through multiple hands before reaching end-users. They show how responsibility is displaced along the chain, as actors assume that harms originating elsewhere fall outside their scope of control.

Cen et al. [2023] provide a complementary perspective focused specifically on AI supply chains as a subject for policy analysis. They distinguish AI supply chains from AI *value chains*, emphasize the proliferation of specialized upstream providers, and highlight the implications for market power and regulation. Their work motivates the need for more empirical mapping to understand how these chains operate in practice. More recent work by Hopkins et al. [2025a] further develops the idea of AI supply chains as a network and provides a theoretical analysis of the effectiveness of upstream constraints/regulations enforce downstream outcomes and the accuracy of disclosures passed down AI supply chains.

Collectively, this literature underscores that AI systems are produced through interdependent networks of actors and that the structure of these networks bears directly on governance questions. Our project builds on these conceptual foundations but shifts the focus from illustrative examples to a systematic, longitudinal, empirical mapping that captures AI supply-chain relationships throughout the ecosystem.

2.2 Empirical Mapping of AI Ecosystems

In parallel with conceptual work, a growing literature has sought to empirically document the structure of AI ecosystems. Bommasani et al. [2023] introduce “ecosystem graphs” for foundation models, constructing graphs that connect models, datasets, tasks, and organizations in order to surface dependencies in foundation model development and use. Laufer et al. [2025] analyze an ecosystem of over nearly million models on the Hugging Face platform by examining interdependencies between models and components at scale. Other work has mapped particular subsectors of the AI ecosystem (e.g., open-source model ecosystems, domain-specific applications) or stakeholder roles in AI supply chains (Hopkins et al. 2025b).

These efforts demonstrate that ecosystem-level representations can provide valuable insight into how AI systems are built and used. However, they are often constrained to a single platform, model family, or snapshot in time. Most rely on static crawls or platform-specific metadata, and therefore capture only a subset of the relationships that structure AI production and deployment.

Our work complements these approaches by constructing a multi-source, continuously-updated dataset that extends beyond a single platform or model class and tracks these relationships dynamically as they appear and evolve. Rather than focusing on technical dependencies alone, we look at organizational relationships that shape access to AI capabilities and constrain downstream behavior. This broader scope enables analysis of how dependencies evolve over time and across sectors.

2.3 Governance, Accountability, and Policy

A third strand of work emphasizes the governance implications of the complex, distributed, modular AI production. Cobbe et al. [2023] discuss obstacles to accountability arising from the distributed responsibility and service-based models of AI, and opacity created by the so-called “accountability horizon,” or the limited visibility actors have into parts of the AI supply chain beyond their contractual partners. Widder and Nafus [2023] show how current responsible AI guidelines, which often assume centralized control, do not align with engineers’ actual roles in AI supply chains.

Legal scholars have begun to explore how supply chain structure complicates liability and compliance with copyright. For example, Lee et al. [2023] analyze generative-AI supply chains to show how upstream training data, model development, and downstream uses create complex chains of potential liability. Policy documents and frameworks—such as the NIST AI Risk Management Framework and the EU AI Act—also increasingly reference supply-chain relationships, emphasizing the shared responsibility and the need for transparency about upstream components and services [NIST AI, 2023].

Recent work has also pointed to the disconnect between proposals to regulate AI and the technical and institutional feasibility of those proposals. Guha et al. [2024] argue that standard regulatory approaches—licensing, registration, disclosure, and auditing—are misaligned with the technical and organizational realities of AI. Our contribution is to connect these governance debates to empirical evidence. Rather than assuming a particular supply chain structure, we empirically document it and show how it can shape the feasibility and efficacy of governance interventions.

3 Layers of the AI Supply Chain

The AI supply chain can be understood as a network of layered, partially modular system in which distinct resources, products, services, capital, and even labor flow betweenx organizations. Conceptually distinguishing these layers clarifies how dependencies are arranged and how they arise, where power is concentrated, and why governance challenges differ across the ecosystem. While boundaries between layers are often porous and blurred by some degree of vertical integration, analytically separating the layers allows for more leverage on important policy questions.

Compute Infrastructure. Compute infrastructure functions as a general-purpose input to AI development and deployment. Upstream providers offer cloud platforms, specialized hardware, and services that downstream actors rely on for both training and inference. Dependence on these providers introduces technical, economic, and strategic constraints that shape the entire supply chain. Scholars have increasingly identified compute as a central bottleneck in AI development [Sastry et al., 2024]. Unlike data or talent, compute is capital-intensive [McKinsey & Company, 2025] and subject to changing export controls [Harithas, 2024].

Data. Data constitutes another core layer of the AI supply chain, encompassing labeled training datasets, fine-tuning corpora, evaluation benchmarks, and feedback from deployment. Data flows through the supply chain in different forms: proprietary datasets, licensed third-party data, synthetic data generated by models, and user interactions that feed back into systems.

Unlike compute, data markets are fragmented and domain-specific. Some data inputs, such as large web crawls, are widely accessible, while others, like medical records or proprietary enterprise data, are tightly controlled. This heterogeneity requires application or domain-specific governance structures to safeguard copyright and privacy. Recent legal scholarship emphasizes that data supply chains introduce layered liability and governance challenges, particularly for generative AI (Lee et al. 2023; Samuelson 2023). Empirical mapping is essential to understanding how data dependencies interact with model deployment at scale.

Chips and Hardware. Underlying compute infrastructure requires specialized chips and manufacturing capacity (Fransoo et al. 2025). Dependence on advanced semiconductors introduces additional chokepoints that are, in turn, shaped by access to capital, supply constraints, and international trade policies. While often treated as separate from AI governance discussions, hardware dependencies play a critical role in shaping who can develop and deploy large-scale models.

Talent. Human labor occupies upstream and downstream positions in the AI supply chain. Upstream, AI systems rely on highly specialized technical talent (e.g., research scientists, engineers, and safety specialists). They also rely on less visible forms of labor, including data annotators and content moderators (Gonzalez-Cabello et al. 2024).

This layer raises distinct governance concerns. Prior work documents difficult working conditions faced by data workers (Gray and Suri 2019, Cant et al. 2024) and asymmetric power relations (Gonzalez-Cabello et al. 2024). At the same time, competition for technical talent concentrates power within a small number of firms and regions. Although labor dependencies are difficult to observe due to limited disclosure, supply chain analysis can reveal patterns of outsourcing and concentration that bear on equity, accountability, and vulnerability.

4 Mapping the Supply Chain

In this section, we describe the procedure for tracking and extracting relationships in the AI industry. We use publicly available information, combining SEC filings with news, press releases, and company blogs. Our pipeline is fully automated and runs, with extensive error handling and retry logic, regularly to create an up-to-date archive of AI supply chain relationships. We note that although our pipeline seeds its current searches with the names of major model providers, it can be used to provide comprehensive coverage of the AI industry; all that is needed is to update the search queries used to seed the searches. Please note that we plan to release the data and code for this project upon publication.

4.1 Data Sources

To construct an empirical representation of the AI supply chain, we rely on publicly available documents that describe relationships between model providers and other organizations with which they contract. We summarize our two primary data sources and explain why they provide complementary views of the ecosystem.

SEC Filings. Publicly traded companies in the United States must disclose material information about their business operations through periodic filings with the U.S. Securities and Exchange Commission (SEC), including Form 10-K (annual reports), Form 10-Q (quarterly reports), and Form 8-K (current reports or significant events). These filings provide audited descriptions of major customers, suppliers, partnerships, and risk factors (Marin et al. 2025, Kirk et al. 2025). SEC filings offer several advantages. First, the filings are reliable and structured. Filings are audited and subject to legal penalties for misrepresentation, which encourages accurate disclosure of material relationships. Second, the filings use a standardized format. While narrative language varies, the overall structure of forms like 10-K and 10-Q is relatively consistent, facilitating automated parsing and section-level retrieval. Third, because disclosures focus on material relationships, the relationships we uncover are likely to reflect economically significant dependencies, including potential chokepoints.

However, SEC filings also exhibit important limitations. First, many firms are privately held and therefore absent from SEC filings. The filings we collect primarily describe how publicly traded firms interact with AI providers and partners. Second, firms may not disclose all relationships, especially if they are not deemed material, or when doing so may reveal sensitive strategic information. For those reasons, SEC filings provide a high-quality but partial view of the AI supply chain, heavily weighted toward public firms and material relationships.

News, Press Releases, and Company Blogs via Google Search. To complement SEC filings, we use Google Search to assemble a broad corpus of publicly accessible documents describing AI-related relationships. Our searches primarily focus on news articles, press releases, and company blogs. Compared to SEC filings, these sources offer several benefits. First, they offer broader coverage. They include both public and private firms, as well as startups and niche providers. Second, partnerships, product launches, strategic partnerships, funding announcements, and other relationships are often reported in the press or on company blogs well before they appear in regulatory filings. Third, they offer rich qualitative descriptions. Press releases and articles often describe the nature and purpose of a relationship (e.g., "co-developing a domain-specific LLM for healthcare") in more detail than filings.

However, they also introduce trade-offs. For one, press releases and marketing materials emphasize positive narratives and may gloss over risks or other issues. Also, unlike SEC filings, these documents are not subject to the same legal and auditing constraints. By combining SEC filings with search-based news and company disclosures, we aim to leverage the strengths of both: the reliability of audited filings and the coverage and recency of the broader web.

4.2 Search Pipeline

Our earliest searches begin November 2021. We seed our search results with model providers. That is, presuming that much of the AI ecosystem is built around providing stronger AI models, we focus on key foundation model providers and their models. The full list is given in Section A.1 and Section A.2. We search from November 2021 onwards. The procedure is described in detail below.

SEC EDGAR Search. The SEC search pipeline uses the config file given in Section A.1 generate search queries for the SEC's search database EDGAR. For each model provider specified in the config file, the pipeline performs targeted searches across multiple SEC filing types (10-K, 10-Q, 8-K, proxy statements, registration statements, etc.) and time windows (from the last search date to the current date). We use Selenium WebDriver to interact with the EDGAR interface. We handcraft several search queries for each model provider, including combinations of the model provider's name and their flagship model names. After executing a search, the pipeline paginates through search results and extracts the filings' metadata including document links, filing dates, form types, and company information. Results are saved incrementally, with status tracking to resume interrupted searches and avoid duplicate processing.

After identifying relevant filings, the pipeline enters a second stage where it visits each filing document URL and extracts text excerpts around mentions of keywords in the search query that led to the filing. For each keyword mention, we extract an excerpt 1250 characters before and after to provide context. The pipeline handles various document formats, manages SEC website popups and rate limiting, and includes robust retry logic for network issues and page loading failures. These text excerpts are saved for downstream relationship extraction.

Google SerpApi Search. The search pipeline uses a specific configuration (included in Section A.2). For each organization, the pipeline performs three types of searches using SerpAPI: (1) Google News searches for the organization name combined "AI partnership"; (2) site-specific searches to ensure coverage of the organization's official blog/news URL; and (3) searches on specified news sites (specifically, TechCrunch and VentureBeat) to achieve wide and consistent coverage. We narrow search to a time window of 1 month, proceeding systematically from the last search date to the current date, with status tracking to avoid duplicate searches. There is some expected variation in search results by location; we use San Francisco as our default location. Although providing multiple locations could help, due to the volume of search queries, we limited our search to a single location.

After search results are collected, the pipeline proceeds to a second stage where it scrapes the full content of each article link found. The site scraping step uses Selenium WebDriver to visit each URL, extract article titles and body text from the main content area, and store the full page source. We use extensive error handling, fallback mechanisms, and retry logic to handle different content formats and network issues.

4.3 Relationship Extraction

The final stage of the SEC pipeline processes the extracted text excerpts to identify AI relationships. The pipeline processes documents from two sources (SEC filings excerpts and Google Search article content) through two LLM stages. First, a relevance filtering stage uses batch LLM queries with GPT-4o-mini at temperature 0.01 to determine whether each document contains relationships between actors and whether those relationships are AI-related (involving data, compute, capital, products, services, infrastructure, or talent). Documents are chunked if they exceed token limits (50,000 tokens), and the system tracks how many times each document has been processed to ensure sufficient coverage (default target of 2 runs per document). Documents that pass relevance filtering (using a relevance threshold of 0.51) then proceed to the relationship extraction stage, where the GPT-4o-mini is prompted to extract pairwise relationships as tuples of two actor names with descriptive text. The extracted relationships are validated to ensure both actors are explicitly mentioned in the description. The pipeline handles retries for failed LLM calls (up to 4 max trials), deduplicates articles by link, and processes documents in parallel batches, before saving the results to the database. We acknowledge that using a language model for this task may introduce errors. To counter this possibility, we manually review the results regularly, and we submit one LLM query per document to ensure minimal contamination.

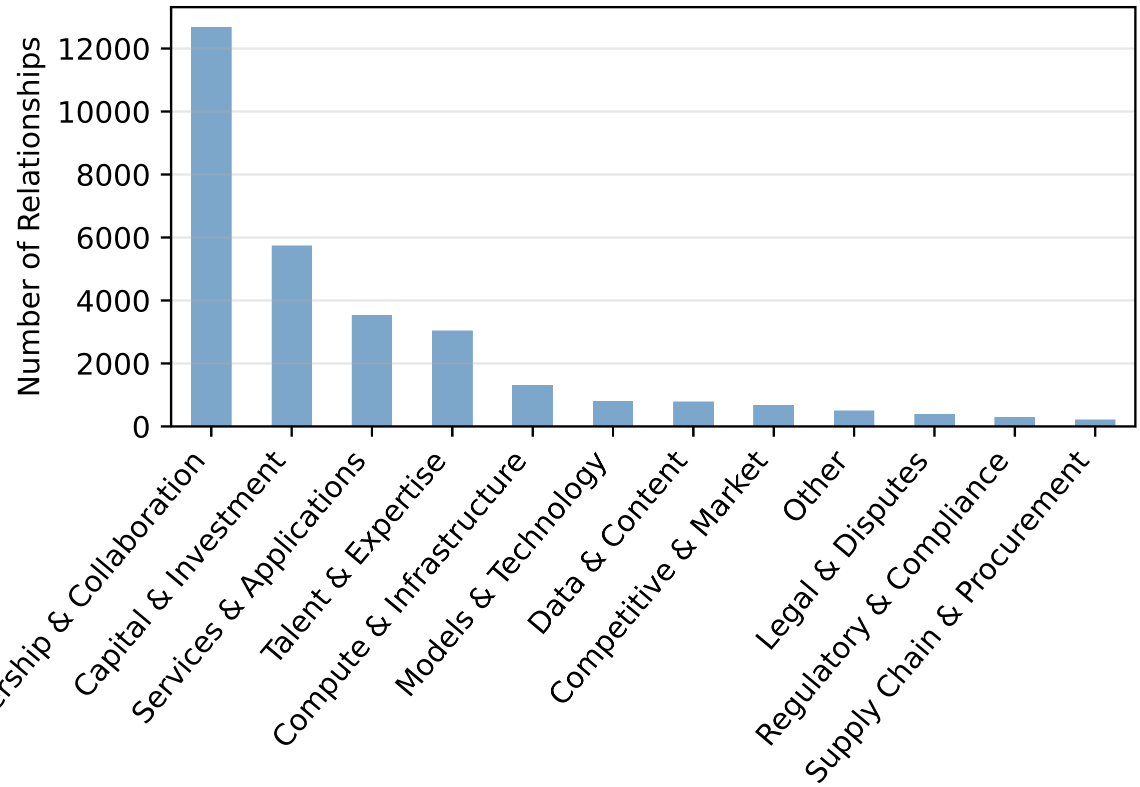

Figure 1: Distribution of relationship categories. Shows the count of unique relationships (after deduplication) for each category.

5 Analysis

In this section, we provide brief illustrations of analyses that can be performed with our tool and dataset.

