On AI Deployment: AI Supply Chains

Introduction:
We need to talk about AI deployment

Sarah H. Cen, Aspen Hopkins, Andrew Ilyas, Aleksander Mądry, Isabella Struckman, & Luis Videgaray
Adapted from a post published 04/03/2023.

AI is a hot topic these days, with everyone from major publications to primetime news to late-night comedy talking about it. In 2023, AI has become mainstream—it is no longer a technology reserved for technical experts and sci-fi enthusiasts.

There is a reason that AI is receiving so much attention right now. We are in the midst of a pivotal moment, marked by the advent of generative AI systems like ChatGPT (and now GPT-4), Bard, DALL·E 2, and Midjourney. These tools are directed using simple natural language prompts, making it easy for anyone to use AI, not just engineers or researchers. Indeed, you can communicate with them much like you would with another person, asking questions like "How does Game of Thrones end?" or assigning tasks like "Design a birthday card for my friend who likes cats." Everyone can now witness firsthand the significant progress that has been made in AI over the past decade.

Still, as much as there is intense discussion around the capabilities, advantages, and dangers of AI systems, there seems to be much less focus on how these systems are put into action and by whom. That is, the issue of AI deployment is mostly absent from the conversation. This is likely to be a critical and dangerous omission.

The forking paths of AI Policy

The furor about AI has not been confined to the media or water cooler conversations. It has also captured the imagination of business leaders and policy makers. ChatGPT was among the most discussed topics in the recent edition of the World Economic Forum and, on the same

day a few weeks ago, there were hearings in both the House and the Senate on the subject of AI. These developments reflect a growing public policy interest in AI—an interest that, admittedly, took time to build up. Although people have taken notice of the advances in AI since the early 2010s, it was around 2017 that policy makers started to realize that AI is a disruptive technology with significant societal implications. Since then, over 50 countries have released AI national strategies, while dozens of documents on "AI principles" have been published by academia, NGOs, and multilateral organizations. There have also been several AI-related legislative developments around the world, including in Europe, the US, and China.

This burgeoning activity in AI policy reflects a general concern that AI impacts society and that this impact must be modulated. At the same time, however, there is disagreement about what risks AI poses and what policy responses are appropriate. We highlight two axes of disagreement: (i) disagreement over AI's capabilities and (ii) disagreement over AI's potential impact.

AI's capabilities

The first axis of disagreement concerns AI's capabilities. For example, consider the deployment of generative "large language models" (LLMs) such as ChatGPT. These models are very impressive, and yet there is significant disagreement as to whether they have the capability to truly reason (either now, or in the future).

In fact, there is no real scientific consensus on what it means for AI to be able to "reason" about the world, how we would know when AI is able to reason, or whether the current paradigm of AI techniques will ever achieve it at all. Even going beyond these somewhat philosophical questions, there are diverging views on more practical matters: when can we expect self-driving cars to be fully autonomous on public roads? Will AI eventually replace radiologists, and if so, when? Can AI foster financial inclusion in developing nations? Understanding AI's capabilities is important because it determines the degree to which AI will

intervene in society and, as a consequence, will also inform how we should prepare ourselves.

AI's societal impact

The second axis of disagreement centers on AI's potential societal impact. There is plenty of debate, for example, over whether AI will empower workers by improving their productivity, or if it will instead drive them into permanent unemployment. Some believe AI has a great potential to deliver essential services to disadvantaged communities (and countries), while others focus on AI's potential role in deepening inequities, discrimination, and social injustice. Many believe the potential for nefarious uses of AI–including for authoritarian surveillance, addictive social media, the spread of misinformation or the deployment of lethal autonomous weapons–outweighs its benefits to society. The control of AI systems, and whether we are on the path towards increased power concentration or democratization of AI, are also topics of controversy.

Takeaways

Where one's opinion lies on these two axes of AI's potential tends to inform their stance on how rapid, aggressive, and far-reaching AI policy and regulation should be (and we will explore some of these issues in our upcoming blog posts).

As important as these two axes are, however, the unique technical and economic characteristics of AI—characteristics that we will discuss in upcoming posts—suggest that AI's impact on society *will largely be shaped by the specifics of its deployment*. And, as we will see in this series, trends in current AI deployment practices do not bode well for the future. The ultimate success of AI policy will depend crucially on changing—or, at least, mounting a proper policy response to—these trends.

While there is ample consensus that AI policy should emerge from diverse and interdisciplinary perspectives, in practice the computer science and policy arenas remain quite isolated from each other. This needs to change. The following discussion, grouped by chapters, outlines a series of policy recommendations infused by science, and research directions in science infused by the reality of policy. To scope our discussion, we focus in particular on the growing trend of *AI supply chains* and their implications for researchers, organizations, consumers, and regulators.

1
AI supply chains (and why they matter)

Sarah H. Cen, Aspen Hopkins, Andrew Ilyas, Aleksander Mądry, Isabella Struckman, & Luis Videgaray
Adapted from a post published 04/03/2023.

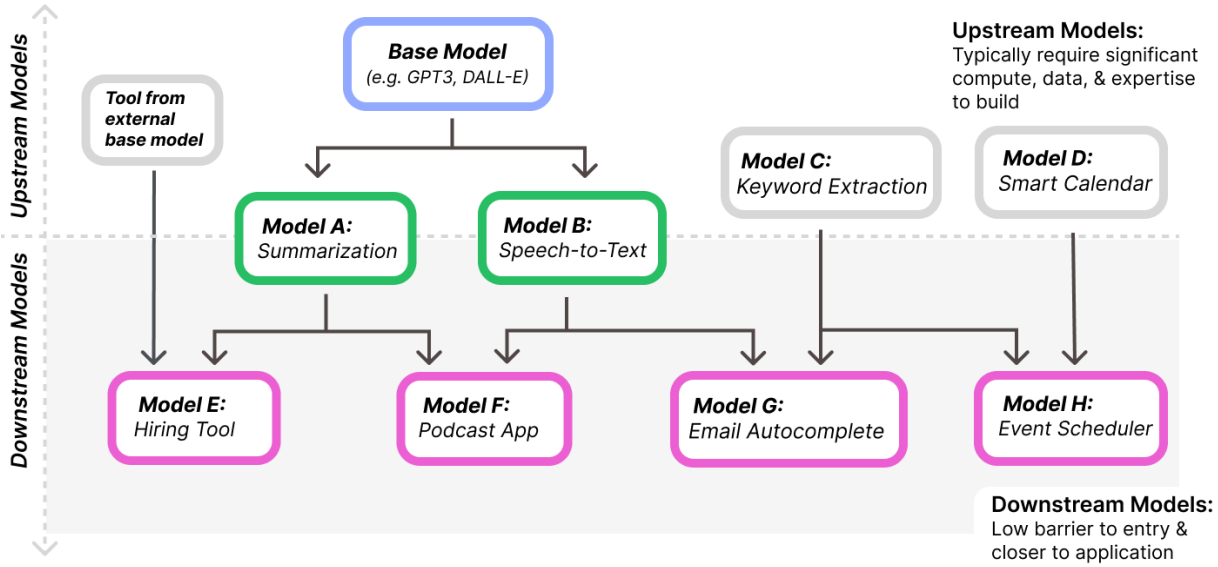
These days, AI development rarely happens "in-house." Rather, the AI system that you interact with is typically a product of many AI components glued together. One company may curate datasets that another company uses to train their AI model, which a third company then fine-tunes for a specific application (like mortgage lending decisions).

The result is a complex network of **AI supply chains.** In the past year, it has become abundantly clear that these AI supply chains are here to stay. Indeed, base models (a.k.a. foundation models) such as GPT-4, PaLM, Midjourney, Stable Diffusion, and Codex have attracted so much attention precisely because *other AI systems can be built on top of them*. GPT alone has generated an almost Cambrian explosion of startups, companies and plugins. While some are hoping to use GPT to more efficiently generate news articles, others are devising GPT-based systems to help doctors diagnose patients.

In this way, base models like GPT will become "upstream links" in AI supply chains, alongside companies that aggregate data[1]. Farther downstream the supply chain will be AI systems—such as writing assistants and diagnostic aids—that directly interface with users. And wedged in between will be various AI models, tools, and services—such as computing platforms and data-labeling tools.

1 Widder & Naffus (2022). Hopkins & Booth (2021). Bommasani, R. et al (2022).

Should we be concerned that AI is being deployed through an intricate, interdependent, and growing network of AI systems? The answer is: yes, for two key reasons.



An example of AI supply chains. AI models are built on top of another one, forming complex, multi-layered supply chains. Base models are typically at the top of a supply chain (upstream). Later in the supply chain (downstream) are AI products built using other (upstream) AI systems.

First, **AI supply chains amplify existing, known problems in AI**. For example, there are already concerns that AI systems are biased and inscrutable. However, up until now, we have only thought about these problems within the context of a *single* AI system. The bad news is: when multiple AI components are combined, these issues are often *exacerbated*. In other words, problems that exist within a single AI system are not only propagated, but often amplified in an AI supply chain.

The second reason is that **AI supply chains are likely to undermine existing efforts to regulate AI**. For example, determining who is liable for damages caused by AI will become even more difficult when there is an AI supply chain. In addition, existing approaches to supply chain regulation in other industries will not necessarily work for AI. Unless we begin unpacking the implications of the AI supply chain, we will be ill-equipped to handle a rapidly approaching future. To illustrate this point, we will spend this chapter examining several

prototypical supply chains (such as those found in the auto industry). We will then discuss three characteristics that make AI supply chains unique and surface gaps in our existing approach to AI governance.

Complex supply chains are the norm in our economy

Let's begin by taking a step back from AI and looking at existing supply chains.

Complex supply chains are not new—they're everywhere. Auto manufacturing, for example, depends on a supply chain that comprises a vast network of companies, from raw material suppliers to parts manufacturers to assembly factories. In fact, understanding the different types, drivers, and consequences of supply chain complexity—as well as how to handle them—is a major area of study in Supply Chain Management (SCM)[2][3].

Each component in a supply chain requires careful coordination and management. For example, the success of the auto industry depends on the ability of each company to deliver parts on time and ensure that each part meets the required specification and quality standards. Similarly complex supply chains can be found in pharmaceuticals, food production, and aerospace; in service sectors such as banking, healthcare, and hospitality; and in other engineering domains such as software development.

What is remarkable about these supply chains is that they (generally) *work*. One factor that contributes to their success is modularity: a supply chain can be broken down into distinct components. For example, when romaine lettuce is recalled due to E.Coli outbreaks[4], restaurants can replace the "module" that failed—that is, to remove the undesirable produce

---

[2] Complexity in the context of supply chains is understood as "a large number of different elements (such as specific technologies, raw materials, products, people, and organizational units) that have many different connections to one another," see Reeves, et. al. (2020)

[3] Campos et al. (2022), Akin Ates et al. (2021), Bode and Wagner (2015), Serdarasan (2013), Ellison et al. (2010)

[4] Why Romaine Lettuce Keeps Getting Recalled From E.coli Contamination. K. Kindy & J. Achenbach. Washington Post, 2019.

from their menu temporarily, substituting it with spinach. And when brakes are recalled in a 2006 Subaru Outback[5], Subaru knows what to provide the local dealer with so they know how, when and what to replace said brakes with. While failures *did* occur in both of these examples, the number of people impacted by it was greatly decreased through quick and explainable responses. *When a misstep inevitably occurs in a modular system, we are able to mitigate further escalation*.

