Exploring Vulnerability in AI Industry

Claudio Pirrone*, Stefano Fricano, Gioacchino Fazio

PRELIMINARY DRAFT

Abstract

The rapid ascent of Foundation Models (FMs), enabled by the Transformer architecture, drives the current AI ecosystem. Characterized by large-scale training and downstream adaptability, FMs (as GPT family) have achieved massive public adoption, fueling a turbulent market shaped by platform economics and intense investment. Assessing the vulnerability of this fast-evolving industry is critical yet challenging due to data limitations. This paper proposes a synthetic AI Vulnerability Index (AIVI) focusing on the upstream value chain for FM production, prioritizing publicly available data. We model FM output as a function of five inputs: Compute, Data, Talent, Capital, and Energy, hypothesizing that supply vulnerability in any input threatens the industry. Key vulnerabilities include compute concentration, data scarcity and legal risks, talent bottlenecks, capital intensity and strategic dependencies, as well as escalating energy demands. Acknowledging imperfect input substitutability, we propose a weighted geometrical average of aggregate subindexes, normalized using theoretical or empirical benchmarks. Despite limitations and room for improvement, this preliminary index aims to quantify systemic risks in AI's core production engine, and implicitly shed a light on the risks for downstream value chain.

Keywords: Artificial Intelligence, Foundation Models, Industrial Vulnerability, Vulnerability Index

1. Introduction

While research on large neural networks has a longer history, going back to work on neural language models and word embeddings (e.g. Bengio et al., 2003), the key technological breakthrough that directly enabled the current generation of Foundation Models (FMs) (Schneider et al., 2024) on which all the present AI ecosystem relies upon was the 'Transformer' architecture by Google Brain team (Vaswani et al., 2017). This paper demonstrated that an architecture relying solely on attention mechanisms, without the recurrent or convolutional layers previously dominant in sequence modeling, could achieve state-of-the-art results in tasks like machine translation, while being more parallelizable and requiring less training time. This scalability was crucial for training the massive models that followed. This advance was so disruptive to gain the reputation of 'foundational', as popularized by the Stanford Institute for Human-Centered Artificial Intelligence (Bommasani et al., 2021). Broadly speaking, FMs are characterized as large-scale models developed through training on extensive datasets, typically employing self-supervised learning at scale. Their key attribute is adaptability, enabling them to be effectively specialized by dedicated apps for a wide spectrum of downstream tasks. To the best of our knowledge, the first FM released for developers was GPT API from Open AI (Radford et al., 2018), and first actual service available for the public was ChatGPT (based on GPT 3.5), released on November 30, 2022, as a "research release". As it is well documented, the adoption of FM based services was, and still is, tumultuous. According to Digital 2026 Report (We Are Social & Meltwater, 2025), 26.5% of internet users aged 16 years of more used ChatGPT at least once in a month in Q2_2025, and this share is growing in double digits quarter over quarter. In absolute numbers, it accounts for 557 million users in August 2025, while ChatGPT competitors were far behind, with 246.5M users cumulating Gemini, DeepSeek, Perplexity, Grok, Copilot, and Claude. Uncountable applications are under continuous development using FMs API, both under commercial license or open-source. For an order of magnitude, estimates from GWI highlights that AI based tools are used by around 54% of workers, share that rises until 75% if people interested in integrating AI tools in their working routine are added. Still, the market is turbulently changing under competitive forces and scale returns, echoing platform economics (Rochet & Tirole, 2003; Alstyne & Parker, 2017; Spulber, 2019). For instance, while we are writing this paper, Open AI is releasing Atlas, its GPT-powered search engine and introducing sponsored content for their free-to-use customers, with the clear intent of breaking the Google's de facto monopoly. From a financial point of view, AI ecosystem seem drawing around half of all venture capital investment in 2025 (Figure 1), raising concerns about the emergence of a new market bubble (Fang et al., 2025). And, for how FMs impact everyday economic activity, new or improved agents are released in a continuous flow. Under such conditions, exploring the vulnerability of the AI industry is both

