Reconsidering Monopolies through Foundation Models

Hamidah Oderinwale

April 2024

Abstract

This paper challenges common assumptions about foundation models (FMs) and how they affect the market of innovation. It questions the beliefs that FMs lead to monopolies and that AI diminishes the need for scientific expertise. While FMs benefit major players like OpenAI and DeepMind, smaller entities can still leverage open-source models. Market dominance in AI may not solely rely on FMs but also on specialized models with unique, and State-of-the-Art (SOTA) architectures. The paper discusses tradeoffs in AI capabilities and its implications for market competition and scientific progress. It proposes a taxonomy for identifying strategic domains for specialized AI. The conclusion suggests an adapted multi-sided platform model, where FM providers offer diverse ”products” but achieving critical mass isn't the primary factor. High-quality training data is advantageous, and firms gain an edge through horizontally ”integrated” product development.

1 Introduction

This paper challenges two prevailing views regarding foundation models in relation to firm dynamics and innovation. The first view is that foundation models will easily create monopolies for the firms that develop them (7). And the second view we contest is the concern that the advancement of AI will diminish the necessity for scientific expertise (12; 29; 30; 20).

Foundation models are large-scale models trained on very large amounts of data. These models, like GPT-4, Claude-3-Opus, Gemini, Meta Llama 3, and Mistral 7B, are trained on extensive and diverse datasets and demonstrate adaptability across a wide range of tasks (14). It is very expensive to acquire training data and subsequently train a model. As such, foundation models are synonymous with large, well-resourced AI firms like OpenAI, DeepMind, and Anthropic. The fear is that these firms will gain unchecked market power, and scientific expertise will have little to show in the market.

State-of-the-art models are the field's most advanced and cutting-edge solutions, incorporating the latest techniques and methodologies. They evolve with new research and innovations, varying by task or domain.

Foundation models (FMs) are a subset of state-of-the-art models, serving as base frameworks for AI applications. These large-scale pre-trained models learn general representations and can be fine-tuned for specific tasks. While all foundation models are state-of-the-art, not all state-of-the-art models are foundation models. State-of-the-art models (SOTAs) include specialized approaches, while foundation models are specifically pre-trained models forming the basis for AI applications.

This paper explores the competitive market structure and introduces a taxonomy to identify strategic areas for specialized AI, specifically SOTAs for strategic domains. Drawing insights from successful projects such as GnoMe and AlphaFold, we explore the tradeoffs in deciding what areas to specialize in. What would be the next Alpha-N? In this world, we assume many tradeoffs, but a firm that excels without sacrificing any feature is one that many large firms are openly vying for. Full generalizability and expert ability are promises of Artificial General Intelligence. Large firms like OpenAI, DeepMind, and Anthropic have also declared their intentions to develop Artificial General Intelligence—a promise of transformitivity—the product of “solving intelligence.” Still, there are ways to go, and this paper aims to expand on our current understanding of AI capabilities.

Additionally, while large firms traditionally dominate discussions around AI deployment, we outline the ability of smaller firms to compete from an industrial organization perspective and technical perspective. One of our arguments is that open-source models trained on specialized data allow niche but small firms to exist in market gaps left by firms developing foundation models.

Large firms developing foundation models will try for market dominance per competition theory. A firm, largely irrespective of domain, will try to monopolize product markets strategically and will use its means to do so. However, for AI firms, it will mostly not be through the foundation models they have developed. Rather, their resources and characteristics can lend themselves to developing specialized models with unique architecture(s) that are capable in a way that a scientific expert would be. Well-resourced firms will apply AI to functions with the most reward, but it is a trade-off against vertical backstreaming and the other functions they choose not to pursue.

1.1 Motivation

AI is nuanced and complex. It is a technology that can not only augment the work of humans but potentially do most of it well and with autonomy. This work considers technical considerations, economic frameworks, and the functions of the human scientist to understand firm-level dynamics. We draw from microeconomics, management theory, industrial organization, and innovation economics to inform models of an AI-augmented innovation pipeline. There is a gap in this kind of literature and a desire for more research in this area, *A Research Agenda for Assessing the Economic Impacts of Code Generation Models* released in March 2022 by researchers at OpenAI (25).