5.1 Analysis Methodology

The results throughout this section are obtained using the same procedure. We first load the raw relationship data obtained using the procedure described in Section 4. Then, we classify and categorize the relationships based on a taxonomy developed following an in-depth review of the data. The full taxonomy is included in Section B.1.1. Using this taxonomy, each relationships is assigned to (i) a category, (ii) a subcategory, (iii) a primary directional category, and (iv) an optional secondary directional category. We do so using OpenAI's GPT-4o-mini. We acknowledge that using a language model for this task may introduce errors. When applicable, we comment on how this may affect the results. For specific relationships that we discuss, we manually review and check the results.

We then de-duplicate actors by creating a set of parent organizations that subsume many actors (e.g., "Google" for "Google AI" and "DeepMind"). Next, unless indicated otherwise, we create an adjacency matrix where entry (i, j) is 1 if there is a (potentially directed) edge from actor i to actor j, and 0 otherwise. Finally, we filter out actors with total (in plus out) degree ≤ 2.

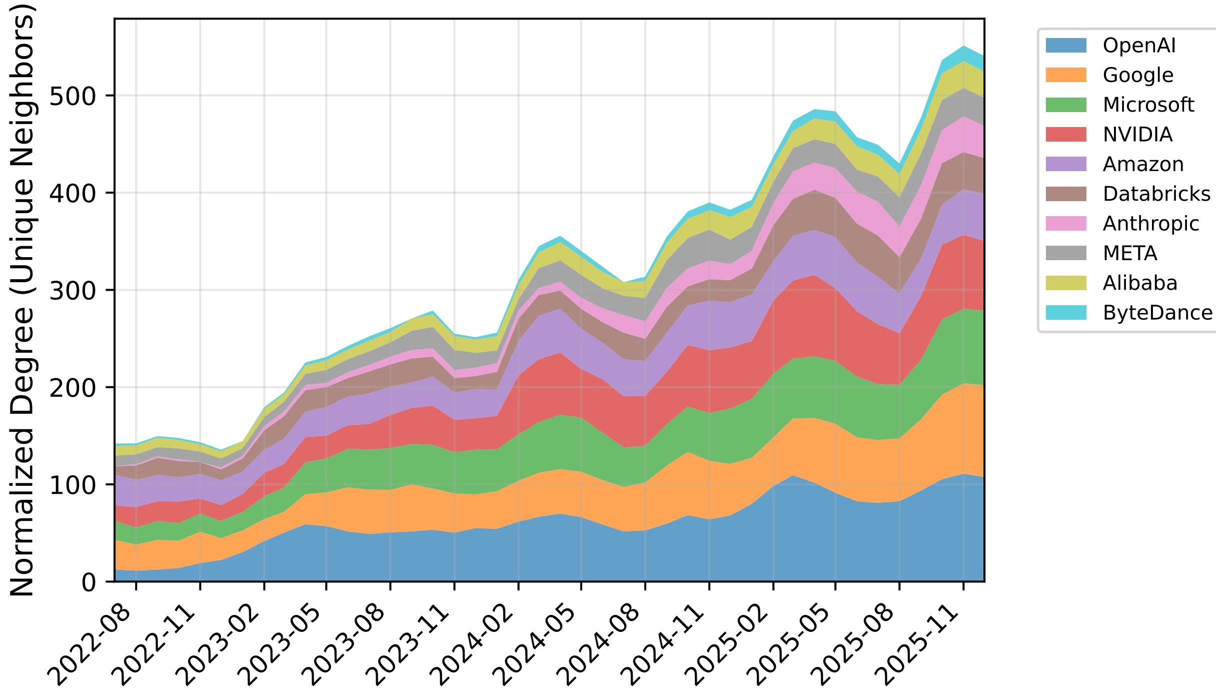

Figure 2: Normalized degree of top 10 actors (by degree across all time) from July 2022 to January 2026

5.2 Quantitative Analyses

Summary Statistics. Our dataset contains 2,775 SEC filings and 34,296 other documents, for a total of 37,071 documents (e.g., news articles and company blogs) as of January 27, 2026. Through the process of deduplicating and filtering, we narrow our attention to 2,930 actors and 13,843 edges (where each edge may correspond to multiple documents and relationships between the same two actors). The network density is 0.00161, meaning that on average, each actor is connected to 4.72 other actors. There are 377 strongly connected components (SCCs), the largest SCC size is 2,506, and the diameter (the longest shortest path within the largest SCC) is 10. Detailed network statistics, including computation procedures and metric definitions, are provided in Section B.1.

Relationship Type. Using our classification of relationships, we compute the frequency of each relationship type, as shown in Figure 1. While this plot shows the frequency of each relationship category, analogous histograms for subcategories and primary directional categories are given in Figure 4 and Figure 5. As show in Figure 1, most relationships are informal, falling under Partnership & Collaboration. The next most common relationships in our dataset are related to Capital & Investment, Services & Applications, Talent & Expertise, and Models & Technology. Note again that these categories are obtained using a language model; while they may contain some errors, we believe the broad trends hold.

Actor Centrality. We now examine the most central actors in the AI supply chain that we extract. Figure 2 shows the top 10 actors by normalized degree (how many unique neighbors they have) across time. Table 4 shows the top 20 actors by degree in the overall network (first column) and by betweenness centrality in the overall network (second column). In the third column, we illustrate how the actor rankings may change based on the relationship category, giving the degree based on the Models & Technology category. To complement Figure 2, we further provide Table 7 shows the percentage growth in degree and betweenness centrality (computed for the whole network rather than for a relationship category) from 2022 to 2025. Additional rankings by category, time series analyses, and detailed centrality values are provided in Section B.1.7 and

Table 1: Top 20 Actors by Centrality Measures (Overall Network)

Rank	Degree	Betweenness	Degree (Models)
1	Google	OpenAI	OpenAI
2	OpenAI	Google	HuggingFace
3	NVIDIA	NVIDIA	Meta
4	Microsoft	Microsoft	Google
5	Amazon	Amazon	DeepSeek
6	Databricks	Databricks	NVIDIA
7	Meta	Meta	Microsoft
8	IBM	Alibaba	Anthropic
9	Anthropic	IBM	Alibaba
10	Alibaba	Anthropic	Stability AI
11	Snowflake	Snowflake	Apple
12	Huawei	Huawei	Mistral AI
13	HuggingFace	HuggingFace	IBM
14	DeepSeek	DeepSeek	Amazon
15	Cohere	ByteDance	xAI
16	ByteDance	Baidu	Databricks
17	Baidu	Cohere	Alphabet
18	xAI	Mistral AI	Snowflake
19	Mistral AI	xAI	Alibaba Group
20	Apple	Stability AI	Huawei

Table 2: Growth Metrics for Top 10 Actors (2022-2025)

Actor	Deg. Growth%	Between. Growth%
OpenAI	896.30	113.79
Huawei	417.65	62.94
xAI	104.00	65.51
Cohere	180.65	125.35
EleutherAI	162.50	22.87
Anthropic	369.39	163.11
Alibaba	413.33	101.61
Google	272.00	-14.01
HuggingFace	291.30	-9.11
Microsoft	266.67	-16.89

Section B.1.6.

Our findings show that Google, OpenAI, NVIDIA, Microsoft, and other expected organizations are among the most central in the AI supply chain. Further, OpenAI, Huawei, and xAI, among others experience the greatest growth in centrality from 2022 to 2025, with actors such as Amazon and Meta falling outside the top 10. Note that our findings are reflective of our search criteria (which used model providers to seed our search) and, due to our data sources, are biased toward US entities as well as relationships that are publicly available (e.g., publicized or for which a declaration is legally mandated).

Actor Importance and Network Resilience. We also characterize the importance of a node by examining the impact of removing it from the network. This node importance analysis examines the percentage reduction in the size of the network's largest connected component (LCC) when the node is removed using the equation:

$$\text{Reduction } \% = \frac{\text{LCC}_{\text{original}} - \text{LCC}_{\text{removed}}}{\text{LCC}_{\text{original}}} \times 100 \tag{1}$$

Figure 3 shows the node importance of the top 20 nodes (that is, we first find the 100 highest-degree nodes, then compute their node importance, plotting the top 20). Google and OpenAI are clearly the two most

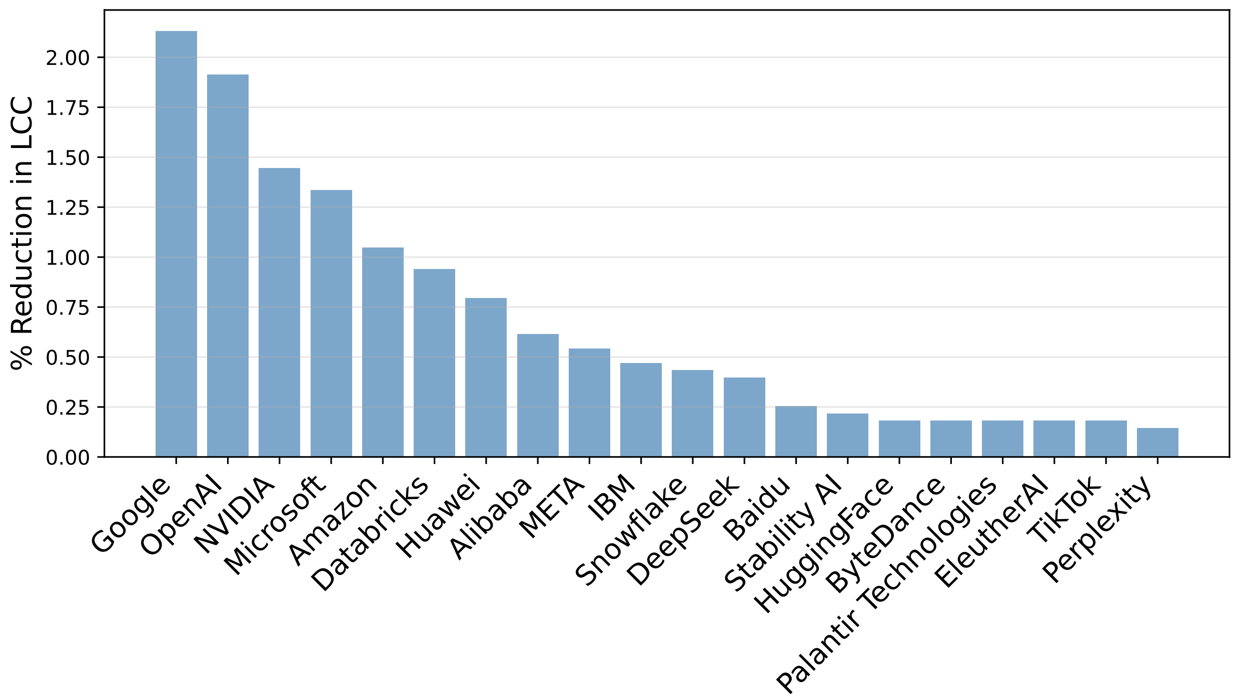

Figure 3: The plot shows the node importance: the percentage reduction in largest connected component size when each actor is removed from the network.

important nodes, with a dropoff in node importance after them. Thereafter, the actors between NVIDIA and Stability AI have node importance that degrades gracefully, after which most actors have relatively even node importance, resulting in a long tail. At the same time, if we split these actors by what they contribute, the dropoff is more pronounced. From the perspective of GPU and compute providers, NVIDIA dominates. Interestingly, Anthropic does not appear in the top 20, despite being a major player in the AI ecosystem.

5.3 Close Relationships In Our Data

Talent Exchange. We find that there is significant talent exchange among actors. Our dataset enables rapid construction of talent exchange timelines, revealing extensive cross-company movement of researchers and executives. For example, it documents how Meta hired Robert Fergus, a former research director at Google DeepMind, to lead its Fundamental AI Research lab [Zeff, 2025b]; Anthropic hired Tom Turvey, the former head of partnerships for Google Books; and Mustafa Suleyman, co-founder of DeepMind (now owned by Google), began leading a Microsoft team created in early 2024 [Mann, 2024]. It tracks how OpenAI hired numerous former Google and Facebook employees in 2023 and 2024 [Sharma, 2023, Dube, 2024], including Caitlin Kalinowski, the former head of Meta's augmented reality glasses efforts in late 2024 [Zeff, 2024]. The dataset describes how OpenAI also hired Chan Park, former senior director of congressional affairs at Microsoft, to head its U.S. and Canada partnerships in 2024 [Wiggers, 2024]; Tim Brooks, a co-lead on OpenAI's video generator Sora, left OpenAI to lead a new AI team at Google DeepMind in January 2025 [Wiggers, 2025]; and Meta has hired a highly influential OpenAI researcher, Trapit Bansal, to work on its AI reasoning models under the company's new AI superintelligence unit [Zeff, 2025a]. Moreso than simply tracking these relationships, our dataset shows how talent flows, often between top companies, illustrating the importance of talent to AI development.

Close Financial Relationships. Our dataset reveals close financial relationships, sometimes uncovering intertwining investments. For example, NVIDIA made significant investments in Databricks [Novet and

Levy, 2023]. NVIDIA also has a close alliance with OpenAI, having committed to a $100 billion investment in exchange for a sizable non-voting stake in the company (that some estimate could amount to 25% of OpenAI) [Campbell et al., 2025]. At the same time, OpenAI has a $100 million deal with Databricks [Columbus, 2025]. Such close relationships are not uncommond. In addition to NVIDIA's close relationship with OpenAI, NVIDIA also invested in CoreWeave, an AI cloud computing startup, and now owns about 7% of the company [Keegan, 2025]. Yet Coreweave and OpenAI are also connected, striking a five-year, $11.9 billion deal with OpenAI holding a $350 million equity stake in the company [Loconsolo et al., 2025].

The are also organizations that do not have direct investment relationships in our dataset, but are connected through indirect financial mechanisms, such as Microsoft and NVIDIA. Our dataset shows that NVIDIA serves as a principal partner for a $30 billion initiative launched by Microsoft and BlackRock designed to mobilize private capital for buildingAI data centers and energy infrastructure Two Seas Capital LP [2025]. They also jointly invested $675 million Series in Figure, signaling that they are strategically in alignment Reuters [2024]. Perhaps most strikingly, Microsoft and NVIDIA also led a collaborative $1.3 billion funding round for Inflection AI, illustrating their roles as "kingmakers" Wire [2023].

Data Partnerships. Our data further illustrates data partnerships. For example, both Google and OpenAI have entered into partnerships with Reddit and The Associated Press [Paul, 2024, David, 2024, Srinivasaragavan, 2025, O'Brien, 2023]. These partnerships suggests that AI developers are competing for the same data sources, and that the data sources are becoming increasingly valuable. There is even some suggestive evidence that OpenAI's ChatGPT Plus version utilizes Google's search index for its paid services even though OpenAI's official documentation states a partnership with Microsoft's Bing [Rijo, 2025].

Our data surfaces other interesting data-related relationships of note. Google is seeking deals with news publishers to obtain training data for its models, actively approaching content providers Ostwal [2025]. Meta is leaning into its competitive advantage by leveraging user data from Facebook and Instagram Silberling [2025]. As the race for data intensifies, companies that have in-house user data may have a significant competitive advantage.

6 Implications for AI Policy and Governance

The AI supply chain is not merely a descriptive artifact, it is the ecosystem in which accountability and power are negotiated. Studying the AI supply chain can yield fruitful insights for market governance and regulation. For instance, policymakers increasingly seek to regulate AI systems through disclosures (El Ali et al. 2024), licensing (Contractor et al. 2022), audits (Ojewale et al. 2025, Raji et al. 2022), and liability rules (Vladeck 2014). Yet these interventions implicitly assume that components of the AI supply chain are easy to trace and capabilities may be attributable to specific model inputs. In practice, AI development and deployment is distributed across many actors. Thus, governance proposals misunderstand the problem (Guha et al. 2024) and risk targeting the wrong actors, imposing obligations that are infeasible in the AI supply chain or overlooking upstream bottlenecks where intervention would be most effective.