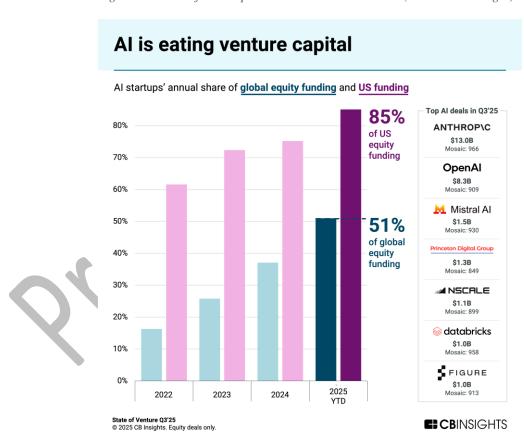


Figure 1 - Share of AI companies in Venture investments (source: CB Insight)

essential, and a conundrum. It is essential, because such surge of interest underline a real demand for productivity and effectiveness. It is also a conundrum, because the industry evolves too fast for allowing to establish consistent data series upon which validate assessment models, as well as microdata availability is scarce, and interest-shaped narratives are opaque. In order to overcome this contingence, we suggest to restrain the focus on the upstream value chain of FMs. Our ambition is to

suggest a Vulnerability Index (VI) for this segment of the industry, well aware of need for continuous refinement and improvement. In conceiving our VI, we prioritize publicly available data to maximize transparency and debate.

2. The Conceptual Framework

As first step, we posit a general production function for FMs, identifying five relevant 'inputs'

$$FM_{output} = f(Compute, Data, Talent, Capital, Energy)$$

As for identifying compute, data, and talent, we rely on CMA (2023, p. 10) analysis of the AI value chain. We add capital, due to high investments required for producing effective FMs (Fang et al., 2025; Maslej et al., 2025). We also add energy, due to raising concerns expressed in the literature (de Vries, 2023; Maslej et al., 2025). Compute indicates the increasing need for computational power in order to run FMs (Maslej et al., 2025, p. 12). Data indicates the need for new data in order to improve FMs training (Villalobos et al., 2024). Talent refers to the availability of high-level competences which are needed to run and improve FMs (Bone et al., 2024). Capital points the highly capital intensity of the FMs industry. Energy relates to increasing energy needs for running FMs inferences and training (Maslej et al., 2025). The hypothesis retained here, is that a vulnerability in the supply of any of these inputs represents a critical danger for the industry prosperity.

2.1. Compute

This is the most analyzed and tangible vulnerability. The 'production' of an FM is an act of massive computation, which relies on an upstream market with extreme concentration. Indeed, the FM 'factory' is a downstream buyer facing a near-monopoly supplier (NVIDIA) and an oligopoly fabricator (TSMC, Samsung). As for NVIDIA, it's market dominance (Maslej et al., 2025) in advanced GPUs (its Blackwell and Rubin architectures are the *de facto* standard) allows it to engage in monopoly pricing and extract rents from the entire value chain. FMs labs have no credible alternative for training frontier models, making their production costs vulnerable to NVIDIA's pricing strategy. In addition, the fabrication of these advanced chips is geographically concentrated, primarily in Taiwan, Singapore, and Hong Kong (Papadopoulos & Magafas, 2025). This exposes the entire supply chain of FMs to geopolitical instability (Miller, 2022). The capital cost and technical expertise required to build a competing semiconductor fabrication plant or design a competing GPU are considerable, reinforcing the incumbents' market power.

2.2. Data

While the performance of FMs steadily grew in last years, there are still request for improvement. So-called hallucinations, as well as complex reasoning and implicit bias, are significant issues. Improving the performance of a trained model, requires a continuous flow of new, high-quality data. However, if data in the internet era were perceived as a common-pool, virtually infinite, resource, digging so deep in the mine transformed data into a scarce, contested input (CMA - Competition and Market Authority, 2023). Several analyses (e.g. Villalobos et al., 2024), including from the FM labs themselves, have concluded that frontier models are approaching the limits of

high-quality public text data available on the internet. Continuous improvement is now facing diminishing marginal returns from data (Maslej et al., 2025), and 'autophagy' is an issue (Xing et al., 2025). This strife for data also leaded to legal issues. Scraping the public web for private profit scope is seen as an act of unfair appropriation that is now being aggressively challenged. Lawsuits (e.g., *The New York Times vs. Open AI & Microsoft*) threaten the legal foundation of this data supply chain. A ruling in favor of content creators could retroactively poison existing models or force licensing fees, fundamentally altering the production cost. Finally, continuous improvement also relies on fine-tuning from users' feedback (Reinforcement Learning with Human Feedback - RLHF). This category of data is perceived as critical for the performance of AI models (Sarikaya, 2025). Of course, RLHF data are privately owned by the FM firm. Bigger players benefit of their market dominance creating a barrier to entry to new companies which cannot rely on comparable RLHF databases.