Modularity is only effective because there are also *redundancies*—the second feature of a well-functioning supply chain. If a given manufacturer or transporter has a failure, others are able to pick up the slack. Beyond this, there are also industry-specific standards encouraging replicability across manufacturers; well-articulated product specifications; certifications for safety quality and environmental compliance; as well as state, federal, and international regulations regarding materials, construction, and safety.

The third reason existing supply chains work is that we can track the *provenance*[6] of each particular component of the final product. This ability makes it possible to both explain failures as well as to fix or work around such failures. Tracking provenance is an incredibly powerful tool: while we do see cases where an entire supply chain fails—as in the COVID-induced supply chain failures of 2021—these failures are still explainable, and thus are largely mitigatable.

What about AI?

So, supply chain complexity is the norm throughout the economy—it is present in car manufacturing and in the software industry. But what's the point of talking about it in the context of AI deployment? After all, policy makers have largely ignored, for example, the software supply chain[7]. So, why should the treatment of AI deployment be any different?

[5] See other Subaru Outback recalls here.
[6] Provenance here refers to tracking the history of an object or manufactured component to understand its development or origin.
[7] While this is an interesting case study on its own, we note that in many ways this was possible due to the tech industry and special interest groups creating standards early on for issues of accessibility, web development, and more. It's unclear whether this approach led to ideal outcomes.

Indeed, when AI deployment came about, it was natural to treat it as just another form of software. But it turns out that AI supply chains (at least in their current form) are very *different* from software supply chains (or any supply chains that have come before them) and will pose a unique set of risks and policy challenges.

Two key elements of AI systems: Data and algorithms

Before we delve into the three characteristics of AI that make the AI supply chain particularly unwieldy, we take a brief detour to introduce the two key elements of an AI system that we'll continually refer to in our discussion: data and algorithms.

Let's consider, as an illustrative example, an AI hiring assistant that decides, based on someone's resume, whether they should be interviewed at a particular company.

The first element that goes into building such an AI system is data (also called training data). That is, before developing a model, algorithm designers must source thousands (or millions) of resumes, as well as information about the corresponding applicants' interview performance. Such data is valuable because it contains patterns. For example, candidates with a particular related job experience or with a specific major on their resume may interview better.

It is the job of the algorithm to extract these patterns—at scales and speeds that outmatch humans—and distill them into a set of rules that can be used to evaluate future resumes. Much of the innovation in AI lies in how these patterns are extracted—a process known as model training. During this process, the model is (repeatedly) exposed to different pieces of training data—each time it is shown a piece of this data, the model updates its internal logic—logic that specifies how the model will decide when deployed whether a given applicant should be interviewed.

Now, let us return to the risks particular to AI supply chains.

AI and its supply chain

To understand just what makes AI supply chains unique, we must unpack three key differences between the AI supply chain and the supply chains that came before it:

1. Poor specification in AI,
2. The lack of modularity in AI supply chains, and
3. Hidden data interactions in AI.

These three properties amplify the risks of AI deployment and create new challenges in AI policy. Together, they highlight why the AI supply chain deserves attention from both business leaders and policy makers.

I: Poor specification

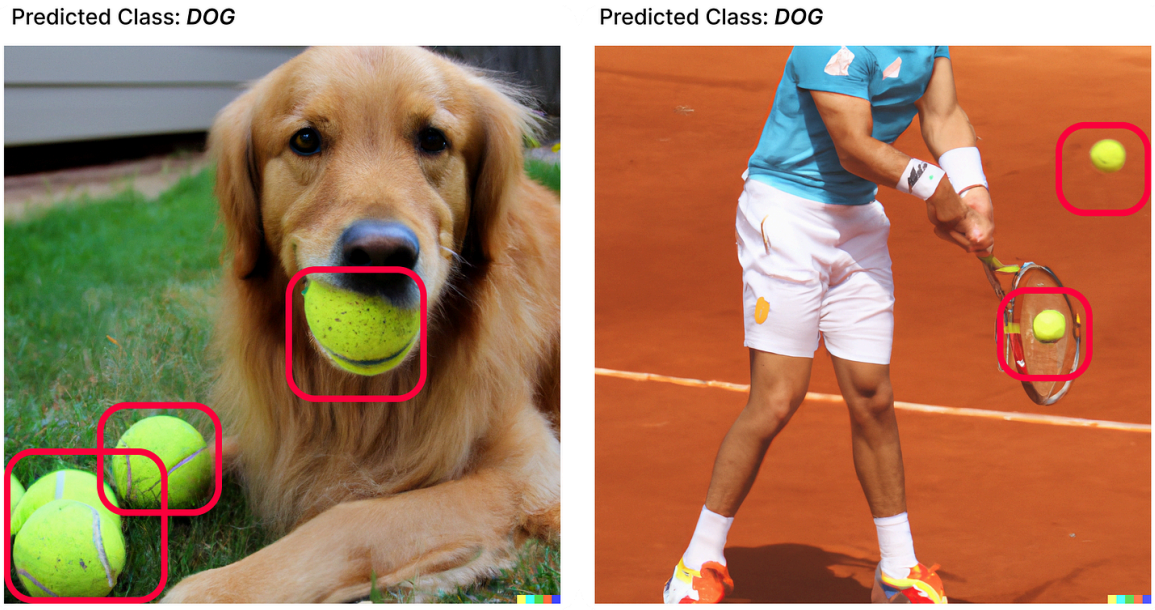
One reason why most supply chains are manageable is that their individual components are *well-specified*. That is, we know what purpose each component in the supply chain serves and how it is expected to perform (both on its own and when combined with other components).

In contrast, AI systems currently don't come with any formal specifications or provide meaningful, standardized performance guarantees. There is simply no industry-wide standard (or even a currently viable way to go about coming up with such a standard) for how to specify or audit AI systems. There is thus a wide gap between the expectations that downstream developers have when using an AI tool and what is actually observed through its deployment—this gap we refer to as "poor specification".

Poor specification[8] can contribute to many of the challenges that AI faces in the context of delivery of robust performance, explainability, and bias. For example, suppose that one company develops an AI model to caption photos, and that this company's training data only

8 Note that "poor specification" here is not to be confused with the—admittedly, related—technical term of "underspecification," which typically refers to having more degrees of freedom in a model than pieces of evidence (i.e., data). D'Amour, Heller Moldovan, et al. (2022),

contains photos captured on people's phones in everyday life. Say, however, that this training data is sourced from another company that never specifies this fact—that is, they never state (i) what type of training data they use or (ii) what criteria they use for labeling. This lack of specification from the data provider prevents the captioning company from specifying when their tool should and should *not* be used.



An example of poor specification. Consider a company building an AI captioning tool. Suppose that every photo of a dog in the tool's training data contains a tennis ball, which causes the model to associate tennis balls with the label "dog" (see the **left** photo, in which a model labels the image a "dog" due to the presence of a tennis ball). When this model is deployed in another context (e.g., a tennis match, **right** image) it will incorrectly believe that a "dog" is present. Lack of specification here can then hurt downstream performance, as the model may be inappropriately applied in the next component of an AI supply chain.

Such poorly specified systems can be problematic. In our photo captioning example[9], other models may leverage the first model as a building block, mistakenly believing that this model performs well on all types of photos. Without specification, subsequent models may propagate unanticipated errors that are difficult to trace and explain.

9 We refer to this particular case as an instance of spurious correlations.

As AI systems become small building blocks in a larger supply chain, the risks associated with poor specification will only grow, and are likely to be transmitted (and even amplified) downstream.

II: Non-modularity

Much like in the car manufacturing and agriculture supply chain examples we describe above, a key to the success of the software supply chain is its *modularity*, in the sense that its components are connected but can be clearly separated. When components are modular, connecting or disconnecting them does not change their individual attributes—just like connecting Lego bricks does not change their shape or color.

AI systems, however, tend to interact with each other in a *non*-modular way. The way that components are combined when building such a system (and even the *order* used to combine them) matters, and disentangling the role of each component after-the-fact is difficult. One can think about it almost like cooking a soup—once all the ingredients have been combined, it's hard for the recipient (or, in the case of AI, the end-user) to determine what went into the soup or how it was made.

This can be true even within a single model. After all, a single AI model is trained on many pieces of data and the way each training sample affects the model's output is often unintuitive. Although researchers have recently begun developing tools that enable one to deconstruct the influence of each sample at scale[10], there's a long way to go before we can confidently attribute model behavior to any one sample.

The problems posed by non-modularity are, however, only further amplified in a multi-layered supply chain. A prime example of this is *transfer learning*: a process in which an AI model that is trained on one task is then adapted (or *fine-tuned*) for another. (Think of an AI model that is trained to translate between English and French being fine-tuned to perform

10 Koh & Liang (2017); Feldman & Zhang (2020); Ilyas et al. (2022); Park et al. (2023)

English-Spanish translation.) This fine-tuning changes the adapted model *permanently*, making it difficult to trace the behavior of the new model back to the behavior of the first model. In some cases, we have knowledge of the data and training context that the downstream model *inherits* from upstream models. But, in many instances, we do not[11] (see our discussion of poor specification above).

III: Hidden data interactions

As we discussed, AI models learn patterns from data. In a way, data is key—without good, accurate information, a model will inevitably make mistakes. However, the problem is not just that each piece of data must be "good" on its own. It must also play well with the other pieces of data that we leverage. For example, a dataset should be *balanced*—that is, it should contain data across different settings of interest.

The interactions between data become more complicated when multiple datasets are combined. For example, two datasets might be "good" and "balanced" on their own, but misleading when combined. Indeed, suppose a weather forecast tells us it will rain tomorrow. If a second weather forecast also tells us it will rain, we become more confident that it will rain. But should we be more confident? What if it turns out that the two forecasts used the *exact same* meteorological measurements to produce their predictions? In this case, we should not have been more certain that it will rain after seeing the second forecast—we've fooled ourselves into being overconfident.

This principle extends to AI decisions, but in a subtler way. Consider a local credit union. Let's say this credit union uses multiple AI models to make lending decisions. It uses one model to predict a mortgage's performance based on an internal dataset. It then feeds those predictions (of who will default on their mortgage) into a downstream model. This downstream model considers this prediction and multiple other factors—like home valuation, income, and the

11 Salman et al. (2022)

type of loan being requested—to generate a range of reasonable interest rates for the borrower.

Here's the problem: the dataset used to train the upstream model (the mortgage default predictor) might be similar to—or even coincide with—the dataset used to train the downstream model. If it does, the downstream model may generate *biased* recommendations and may also be *overconfident* in its recommendations (similarly to the weather forecast scenario we discussed above). And this is just one of many possible hidden interactions between data that can be misleading to the model—we'll delve into more in future posts.