Below, we outline how supply chain mapping can inform governance to promote accountability, transparency, competition, and resilience, using examples in our dataset.

Accountability and Responsibility. As Cobbe et al. [2023] argue, algorithmic accountability debates must reckon with supply chains because harms often arise from interactions between different layers of the supply chain. Our supply-chain graph provides a structural lens on this problem. By mapping which actors are upstream of which systems, we can identify where responsibility for particular risks lies under different paradigms. For example, several dominant AI players are entering into healthcare partnerships, including Microsoft Health Catalyst, Inc. [2024] and OpenAI OpenAI [2024]. Other actors in the healthcare sector like the Centers for Disease Control and Prevention (CDC) [Centers for Disease Control and Prevention, 2025], HealthLynked [HealthLynked Corp., 2025], and Zepp Health [Zepp Health Corporation, 2025] are integrating AI products and services into their workflows. In both cases, there is increasing reliance on AI in healthcare and medicine. Empirically tracking the AI supply chain can illuminate these dependencies clearly. Moreover, by examining repeated patterns of relationships (e.g., model providers licensing to multiple downstream sectors with similar risk profiles), we can identify actors whose decisions have systematic consequences across sectors.

This does not resolve the normative question of who *should* be held responsible; that depends on the legal, ethical, and political judgments of relevant stakeholders. But it provides the empirical infrastructure necessary to make those debates more concrete and to design accountability regimes that reflect the actual patterns of use and deployment.

Disclosure and Transparency. Several emerging regulatory frameworks call for enhanced transparency about AI supply chains, including mandates for providers to disclose training data sources, performance metrics, and downstream deployments (Cobbe et al. 2023). In addition, there are concerns around disclosures between partners, e.g., what should be disclosed in contractual agreements.

Our empirical analysis finds that public data about AI relationships are fragmented and vary in scope. For example, in early 2024, Reddit and Google announced a $60 million deal to license Reddit's data to Google to help train its AI models [CBS News and Associated Press, 2024]. However, this deal is not reported in SEC annual or quarterly reports by either Google or Reddit and thus becomes more difficult for regulators to track. Moreover, Google's press release about the deal conveniently sidesteps the significance of obtaining Reddit data to train its models, stating that "this partnership will facilitate more content-forward displays of Reddit information that will make our products more helpful for users and make it easier to participate in Reddit communities and conversations" Patel [2024]. Recent disclosures by Reddit state that they view Google as a competitor who may reduce traffic to their site; this discrepancy indicates a potential inconsistency between what happens behind closed doors and information revealed to the public.

A policy agenda focused on AI supply chains could create standardized reporting requirements, analogous to financial disclosures, to make supply chain monitoring an object of routine governance. We suggest that such measures should consider the *scope of disclosure* (what relationships and how far upstream/downstream obligations should extend); *standardization* (how to structure disclosures so they are machine-readable and amenable to relational analysis); and *interoperability and interpretability* (how to link disclosures across actors so that regulators, auditors, and researchers can construct supply-chain graphs similar to ours but with other sources of data).

Competition, Market Power, and Concentration. Concerns about concentration and market power in AI often focus on a small number of highly visible actors (e.g., foundation model developers or cloud providers). While these actors are clearly important, a supply-chain perspective reveals a more complete picture.

Supply chain graphs can help illuminate these issues. First, they can identify where bottlenecks and dependencies concentrate (e.g., in computation, specialized hardware, proprietary datasets, or particular models). They can also help distinguish between horizontal concentration (i.e., dominance within a layer, such as cloud compute) and vertical integration (i.e., dominance across multiple layers, such as compute, models, and platforms). Finally, supply chain graphs can help highlight emergent dependencies on specialized upstream providers that may not be widely recognized. We provide both quantitative and qualitative analyses market concentration and dealmaking patterns in Section 5.

Systemic Risk and Resilience. Supply-chain mapping is also central to understanding systemic risk—the risk that failures in one part of the AI ecosystem propagate to others. For example, reliance on a small number of upstream compute providers or chip manufacturers may create coordinated vulnerabilities, including failures in response to cyberattacks or geopolitical shocks. In addition, reliance on a small number of models may lead to "algorithmic monocultures" (Kleinberg and Raghavan 2021) that create shared flaws downstream. Key intermediaries may become targets or present opportunities for safety interventions. Network analyses of the kind we describe in Section 5 can be used to design simulations that test the ability of nodes or edges to withstand failures (Ivanov 2018). These stress tests could inform policies akin to those used in financial regulation, like reliance requirements, diversification incentives, or contingency planning for critical providers.

Future work. There are various aspects of AI policy and governance to further explore, including global competitive dynamics and geopolitical risks (e.g., where do components in the AI supply chain reside) and the presence of an AI bubble. One could even use our data to study the effects of government policy on AI industry growth, adoption, and risk mitigation, leveraging policies as natural experiments from which one could estimate its causal effects.

7 Conclusion and Limitations

This paper has outlined a framework and an empirical project for mapping the AI supply chain as a dynamic network of upstream-downstream relationships. As AI systems become increasingly embedded across the economy, both their risks and benefits derive from the complex economic relationships built around foundation models.

By systematically mapping these dependencies, we provide a foundation for governance that is both more realistic and more effective. Through our approach to combine SEC filings with news, press releases, and company blogs, and apply natural language processing and network analysis, we can begin to see the contours of an ecosystem that has so far been largely obscured from public view.

Our approach has several important limitations and implications that point toward future research directions.

First, since our dataset is derived from public documents, it inevitably omits undisclosed or confidential relationships, and may over-represent high-profile firms and deals. Future work could explore ways to integrate other sources of data or other ways to identify relationships between firms—for instance, including co-authorship networks in industry (Liang 2024).

Second, to be able to get traction on this problem, we identify relationships that run through major model developers. This is obviously only one segment of the larger supply chain. Future work could build to more comprehensive coverage across all segments of the AI supply chain.

More broadly, this work contributes to a growing recognition that governing AI requires governing the ecosystem not just model artifacts. We argue that this kind of empirical mapping should inform AI governance. And it can provide a common empirical ground for scholars across disciplines—from law to computer science to economics—to reason about how the AI ecosystem is organized and how it ought to be governed.

Impact Statement

This paper has outlined a framework and an empirical project for mapping the AI supply chain as a dynamic network of upstream-downstream relationships. As such, we do not develop a new technology and thus do not envision direct harm due to the use of our tool.

At the same time, our approach has several important limitations. First, since our dataset is derived from public documents, it inevitably omits undisclosed or confidential relationships, and may over-represent high-profile firms and deals. Second, to be able to get traction on this problem, we identify relationships that run through major model developers. This is obviously only one segment of the larger supply chain, and our results provide an incomplete picture of the AI supply chain. Third, we use LLMs as a part of our analysis, which introduces risk due to other errors such as hallucination. To mitigate this risk, we use public data sources so that results are reproducible and verifiable (our tool, which we plan to release upon publication, always includes the original data source to allow for independent verification).

References

Richard E Baldwin. Global supply chains: why they emerged, why they matter, and where they are going. 2012.

Rishi Bommasani, Dilara Soylu, Thomas I. Liao, Kathleen A. Creel, and Percy Liang. Ecosystem graphs: The social footprint of foundation models, 2023. URL https://arxiv.org/abs/2303.15772.

Ricardo J Caballero. Speculative growth and the ai "bubble". Technical report, National Bureau of Economic Research, 2026.

Charlie Campbell, Andrew R. Chow, and Billy Perrigo. The architects of ai are time's 2025 person of the year, dec 2025. URL https://time.com/7339685/person-of-the-year-2025-ai-architects/.

Callum Cant, James Muldoon, and Mark Graham. *Feeding the machine: The hidden human labor powering AI*. Bloomsbury Publishing USA, 2024.

CBS News and Associated Press. Google strikes $60 million deal with reddit, allowing search giant to train ai models on human posts. CBS News, February 2024. URL https://www.cbsnews.com/news/google-reddit-60-million-deal-ai-training/. Web article.

Sarah Huiyi Cen, Aspen Hopkins, Andrew Ilyas, Aleksander Madry, Isabella Struckman, and Luis Videgaray Caso. Ai supply chains. 2023.

Centers for Disease Control and Prevention. CDC's Vision for Using Artificial Intelligence in Public Health. Web page, August 2025. URL https://www.cdc.gov/data-modernization/php/ai/cdcs-vision-for-use-of-artificial-intelligence-in-public-health.html.

Jennifer Cobbe, Michael Veale, and Jatinder Singh. Understanding accountability in algorithmic supply chains. In *2023 ACM Conference on Fairness Accountability and Transparency*, FAccT '23, page 1186–1197. ACM, June 2023. doi: 10.1145/3593013.3594073. URL http://dx.doi.org/10.1145/3593013.3594073.

Louis Columbus. How runtime attacks turn profitable ai into budget black holes, jun 2025. URL https://venturebeat.com/security/how-runtime-attacks-turn-profitable-ai-into-budget-black-holes.

Danish Contractor, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. Behavioral use licensing for responsible ai. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 778–788, 2022.

Emilia David. Openai strikes reddit deal to train its ai on your posts, may 2024. URL https://www.theverge.com/2024/5/16/24158529/reddit-openai-chatgpt-api-access-advertising.

Mudit Dube. Openai poaches 3 ai experts from rival google deepmind, dec 2024. URL https://www.newsbytesapp.com/news/business/openai-hires-ai-engineers-from-google-deepmind-appoints-first-cmo/story.

Abdallah El Ali, Karthikeya Puttur Venkatraj, Sophie Morosoli, Laurens Naudts, Natali Helberger, and Pablo Cesar. Transparent ai disclosure obligations: Who, what, when, where, why, how. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2024.

Luciano Floridi. Why the ai hype is another tech bubble. *Philosophy & Technology*, 37(4):128, 2024.

Jan C Fransoo, Robert Peels, and Maximiliano Udenio. Navigating supply chain dynamics for sustained ai growth. In *AI in Supply Chains: Perspectives from Global Thought Leaders*, pages 37–54. Springer, 2025.

Gary Gereffi, John Humphrey, and Timothy Sturgeon. The governance of global value chains. *Review of international political economy*, 12(1):78–104, 2005.

Martin Gonzalez-Cabello, Auyon Siddiq, Charles J Corbett, and Catherine Hu. Fairness in crowdwork: Making the human ai supply chain more humane. *Business Horizons*, 2024.

Mary L Gray and Siddharth Suri. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business, 2019.

Neel Guha, Christie M Lawrence, Lindsey A Gailmard, Kit T Rodolfa, Faiz Surani, Rishi Bommasani, Inioluwa Deborah Raji, Mariano-Florentino Cuéllar, Colleen Honigsberg, Percy Liang, et al. Ai regulation has its own alignment problem: The technical and institutional feasibility of disclosure, registration, licensing, and auditing. *George Washington Law Review*, 92(6), 2024.

Umit G Gurun, Zheng Liu, and Steven Chong Xiao. Gatekeepers and kingmakers: Do superplatforms distort digital competition? 2025.

Barath Harithas. Securing the agi laurel: Export controls, the compute gap, and china's counterstrategy. *CSIS Analysis*, 2024. URL https://www.csis.org/analysis/securing-agi-laurel-export-controls-compute-gap-and-chinas-counterstrategy.

Health Catalyst, Inc. Form 10-k (annual report) for the fiscal year ended december 31, 2023. U.S. Securities and Exchange Commission (SEC) EDGAR filing, February 2024. URL https://www.sec.gov/Archives/edgar/data/1636422/000163642224000016/hcat-20231231.htm. Commission File No. 001-38993. Accession No. 0001636422-24-000016.

HealthLynked Corp. Form 10-k (annual report) for the fiscal year ended december 31, 2024. U.S. Securities and Exchange Commission (SEC) EDGAR filing, March 2025. URL https://www.sec.gov/Archives/edgar/data/1680139/000121390025026262/ea0235172-10k_health.htm. CIK 0001680139. Commission File No. 000-55768. Accession No. 0001213900-25-026262.

Aspen Hopkins, Sarah H. Cen, Isabella Struckman, Andrew Ilyas, Luis Videgaray, and Aleksander Mądry. Ai supply chains: An emerging ecosystem of ai actors, products, and services. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2):1266–1277, October 2025a. ISSN 3065-8365. doi: 10.1609/aies.v8i2.36628. URL http://dx.doi.org/10.1609/aies.v8i2.36628.

Aspen Hopkins, Isabella Struckman, Kevin Klyman, and Susan S Silbey. Recourse, repair, reparation, & prevention: A stakeholder analysis of ai supply chains. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 209–227, 2025b.

Dmitry Ivanov. Revealing interfaces of supply chain resilience and sustainability: a simulation study. *International Journal of Production Research*, 56(10):3507–3523, 2018.

Jon Keegan. How nebius emerged from russia and became one of nvidia's top-performing investments, jun 2025. URL https://sherwood.news/tech/nebius-nvidia-gpus-ai-startup/.

Marcus Kirk, Peter Molk, and Evan Pondel. Ai in corporate disclosure: Ir survey evidence, legal risks, and research opportunities. *Accounting Horizons*, pages 1–16, 2025.

Jon Kleinberg and Manish Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22):e2018340118, 2021. doi: 10.1073/pnas.2018340118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2018340118.

Benjamin Laufer, Hamidah Oderinwale, and Jon Kleinberg. Anatomy of a machine learning ecosystem: 2 million models on hugging face, 2025. URL https://arxiv.org/abs/2508.06811.

Katherine Lee, A Feder Cooper, and James Grimmelmann. Talkin"bout ai generation: Copyright and the generative-ai supply chain. *arXiv preprint arXiv:2309.08133*, 2023.

Lizhen Liang. Investigating co-authorship networks of academic and industry researchers in artificial intelligence. *Proceedings of the Association for Information Science and Technology*, 61(1):559–563, 2024.

Theresa Loconsolo, Maxwell Zeff, Anthony Ha, Kirsten Korosec, and Rebecca Bellan. Openai and microsoft's 'frenemies relationship,' and what you missed at sxsw, mar 2025. URL https://techcrunch.com/podcast/openai-and-microsofts-frenemies-relationship-and-what-you-missed-at-sxsw/.

Jyoti Mann. Read the memo satya nadella sent to employees announcing google deepmind cofounder's move to microsoft, mar 2024. URL https://www.businessinsider.com/satya-nadella-mustafa-suleyman-microsoft-memo-2024-3.

Lucas G Marin, Bram Rijsbosch, Gerasimos Spanakis, and Konrad Kollnig. Are companies taking ai risks seriously? a systematic analysis of companies' ai risk disclosures in sec 10-k forms. *arXiv preprint arXiv:2508.19313*, 2025.

McKinsey & Company. The cost of compute: A $7 trillion dollar race to scale data centers. *McKinsey Insights*, April 2025. URL https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-cost-of-compute-a-7-trillion-dollar-race-to-scale-data-centers#/.

James Muldoon, Ana Valdivia, and Adam Badger. The politics of artificial intelligence supply chains. *AI & SOCIETY*, pages 1–13, 2025.

NIST AI. Artificial intelligence risk management framework (ai rmf 1.0). *URL: https://nvlpubs. nist. gov/nistpubs/ai/nist. ai*, pages 100–1, 2023.

Jordan Novet and Ari Levy. Nvidia and capital one back databricks at $43 billion valuation in latest funding round, sep 2023. URL https://www.cnbc.com/2023/09/14/nvidia-and-capital-one-invest-in-databricks-at-43-billion-valuation.html.

Matt O'Brien. Chatgpt-maker openai signs deal with ap to license news stories, jul 2023. URL https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a.

Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. Towards ai accountability infrastructure: Gaps and opportunities in ai audit tooling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, page 1–29. ACM, April 2025. doi: 10.1145/3706598.3713301. URL http://dx.doi.org/10.1145/3706598.3713301.

OpenAI. Oscar. OpenAI customer story, 2024. URL https://openai.com/index/oscar/. Web page.

Trishla Ostwal. Google eyes publisher deals to train ai, following openai and perplexity's lead, 2025. URL https://www.adweek.com/media/google-publisher-deals-train-ai-openai-perplexity/. ADWEEK, July 23, 2025.

Rajan Patel. An expanded partnership with reddit. The Keyword (Google), February 2024. URL https://blog.google/company-news/inside-google/company-announcements/expanded-reddit-partnership/. Web post.

Paul. Google and reddit announce ai training data partnership, mar 2024. URL https://www.okoone.com/spark/industry-insights/google-and-reddit-announce-ai-training-data-partnership/.

Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg, and Daniel E. Ho. Outsider oversight: Designing a third party audit ecosystem for ai governance, 2022. URL https://arxiv.org/abs/2206.04737.

Reuters. Robotics startup figure raises $675 mln from microsoft, nvidia, other big techs. *Reuters*, February 2024.

Luis Rijo. Chatgpt uses google search despite official bing partnership claims, aug 2025. URL https://ppc.land/chatgpt-uses-google-search-despite-official-bing-partnership-claims/.

Pamela Samuelson. Generative ai meets copyright. *Science*, 381(6654):158–161, 2023.

Girish Sastry, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage, Julian Hazell, Cullen O'keefe, Gillian K Hadfield, Richard Ngo, Konstantin Pilz, et al. Computing power and the governance of artificial intelligence. *arXiv preprint arXiv:2402.08797*, 2024.

Divyanshi Sharma. Former facebook and google employees are working for chatgpt's parent company openai, details here, feb 2023. URL https://www.indiatoday.in/technology/news/story/former-facebook-and-google-employees-are-working-for-chatgpts-parent-company-openai-details-here-2338

Amanda Silberling. Meta launches a stand-alone ai app to compete with chatgpt, 2025. URL https://techcrunch.com/2025/04/29/meta-launches-a-standalone-ai-app-to-compete-with-chatgpt/. TechCrunch, April 29, 2025.

Stephen C Slota, Kenneth R Fleischmann, Sherri Greenberg, Nitin Verma, Brenna Cummings, Lan Li, and Chris Shenefiel. Many hands make many fingers to point: challenges in creating accountable ai. *Ai & Society*, 38(4):1287–1299, 2023.

Suhasini Srinivasaragavan. Google signs deal with associated press for gemini news content, jan 2025. URL https://www.siliconrepublic.com/machines/google-signs-deal-with-associated-press-for-gemini-news-content.

Two Seas Capital LP. Two seas capital lp definitive additional materials (schedule 14a) regarding core scientific, inc., oct 2025. URL https://www.sec.gov/Archives/edgar/data/1839341/000090266425004458/p25-2197dfan14a.htm.

David C Vladeck. Machines without principals: liability rules and artificial intelligence. *Wash. L. Rev.*, 89:117, 2014.

Kevin Wei, Carson Ezell, Nick Gabrieli, and Chinmay Deshpande. How do ai companies "fine-tune" policy? examining regulatory capture in ai governance. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, pages 1539–1555, 2024.

David Gray Widder and Dawn Nafus. Dislocated accountabilities in the "ai supply chain": Modularity and developers' notions of responsibility. *Big Data & Society*, 10(1):20539517231177620, 2023.

Kyle Wiggers. Ai startups ramp up federal lobbying efforts, jul 2024. URL https://techcrunch.com/2024/07/31/ai-startups-ramp-up-federal-lobbying-efforts/.

Kyle Wiggers. Google is forming a new team to build ai that can simulate the physical world, jan 2025. URL https://techcrunch.com/2025/01/06/google-is-forming-a-new-team-to-build-ai-that-can-simulate-the-physical-world/.

Business Wire. Inflection ai announces $1.3 billion of funding led by current investors microsoft and nvidia. https://www.businesswire.com/news/home/20230629810313/en/Inflection-AI-Announces-1.3-Billion-of-Funding-Led-by-Current-Investors-Microsoft-and-NVIDIA, June 2023.

Maxwell Zeff. Meta's former hardware lead for orion is joining openai, nov 2024. URL https://techcrunch.com/2024/11/04/metas-former-hardware-lead-for-orion-is-joining-openai/.

Maxwell Zeff. Meta hires key openai researcher to work on ai reasoning models, jun 2025a. URL https://techcrunch.com/2025/06/26/meta-hires-key-openai-researcher-to-work-on-ai-reasoning-models/.

Maxwell Zeff. Meta taps former google deepmind director to lead its ai research lab, may 2025b. URL https://techcrunch.com/2025/05/08/meta-taps-former-google-deepmind-director-to-lead-its-ai-research-lab/.

Zepp Health Corporation. Exhibit 99.1: Zepp health corporation reports first quarter 2025 unaudited financial results. U.S. Securities and Exchange Commission (SEC) EDGAR filing, May 2025. URL https://www.sec.gov/Archives/edgar/data/1720446/000110465925050677/tm2515579d1_ex99-1.htm. CIK 0001720446. Accession No. 0001104659-25-050677.

A Addition Details About Information Gathering Methodology

A.1 SEC EDGAR Search Configuration


```
{
    "sec_keywords": {
        "Bytedance AI": [
            "\"Bytedance AI\"",
            "\"Bytedance\" \"Doubao\"",
            "\"Bytedance\" \"Skylark\"",
            "\"Bytedance\" \"Seedance\""
        ],
        "Huawei": [
            "\"Huawei AI\"",
            "\"Huawei\" \"Pangu\""
        ],
        "IBM AI": [
            "\"IBM AI\"",
            "\"IBM\" \"watsonx\"",
            "\"IBM Granite\""
        ],
        "Midjourney": [
            "\"Midjourney\""
        ],
        "Snowflake": [
            "\"Snowflake AI\"",
            "\"Snowflake Arctic\"",
            "\"Snowflake\" \"Cortex\""
        ],
        "Meta AI": [
            "\"Meta AI\"",
            "\"Meta\" Llama"
        ],
        "xAI": [
            "\"x.AI\"",
            "\"xAI\" \"Grok\"",
            "\"x.AI\" \"Grok\""
        ],
        "Anthropic": [
            "\"Anthropic\""
        ],
        "Microsoft AI": [
            "\"Microsoft AI\"",
            "\"Microsoft\" \"Phi\"",
            "\"Microsoft\" Magma",
            "\"Microsoft\" \"Azure AI\"",
            "\"Microsoft\" \"Copilot\""
        ],
        "Inflection": [
            "\"Inflection AI\"",
            "\"Inflection\" \"Pi\""
        ],
        "AI21 Labs": [
            "\"AI21 Labs\"",
            "\"AI21\" \"Jamba\"",
            "\"AI21\" \"Jurassic\""
        ],
        "Adept": [
            "\"Adept AI\"",
            "\"Adept\" \"Fuyu\"",
            "\"Adept\" \"ACT-1\"",
            "\"Adept\" \"ACT-2\""
        ],
        "Krutrim": [
            "\"Krutrim\""
        ],
        "Mistral AI": [
            "\"Mistral AI\"",
```

```
    "\"Mistral\" \"Mixtral\"",
    "\"Mistral\" \"Large\""
  ],
  "Reka": [
    "\"Reka AI\"",
    "\"Reka Core\"",
    "\"Reka Flash\"",
    "\"Reka Edge\""
  ],
  "01.ai": [
    "\"01.ai\""
  ],
  "Google AI": [
    "\"Google AI\"",
    "\"Google\" \"Gemini\"",
    "\"Google\" \"Gemma\""
  ],
  "DeepMind": [
    "\"DeepMind\"",
    "\"DeepMind\" \"Gemini\"",
    "\"DeepMind\" \"Gemma\""
  ],
  "LightOn": [
    "\"LightOn\""
  ],
  "OpenAI": [
    "\"OpenAI\"",
    "\"ChatGPT\""
  ],
  "Aleph Alpha": [
    "\"Aleph Alpha\""
  ],
  "Numenta": [
    "\"Numenta\""
  ],
  "4Paradigm": [
    "\"4Paradigm\""
  ],
  "Contextual AI": [
    "\"Contextual AI\"",
    "\"Contextual Language Model\"",
    "\"CLM\" \"Contextual\""
  ],
  "Hugging Face": [
    "\"Hugging Face\""
  ],
  "NVIDIA": [
    "\"NVIDIA AI\"",
    "\"NVIDIA\" \"Nemotron\""
  ],
  "EleutherAI": [
    "\"EleutherAI\""
  ],
  "Together AI": [
    "\"Together AI\"",
    "\"Together.AI\""
  ],
  "Stability AI": [
    "\"Stability AI\"",
    "\"Stability\" \"Stable Diffusion\""
  ],
  "AI2": [
    "\"AI2\"",
    "\"Allen Institute for AI\""
  ],
  "Cohere": [
    "\"Cohere\"",
    "\"Cohere Command\""
```

```
        ],
        "Databricks": [
           "\"Databricks\"",
           "\"DBRX\""
        ],
        "Amazon": [
           "\"Amazon AI\"",
           "\"Amazon\" \"Titan\""
        ],
        "Baidu": [
           "\"Baidu\" \"Ernie\""
        ],
        "Alibaba": [
           "\"Alibaba AI\"",
           "\"Alibaba\" \"Qwen\""
        ],
        "DeepSeek": [
           "\"DeepSeek\"",
           "\"DeepSeek\" \"DeepSeek-R1\""
        ],
        "Cartesia": [
           "\"Cartesia\"",
           "\"Cartesia\" \"Cartesia-1\"",
           "\"Cartesia\" \"Sonic\""
        ]
   },
   "max_trials": 5,
   "num_chars": 1250
}
```


A.2 Search Pipeline Configuration


```
{
    "max_num_results_orgsite": 10,
    "max_num_results_newssite": 5,
    "max_num_results_gnews": 15,
    "time_duration": "1M",
    "org_sites": {
        "DeepMind": ["https://deepmind.google/discover/blog/", "https://deepmind.google/blog/", "https://blog.google/technol
        "Meta AI": ["https://ai.meta.com/blog/"],
        "xAI": ["https://x.ai/news"],
        "Anthropic": ["https://www.anthropic.com/news"],
        "Microsoft AI": ["https://blogs.microsoft.com/ai/"],
        "Inflection": ["https://inflection.ai/blog"],
        "AI21 Labs": ["https://www.ai21.com/blog", "https://www.ai21.com/newsroom/"],
        "Adept": ["https://www.adept.ai/blog"],
        "Krutrim": ["https://blog.olakrutrim.com/", "https://tech.olakrutrim.com/"],
        "Mistral AI": ["https://mistral.ai/news/"],
        "Reka": ["https://reka.ai/news"],
        "01.ai": ["https://www.01i.ai/news"],
        "Google AI": ["https://blog.google/technology/ai/"],
        "LightOn": ["https://www.lighton.ai/blog/", "https://www.lighton.ai/newsroom"],
        "OpenAI": ["https://openai.com/news/"],
        "Aleph Alpha": ["https://www.aleph-alpha.com/newsroom"],
        "Numenta": ["https://numenta.com/blog/"],
        "Contextual AI": ["https://contextual.ai/news"],
        "Hugging Face": ["https://huggingface.co/blog"],
        "NVIDIA": ["https://blogs.nvidia.com/"],
        "EleutherAI": ["https://blog.eleuther.ai/"],
        "Together AI": ["https://www.together.ai/blog"],
        "Stability AI": ["https://stability.ai/news"],
        "AI2": ["https://allenai.org/blog"],
        "Cohere": ["https://cohere.com/blog"],
        "Databricks": ["https://www.databricks.com/blog", "https://www.databricks.com/company/newsroom"],
        "Amazon": ["https://www.aboutamazon.com/amazon-news-today", "https://www.aboutamazon.com/artificial-intelligence-ai-
        "Baidu": ["https://research.baidu.com/blog", "https://ernie.baidu.com/blog/"],
        "Alibaba": ["https://www.alibabacloud.com/blog", "https://www.alibabacloud.com/en/news/"],
```

```
        "Cartesia": ["https://cartesia.ai/blog"],
        "IBM AI": ["https://newsroom.ibm.com/", "https://research.ibm.com/blog"],
        "Huawei": ["https://www.huawei.com/en/news"],
        "DeepSeek": ["https://deepseek.ai/blog"],
        "Snowflake": ["https://www.snowflake.com/en/news/"],
        "Midjourney": ["https://www.midjourney.com/updates"],
        "Bytedance": ["https://www.bytedance.com/en/news"]
    },
    "news_sites": [ "techcrunch.com", "venturebeat.com"],
    "search_strings": [
        "AI partnership"
    ],
    "search_locations": [
        "San Francisco, California, United States"
    ]
}
```

B Additional Results

In this appendix, we expand on the results in Section 5.

B.1 Additional Methodology for Analysis

As discussed in Section 5, we compute network statistics for the AI supply chain network, where all statistics are computed on the deduplicated network, where child actors have been consolidated into their parent organizations using the actor deduplication mapping (e.g., "DeepMind" and "Google AI" are consolidated into "Google", "Facebook" is consolidated into "Meta", "Azure" and "ChatGPT" are consolidated into "Microsoft" and "OpenAI" respectively). Actors with a total degree (in-degree plus out-degree) that is less than 2 are excluded from the analysis to avoid spurious results.