2.3. Talent

FMs are not just produced by capital and data; they are designed by a relatively small, elite group of researchers. Exact information about talent availability is scarce. For instance, Stanford University tried estimating the AI talents, without distinguishing whether they relate to FMs or to agents and applications benefitting of FMs via API. Yet, they had to use LinkedIn profiles as a proxy, information which is affected by selection and declaration biases (Maslej et al., 2025). Even with these limitations, it is clear that demand for AI talents grows at higher rate than their availability on the job market. Also, the global demand for AI talents by industries and developers of expected high-value applications, reduce the availability for FMs industry. This can indirectly be observed by the geographical distribution of hiring offers worldwide, which more dispersed that FMs laboratories. Other indirect witness of scarcity can be found in the level of salary paid to top AI engineers (up to 10m/y, according to the Financial Times). This a hint for a 'superstar' race (Rosen, 1981) which happens when the specific talent required is a narrow niche. Finally, estimation from MacroPolo (The Global AI Talent Tracker 2.0, data relatives to 2022) note that 57% of top-tier AI researchers work in the US, followed by China (12%), and UK (8%). Also, according to this estimation, top-tier AI talents are less and less internationally mobile, which can generate bottlenecks and make difficult for foreign newcomers in AI industry to attract well-skilled talents.

2.4. Capital

Frontier FM production is one of the most capital-intensive R&D processes in human history. This creates a vulnerability in the supply of capital. The case of Open AI is paradigmatic: despite tumultuous gains in revenues, it still burns millions (2.5 M\$ in H1 2025, according to Reuters, compared to 4.3M\$ in revenues, up 16%), due to investments and operational costs of their FM. According to Maslej et al. (2025) citing the CEO of Open AI, it costed more than 100M\$ to train GPT4. Because the cost is so high, capital cannot be sourced uniquely from traditional venture capital. The role of strategic partners, especially the cloud hyperscalers (Microsoft, Google, Amazon, ...) is highly important. From publicly available sources, Microsoft invested over 10B\$ to support Open AI in first stage of its life, while Google and Amazon heavily supported Anthropic. One relative exception to this path is Mistral AI which, according to Tracxn, received small financing from Microsoft, and rely more on partnership with manufacturers as Samsung, NVIDIA,

and Cisco. These strategic dependencies can make FMs companies vulnerable to the strategy of their partners which, on reverse, have more flexibility. From this point of view, the choice of Open AI to launch its own service of search engine, as well as to attack Google on its primary market of online ads can be seen as a do-or-die move.

2.5. Energy

Analysis from De Vries (2023) analyses project that, to maintain the current trajectory of artificial intelligence (AI) capacity expansion and adoption rates, NVIDIA would be required to produce approximately 1.5 million AI server units annually by 2027. Operating at full computational capacity, these systems are estimated to demand between 85.4 and 134.0 terawatt-hours (TWh) of electricity per year, which is a significant share of the worldwide consumption. This rapid escalation in infrastructure requirements underscores the growing energy and resource intensiveness of large-scale AI systems. As foundation models (FMs) continue to increase in scale, parameter count, and architectural complexity, their corresponding computational and environmental costs are expected to rise disproportionately, posing substantial challenges for the sustainable development, deployment, and governance of next-generation AI technologies. This evolution is confirmed by Maslej et al. (2025). Indeed, they acknowledge that energy efficiency is improving, but the total energy consumption and related CO₂ emission are growing at a steady pace.

3. The Assessment Model

Once the main potential sources of vulnerability, before establishing a computable model we have to understand in which manner the five inputs interact in our general production function. Broadly speaking, we have if they are substitutes, complements, or imperfect substitutes. Respectively, this would lead to additive, min-max, or combinatory assessment of each dimension. This also allow for modularity, which is a desirable feature when discussing about a changing industry.