Takeaways and what's next

Overall, as AI systems move out of research labs into deployment, we must pay a very close attention to the emerging AI supply chains. In particular, AI supply chains will be highly problematic for (at least) three reasons: (I) AI suffers from poor specification, (II) AI systems are non-modular, and (III) there are hidden data interactions between AI systems. And these three reasons bring with them a host of other problems: AI systems will become more difficult to audit and therefore to trust. Questions of accountability will arise too. To what extent will already known problems of AI, such as hidden biases and inability to explain AI-driven decisions, compound (and get exacerbated)? Who is liable for a harmful decision made by a downstream AI model composed of many non-modular parts? All in all, it is clear that the regulatory and policy initiatives intended to ensure that AI is safe, fair, and trustworthy cannot afford to overlook the AI supply chain. We will explore all these questions in future posts.

Deploying AI safely requires careful, comprehensive, and end-to-end consideration of the AI supply chain. At the moment, there is little discussion of the AI supply chain, perhaps because complex supply chains are such an ingrained part of the software industry (the industry spearheading AI development). However, as new AI systems (including generative AI) enter the picture, the issues highlighted in this post will continue to grow.

2

Who will provide AI to the world?

Aspen Hopkins, Andrew Ilyas, Aleksander Mądry, Isabella Struckman, & Luis Videgaray
Adapted from a post published 05/01/2023.

On March 14, 2023 ("Pi Day"), three prominent developers of generative AI showcased their latest products. Google revealed it is now integrating chatbot capabilities into its Workspace apps (including Gmail and Docs); Anthropic announced a rollout of its new large language model Claude; and OpenAI revealed its highly anticipated new multi-modal large model, GPT-4. The fact that these announcements occurred on the same day illustrates the intensity of the competition in the field of generative AI[12]. Even before the "*Pi* Day" announcements, there was already a surge of venture capital interest in generative AI.

Amidst this whirlwind of AI euphoria, critical questions regarding the technology's implications remain unanswered. Some of these questions stem from our direct interaction with these novel AI tools: how will they change our cognition, our creativity, and the very fabric of our social relationships? Other questions concern the broader social and economic ramifications of AI, such as the impact on labor markets or intellectual property rights.

Today, we dissect another question, one that will be pivotal for both business strategy and economic policy: *will the global supply of base (or foundation) models be monopolized by a handful of titans, or will it flourish in a competitive market structure?*

In other words, *who will hold the reins as AI shapes the world?*

12 And possibly a general appreciation for *Pi* by the tech community!

In this chapter, we will explore two possible futures of base AI models: *healthy competition* (where there are a variety of AI providers—good for users), and *market concentration* (where only a handful of "big players" provide AI models—bad for users). We consider three factors that affect the likelihood of the latter, namely scale effects, data network effects, and platform effects.

Two Futures

As base AI models are poised to become an important component of the global economy, the evolution of their market structure will be vital to companies, investors, policymakers, and consumers. However, a future where only a few dominant players control the market looks dramatically different from one where many developers engage in intense competition.

Healthy Competition

Let us first consider the ideal–a market for base models with healthy competition between diverse actors. In this scenario, we expect *downward pressure* on the prices that downstream developers and end users are charged. Upstream actors are held to a competitive standard, reducing the possibility of rapid vertical integration[13] or harmful price coordination. It's certainly possible that in such a climate, resource-intensive base models become a low-margin commodity, as some propose. Or, instead, perhaps the competitive environment will lead to "numerous, high-utility AI systems ... [that] emerge, distinct from [single] general AI models"[14].

Market Concentration

In contrast, if the supply of base models is concentrated in the hands of only a handful of upstream suppliers, we may observe the consequences of oligopolistic or monopolistic

13 Vertical integration refers to the case where one company combines two or more stages of production normally operated by separate companies.

14 Does One Large Model Rule Them All?

behavior: price inflation, artificial availability or access limits, and an accepted norm of poorer-quality offerings. This is suboptimal for both consumers and downstream actors.

Beyond anti-competition, concentration can lead to concerns regarding reliability. First, if a sizable portion of society or an industry is dependent on only a small number of base models (controlled by two or three companies), what happens if one model malfunctions? A single failure can have disastrous downstream consequences. Second, when these failures do happen (whether they correspond to outright outages or just undesirable behavior), market concentration means that the free market will do little to compel base model developers to restore functionality quickly.

In short: ***market concentration is never good for users.*** Knowing this, what should we pay attention to in order to reduce its likelihood?

Possible Drivers of Market Concentration

For the remainder of this post, we focus on three historically relevant sources of persistent, technologically-driven market concentration: ***scale*** effects, ***network*** effects, and ***platform*** effects. We discuss how each of these effects may (or may not) present themselves in upstream AI products. The mere possibility of these effects arising suggests that AI market concentration is a genuine concern that should not be underestimated by analysts and policymakers.

Driver I: Scale Effects

The first potential driver of market concentration of AI is the ever-improving returns from scaling machine learning models and datasets.

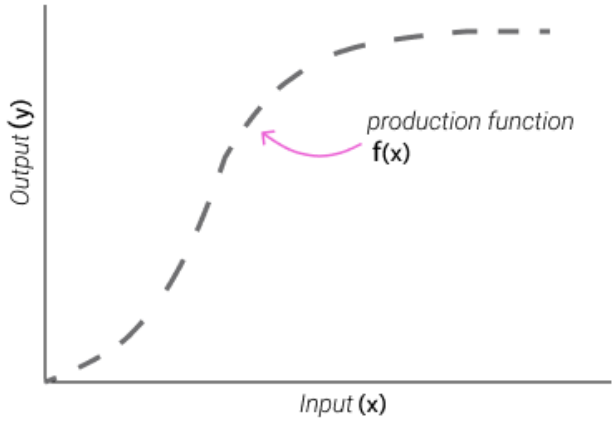
Scale has had a profound impact on AI technology. The technical underpinnings of tools like GPT-4 or Stable Diffusion were (for the most part) available five years ago. It was, however, the

enormous scale of data and compute[15] poured into these tools that brought them into existence, enabling AI companies to reap immense rewards. And these rewards do not seem to be slowing down—training larger models with more data *continues* to lead to performance improvements, defying the expectations set by analysts, economists, and even AI experts themselves.

AI's production function

Understanding the broader implications of this phenomenon requires understanding the relationship between scale and model behavior. In economists' parlance, we need to know the "production function" of AI: the relationship between economic inputs (like labor and capital) invested into an AI system and economic outputs (like quantity of goods produced) driven by the system.

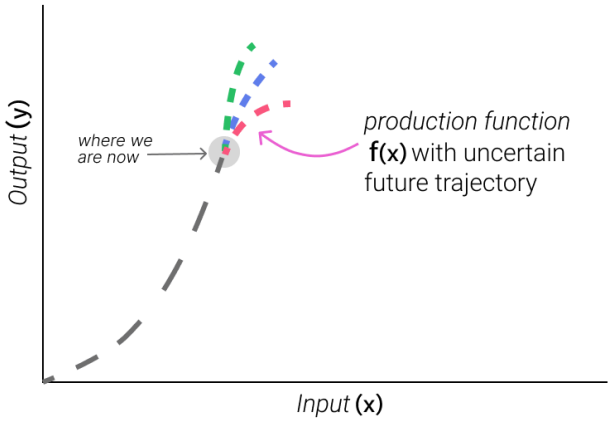
In a typical production function (like the one visualized below), *diminishing returns* naturally mitigate market concentration. That is, as a single company expends more resources on its product, the improvement in output (e.g. number of products) typically decreases. Think of a toy factory: early on, hiring more workers allows the factory to produce more toys. But as the factory gets bigger, the floor gets more cluttered and the company starts needing to instate bureaucracy to keep track of the workers—eventually, hiring additional workers does not help the factory's bottom line at all, and the production function "flattens out."



15 By compute, we are referring t

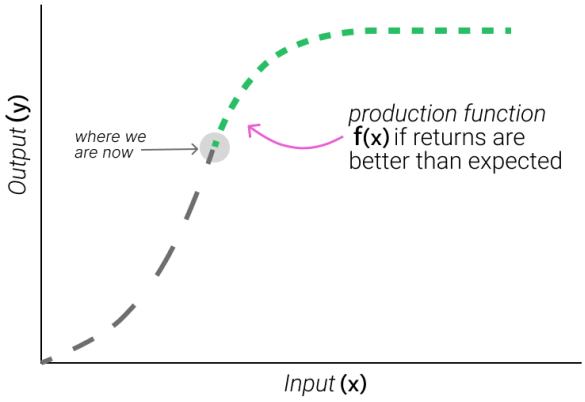
The traditional production function.

So far, these diminishing returns have not emerged in the context of AI systems. Instead, extraordinarily large-scale bets by AI companies have been paying off–improving system performance and ensuring new capabilities—incentivizing developers to continue pushing the boundaries of scale. In terms of the production function, all we've been able to observe so far is this:

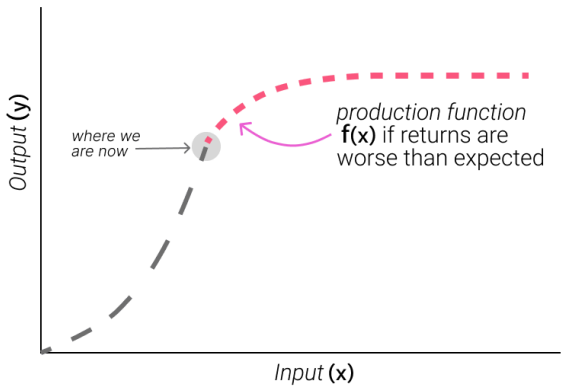


Production function showing (1) where we are in the current state of AI and base models, and (2) the uncertainty of what the future returns will look like.

We currently do not know which of two following figures the future will look like:



Production function where returns are better than analysts forecast.



Production function where returns are worse than analysts forecast.

Investing resources according to any of these future curves is a risky, expensive gamble, and one that big AI companies (already operating at the tip of the existing curve) are by far best equipped to tackle.

In other words, playing the scaling game may only be possible for a few players with both significant capital and the conviction that scale will continue to lead to better performance.

Note that market concentration here does not require AI to continue to scale infinitely, or to never see diminishing returns. In fact, we are not ruling out a scenario where diminishing returns have already started. Instead, the issue is that we simply do not know how AI will continue to scale from an economic perspective. Because of the uncertainty surrounding the AI production function, new players will (rightfully) be hesitant to enter the space, likely further entrenching existing advantages around data and computing power.

Data and compute

A common argument against market concentration appeals to *fundamental limits* around data and computing power—e.g., that there is only so much data on the internet, and as a result, the production functions above are bound to "flatten out" (and, given the estimated size of modern training sets, do so soon). However, recent developments have challenged this

presumption. Advances in speech-to-text will allow large language models to leverage audio and video data; the emergence of multi-modal models such as GPT-4 and BLIP open the door to massive new data sources; and the quality of *synthetic* data (i.e., fake data generated only to train models) is improving at an impressive pace[16].

In fact, the importance of data and compute suggests that existing moats around both might make market concentration *more* likely. Google and Meta for example (and perhaps now OpenAI, given its recent wave of hiring data-generating contractors) have swaths of private data that they've collected from other products. Similarly, cloud compute providers like Amazon (through AWS), Microsoft (through Azure), and Google (through GCP) are also naturally positioned to explore the frontier of AI scaling, having accrued years of high-performance computing equipment and expertise.