B.1.1 Categorization of Relationships

We begin by taking the relationships extracted from our main methodology. Recall that each relationship is identified as a pair of actors (Actor1 and Actor2) along with a textual description of their relationship. To infer the directionality of these relationships, we employ OpenAI's GPT-4o-mini to analyze the relationship descriptions. For each relationship, the LLM provides (1) the relationship category (e.g., "Capital & Investment", "Talent & Expertise", "Compute & Infrastructure"); (2) the relationship subcategory (e.g., "Venture Investment", "Hiring/Recruitment", "Data Partnerships"); (3) the primary and secondary directional actions from a predefined taxonomy (e.g., "invests in", "provides data to", "licenses model to"); (4) the direction as either "forward" (Actor1 $\rightarrow$ Actor2), "backward" (Actor2 $\rightarrow$ Actor1), or "bidirectional" (Actor1 $\leftrightarrow$ Actor2), and (5) confidence scores for each classification. The full taxonomy is:

```
{
  "relationship_categories": [
    "Compute & Infrastructure",
    "Data & Content",
    "Models & Technology",
    "Capital & Investment",
    "Talent & Expertise",
    "Services & Applications",
    "Partnership & Collaboration",
    "Regulatory & Compliance",
    "Competitive & Market",
    "Supply Chain & Procurement",
    "Legal & Disputes",
    "Other"
  ],
  "relationship_subcategories": {
    "GPU/Accelerator Supply": "Compute & Infrastructure",
    "Cloud Compute Services": "Compute & Infrastructure",
    "Data Center Infrastructure": "Compute & Infrastructure",
```

```
"Edge Computing": "Compute & Infrastructure",
"Network Infrastructure": "Compute & Infrastructure",
"AI-Specific Infrastructure": "Compute & Infrastructure",
"Training Data": "Data & Content",
"Content Licensing": "Data & Content",
"Data Partnerships": "Data & Content",
"Data Services": "Data & Content",
"Data Annotation": "Data & Content",
"Model Licensing": "Models & Technology",
"API Access": "Models & Technology",
"Technology Transfer": "Models & Technology",
"Open Source": "Models & Technology",
"Algorithm Development": "Models & Technology",
"Model Weights Sharing": "Models & Technology",
"Venture Investment": "Capital & Investment",
"Strategic Investment": "Capital & Investment",
"Acquisition/M&A": "Capital & Investment",
"Mergers": "Capital & Investment",
"Funding Rounds": "Capital & Investment",
"Debt Financing": "Capital & Investment",
"Equity Stakes": "Capital & Investment",
"Hiring/Recruitment": "Talent & Expertise",
"Research Collaboration": "Talent & Expertise",
"Advisory Relationships": "Talent & Expertise",
"Academic/Educational Partnerships": "Talent & Expertise",
"Talent Exchange": "Talent & Expertise",
"Consulting Services": "Talent & Expertise",
"Training/Education Programs": "Talent & Expertise",
"SaaS Integration": "Services & Applications",
"Consumer Products": "Services & Applications",
"Enterprise Solutions": "Services & Applications",
"Platform Services": "Services & Applications",
"Platform Hosting/Marketplace": "Services & Applications",
"White-label Solutions": "Services & Applications",
"API Services": "Services & Applications",
"Strategic Partnerships": "Partnership & Collaboration",
"Joint Ventures": "Partnership & Collaboration",
"Ecosystem Partnerships": "Partnership & Collaboration",
"Industry Consortia": "Partnership & Collaboration",
"Co-marketing": "Partnership & Collaboration",
"Research Partnerships": "Partnership & Collaboration",
"Event/Organization Partnerships": "Partnership & Collaboration",
"Safety Standards": "Regulatory & Compliance",
"Regulatory Compliance": "Regulatory & Compliance",
"Audit Relationships": "Regulatory & Compliance",
"Governance Frameworks": "Regulatory & Compliance",
"Certification": "Regulatory & Compliance",
"Standards Development": "Regulatory & Compliance",
"Market Competition": "Competitive & Market",
"Market Positioning": "Competitive & Market",
"Competitive Intelligence": "Competitive & Market",
"Antitrust/Regulatory": "Competitive & Market",
"Market Share Dynamics": "Competitive & Market",
"Competitive Analysis": "Competitive & Market",
"Component Supply": "Supply Chain & Procurement",
"Manufacturing": "Supply Chain & Procurement",
"Logistics": "Supply Chain & Procurement",
"Procurement": "Supply Chain & Procurement",
"Vendor Relationships": "Supply Chain & Procurement",
"Supply Agreements": "Supply Chain & Procurement",
"Lawsuits": "Legal & Disputes",
"IP Disputes": "Legal & Disputes",
"Regulatory Actions": "Legal & Disputes",
"Legal Settlements": "Legal & Disputes",
"Contract Disputes": "Legal & Disputes",
"Antitrust Actions": "Legal & Disputes",
"Unspecified": "Other",
"Miscellaneous": "Other"
```

```
  },
  "directional_categories": [
    "provides compute to",
    "provides hardware to",
    "shares infrastructure with",
    "provides data to",
    "licenses content to",
    "shares data with",
    "licenses model to",
    "provides API access to",
    "transfers technology to",
    "open sources to",
    "develops technology with",
    "makes available on platform",
    "adapts for",
    "invests in",
    "acquires",
    "merges with",
    "provides funding to",
    "hires talent from",
    "collaborates on research with",
    "provides advisory services to",
    "partners on research with",
    "integrates services of",
    "provides services to",
    "powers services of",
    "hosts on platform",
    "partners on services with",
    "forms partnership with",
    "forms joint venture with",
    "collaborates with",
    "co-organizes with",
    "partners to develop with",
    "co-markets with",
    "complies with standards of",
    "audits",
    "certifies",
    "develops standards with",
    "competes with",
    "positions against",
    "supplies components to",
    "manufactures for",
    "sues",
    "files complaint against",
    "settles dispute with",
    "licenses IP to",
    "has relationship with"
  ]
}
```

B.1.2 Constructing Network Graph

We then use the relationships categorizations to create a network and its adjacency matrices for analysis. We first load the raw relationship data from the directional relationships CSV file and apply actor deduplication using the actor deduplication mapping, which consolidates child actors (e.g., "Google AI", "DeepMind") into their parent organizations (e.g., "Google"). This ensures that relationships involving different names for the same organization are properly aggregated. Self-loops (relationships where actor1 equals actor2 after deduplication) are removed. We then create a binary adjacency matrix (indicator matrix) from the deduplicated relationship data, where a value of 1 indicates that a relationship exists between two actors, and 0 indicates no relationship. The matrix handles directional relationships: forward relationships create edges from actor1 to actor2, backward relationships create edges from actor2 to actor1, and bidirectional relationships create edges in both directions. Next, we filter out actors with total degree (in-degree + out-degree) ≤ 2 from the network, which removes actors with very few connections and focuses the analysis on actors that are more central to the network structure. The filtering uses a minimum degree threshold of 2,

Table 3: Overall Network Statistics

Metric	**Value**
Number of filtered actors	2,930
Number of filtered edges	13,843
Network density	0.00161
Average in-degree	4.72
Average out-degree	4.72
Average total degree	9.45
Number of strongly connected components	377
Largest SCC size	2,506
Number of weakly connected components	126
Largest WCC size	2,770
Average path length	3.04
Diameter	10
Average clustering coefficient	0.389
Transitivity	0.0291

meaning only actors with total degree > 2 are retained (actors with degree ≤ 2 are excluded).

B.1.3 Network Statistics Table

Table 3 presents the overall network statistics. The following definitions explain each metric in the network statistics table.

Number of actors The total number of unique organizations (nodes) in the network after deduplication and degree filtering.

Number of edges The total number of directed relationships (edges) in the network. Each relationship represents a connection between two actors.

Network density The ratio of actual edges to possible edges in a directed network, where 0 implies no connections and 1 implies fully connected, as given by: $\frac{E}{n(n-1)}$ where E is the number of edges and n is the number of nodes.

Average in-degree The average number of incoming edges per actor.

Average out-degree The average number of outgoing edges per actor.

Average total degree The average of the sum of in-degree and out-degree for all actors. This represents the average total number of relationships (both incoming and outgoing) per actor.

Number of strongly connected components A strongly connected component (SCC) is a maximal subgraph where there exists a directed path from any node to any other node in the subgraph. This metric counts how many such components exist in the network.

Largest SCC size The number of actors in the largest SCC. This indicates the size of the largest group of actors that are mutually reachable via directed paths.

Number of weakly connected components A weakly connected component (WCC) is a maximal subgraph where there exists a path (ignoring edge direction) from any node to any other node. This metric counts how many such components exist when edge directions are ignored.

Largest WCC size The number of actors in the largest WCC. This indicates the size of the largest group of actors that are connected when edge directions are ignored.

Average path length The average length of the shortest path between all pairs of nodes in the largest strongly connected component. Only computed for the largest SCC if the network is not fully strongly connected.

Diameter The longest shortest path between any two nodes in the largest SCC. Only computed for the largest SCC if the network is not fully strongly connected.

Average clustering coefficient The local clustering coefficient of a node measures the proportion of its neighbors that are also connected to each other. For a node i with k_i neighbors, it is calculated as: $C_i = \frac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of i. The average clustering coefficient is the average across all nodes and is computed on the undirected version of the network.

Transitivity The ratio of triangles (three nodes all connected to each other) to the number of connected triples (three nodes with at least two edges). Also known as the global clustering coefficient. It measures the tendency of nodes to form clusters. Calculated on the undirected version of the network. Transitivity ranges from 0 (no triangles) to 1 (all triples form triangles).

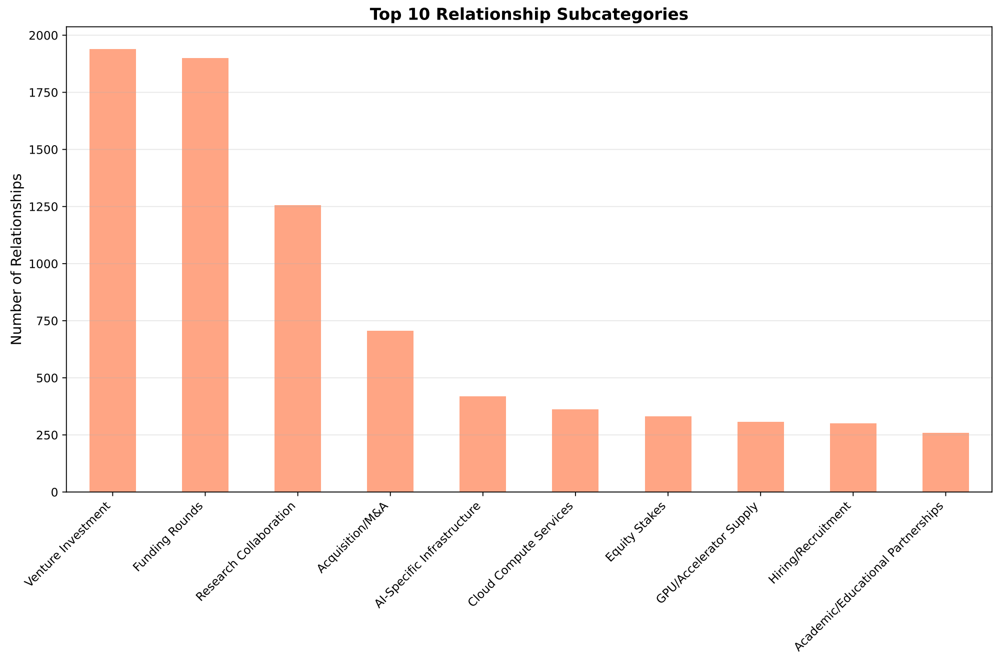

Figure 4: Top 10 relationship subcategories. Shows the count of unique relationships (after deduplication) for the most common subcategories.

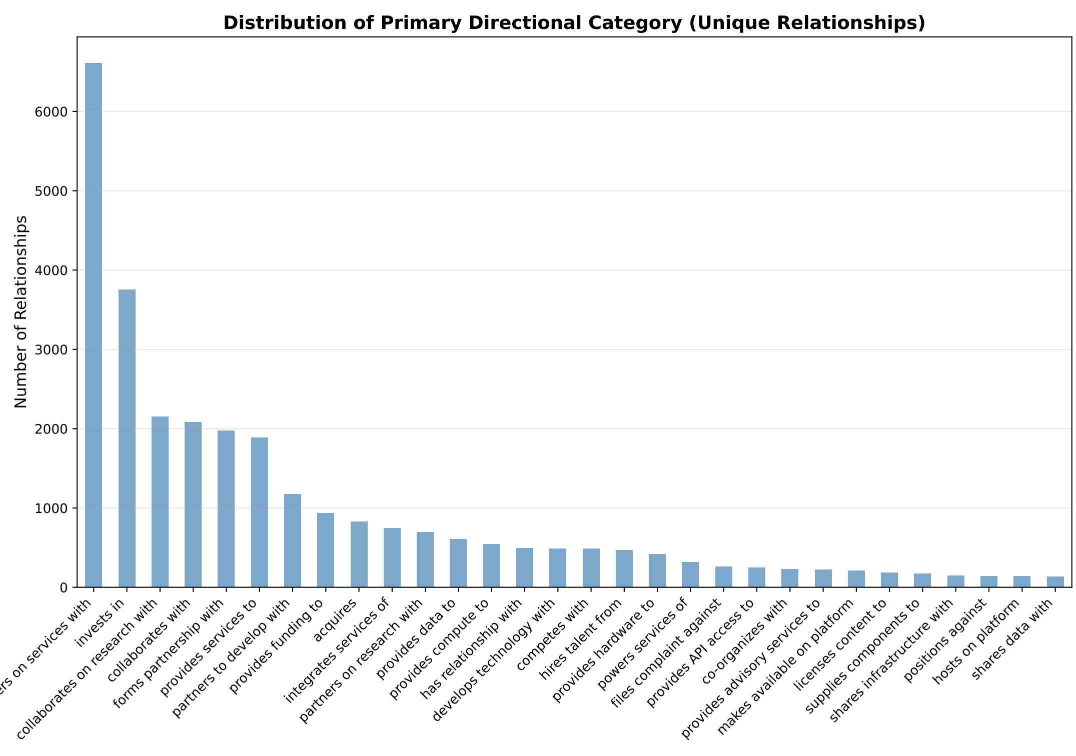

Figure 5: Top 30 primary directional categories. Shows the count of unique relationships (after deduplication) for the most common directional relationship types.

B.1.4 Relationship Statistics

The relationship category distribution is given in the main text in Figure 1. Figure 4 shows the top 10 relationship subcategories, which provide more granular classification of relationships within categories. Figure 5 shows the top 30 primary directional categories, which indicate the direction and nature of relationships between actors.

B.1.5 Word Frequency Analysis

This section presents word frequency histograms across different time periods, showing the distribution of key terms in SEC filings related to AI supply chain relationships. Figure 6 shows the overall word frequency distribution across all time periods in the dataset. Figures 7, 8, 9, and 10 show word frequency distributions for individual years from 2022 to 2025. The word analysis is done on the relationship descriptions extracted from the pipeline described in Section 4.

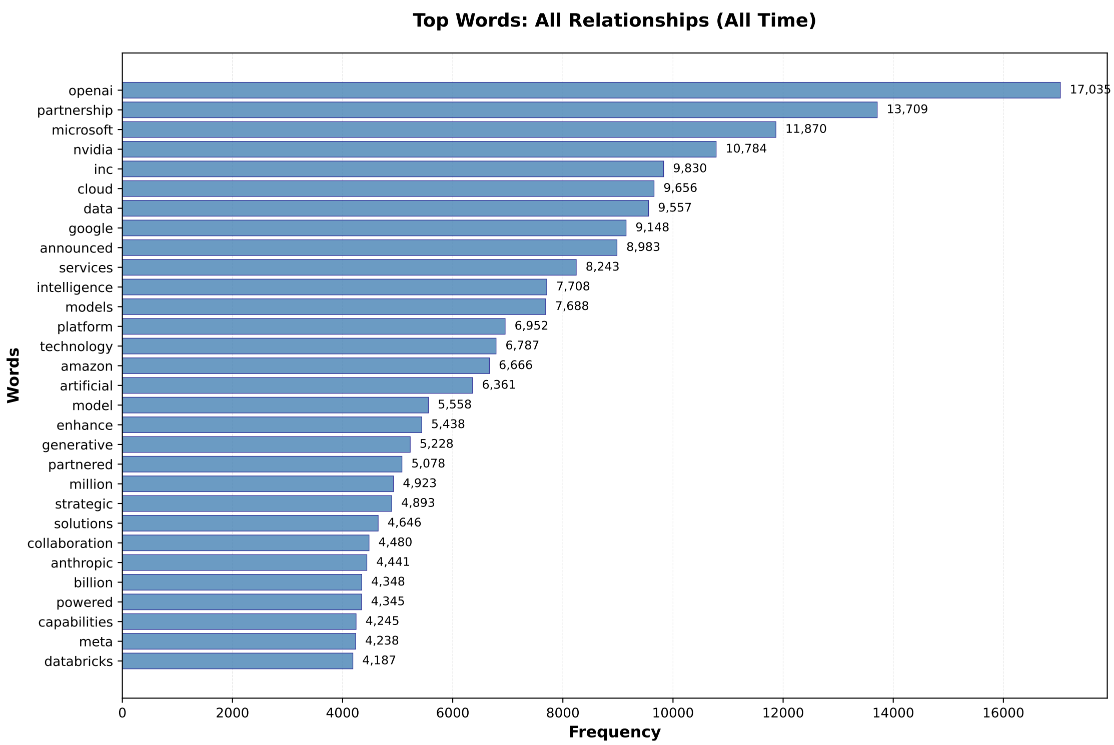

Figure 6: Overall word frequency histogram (all time periods)