3.1. How inputs interact

From previous discussion, it emerges with clarity that our selected inputs cannot be considered perfect substitutes. Each of them has a specific role in the production function, so additive modality of aggregation, even weighted, cannot be applied. In trivial words, we cannot accept the simplistic hypothesis that, for example, a unit of talent can, under any condition, replace its weighted equivalent in unit of computation, and reversely. The autophagy issue discussed before is an example of the tradeoffs between compute and data inputs. However, it is also evident that, to some unknown extent, input can partially substitute each other. Talent can provide improved algorithms, reducing computational need. Research can provide new processors, more efficient in energy, partially reducing energy needs. And so on. Thus, we will accept the two following hypothesis. Inputs cooperate in a combinatory way to determine the output of the system. If any input fall to zero, output will equally become null.

3.2. Consequences for index mathematics

A standard index with partially substitute inputs will have the following general form

$$Index = \prod_{i=1}^{n} Sub-Index_i^{w_i}$$

where: i represents each input of the production function; Sub-Index_i is the specific index for the input i; and w_i is the weight assigned to the input under the following constraint

$$1 = \sum_{i=1}^{n} w_i$$

Applying naively this approach will lead to nonsense. Indeed, if vulnerability for one input is null, the global index will be equally null. Consequently, we define

Vulnerability =
$$1 - Potential$$

Where potential denotes the inherent capacity of the input to sustain the production function. Consequently, the aggregate AI Vulnerability Index is defined as follow

$$AIVI = 1 - \prod_{i=1}^{n} Pot_{Sub-Index_i}^{w_i}$$

with

$$0 \leq \left(\prod_{i=1}^{n} Pot_{\text{Sub-Index}_{i}}^{w_{i}} \right) \leq 1$$

which better fit our goals and the nature of the described production function.

3.3. Constructing *Pot*_{Sub-Index};

Last condition is assured if all subindexes are also fractional values between 0 and 1. Weighted average of normalized components will ensure the condition is respected, and allows for merging information expressed in different units of measure. In this exploratory step, we adopt a standard normalization algorithm, where

$$N(x) = \frac{x - \min(x)}{\max(x) - \min(x)}$$

max(x) and min(x) coming from theoretical limit values, as in Herfindahl-Hirschman Index (HHI) for market concentration, or from empirically defined limits, when data are available. Here again, we have to choose between arithmetic or geometric averaging the raw components in subindex. While questionable, our opinion is that adopting weighted simple averages is preferable for this layer of analysis. First, within the limits of available information and critical points awareness, it is possible to select components which are less entwined than our general 'inputs'. Second, arithmetic averages are less sensitive to errors in weights than geometric averages, which is a useful feature.

3.3.1. Compute $Pot_{Sub-Index_i}(I_C)$

Relying on what is expressed in §2.1, we identify four indicators which are both meaningful and likely to be computable with publicly available data

Indicator	Topic	Description	Expected Data Source
(HHI_{fab})	Market	Herfindahl-Hirschman	Company annual reports
, , , , , ,	Concentration	Index for the leading-edge	(revenue for TSMC, Samsung,
		semiconductor fabrication	Intel). Market share data from
		market.	industry reports.
$(GeoC_{fab})$	Geographic	Percentage of leading-edge	Industry analysis reports,
	Concentration	semiconductor	government reports (e.g., from
		manufacturing capacity	the U.S. Department of
		located in a single	Commerce), and reputable
		geographic region	news outlets often cite these
			figures.
(HHI_{design})	Design	HHI for the GPU/AI	Company revenue reports for
	Chokepoints	accelerator design market	relevant segments.
(TD_{chips})	Trade	Total volume of chips	UN COMTRADE database
	Dependency	crossing borders from a few	(using relevant HS codes for
		key locations.	integrated circuits).

Components are aggregated as weighted average:

$$I_{C} = 1 - \left(w_{C1} \cdot N(HHI_{fab}) + w_{C2} \cdot N(GeoC_{fab}) + w_{C3} \cdot N(HHI_{design}) + w_{C4} \cdot N(TD_{chips})\right)$$

$$3.3.2. \text{ Data } Pot_{Sub-Index_{i}}(I_{D})$$

Relying on what is expressed in §2.2, we identify two indicators which are both meaningful and likely to be computable with publicly available data

Indicator	Topic	Description	Expected Data Source
(S_{data})	Data Scarcity	A proxy for data exhaustion,	AI Index Report, academic
		measured by the deceleration in	papers.
		the growth of training datasets	
		for frontier models.	
(LV_{data})	Data Input Cost	The rising cost of data inputs,	AI Index report, academic
		proxied by the total publicly	papers, company
		announced value of data	information.
		licensing deals for AI training.	
		We use this parameter also an	
		indirect manifestation of legal	
		issues on data access.	