Driver II: Data network effects

A second potential source of market concentration in the upstream of the AI supply chain is the presence of *data network effects*: a self-perpetuating cycle whereby platforms with more users accrue more training data, which in turn fuels better models, thus drawing more users.

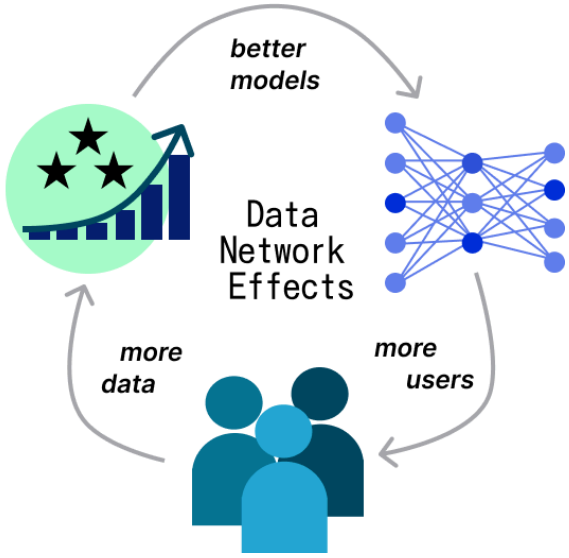
Why do data network effects matter?

Generally, a *network effect* refers to when the value of a product, platform, or service is dependent on the number of people leveraging it. Social media platforms provide good examples of network effects: if all your friends join Instagram, the platform becomes more attractive to you—even if it hasn't changed or innovated at all. Data network effects are a special instance of network effects that arise because a system *learns* from the data collected about its users[17]: the more people use a product, the more data they provide; this additional data helps the product improve (e.g. better recommendations, more accurate classification, or

16 Azizi, S., Kornblith, S., Saharia, C., Norouzi, M., & Fleet, D. J. (2023).

17 Gregory et al. (2020).

more natural chat experiences). This cycle continues and over time competitors are unable to serve users as well.



Illustrative diagram of data network effects. As increasing numbers of users use a product, more data is collected about these users. Model performance improves as a result.

Such a feedback loop can unleash "winner-take-all" dynamics resulting in one or very few companies dominating a market.

Will AI supply chains be shaped by data network effects?

There are two main conditions for base models to exhibit data network effects strong enough to drive concentration:

1. The system should learn from data gathered through interactions with users
2. The performance improvement gained from this learning should become so quickly apparent that existing and new users continue to be attracted to the system.

Once these two conditions are met, data network effects will emerge: more users will lead to better performance, in turn attracting further users, and so on.

Early base models (e.g., the very first LLMs) did not satisfy *either* of these conditions. They were trained on only historical data like Wikipedia or archived versions of pages on the internet. The inaccessibility of base models to the general public also made it difficult to obtain meaningful user data. Moreover, model improvements came from re-training the model on new *external* data, and were fairly infrequent.

Now, however, the landscape is changing. As more people start to interact with base models, companies can collect useful information, like how users rate a ChatGPT output or whether a user edits an email reply suggested by Gmail. Furthermore, newly-developed techniques like "reinforcement learning from human feedback" (RLHF) enable developers to adapt models more rapidly to this user data.

For now, as far as we can tell, companies are *not* improving base models in real time based on user data. This may be because companies (rightfully) fear the relatively new method will be prone to manipulation and abuse (e.g., users might try to coerce ChatGPT into outputting racist content by "upvoting" and "downvoting" corresponding answers). Instead, updates to popular base models like ChatGPT and Bard have come every few weeks, which—although rapid compared to previous generations of base models—may not yet be fast enough to trigger data network effects.

And the status quo may already be changing, given cases like the newly released RLHF chatbot from Stability AI. The growing availability of useful user interaction data, coupled with pressures from an increasingly competitive landscape, may provide the necessary incentives for higher frequency model updates and real-time user data integration, making data network effects–and the resulting feedback loop that leads to market concentration–a real possibility.

In summary, although the two conditions for data network effects to drive market concentration do not appear to have been met by *current* base models, this could change very soon.

Driver III: (Innovation) Platform Effects

Finally, another potential driver of concentration in the AI ecosystem is so-called *platform effects* (referred to in economics literature as *innovation platform effects*[18]), where downstream developers of AI are "locked in"—implicitly or explicitly—to a given upstream AI supplier.

What are platform effects?

Platform effects are widespread across the technology sector. Apple's iPhone, for example, benefits from platform effects through its tightly-integrated operating system iOS. Features such as one-click sharing or in-app purchases create incentives for app developers to innovate on features that adapt to the idiosyncrasies of both iOS and other existing apps in the iOS ecosystem. These incentives drive innovation platform effects: developers prefer to stay in a platform because of the synergies they find when interacting with the platform and with other applications also developed in the same platform[19].

Importantly, not every platform induces the same extent of platform effects. The strongest platform effects emerge when there are incentives to create *complementarities*, i.e., applications that work together seamlessly because of (and through) the platform.

Will AI supply chains see platform effects?

To what extent do we have to worry about platform effects in AI? Well, the recent wave of AI systems *are* platforms—they provide a common resource (a base model) on which others can innovate. Given the new announcement of Amazon Bedrock and recent rollouts of ChatGPT plugins, platform effects are likely to soon emerge. The question, then, is whether users of upstream models will create new *complementary* products and services, adapting to the idiosyncrasies of other products and the upstream models themselves.

---

18 There are multiple types of platform effects, *innovation* platform effects are most relevant to our current discussion.

19 Cusumano et al (2019).

There is potential for complementarity among the downstream uses of AI base models stemming from some of the attributes of AI systems. For example, as we discussed in our previous post, AI systems tend to be non-modular and underspecified. Such attributes imply that downstream AI systems may need to adapt to the idiosyncrasies of the base model on which they are built (and that "transplanting" an application to a competing model may be costly). It is then possible that applications developed to deal with and exploit the idiosyncrasies of the same base model end up showing some degree of complementarity among themselves. If such complementaries are relevant enough (and if the right business strategy is adopted), it would support a future where upstream base model providers become (innovation) platforms.

In summary, our analysis of platform effects leads us to the same conclusion as our analysis of data network effects. Even if the effects themselves are not yet fully realized, they have the potential to emerge, laying the groundwork for market concentration.

Takeaways

Pioneers in new industries often strive to maintain their edge through a combination of business strategies such as safeguarding trade secrets, attracting and retaining top talent, or securing preferential access to key resources. These moats are already visible among base model developers at the top of the AI supply chain. OpenAI's unwillingness to share technical details about GPT-4 is an example of this. However, business strategies alone typically don't result in lasting market concentration. Indeed, early market leaders OpenAI and Google are learning by experience that talent can migrate elsewhere (with competitors Cohere, Anthropic, and Adept founded by their former employees), and trade secrets can't be flawlessly protected[20]. While effective business strategies can lead to success and profitability,

20 Note that researchers were already able to fine-tune an open-sourced language model using text generated by an OpenAI model to get an approximate "copy" of ChatGPT.

they alone do not result in sustainable market concentration. As economic history has shown, enduring market concentration is primarily driven by technology[21].

It is too early to tell if the supply of base AI models will be highly competitive or concentrated by only a few big players. However, the possibility of better-than-expected returns to scale, data network effects, and the innovation platform potential of base model systems make it hard to rule out a concentration scenario. And what of its consequences? Of course, developers of downstream AI applications would bear the brunt of upstream concentration by having a very asymmetric relationship with their main supplier—a topic we'll explore further in our next post. But, also, the far-reaching influence of powerful upstream AI developers may span industries and nations and give rise to unparalleled concerns in economic policy, geopolitics, and national security. We thus cannot simply dismiss these scenarios and trust market forces to foster competition, particularly when the technological groundwork for market concentration may already be in place.

21 Other sources of enduring market concentration are anti-competitive practices (which are illegal in the US by virtue of antitrust law) and government-awarded monopolies (as, e.g., with local electric utilities), but these are not (yet) relevant to the context of AI supply chains.

3

Downstream AI products benefit (and suffer) from access to upstream AI

Sarah H. Cen, Aspen Hopkins, Andrew Ilyas, Isabella Struckman, & Luis Videgaray
Adapted from a post published 06/18/2023.

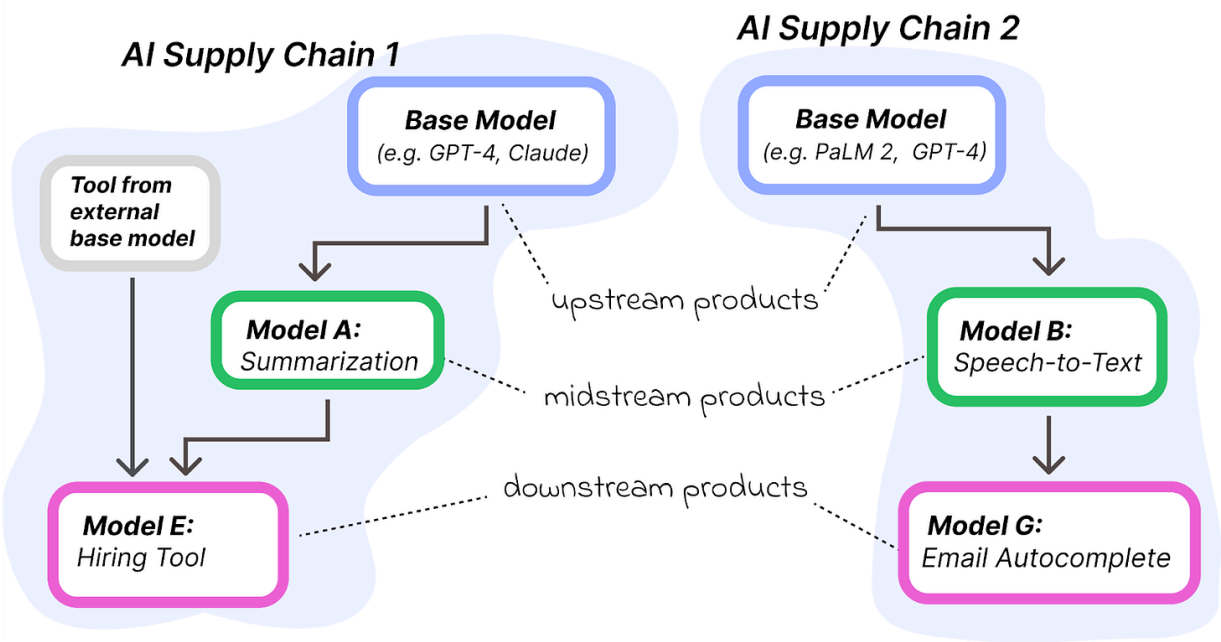
In 2023, AI is everywhere. Up until now, most of the public's attention has been on recent advances in *base models*, like GPT-4 and Stable Diffusion. But base models are only part of a much larger, more complex AI ecosystem. So, what else should we be paying attention to?

In this post, we'll argue that much of the explosive development in AI is actually happening farther along the **AI supply chain**[22]. In an AI supply chain, base models and datasets constitute "upstream" AI: they supply goods and services that are used by other companies. These other companies—that utilize upstream AI for specific domain applications—make up what we call the "midstream" and "downstream" layers of the AI supply chain.