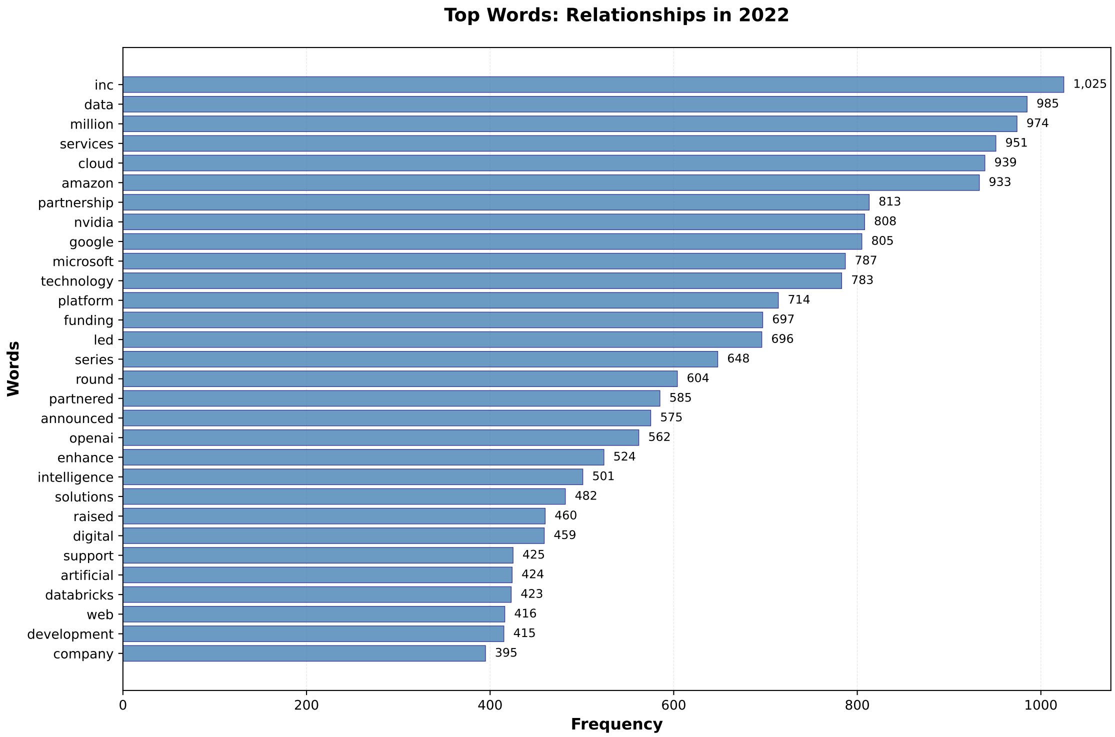

Figure 7: Word frequency histogram for 2022

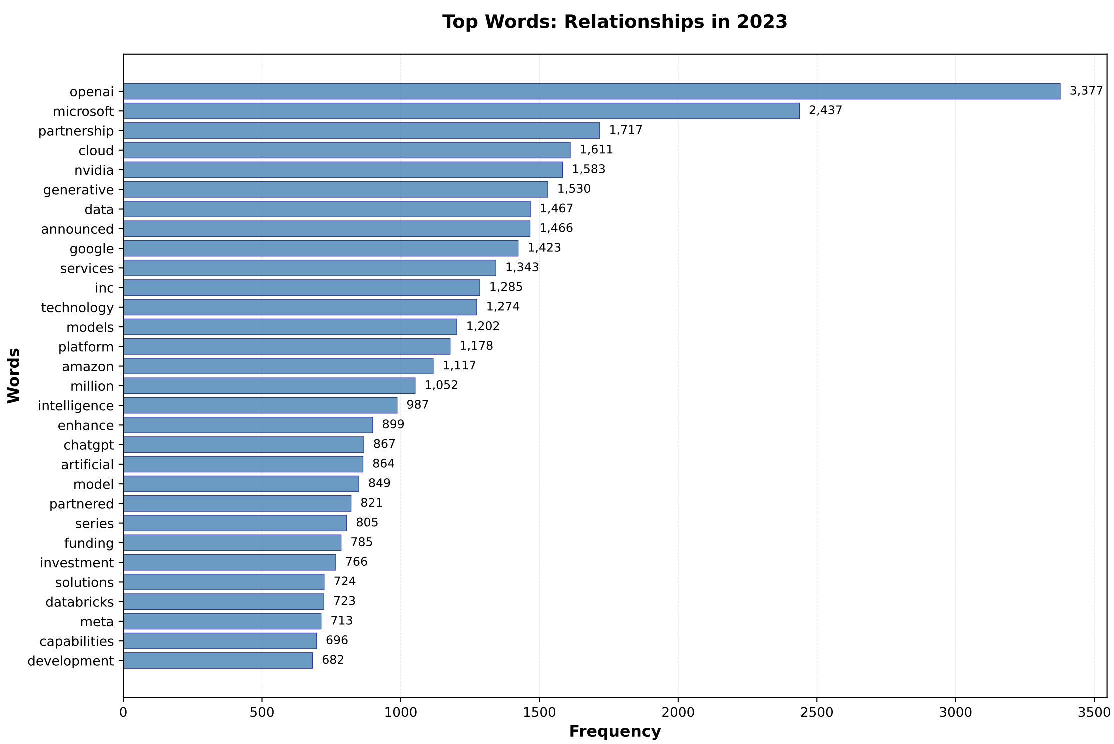

Figure 8: Word frequency histogram for 2023

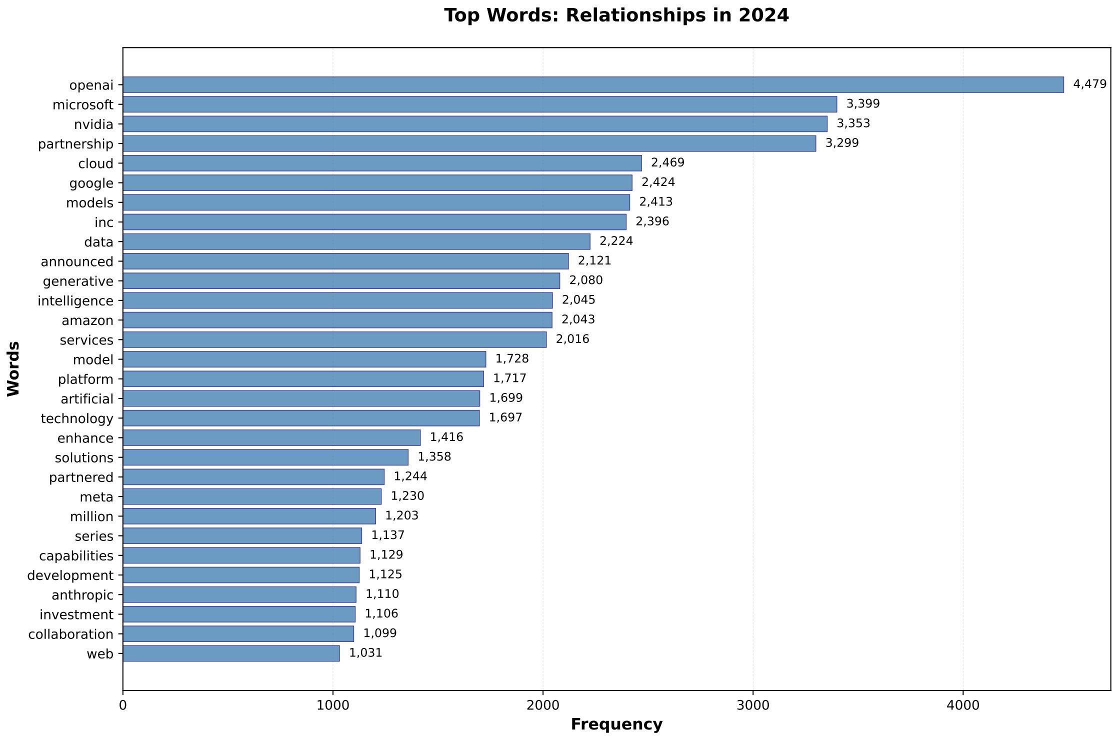

Figure 9: Word frequency histogram for 2024

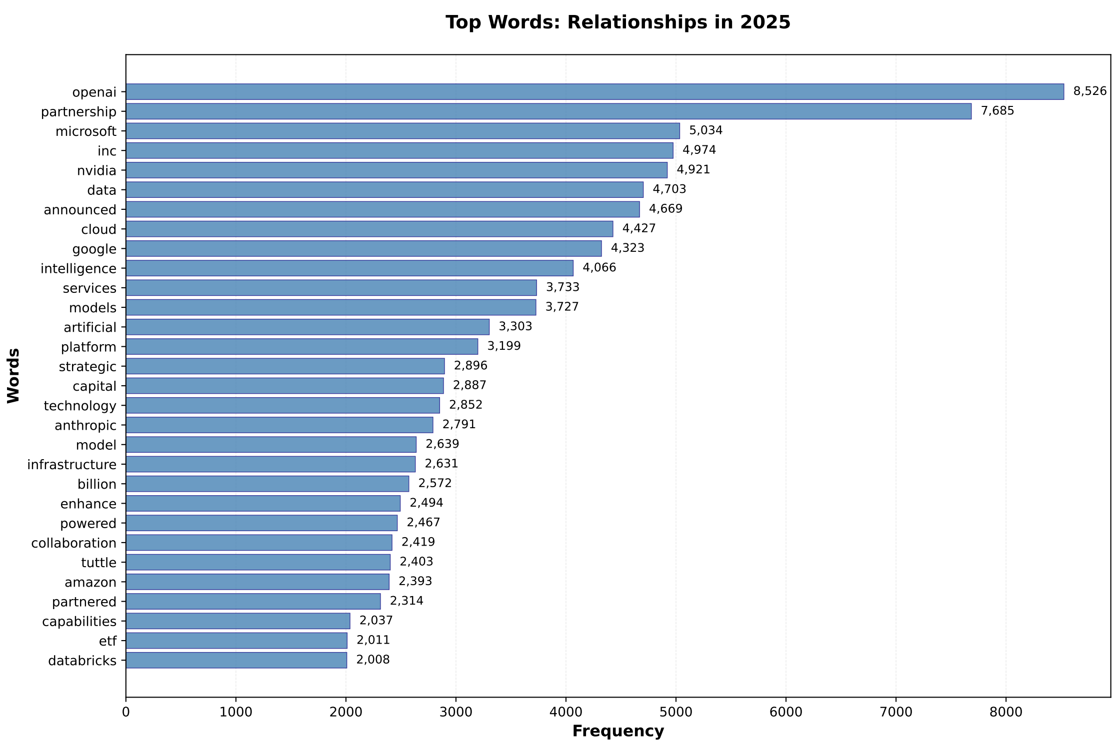

Figure 10: Word frequency histogram for 2025

B.1.6 Time Series Analysis

This section presents time series plots showing how actor centrality change over time. When we say "unnormalized degree", we mean the total number of relationships (i.e., total number of mentions), which may refer to the same relationship in different articles. Thus, in order to avoid inflating the strength of relationships by accidentally double counting (e.g., when two articles that we scrape refer to the same relationship), we also normalize the degree by the number of unique neighbors. When we refer to the "top 10", we mean the highest degree actors under the given metric. When we refer to the "model providers from our targeted search list", we only consider the model providers from our initial targeted search list, as given in Section A. Figure Figure 11 shows the top 10 actors by unnormalized degree (total relationship count) across all sources as a time series, while Figure 12 presents the same data as a stacked area chart. Figure 13 shows the top 10 actors by normalized degree (unique neighbors) across all sources as a time series, with Figure 2 in the main text providing the stacked area chart view. For model providers from our targeted search list, Figure 14 shows the top 10 by unnormalized degree as a time series, with Figure 15 showing the stacked area chart. Figure 16 shows the top 10 model providers by normalized degree as a time series, and Figure 17 presents the stacked area chart view.

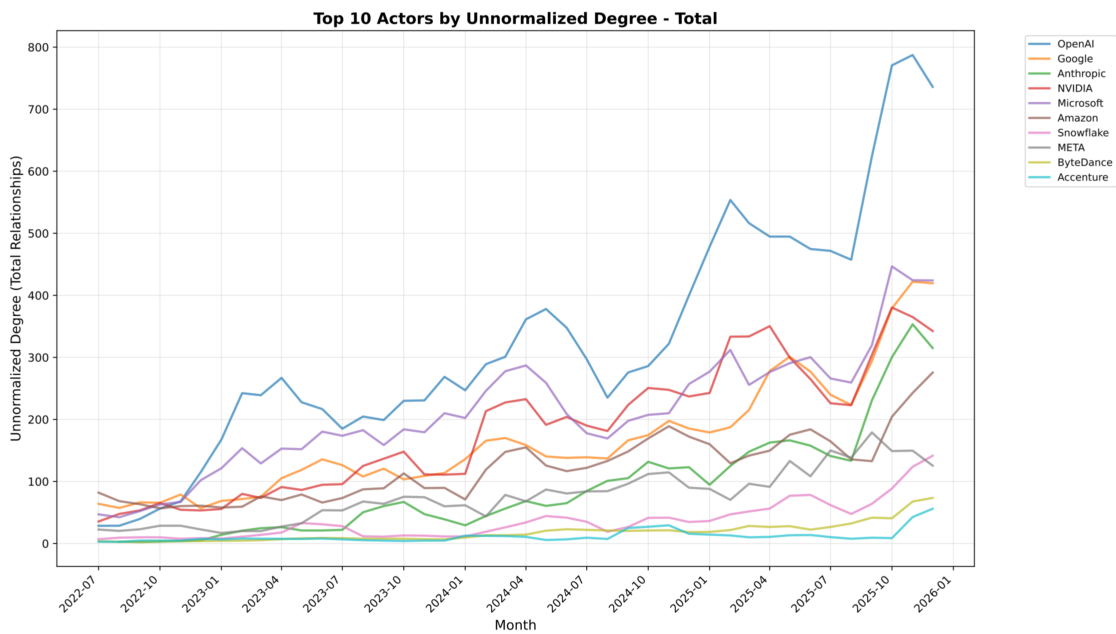

Figure 11: Top 10 actors by unnormalized degree (total relationships) - Time series

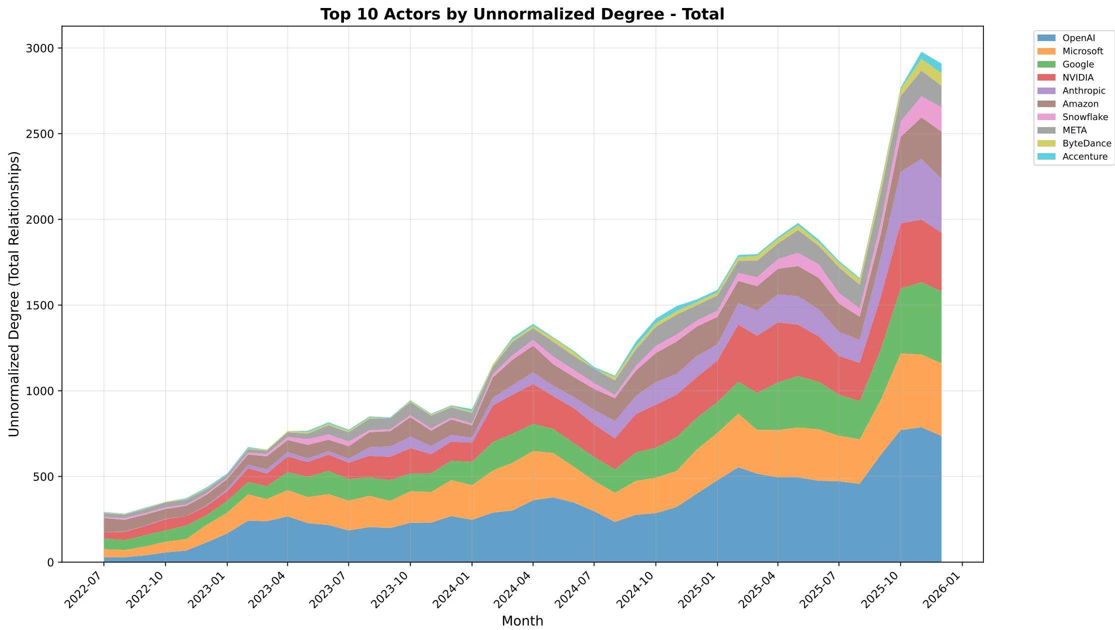

Figure 12: Top 10 actors by unnormalized degree (total relationships) - Stacked area chart