Components are aggregated as weighted average:

$$I_D = 1 - (w_{D1}(S_{data}) + w_{D2} \cdot N(LV_{data}))$$

3.3.3. Talent
$$Pot_{\text{Sub-Index}_i}(I_T)$$

Relying on what is expressed in §2.3, we identify three indicators which are both meaningful and likely to be computable with publicly available data

Indicator	Topic	Description	Expected Data Source
(HHI_{talent})	Elite Talent	The HHI of elite AI	AI Index Report, MacroPolo,
	Concentration	researchers by their	academic papers.
		corporate affiliation.	
(RC_{pubs})	Research	Percentage of top-tier AI	Conference proceedings,
	Concentration	conference publications	arXiv, and university reports
		affiliated with the top 5	like the AI Index.
		corporate labs.	
$(HHI_{patents})$	IP Concentration	HHI of granted patents in	Google Patents, USPTO bulk
		key AI categories	data.

Components are aggregated as weighted average:

$$I_{T} = 1 - \left(w_{T1} \cdot N(HHI_{models}) + w_{T2}(RC_{pubs}) + w_{T3} \cdot N(HHI_{patents})\right)$$

3.3.4. Capital
$$Pot_{Sub-Index_i}(I_K)$$

Relying on what is expressed in §2.4, we identify two indicators which are both meaningful and likely to be computable with publicly available data

Indicator	Topic	Description	Expected Data Source
$(B_{capital})$	Capital Barrier	The estimated monetary cost of	AI Index Report,
		training a single state-of-the-art	academic papers, and
		(SOTA) foundational mode	industry announcements.
$(HHI_{dependency})$	Strategic	The HHI of strategic capital	Crunchbase, SEC filings,
	Dependency	invested in the top FM labs,	and major news reports.
		measuring the concentration of	
		funders, who are also cloud	
		providers	

Components are aggregated as weighted average:

$$I_K = 1 - (w_{K1} \cdot N(B_{capital}) + w_{K2} \cdot N(HHI_{dependency}))$$

3.3.5. Energy $Pot_{Sub-Index_i}(I_E)$

Relying on what is expressed in §2.5, we identify three indicators which are both meaningful and likely to be computable with publicly available data

Indicator	Topic	Description	Expected Data Source
(GR_{energy})	Energy	Growth rate of AI related	Reports from IEA, IDC and
	Consumption	energy consumption	Webscale AI
(GR_{co2})	Emissions	Growth rate of AI related	IEA (proxy with data centers)
		CO_2	
(EB_{energy})	Energy efficiency	A proxy for the efficiency	Technical report and analysis
		barrier, measured by the	(ex. Epoch AI)
		deceleration in	
		improvements in AI related	4.0
		energy efficiency	

Components are aggregated as weighted average:

$$I_E = 1 - (w_{E1}(GR_{energy}) + w_{E2}(GR_{co2}) + w_{E3}(EB_{energy}))$$

Because any component is expressed as a rate, no normalization is required

4. Limitations and Perspectives

To the best of our knowledge, this is the first attempt to provide a synthetic index of vulnerability in AI industry. However, our exploration shows some limits that have to be addressed in future developments.

4.1. Modelling

In order to build our assessment model, we relied on most recent academic and non-academic sources on information. Nonetheless, AI industry is a fast evolutionary industrial sector. Consequently, there is significant room for improving and consolidating the model as knowledge about specific vulnerabilities progress. This applies with force to the components of subindexes. We are confident that running factual estimations, linked with a constant attention to the literature and to the comments from the scientific community will give us the information we need to achieve the goal to provide a useful, meaningful, and consistent index.

4.2. Data quality

As we stated in introduction, good data for our purpose are scarce. In such condition, privileging publicly available data allows the community for control and suggestion for improvement. As for us, we are well aware of the need to constantly search for better data, in order to improve and consolidate the database of reference. Only this consolidation will allow for intertemporal comparison and will open room for more geographically specific analysis.