The surge in midstream and downstream AI products matters. As companies rush to harness the capabilities of powerful, general-purpose base models and datasets for their own purposes, the AI supply chain will continue to expand. Although there will almost certainly be growth at midstream and downstream layers, the situation might be different upstream. In

22 In one of our previous posts, we introduced the notion of a complex "AI supply chain" as the prevalent form of AI deployment into the world. That is, most of the AI systems that we interact with are typically a product of many AI components glued together, therefore dividing AI companies into different layers of the AI supply chain. Moreover, in our latest installment, we published a database that illustrates the current complexity of the AI supply chain.

this post, we'll explore the implications of AI supply chain growth and, in particular, what may happen if the number of upstream players remains stagnant.



AI supply chains may include upstream, midstream, and downstream products or models.

What do we mean by "Downstream" AI?

Before we unpack the dynamics between upstream and downstream AI, let's define what we mean by "upstream," "midstream," and "downstream" AI.

While every component in the AI supply chain offers goods and services, there are some components that are higher—or more *upstream*—in the supply chain. These include base (or foundation) models—like OpenAI's GPT-4, Anthropic's Claude, and Google's PaLM 2—as well as large datasets on which base models are trained. Generally speaking, upstream components have the key characteristic that they can be repurposed in many different ways.

Farther down the AI supply chain are components that we refer to as *downstream* AI. These components are typically products that directly interface with users, such as an AI-driven

hiring tool or shopping app. There are, in addition, many components in between that we refer to as *midstream* AI. A hiring tool, for example, may be built on top of a midstream AI model that summarizes resumés. Or a shopping app may solicit a customer's opinions on a midstream dataset of AI-generated outfits[23]. For a primer on AI supply chains, see our second post.

The relationship between upstream, midstream, and downstream AI is the focus of this post. While most of the public's attention has been about upstream AI (like GPT-4), we want to discuss what will happen midstream and, in particular, downstream.

Downstream AI is becoming easier to develop

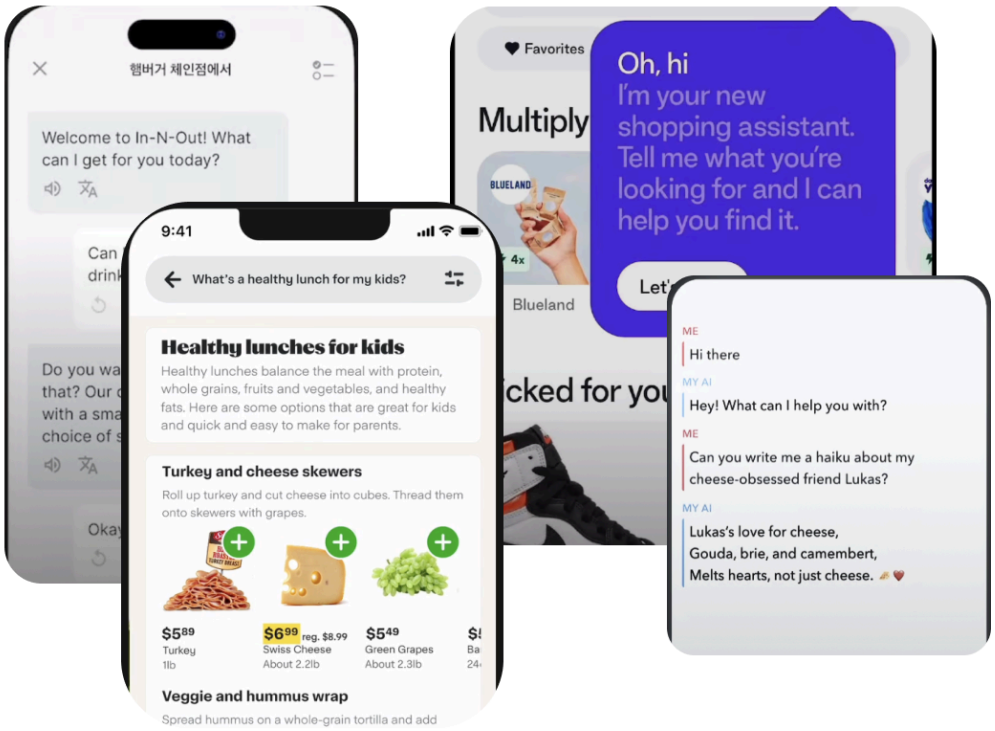
AI supply chains are making AI more accessible, so much so that AI-driven products will soon become ubiquitous. This is partially due to what's now emerging at the upstream layers. AI has previously been difficult to harness, requiring significant resources as well as expertise. However, the rise of base (upstream) models and datasets drastically reduced the cost of AI development and lowered barriers to entry. Upstream components are now doing a lot of the leg work for us. Base models are trained in a general-purpose way that allows downstream models to learn new concepts with significantly less data and compute. Similarly, base datasets provide a raw material that would be expensive to curate from scratch.

Midstream components make upstream AI even more accessible. They help to bridge the advanced capabilities of upstream models with the specific, practical applications found downstream. For example, a hiring tool may wish to use midstream models that perform intermediate tasks (like summarizing resumés or generating interview questions) rather than figure out how to use GPT-4 for the same purposes (which might require additional expertise, like prompt engineering).

---

23 An AI component does not need to fall neatly into one of the three categories. Take ChatGPT, for example. (For a primer on AI supply chains, see our second post.)

Together, these factors have made it increasingly easy to develop downstream AI. People can now harness the power of AI without any technical training, and the cost of developing an AI-driven product has dropped dramatically. So, what's next?



Early adopters using ChatGPT or Whisper APIs to develop consumer-facing products. From left to right: Speak, Shop App, Instacart, & Snapchat.

AI supply chains have a strong gravitational pull

As AI becomes easier, cheaper, and faster to harness, we will see an explosion of AI-driven products. Think of an insurance company that is limited by its ability to correctly process claims or a law firm whose main time sink is bookkeeping and administrative tasks. AI-driven software can automate many of these processes, increasing productivity and driving creativity in a way that seemed like science-fiction not too long ago.

The observation that AI will drive productivity and creativity across various industries isn't new. What's particularly noteworthy is that these AI-driven products will likely *emerge from an ever-growing AI supply chain*. That is, the vast majority of AI-driven products won't be developed independently—they'll be built on top of a long chain of AI components.

Why? For one, the rise of base models and datasets presents AI companies with a trade-off.

Barriers that previously prevented startups from succeeding will be reduced with the rise of the AI supply chain–this means that AI is leveling the playing field, allowing startups to thrive in greater numbers than before. These organizations must either invest in expensive in-house AI development or join the AI supply chain and build their products on top of upstream AI components. Many will choose the latter option, seizing the low-cost opportunities offered by the AI supply chain.

For another, new quality standards will emerge as competitors incorporate upstream AI models into their products. These standards will incentivize outsiders to opt into the AI supply chain rather than develop in-house models (as these are likely to have poorer performance). Companies that resist joining the AI supply chain will be compelled to invest in AI to keep up with the rising standards of goods and services.

The AI supply chain thus has a strong gravitational pull, leading it to expand and grow along the mid and downstream layers.

Dependence on upstream AI has downstream consequences

It's clear that the AI supply chain will grow at the midstream and downstream layers. As for the upstream layers, there are two possibilities: *competition* or *concentration*. We must pay careful attention to these two possible worlds because the AI supply chain creates what we

call a “dependency”—countless midstream and downstream players will rely on the goods and services provided by upstream components.

One possibility is **healthy competition**. A multiplicity of upstream components—including open-source alternatives—would benefit downstream AI companies for several reasons, including:

1. **Choice of upstream providers**. Under healthy competition at the upstream layers, midstream and downstream developers would have more flexibility in what they choose. If one upstream component does not suit their needs for any reason, they could turn to another.

2. **Healthy downstream competition**. Upstream competition helps facilitate better downstream competition—more alternatives upstream can lead to more alternatives downstream (if all downstream companies got their models and data from the same sources, there would be little room for downstream companies to differentiate themselves).

3. **End user experience.** Upstream competition is also good for users, as competition across the board fosters innovation, higher quality products, and lower prices[24].

The other possibility is upstream **concentration**. As explained in our previous post, there are technology-driven reasons concentration at the top of the supply chain may occur, including uncertain returns to scale, data network effects, and innovation platforms effects. If such factors result in upstream concentration, then midstream and downstream AI companies may find themselves in a *highly asymmetric relationship* with their upstream providers.

There are many possible implications of upstream concentration, including:

---

24 Even if the upstream exhibits harmful concentration, users would still benefit from competition at the downstream levels by avoiding the “double marginalization” (or double monopoly markup) problem, in which the price to the user is even higher than the price that a vertically integrated monopoly would charge (in short two monopolies across a supply chain are worse than only one). Tirole (1988), Staal Gabrielsen et al (2018)

1. **Economic consequences**. Concentration grants upstream AI providers a technological monopoly. While there are a number of economic consequences for the downstream market, three possibilities stand out.

    a. **Monopolistic rent extraction.** Base model suppliers might rely on unfair pricing strategies to "squeeze" downstream providers' profits, ensuring they fully capture the "rent" that their monopolistic power creates.

    b. **Choosing winners.** With power asymmetries, providers may give preferential treatment to select customers, raising the barrier to entry and reducing competition downstream.

    c. **Selective vertical integration.** As upstream AI providers grow, they may build competitors to existing verticals, forcing out downstream competitors.

2. **Performance consequences.** Beyond economics, there are technical implications of upstream concentration for downstream AI. First, redundancy is decreased—base models frequently experience unexpected or undocumented updates that developers must account for to maintain their product's performance. There are even instances of providers discontinuing base model services (see Codex). If there are only a few upstream providers, the effects of performance degradation will propagate to end users. Further, downstream developers currently must adapt their products to the idiosyncrasies of a single upstream model. Any modifications to an upstream model can lead to product malfunctions or unexpected behaviors that are challenging to manage.

3. **Policy consequences.** Finally, there are consequences for AI governance. For example, power asymmetries enable upstream providers to allocate liability downstream (e.g., through non-negotiable terms of service). The existing (and largely undiscussed) challenges of uncovering liability when AI is deployed through complex supply chains further encourages this outcome. Other issues such as algorithmic monoculture and

bias transfer also arise from upstream concentration, and might require dedicated policy solutions.

Generally, monopolistic practices lead to negative impacts on the availability and cost of AI capabilities that originate at the top of the supply chain. Although the emergence of large base models and datasets comes with the promise of making AI easier, faster, and cheaper, upstream market concentration would (at least partially) undermine the benefits of this promise.

What we don't know

As we discussed in our prior posts, we don't yet know what the future holds. The case has been made that developers of base models (including Google and OpenAI) do not have the technological moats to protect themselves from open-source competition (although it is unclear if this trend can continue once upstream developers stop sharing their models with the open source community). We *do* know, however, that market concentration upstream will have significant downstream effects.

It's similarly too early to tell if downstream AI markets could experience concentration in the long term. If downstream concentration does occur, we don't think the driving factors will be introduced by the AI supply chain but rather from existing, non-technical factors of existing market ownership reinforcing highly profitable or defensible positions (think domain-specific data or exclusive relationships with valuable clients).