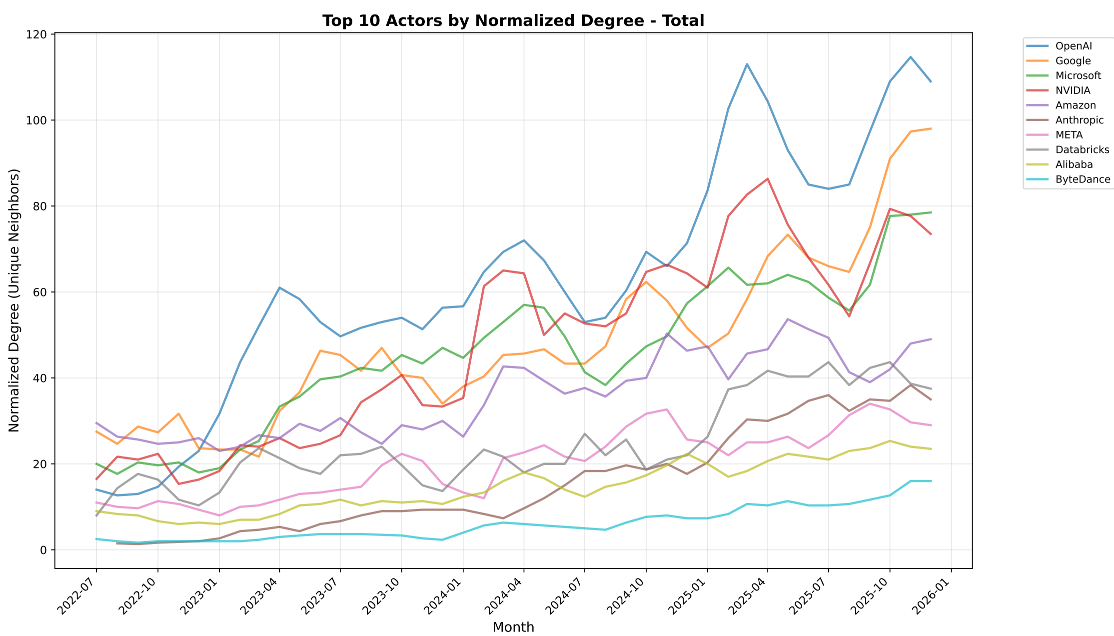

Figure 13: Top 10 actors by normalized degree (unique neighbors) - Time series

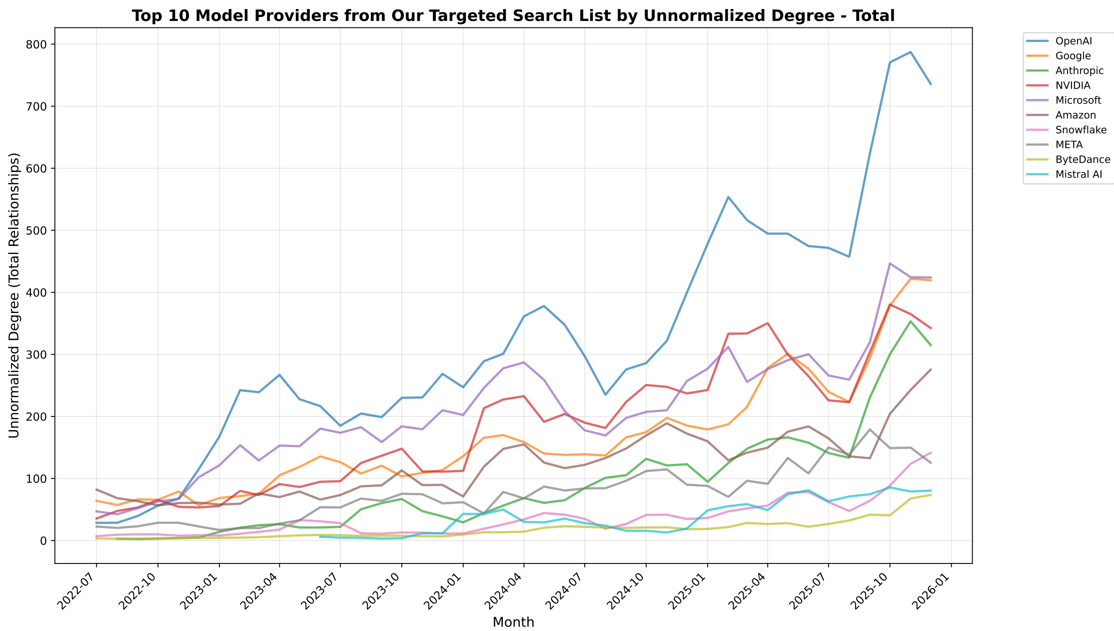

Figure 14: Top 10 model providers from our targeted search list by unnormalized degree (total relationships) - Time series

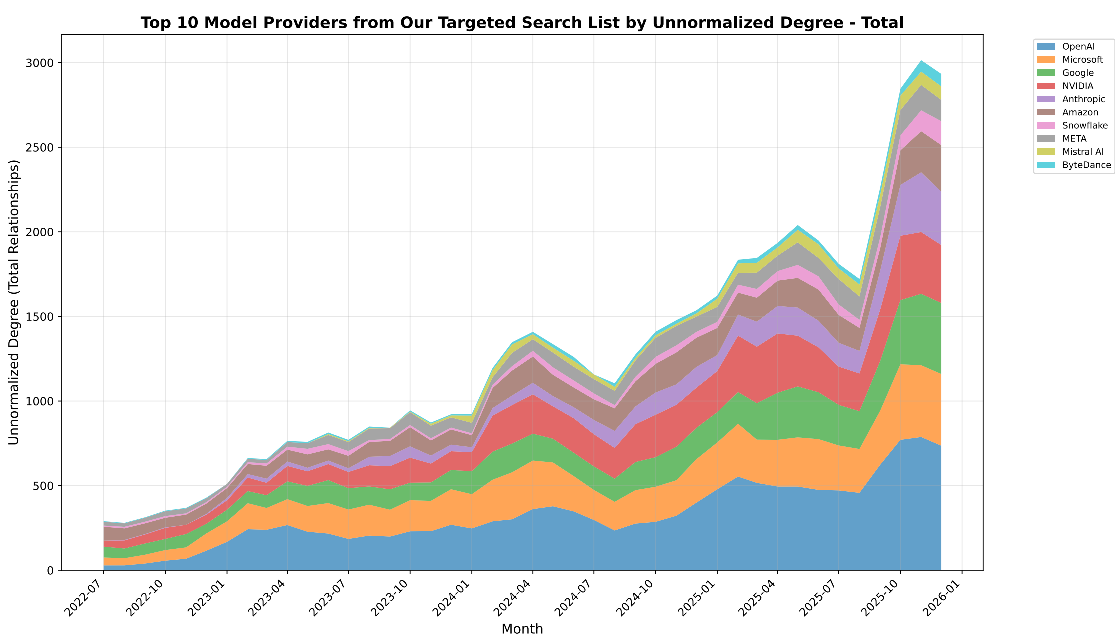

Figure 15: Top 10 model providers from our targeted search list by unnormalized degree (total relationships) - Stacked area chart

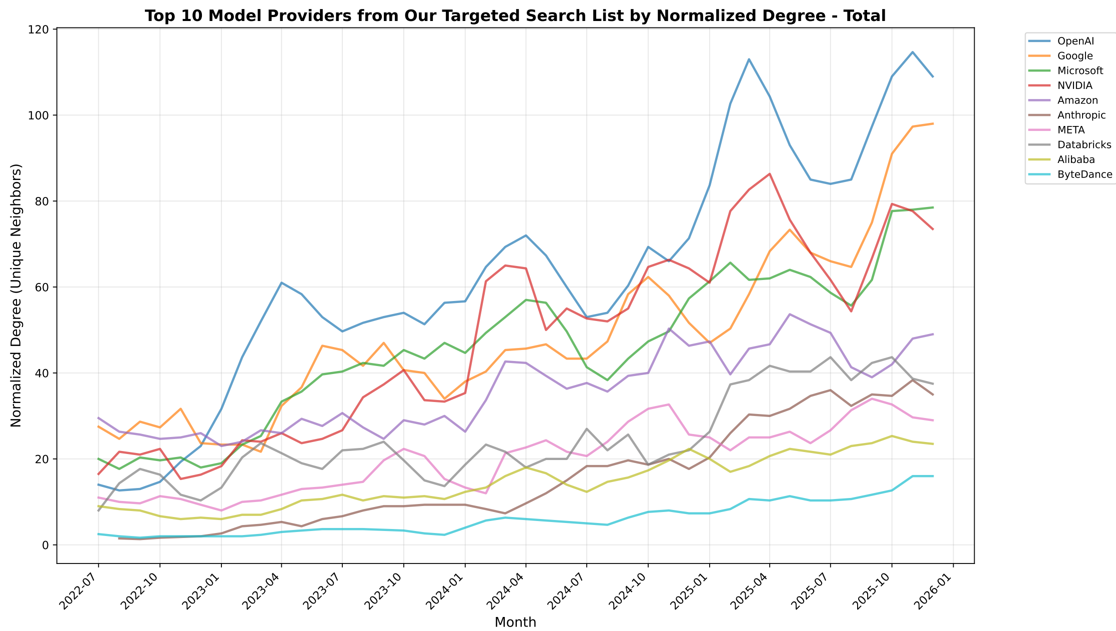

Figure 16: Top 10 model providers from our targeted search list by normalized degree (unique neighbors) - Time series

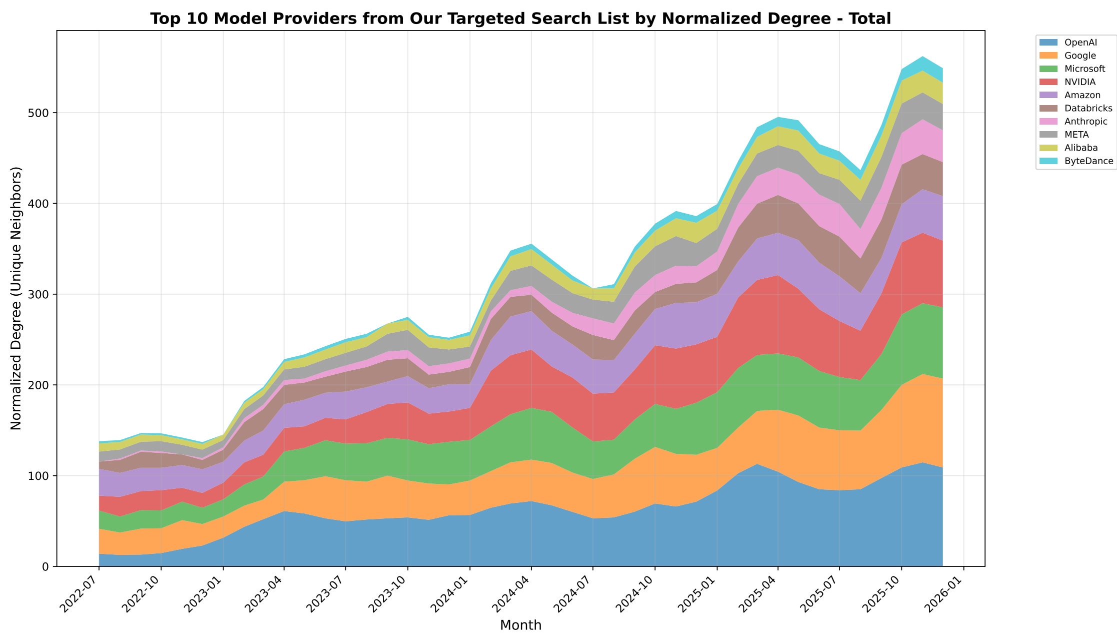

Figure 17: Top 10 model providers from our targeted search list by normalized degree (unique neighbors) - Stacked area chart

B.1.7 Actor Rankings

This section presents rankings of actors by various centrality measures. Rankings are computed on the deduplicated network (where child actors are consolidated into parent organizations) after filtering out actors with total degree ≤ 2. Four centrality measures are computed for each actor (one degree-based measure and three centrality measures):

Total Degree Sum of in-degree and out-degree (total number of relationships)

Betweenness Centrality Measures how often an actor lies on shortest paths between other actors (computed using NetworkX's `betweenness_centrality`)

Eigenvector Centrality Measures importance based on connections to other important actors (computed using NetworkX's `eigenvector_centrality` with max_iter=1000)

PageRank Measures importance using a recursive algorithm that considers both direct connections and the importance of connected actors (computed using NetworkX's `pagerank`)

Table 4 shows the top 10 actors by each centrality measure across the entire network. Table 5 shows the top 10 actors ranked by total degree (sum of in-degree and out-degree) for each year from 2022 to 2025, with each column showing the actor names ranked 1st through 10th for that year, allowing for comparison of how the most connected actors changed over time. Table 6 shows all model providers from our targeted search list, sorted by total degree, with their actual centrality measure values (not ranks), providing a comprehensive view of how these specific actors compare across different centrality measures. Table 7 shows how actor rankings changed from 2022 to 2025, with growth metrics including rank change (negative values indicate improvement), centrality change, growth rate, and best/worst performing years. Rankings are also computed separately for each relationship category, showing which actors are most central within specific types of relationships: the category-specific ranking tables show the top 10 actors by each centrality measure for Capital & Investment (Table 8), Competitive & Market (Table 9), Compute & Infrastructure (Table 10), Data & Content (Table 11), Legal & Disputes (Table 12), Models & Technology (Table 13), Partnership & Collaboration (Table 14), Regulatory & Compliance (Table 15), Services & Applications (Table 16), Supply Chain & Procurement (Table 17), and Talent & Expertise (Table 18).