4.3. Weighting

We acknowledge that AIVI is highly sensible to the way different subindexes and components are weighted. In particular, given the dynamic nature of AI industry, we expect 'true' weights to change over time. From our point of observation, ensure consistent weighting is more an open field of research than a mere technical exercise, and it is the priority for next months in order to provide a preliminary estimation of AIVI. Hypothesis for first draft may be equal weighting, expert weighting, AI driven weighting, and model-based endogenous weighting, if data will allow for it.

5. Conclusion

In the transformation of current global productive system, AI is increasingly becoming the core innovation that pushes evolution. Its direct and indirect impacts are consensually considered as transformative, leveraging on computational capacity and stimulating radical innovation in computing systems, in services, and in industrial process organization. However, many voices are raising concerns about bottlenecks and chokepoints that may hinder this transformative process. In this preliminary paper we build up a synthetic index which assess the vulnerability of the core engine of present AI industry, the foundation models. With all limitations expressed before, this is a first step to better understanding the risks for AI industry and, given its downstream value chain, for the global economy as well. Future step will be to release a first numerical estimation of AIVI and to propose it to the scientific community.

REFERENCES

- Alstyne, M. W. V., & Parker, G. G. (2017). Platform Business: From Resources to Relationships. *NIM Marketing Intelligence Review*, 9, 24–29.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3(Feb), 1137–1155.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). *On the Opportunities and Risks of Foundation Models* (No. arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258
- Bone, M., Ehlinger, E., & Stephany, F. (2024). *Skills or Degree? The Rise of Skill-Based Hiring for AI and Green Jobs* (No. arXiv:2312.11942; Version 2). arXiv. https://doi.org/10.48550/arXiv.2312.11942
- CMA Competition and Market Authority. (2023). *AI Foundation Models: Initial Report* (p. 129). https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf
- de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004
- Fang, X., Tao, L., & Li, Z. (2025). Anchoring AI Capabilities in Market Valuations: The Capability

 Realization Rate Model and Valuation Misalignment Risk (No. arXiv:2505.10590; Version 2). arXiv. https://doi.org/10.48550/arXiv.2505.10590

- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., ... Oak, S. (2025). *The AI Index 2025 Annual Report* (No. arXiv:2504.07139). AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.
- Miller, C. (2022). Chip War: The Fight for the World's Most Critical Technology. Simon and Schuster.
- Papadopoulos, G. D., & Magafas, L. (2025). The global arms and integrated circuits trade through complex network analysis. *Social Network Analysis and Mining*, 15(1), 69. https://doi.org/10.1007/s13278-025-01490-3
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. San Francisco, CA, USA. https://www.mikecaptain.com/resources/pdf/GPT-1.pdf
- Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 1(4), 990–1029. https://doi.org/10.1162/154247603322493212
- Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, 71(5), 845–858.
- Sarikaya, R. (2025). Path to Artificial General Intelligence: Past, present, and future. *Annual Reviews in Control*, 60, 101021. https://doi.org/10.1016/j.arcontrol.2025.101021
- Schneider, J., Meske, C., & Kuss, P. (2024). Foundation Models. *Business & Information Systems Engineering*, 66(2), 221–231. https://doi.org/10.1007/s12599-024-00851-0
- Spulber, D. F. (2019). The economics of markets and platforms. *Journal of Economics & Management Strategy*, 28(1), 159–172. https://doi.org/10.1111/jems.12290
- Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, June 12). *Attention is All you Need*. Neural Information Processing Systems. https://www.semanticscholar.org/paper/204e3073870fae3d05bcbc2f6a8e263d9b72e776
- Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Will we run out of data? Limits of LLM scaling based on human-generated data (No. arXiv:2211.04325; Version 2). arXiv. https://doi.org/10.48550/arXiv.2211.04325
- We Are Social, & Meltwater. (2025). *Digital 2026. Global Overview Report*. https://datareportal.com/reports/digital-2026-global-overview-report
- Xing, X., Shi, F., Huang, J., Wu, Y., Nan, Y., Zhang, S., Fang, Y., Roberts, M., Schönlieb, C.-B., Del Ser, J., & Yang, G. (2025). On the caveats of AI autophagy. *Nature Machine Intelligence*, 7(2), 172–180. https://doi.org/10.1038/s42256-025-00984-1