Instead, we expect competition at the mid and downstream levels of AI supply chains will be most influenced by the decisions that companies and policy makers are making **right now** to shape the upstream market. If steps are taken to mitigate upstream concentration, competition at the mid and downstream layers of AI supply chains will be maintained.

Where we go from here

For entrepreneurs and investors, the current state of AI means new business models over-relying on the easy availability and low cost of upstream AI may not be sustainable. For policy makers and regulators around the world, this should translate to a call to action to closely monitor pricing, liability allocation, technical updates, and preferential treatment practices of upstream AI providers (and for antitrust authorities to be rather skeptical about vertical integration acquisitions in AI.

4

The Diverse Landscape of AI Supply Chains: The AIaaS Supply Chain Dataset

Aspen Hopkins, Isabella Struckman, Aleksander Mądry, & Luis Videgaray
Adapted from a post published 05/21/2023.

The deployment of base–or foundation–models and their influence on industry has grown substantially over the last few months. Yet it remains unclear the extent to which these systems will be embedded into new or existing products.

As we previously discussed, AI supply chains have existed for some time already. Even before large language models (LLMs) such as ChatGPT entered the scene, products like Google Translate could underpin different services (think automatically translating words on a webpage to encourage language acquisition, or to learn a new language while watching Netflix or Youtube). Similarly, a few years ago, Apple introduced libraries allowing iOS developers to easily embed AI tools in their apps, and now many iPhone experiences are powered by Apple's upstream AI models.

It's remained largely unclear, however, what kinds of supply chain configurations exist in the current AI ecosystem, making it challenging to implement and evaluate policy efforts in this space.

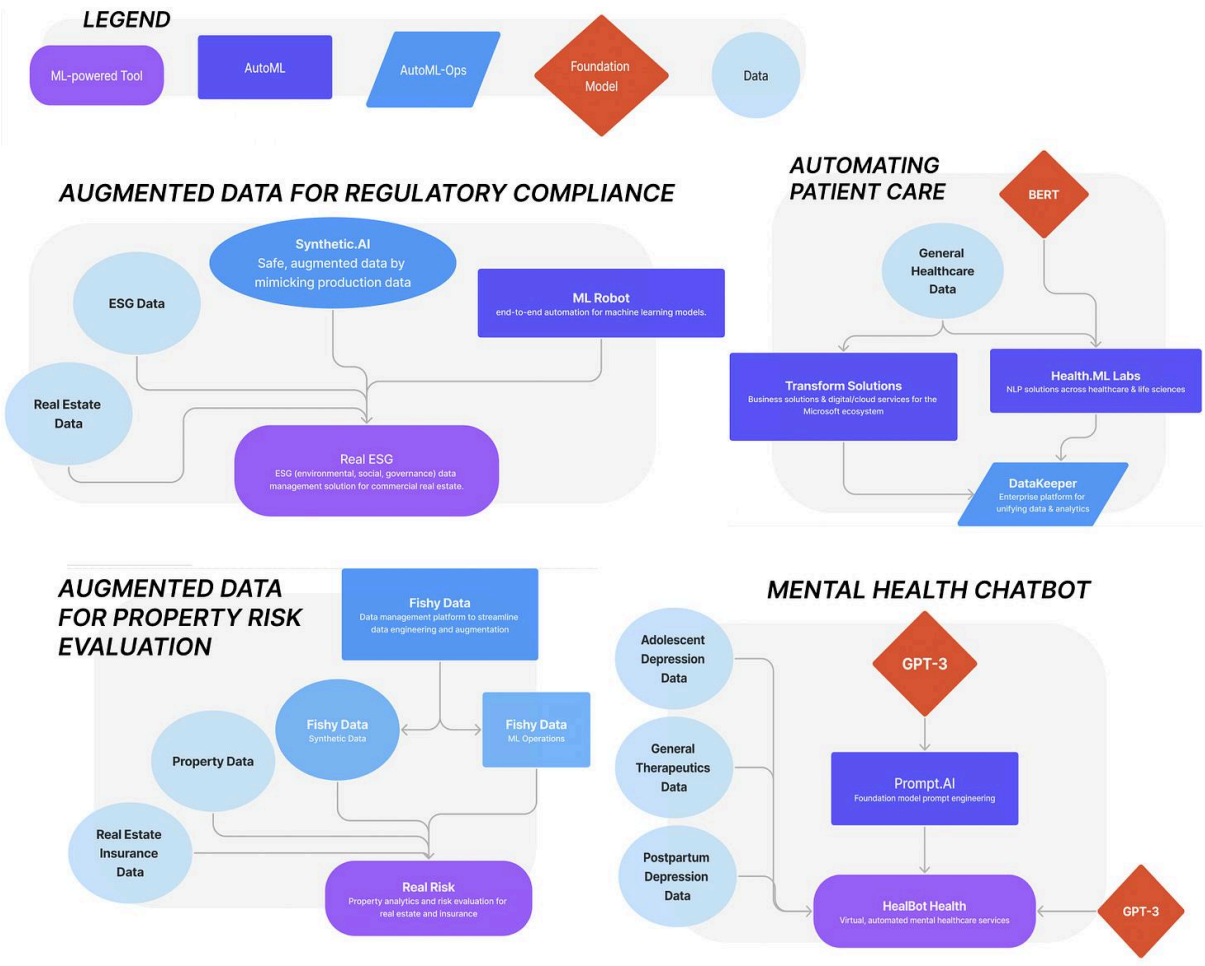
Fortunately, work like the AI Ecosystems Graphs has started to fill in some of these gaps by describing where base models are being used in larger, corporate settings. But we also need to map these supply chains out on a finer-grain level, going beyond the major players like Coca

Cola or OpenAI. Long-tail players are an important, yet often overlooked, part of the AI industry ecosystem.

To this end, the AIaaS Supply Chain Dataset was released to also encompass less prominent actors, enabling a more comprehensive view of AI supply chains. The dataset includes a wide array of both upstream organizations and products–such as base models–and downstream actors–such as AI start-ups, along with their contribution to the development of AIaaS products.

Let's delve into the dataset's specifics.

The AIaaS Supply Chain Dataset



Selected examples of AI supply chains across diverse industries. While several use common base models (such as GPT-3), others build on proprietary upstream models, datasets, and products.

Sourced largely from existing datasets built by venture capital efforts to track the AIaaS industry, the dataset categorizes companies by their contribution to the learning process of ML-powered tools. It includes AIaaS positions in the stack, categories, subcategories, industry-specific labels, company descriptions, and, of course, URLs:

1. ***AIaaS Stack Position***: Company's position within a (four-tier) AIaaS stack:

   Infrastructure, Machine-Learning Operations (MLOps), Automated Machine Learning (AutoML), and ML-Powered Tools (descriptions of each can be found in this git repository).

2. ***Category***: The type of product or services the company provides. Should offer further insight into the company's position in the AI supply chain.

3. ***Industry Labels (optional)***: Assigned to companies that are focused on catering to a particular non-AI industry (e.g., healthcare, finance, or legal).

4. ***Subcategory***: Additional details or keywords related to the company's offerings. This information is useful for quickly differentiating companies in similar roles within the ecosystem.

5. ***Descriptions***: Longer summaries of company services, built with a combination of web scraping and manual data entry.

NAME	AIaaS Stack Level	Category	Industry	Sub-Category	Description	URL	Source(s)
1010data	**AutoML**	Data Analytics fo	N/A	Large Scale Solutions (Multi	1010data's insights platform lets b market intelligence, enterprise ana	https://www.1010	FirstMark
23andMe	**ML Powered Tool**	Process Automat	Healthcare	Genetics Analytics	23andMe is a personal genetics c	https://www.23ar	Firstmark 2023
42Layers	**Infrastructure**	Infrastructure	Marketing/Customer	Data Processing	Osmos eliminates the headaches data with no-code ETL pipelines a	https://42layers.i	FirstMark

Example rows from our AI Supply Chain dataset.

Users can access the dataset using the git repo, or by accessing a Google Sheet. For more information on the dataset or how to use it, please see the Github readMe page. For readers who wish to recommend additions, please submit using the following form: https://forms.gle/gsccWsSKpjsg9UKq5.

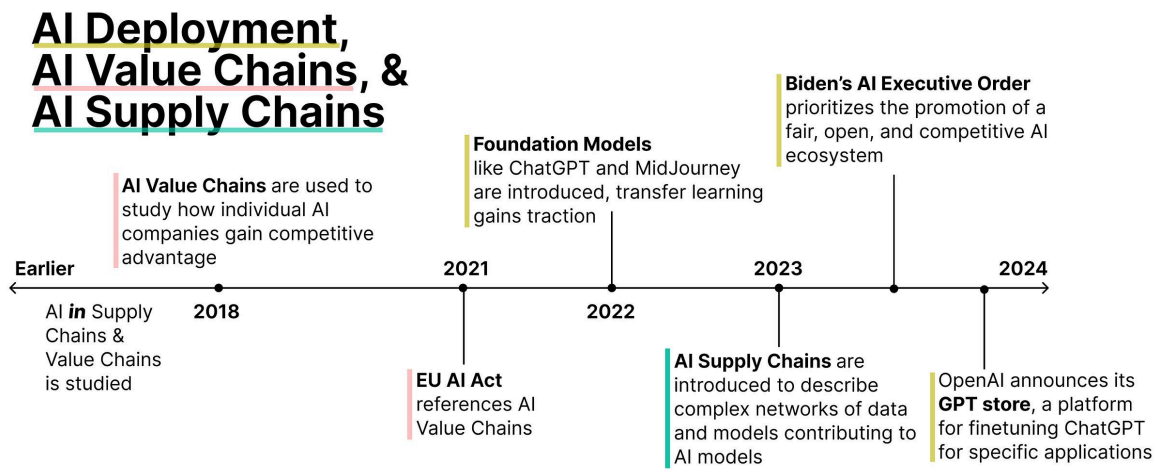
5

AI Supply Chains aren't AI Value Chains

Aspen Hopkins, Isabella Struckman, Aleksander Mądry, & Luis Videgaray
Adapted from a post published 01/19/2024.

AI development has shifted fundamentally. Before 2020, most AI systems were produced in-house with very little outsourcing. That's no longer the case. The rising influence of Base (or Foundation) Models and AI-as-a-service (AIaaS) has created an ecosystem where any organization in need of AI has easy access to the services and pre-trained models they desire. As a result, new AI experiences are often the product of a network of outsourced models, data, and tooling.

In late 2022, technologists and policy-makers alike began to pay close attention to these complex networks. The new phenomenon introduces pressing questions about monopolies, responsibility allocation, and system reliability. We discussed it at length in previous posts and called such networks "AI Supply Chains".