Table 5: Top 10 Actors by Total Degree (2022-2025)

Rank	2022	2023	2024	2025
1	Amazon	OpenAI	OpenAI	OpenAI
2	Google	Google	NVIDIA	Google
3	Microsoft	Microsoft	Google	NVIDIA
4	NVIDIA	Amazon	Microsoft	Microsoft
5	Databricks	NVIDIA	Amazon	Amazon
6	IBM	Databricks	Databricks	Databricks
7	Meta	Meta	Meta	Anthropic
8	OpenAI	IBM	IBM	Meta
9	Snowflake	Alibaba	Anthropic	IBM
10	Alibaba	Anthropic	Huawei	Alibaba

Table 4: Top 20 Actors by Centrality Measures (Overall Network)

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	Google	OpenAI	OpenAI	OpenAI
2	OpenAI	Google	Google	Google
3	NVIDIA	NVIDIA	Microsoft	NVIDIA
4	Microsoft	Microsoft	NVIDIA	Microsoft
5	Amazon	Amazon	Amazon	Amazon
6	Databricks	Databricks	Meta	Databricks
7	Meta	Meta	IBM	Meta
8	IBM	Alibaba	Databricks	IBM
9	Anthropic	IBM	Anthropic	Anthropic
10	Alibaba	Anthropic	HuggingFace	Alibaba
11	Snowflake	Snowflake	Snowflake	Huawei
12	Huawei	Huawei	Alibaba	Snowflake
13	HuggingFace	HuggingFace	DeepSeek	HuggingFace
14	DeepSeek	DeepSeek	Apple	DeepSeek
15	Cohere	ByteDance	Cohere	Baidu
16	ByteDance	Baidu	Mistral AI	Cohere
17	Baidu	Cohere	xAI	xAI
18	xAI	Mistral AI	Oracle	ByteDance
19	Mistral AI	xAI	ByteDance	Mistral AI
20	Apple	Stability AI	SAP	Stability AI

Table 6: Model Providers from Our Targeted Search List: Centrality Values

Actor	Total Degree	Betweenness	Eigenvector	Pagerank
Google	1179.00	0.1748	0.2975	0.0363
OpenAI	1153.00	0.1796	0.3017	0.0383
NVIDIA	1003.00	0.1421	0.2684	0.0309
Microsoft	985.00	0.1302	0.2745	0.0303
Amazon	890.00	0.1196	0.2412	0.0284
Databricks	610.00	0.1011	0.1378	0.0189
Meta	476.00	0.0505	0.1691	0.0139
IBM	408.00	0.0373	0.1402	0.0116
Anthropic	372.00	0.0345	0.1352	0.0111
Alibaba	328.00	0.0425	0.1010	0.0096
Snowflake	295.00	0.0307	0.1071	0.0090
Huawei	292.00	0.0303	0.0655	0.0092
HuggingFace	201.00	0.0178	0.1127	0.0064
DeepSeek	170.00	0.0175	0.0954	0.0053
Cohere	161.00	0.0151	0.0861	0.0042
ByteDance	154.00	0.0166	0.0770	0.0039
Baidu	142.00	0.0154	0.0539	0.0046
xAI	133.00	0.0097	0.0780	0.0040
Mistral AI	132.00	0.0113	0.0801	0.0036
Stability AI	109.00	0.0090	0.0638	0.0035
EleutherAI	88.00	0.0084	0.0597	0.0030
Midjourney	77.00	0.0062	0.0459	0.0030
AI21 Labs	43.00	0.0029	0.0293	0.0014
Aleph Alpha	34.00	0.0021	0.0207	0.0012
Inflection	33.00	0.0013	0.0380	0.0010
Together AI	30.00	0.0015	0.0240	0.0009
Krutrim AI	23.00	0.0007	0.0213	0.0006
Reka	21.00	0.0005	0.0297	0.0007
01.AI	18.00	0.0001	0.0296	0.0006
Cartesia	16.00	0.0001	0.0108	0.0005
Contextual AI	16.00	0.0003	0.0196	0.0005
Numenta	12.00	0.0007	0.0089	0.0005
Adept	9.00	0.0002	0.0143	0.0003
AI2	7.00	0.0001	0.0031	0.0003
LightOn	5.00	0.0000	0.0009	0.0002

Table 7: Growth Metrics for Top Actors (2022-2025)

Actor	Total Degree Growth%	Betweenness Growth%	Eigenvector Growth%	Pagerank Growth%
OpenAI	896.30	113.79	83.44	145.16
Huawei	417.65	62.94	—	37.97
xAI	104.00	65.51	2.68	33.89
Cohere	180.65	125.35	-13.38	21.61
EleutherAI	162.50	22.87	—	5.09
Anthropic	369.39	163.11	58.35	89.45
Alibaba	413.33	101.61	46.52	11.61
Google	272.00	-14.01	-11.44	-11.51
HuggingFace	291.30	-9.11	-19.13	-9.42
Microsoft	266.67	-16.89	-12.59	-0.45
NVIDIA	300.00	17.48	-7.31	-3.14
Databricks	239.29	6.18	-1.22	-19.27
Meta	229.63	-23.77	-30.57	-41.64
Baidu	47.06	-51.30	—	-20.40
Snowflake	169.77	-15.57	15.62	-36.96
Amazon	122.97	-59.90	-45.44	-54.29
Apple	118.18	—	17.59	-35.44
Intel	166.67	19.83	19.78	-16.34
IBM	189.66	-47.14	-23.81	-47.37
Stability AI	30.77	-56.43	—	-47.70
Adobe	—	—	21.12	—
Alphabet	294.12	—	-19.83	—
ByteDance	584.62	13.38	20.79	—
Mistral AI	—	63.44	2.28	—
Oracle	—	—	-5.66	—
SAP	73.68	-15.33	—	—
Salesforce	—	—	8.83	—
T. Rowe Price Associates	—	—	—	—

Table 8: Top 10 Actors in Relationships in Capital & Investment Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	Databricks	Databricks	Databricks	Databricks
2	NVIDIA	NVIDIA	OpenAI	OpenAI
3	OpenAI	OpenAI	NVIDIA	NVIDIA
4	Microsoft	Anthropic	Anthropic	Microsoft
5	Anthropic	Microsoft	Microsoft	Anthropic
6	Amazon	ByteDance	Amazon	Amazon
7	Google	Amazon	CoreWeave	Google
8	xAI	Google	xAI	Pony.AI
9	ByteDance	xAI	Meta	xAI
10	T. Rowe Price Associates	Stability AI	Mistral AI	Cohere

Table 9: Top 10 Actors in Relationships in Competitive & Market Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	OpenAI
2	Google	Google	Google	Google
3	Meta	Microsoft	Meta	Meta
4	Microsoft	NVIDIA	Microsoft	Microsoft
5	NVIDIA	DeepSeek	NVIDIA	NVIDIA
6	DeepSeek	Meta	DeepSeek	DeepSeek
7	Amazon	Amazon	Amazon	Amazon
8	Anthropic	Baidu	Anthropic	Anthropic
9	Baidu	IBM	ByteDance	Baidu
10	IBM	Apple	Baidu	IBM

Table 10: Top 10 Actors in Relationships in Compute & Infrastructure Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	NVIDIA	NVIDIA	NVIDIA	NVIDIA
2	Amazon	Amazon	Microsoft	Amazon
3	Microsoft	Microsoft	Amazon	Microsoft
4	Google	Alibaba	Google	Google
5	OpenAI	Google	OpenAI	OpenAI
6	CoreWeave	OpenAI	Oracle	xAI
7	Oracle	CoreWeave	Advanced Micro Devices	Oracle
8	Advanced Micro Devices	Oracle	CoreWeave	CoreWeave
9	Meta	Databricks	Meta	Alibaba
10	xAI	xAI	Israeli Ministry of Defense	Advanced Micro Devices

Table 11: Top 10 Actors in Relationships in Data & Content Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	OpenAI
2	Google	Google	Google	Google
3	Meta	Meta	Meta	Meta
4	Amazon	Amazon	VentureBeat	EleutherAI
5	HuggingFace	HuggingFace	TechCrunch	CommonCrawl
6	Snowflake	X	Reddit	xAI
7	Universal Music Group	xAI	Midjourney	Universal Music Group
8	DeepSeek	Universal Music Group	Stability AI	Midjourney
9	Microsoft	Tesla	The Associated Press	TechCrunch
10	Stability AI	Databricks	Microsoft	VentureBeat

Table 12: Top 10 Actors in Relationships in Legal & Disputes Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	OpenAI
2	Google	Midjourney	Google	Google
3	Midjourney	Anthropic	NVIDIA	Microsoft
4	Microsoft	Disney	Microsoft	NVIDIA
5	Meta	Microsoft	Character AI	Stability AI
6	NVIDIA	Google	DeepSeek	DeepSeek
7	Disney	Stability AI	Meta	The New York Times
8	Anthropic	Universal Music Group	Midjourney	Midjourney
9	Apple	Meta	Apple	Character AI
10	Stability AI	The New York Times	The New York Times	Meta

Table 13: Top 10 Actors in Relationships in Models & Technology Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	HuggingFace
2	HuggingFace	HuggingFace	HuggingFace	OpenAI
3	Meta	Google	Google	Google
4	Google	Meta	Meta	NVIDIA
5	DeepSeek	NVIDIA	Microsoft	Meta
6	NVIDIA	DeepSeek	Anthropic	Anthropic
7	Microsoft	Alibaba	NVIDIA	DeepSeek
8	Anthropic	Anthropic	Amazon	Microsoft
9	Alibaba	Stability AI	DeepSeek	Stability AI
10	Stability AI	Huawei	Stability AI	Alibaba

Table 14: Top 10 Actors in Relationships in Partnership & Collaboration Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	Google	Google	Microsoft	Google
2	NVIDIA	NVIDIA	Google	Microsoft
3	Microsoft	Microsoft	NVIDIA	NVIDIA
4	Amazon	Amazon	Amazon	Amazon
5	OpenAI	OpenAI	OpenAI	OpenAI
6	Databricks	Databricks	IBM	Databricks
7	IBM	Alibaba	Meta	IBM
8	Meta	IBM	Databricks	Meta
9	Alibaba	Meta	Anthropic	Alibaba
10	Anthropic	Huawei	Snowflake	Huawei

Table 15: Top 10 Actors in Relationships in Regulatory & Compliance Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	OpenAI
2	Google	Google	Microsoft	Microsoft
3	Meta	Meta	Google	Google
4	Microsoft	Microsoft	Meta	European Commission
5	European Commission	European Union	European Commission	Meta
6	European Union	Competition and Markets Authority	Alphabet	Alphabet
7	Alphabet	Apple	European Union	Chinese government
8	Apple	US Government	Government of Canada	Baidu
9	Competition and Markets Authority	TikTok	Competition and Markets Authority	European Union
10	DeepSeek	Alphabet	Partnership on AI	US Government

Table 16: Top 10 Actors in Relationships in Services & Applications Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	OpenAI	OpenAI	OpenAI	OpenAI
2	Microsoft	Microsoft	Microsoft	Microsoft
3	Google	Google	Amazon	Google
4	Amazon	Amazon	Google	Amazon
5	NVIDIA	NVIDIA	NVIDIA	Databricks
6	Anthropic	Databricks	HuggingFace	NVIDIA
7	Databricks	Anthropic	Databricks	Anthropic
8	Snowflake	Alibaba	Anthropic	HuggingFace
9	Meta	DeepSeek	Meta	DeepSeek
10	HuggingFace	Snowflake	Apple	xAI

Table 17: Top 10 Actors in Relationships in Supply Chain & Procurement Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	NVIDIA	NVIDIA	NVIDIA	TSMC
2	TSMC	TSMC	TSMC	NVIDIA
3	Advanced Micro Devices	Microsoft	Intel	Intel
4	OpenAI	Dell Technologies	Apple	Google
5	Microsoft	xAI	Amazon	Microsoft
6	Intel	Super Micro Computer	Microsoft	Amazon
7	Dell Technologies	OpenAI	ByteDance	Dell Technologies
8	xAI	Intel	xAI	Apple
9	Google	Advanced Micro Devices	Google	xAI
10	ByteDance	Huawei	OpenAI	ByteDance

Table 18: Top 10 Actors in Relationships in Talent & Expertise Category by Centrality Measures

Rank	Total Degree	Betweenness	Eigenvector	Pagerank
1	Google	Google	Google	Google
2	OpenAI	OpenAI	OpenAI	OpenAI
3	Microsoft	Microsoft	Microsoft	Microsoft
4	NVIDIA	NVIDIA	Meta	NVIDIA
5	Meta	IBM	NVIDIA	Meta
6	Anthropic	Meta	Anthropic	IBM
7	Amazon	Amazon	EleutherAI	Anthropic
8	IBM	Anthropic	Amazon	Amazon
9	EleutherAI	Cohere	Stanford	Cohere
10	Stanford	Stanford	IBM	EleutherAI

B.1.8 Node Importance

Node importance measures the impact of removing specific actors on the network's largest weakly connected component (LCC). We first identify the 100 highest-degree nodes, then compute node importance for each of these actors by removing them from the network (along with all their incident edges) and calculating the percentage reduction in LCC size: Reduction % $= \frac{\text{LCC}_{\text{original}} - \text{LCC}_{\text{removed}}}{\text{LCC}_{\text{original}}} \times 100$. Higher reduction percentages indicate greater importance, as removing these actors causes more fragmentation of the network. The top 20 nodes by node importance are then plotted.

Repeating the analysis in Figure 3 for our targeted search list of model providers, Figure 18 shows the node importance for the model providers from our targeted search list.

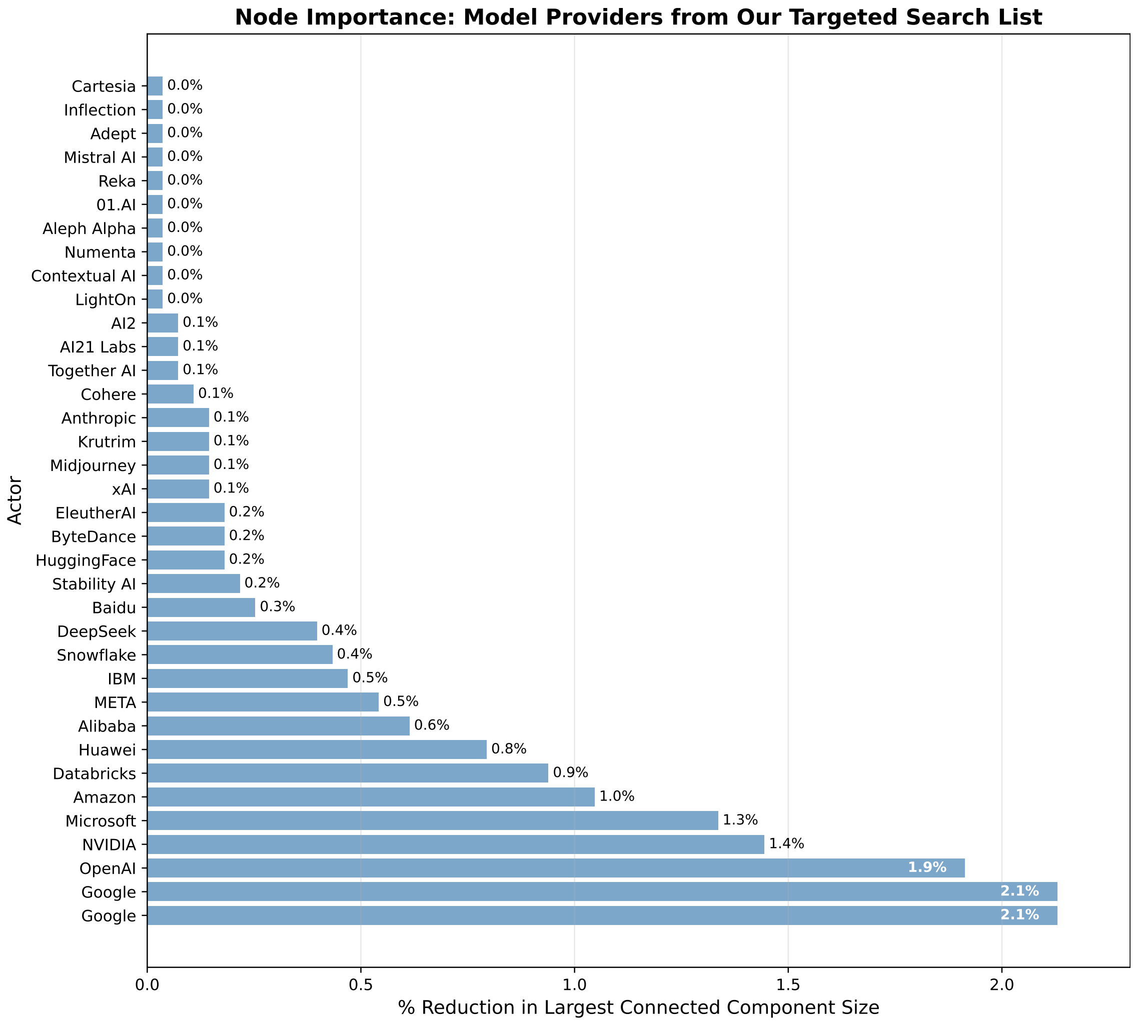

Figure 18: Node importance: Model providers from our targeted search list. The plot shows the percentage reduction in largest connected component size when each model provider is removed from the network.