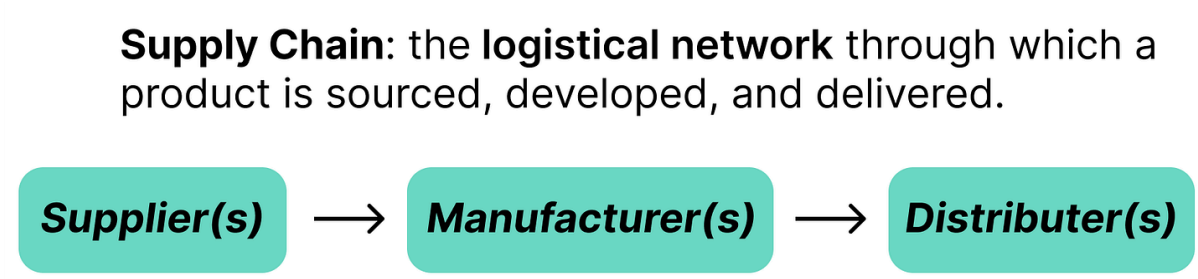
A brief history of AI deployment and value chain and supply chain terminology.

Before this, the only meaningful reference to these networks came from the EU AI Act—the first comprehensive AI legislation proposed by a major regulator. The hundred page document, introduced before AI development networks exploded in complexity, only referenced them in passing and referred to them as "AI Value Chains". Subsequent amendments have since more explicitly described "AI Value Chains" in an effort to shape regulation. But the use of *AI Value Chains* to describe the complex modes of deploying modern AI may not be ideal for the current climate (in regulation or in research). Today, we'll discuss the pros and cons of using *AI Supply Chain* versus *AI Value Chain*, distinguishing the bodies of work from which they draw upon to articulate why the **choice in language matters**[25]**.**

What is an "AI Supply Chain"?

A supply chain refers to the full network of entities involved in producing and delivering a product or service to the consumer. Supply chains are primarily concerned with the operational aspects of getting a product from the initial supplier to the end customer efficiently and effectively.

25 Another emerging term is "AI Stack". Unlike the AI Value and Supply Chains, which evoke economic and operational concepts like value-add and logistical efficiency, "AI Stack" is analogous to the software stack and simply refers to the layers of technology and processes contributing to AI systems. By referencing more neutral technical ideas, the term AI Stack avoids economic implications and fails to benefit from the wealth of literature supporting supply chain and value chain usage, which can help capture and shape the complexities of AI development networks. Additionally, alluding to the highly modular software stack may inadvertently obscure unique challenges stemming from AI's inherent non-modularity.

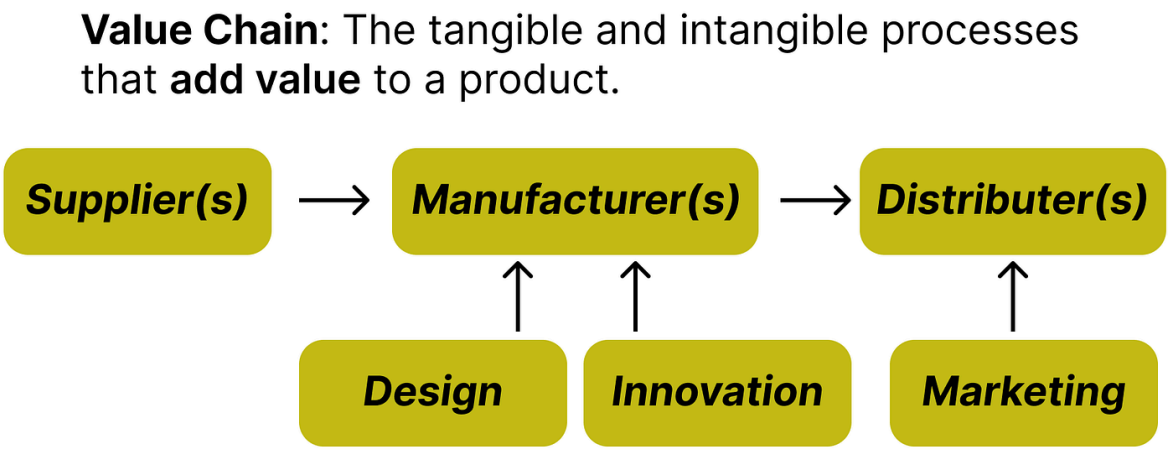


Defining supply chains.

In other words, **AI Supply Chains** model the life cycle of AI development *functionally*, focusing on the sequence of operations—the logistical network and processes involved in the sourcing, development, and delivery of AI technologies and services.

What is an "AI Value Chain"?

While ***supply chains*** are concerned with logistics and operational efficiency, ***value chains*** focus on maximizing value creation and competitive advantage.**Value chains** represent the full range of activities that a business completes in order to bring a product or service to consumers. The concept, introduced by Michael Porter in the 1980s, focuses on ***adding value*** with each step of the product life cycle within a single firm, from design, production, and marketing all the way through after-sales service. With time, the term was expanded to consider multiple firms contributing to the same product, but the emphasis on a local (not global) view of production remains.



Defining value chains.

AI Value Chains offer insight into how **various processes add monetary value** to an AI product, expanding supply chain's tighter scope to include intangible processes (e.g. design and innovation)**.** The supporting value chain literature views operations through this lens and is not specifically tailored to studying complex logistical processes.

Consequently, *AI Value Chains* entered business vernacular long before policy was needed, or before the complex AI development networks of the last year became prominent.

Why the AI Act used value chains

For many years, AI was not developed through complex networks, and language to describe the "chain" of development was not needed on a large scale. In the late 2010's, major consulting firms and think tanks began to talk about the *AI Value Chain*. Such articles explored how companies developing AI gained competitive advantage and where the market offered opportunities for growth. From a financial perspective, *AI Value Chain* was a natural term to use when studying the market.

When the EU AI Act was first proposed in 2021, AI development in industry was just becoming a collaborative, less siloed process. The proposed law refers to the *AI Value Chain* when describing the obligations and rights of participants in that collaborative process. Precedent and knowledge at the time suited this choice—it was not yet clear that AI development would expand *across* firms in such a technologically complex way as to warrant a different scope than value-add.

Later work elaborated on the EU AI Act's *AI Value Chain*, and several parties began examining and mapping the growing networks of AI development. Because AI has grown so complex, much of this recent work, still referencing *AI Value Chains*[26], was devoted to *simply understanding how AI systems are developed*—regardless of "value-add".

Here, the ideas behind the *AI Value Chain* became disadvantageous, and the term grew overloaded. As researchers work to understand the logistics of AI development networks, the monetary value of individual processes in the market becomes secondary to a technical or functional understanding of how these processes work together. Relative to supply chains, value chain literature is not well-suited to developing such a logistics-forward understanding. As a result, increasing numbers of publications are conflating value chain with supply chain—and reintroducing known concepts from supply chain research.

While the *AI Value Chain's* initial adoption within the EU AI Act may have been a natural choice, successive usage has since muddied its definition, making it more difficult to effectively draw from their respective bodies of supporting literature.

Where AI Supply Chains come in

Value chains are most helpful when their user has reasonable insight into the details of each process's value-add. Value-add is critically affected by how it interacts with related processes.

---

26 For work directly related to the EU AI Act, it might have seemed unnatural to introduce different terminology. For other standalone literature, it might have been instinctive to follow an existing precedent.

Individual organizations can understand this for their own value chains, and it may have been practical to map it at a large scale a few years ago.

But work mapping how complex AI systems flow through these development networks, how AI products work, and how they are repurposed has become the focus of researchers and policy makers alike. And, fundamentally, this interest in logistics is the bread and butter of supply chains, in contrast to the (narrow) lens of value-add. Given how early we are in the AI boom, this framing provides richer tooling for understanding complex AI systems.

To illustrate this, we've previously mapped supply chain literature to best practices and priorities for AI development. For example, **modularity in supply chains** and **redundancy in suppliers or manufacturers** are low hanging goals that we believe future regulations should adopt from supply chain literature. And this is only the tip of the iceberg. The existing body of supply chain literature—and its close ties to current regulation—is a well-established interdisciplinary field that extends from business sciences into ethics, operations research, and beyond[27][28].

Looking forward

As we begin to understand, criticize, and eventually tweak the processes of AI development and deployment, the scope of *AI Supply Chains* and the large body of related supply chain literature should serve technologists and policy makers alike.

AI systems are becoming more sophisticated and integrated across sectors. As a result, researchers are studying unique challenges in accountability attribution, security

---

27 Ellram, L.M. (1991), "Supply-Chain Management: The Industrial Organisation Perspective", International Journal of Physical Distribution & Logistics Management, Vol. 21 No. 1, pp. 13-22. https://doi.org/10.1108/09600039110137082

28 This isn't to say *AI Value Chains* aren't useful—studying value-add in AI deployment networks *is* appropriate for many contexts. Venture capitalists can use it to understand investment value, while individual organizations might gain competitive advantage by optimizing their individual value chain.

maintenance, and much more in *AI Supply Chains*. This work benefits from borrowing abundantly from language and ideas in supply chain literature.

And as the field evolves, value chain literature will inevitably continue to grow more useful in novel ways. It's critical to take advantage of both terms (without overloading either). We think standardized terminology and modes to translate between terms must be established—both for the sake of collaboration, and the future of AI policy.

6

Three proposals for regulating AI

Sarah H. Cen, Aspen Hopkins, Isabella Struckman, & Luis Videgaray
Adapted from a post published 08/07/2023

AI supply chains are quickly growing ubiquitous. Upstream providers are introducing models and datasets that are increasing access to AI—a technology that historically required great expense and expertise. Although the complex systems powering AI supply chains are poorly understood, there are several concrete and persistent concerns we must be aware of:

1. First**, upstream developers may gain substantial market power, and there are a multitude of ways that this power might be abused.** We outline several in our prior post.
2. Second**, AI products deployed through AI supply chains are difficult to audit or explain.** See our second post for more information.
3. Finally**, who bears liability in an AI supply chain is currently unclear. As a result, responsibility may be shifted down the supply chain, leaving upstream players unaffected.**

In this chapter, we argue that there are three aspects of the AI industry where early intervention is both possible *and* helpful for ensuring sustainable growth: ***fostering competition***, ***allocating liability***, and ***standardizing disclosures***. Out of the many potentially impactful steps that policymakers might take, targeting these three factors is both tractable—meaning a solution is possible—and conceptually simple, with a diverse spread of existing regulations across other industries to take inspiration from.

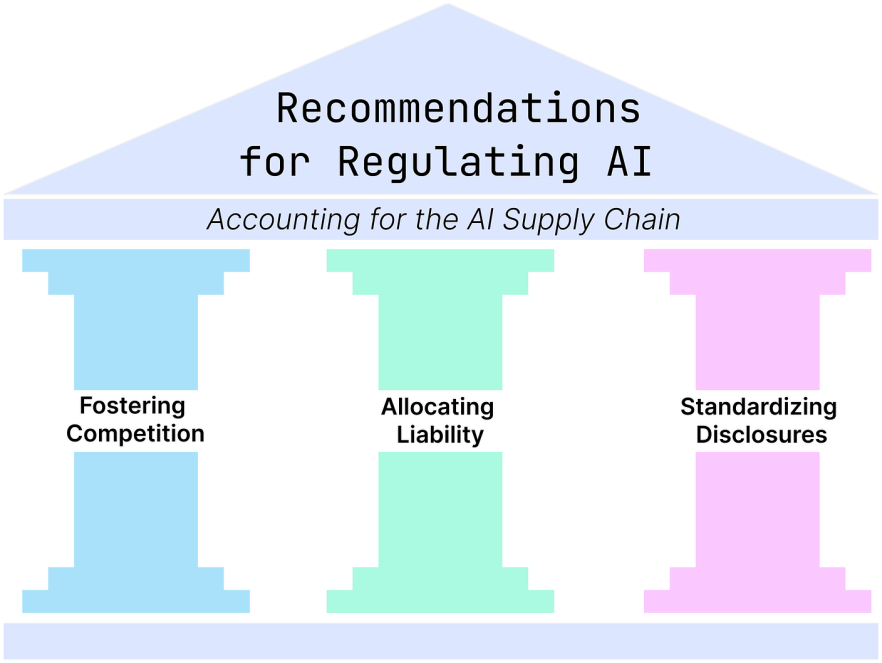
Our discussion of ***what*** to regulate (competition, liability, and disclosures) further considers ***who*** to regulate (it's complicated, but upstream providers at the very least) such that we best respond to the concerns we list above.

What to regulate

Regulating a rapidly changing technology that we don't fully understand is challenging. While we aren't certain what an ideal future for AI is, society does have clarity on what we hope to prevent, including violations of personal privacy, unexplainable decisions, and inequitable outcomes. Yet we struggle to mitigate these issues *even within a single AI system*. As a result (and accounting for differences in cultural and economic priorities), AI policy is internationally quite disparate.

On one hand, the EU has invoked a number of requirements regarding transparency, copyright, and privacy. Italy temporarily banned ChatGPT (citing concerns of privacy and underage access to inappropriate material). In contrast, Japan's focus has been on supporting the budding industry, going so far as to propose removing copyright restrictions for material used to train AI models. China has proposed targeted regulations emphasizing requirements for truthfulness and accuracy, while the UK is expanding regulations on human rights, health and safety, and competition rather than creating a new regulatory body. Most policy makers agree that both cultivating *and* moderating this evolving industry is critical, but threading the needle remains a challenge. How can policy help in scenarios where products are deployed via multiple complex AI components?

Today, we argue that there are three topics that policymakers should focus on: fostering competition, allocating liability, and standardizing disclosures.



Three pillars of policy we recommend policymakers address when regulating AI deployment: (1) fostering competition, (2) allocating liability, and (3) standardizing disclosures.

Fostering competition

In our previous posts (here and here), we considered how healthy competition in upstream providers would shape the AI industry and how the alternative (market concentration) could harm an otherwise robust mid- and downstream ecosystem. Market concentration produces a climate where upstream providers may choose to wield economic and performance pressures indiscriminately. These power dynamics make it difficult for midstream and downstream users to challenge the status quo, whether that's in regards to unfair pricing, uneven distribution of liability, or requesting transparency in upstream product updates. Market concentration also exacerbates the question of what values AI models should uphold. If only one language model is dominant, for example, all subsequent downstream models will be shaped in its image, reflecting a limited set of "beliefs".

On each of these fronts, competition can help. Currently, consumers and mid/downstream developers are limited by their options. In many cases, it's impossible to compete with the

quality of products companies like OpenAI, Google, Anthropic and others are offering. In the long-term, this means opting into a set of values or practices that only a few key players are able to shape. Instead, a robust industry is one where there are multiple options to choose from, particularly upstream. We believe policymakers can support competition in three ways.

1. **Subsidizing access to computing for small businesses.** At the moment, compute (which is necessary for model training and development) is expensive.Moreover, it is not always allocated fairly across those who need it (for instance, priority is often given to large corporations). Although there are efforts to fix this, subsidizing compute from cloud providers (like AWS) for small businesses serves to benefit the supply chain as a whole.

2. **Incentivizing the production of open-source models and datasets**, which would allow independent developers to use and modify existing code rather than start from scratch.

3. **Subsidizing data marketplaces, subject to strict terms of use.** Data marketplaces allow individuals and businesses to buy and sell data. When they are executed responsibly (e.g., while protecting user privacy), data marketplaces have several benefits. For one, individuals selling their data not only have more say over who gets their data, but are also compensated for their data. Importantly, such marketplaces can allow small businesses to source data—a resource that, for the most part, falls in the hands of tech giants like Meta and Google.

Allocating liability

Simply put, product liability[29] determines what party is held legally responsible when a failure or defect occurs in an item, allowing those impacted to seek recompense. Modern laws

29 This is just one of several types of liability, though most appropriate for the topic of AI. The EU's AI Act similarly aligns definitions of liability in AI to product liability. See the AI Act Exploratory Memorandum for more details.

regarding product liability extend purview outside of physical (tangible) consumer products to include intangibles such as gas, naturals (i.e. pets), and even writing (i.e. navigational charts).

Product liability[30] makes it possible for any party within a supply chain to be held liable, from the manufacturer of a product, manufacturers of its individual components, product assemblers, or even the wholesaler or the retailer. In exchange, affected parties must prove negligence or wrongdoing occurred.

This is where AI introduces a unique challenge. As the 2022 EU AI Liability Proposal motions to, characteristics of AI make it challenging to identify liable parties and gather the proof needed for a successful liability claim. As a result, new bodies of AI policy seem to emphasize risk prevention and management, accompanied by a favorable attitude towards potential claimants in the event of damage.

Without careful attention, this means the burden of responsibility will often fall on the "last mile"—the organization that interfaces most immediately with consumers. This allocation of responsibility is further enabled by stringent terms of service by upstream players (e.g., those that provide base/foundation models). But consumer protection and antitrust authorities have an opportunity to change this before it becomes established practice. In particular, regulators with pro-competition purview (such as the US Federal Trade Commission) can proactively use their existing authority to prevent one-sided terms of service which fully shift liability downstream. To do so effectively, policymakers must also ensure that developers—particularly, midstream and downstream players—are aware of the risks and responsibilities that they take on, as we discuss next.

30 Products liability can fall under negligence, but is generally associated with strict liability, meaning that defendants can be held liable regardless of their intent or knowledge.

Standardizing disclosures

To complement these regulatory steps, standardizing **disclosures** can protect the interests of both upstream providers *and* mid/downstream developers. Disclosures involve distributing information—including negative details—about products, corporations, individuals, investors, and legal cases to all involved parties to ensure a common set of facts are used during a decision-making process (for example, public companies typically disclose financial data to investors). Disclosures are common across industries and include, for instance, nutrition facts, warning labels, and product specifications.

More concretely, in a scenario where liability is shifted down the AI supply chain, providing appropriate context (such as when a model or dataset is updated) protects foundation/base model providers while informing mid/downstream development.

There's precedent for such a move. AI fairness research has long espoused the value of documentation to calibrate dataset and model use[31]. Metadata has become relatively common, particularly in the public sector[32], to describe how data was collected along with various facets of relevant information. And disclosures, the mode through which many industries communicate information relevant to a given scenario, are an integral aspect of day-to-day efforts. The challenge isn't in determining *if* disclosures should be incorporated into AI compliance requirements, it's *how.*

Disclosures *should* provide protection to all participating parties, but there are many modes in which they might be constructed. As a starting point, we can borrow elements from the ways that they are applied across various domains.

1. **Disclosures should maintain consistency in structure, legal requirements and language.** Such consistency means that people know what to expect, what to produce,

31 For popular examples of model and dataset documentation, see Mitchel et al (2020)'s Model Cards and Gebru et al (2021)'s Datasheets for Datasets.

32 See New York City's Metadata For All Guide as an example of municipal implementation of metadata to inform data users of relevant context.

and naturally allows for easier auditing. Consistency minimizes uncertainty for all parties—a win-win.

2. **Disclosures must balance the act of informing users with oversharing (proprietary information), prioritizing safety throughout.** To support midstream and downstream developers, and of course consumers, upstream providers should be transparent about what they know about the performance and risks of their own base models (including what they don't know). Midstream and downstream developers might similarly be asked to share context about application dependencies.

3. Finally, with the introduction of AI supply chains, **disclosures must also account for the interactions between various layers of AI supply chains.** This is entangled with the above recommendations, as it is yet unclear what information upstream and lower layers should have a right to.

We frame pros and cons of placing disclosure requirements on different aspects of AI supply chains below.

Upstream disclosures

The idea of upstream players informing mid/downstream users (e.g. through disclosures) isn't particularly surprising. The EU's proposed AI Act acknowledges this by requiring providers include performance guarantees for their models, though it's unclear what such guarantees entail. Given the state of the AI ecosystem, disclosures should include information about model performance, various dataset and training characteristics, and perhaps even performance on a set of known, published benchmarks to give approximate details that can later be used to calibrate downstream expectations.

Further, disclosures could account for any modifications to a model or dataset which influence the model's behavior. If updates to models are rolled out once a week, then there should be a notification of that change, along with a detailed comparison between previous

and updated performance. Further, upstream providers should enable access to older versions of models for some set period of time to allow downstream participants the time needed to robustly evaluate changes and transition over. These steps are critical because developers and users *build expectations of model performance*. Without appropriate notification, discrepancies between expectations and reality can lead to harmful (and avoidable) outcomes.

Midstream and downstream disclosures

While it's clear that upstream providers should inform those that use their products, it's unclear if the reverse should also occur. Should mid/downstream organizations provide disclosures to upstream providers? Should such disclosures be held to a similar level of transparency and stringency? These questions remain relatively unexplored but warrant consideration.

If disclosures act as a two-way mode of communication (where both upstream and mid/downstream organizations share salient information), then they may be doubly protective of consumers. For example, upstream providers may ask for assurances that the downstream takes appropriate precautions ensuring equitable products[33]. This reduces the risk of litigation across the entire supply chain. And, in a world where upstream market concentration may become the norm, asking mid/downstream developers to provide context upstream enables easier auditing for only a handful of companies. As a result, we might expect these upstream providers to hold their contracts accountable to some standard of safety (whether in issuing recommendations and warnings, or taking corrective actions if needed).

The challenge with this formulation is that providing information to upstream providers may unintentionally encourage vertical integration or put an unfair burden on resource-constrained organizations. While neither scenario is desirable, with careful

33 Note that we are not referring to consumers but to businesses and organizations actively participating in the AI supply chain.

consideration we might be able to thread the needle, encouraging transparency while helping to balance liability across the AI supply chain.

A side note on midstream organizations. For most of our prior posts, we've combined the midstream and downstream. This is because the midstream shares the downstream's burden of being dependent on the beneficence of upstream providers. But in the case of liability, midstream products may also face many of the concerns indicated for the upstream. How this plays out in the long term is unclear, but lawmakers should be aware of this potential conflict in status moving forward.

Looking forward: how to regulate

Our comments today focus on *competition*, *liability,* and *disclosures*. We expect that the regulation of each will depend on the domain. Moving forward, AI will largely be governed by legislation that existed before AI rose to prominence and through existing domain-specific regulatory bodies (FDA, FTC, SEC, etc). However, there are gaps that surface when adapting existing regulations and their governing agencies to AI systems, especially in the presence of complex AI supply chains. The three policy directions we introduce in this post compensate for these gaps, supporting the burgeoning AI industry while complementing existing regulations designed to ensure AI models are developed and deployed safely. While regulating AI is an enormous task, failure to consider the complexities of AI—particularly through the lens of the AI supply chain—will challenge an otherwise robust AI ecosystem, and more importantly, harm consumers.