Literature that examines the effects of AI on the macroeconomy is growing. These works are also becoming more refined, looking to give a more accurate picture of what to expect regarding economic growth, especially what to expect in the long term. For example, the breadth and competence of AI capabilities are not uniform, so more recent works have looked at the economic implications of imbalanced automation. What happens when certain tasks in a workflow are automated, but not all of them(9)?

Major firms like OpenAI, DeepMind and Anthropic are building the world’s most capable public models and working towards Artificial General Intelligence: solving intelligence, building models with agency, highly-capable in more mundane tasks but also, highly-specialized, complex ones. There is a trade-off for firms on generalizabilty and competence across tasks.

2 Industrial Organization Analysis

2.1 Upstream and Downstream

First, it is valuable to understand vertical integration and market power within the context of the digital economy. Companies can either vertically integrate upstream or backstream. Companies will only vertically integrate, especially in high-tech industries and in a digital economy, if the innovations generated from upstream and downstream integration are valuable to them (1).

Liu et al. (2016) look at the pharmaceutical industry to describe upstream and downstream integration. Upstream integration would involve developing new drugs, while downstream integration would involve investment in manufacturing and development technologies. As such, upstream integration for R&D firms involves developing applications while downstream integration is an investment in infrastructure: the means to develop new applications (products) and optimize the processes for development (lower costs, quality of product) for existing technologies (2).

2.1.1 R&D Intensity

R&D intensity for foundation models is high. Garg et al. find that products with higher R&D intensity are more likely to be sourced from within (3). Machine learning models, namely foundation models, have high R&D intensity and are sourced from within. A firm, largely with its resources, has to decide between product and process innovation R&D (PIRD) (2).

We can assume that AI firms have costless vertical integration (21) and are more likely to be sourced from within, meaning there are no internal organizational costs. Instead, the costs incurred are (directly) from the costs of development at that level of the supply chain. As such, the

Apparent technical limitations open up opportunities for scientists and human innovators to enter one or more functions served by a scaled foundation model: natural vertical disintegration, as opposed to natural vertical integration.

Acemoglu et al. (2010) show that increased technology intensity tends to discourage vertical integration (VI), prompting firms to prioritize research and development (R&D) and vertical disintegration (VD). Scientific discovery has a high level of technological sophistication, and technologies are leveraged to a high degree. Software, a sophisticated technology, has quickly diffused itself into life, physical and mathematical, while staying relatively more constant in the humanities and arts, exemplifying the disparity in technical complexity across disciplines (10). As such, in the sciences, a technically intensive field, we can conclude from Acemoglu's work that vertical integration is discouraged in this regime.

2.1.2 Discontinuous Vertical Integration

Variants of vertical integration that would still conclude with market dominance by monopoly for foundation model developers are also unlikely. For example, vertical leapfrogging is where an established firm is in one part of the supply chain, and another is established in the following part. Leapfrogging would mean the firm with the lowest in the supply chain can leapfrog on the innovation and investments of the other (15). This is unlikely in the context of models specializing in upstream.

First, developing and improving foundation models often requires involvement in the entire process, from data collection and pre-processing to model training and evaluation. This continuous involvement makes it challenging for firms to leapfrog over others in the supply chain because each step is intricately linked and builds on previous work.

Additionally, while there are possibilities for AI models to train one another, as seen in techniques like transfer learning or federated learning, these processes can be technically complex and require full access to all models involved. For example, federated learning is very involved with all stakeholders but does technically allow for a specialized model in one domain to enable the specialization of a foundation model in another specialization. Still, this would imply more horizontal integration than vertical integration (23).

Overall, the technical complexity, continuous scaling nature, and interdependence of foundation models make it challenging for firms to engage in discontinuous vertical integration or leapfrogging to achieve market dominance independently.

This makes it difficult for a firm to leapfrog on the innovation and investments of another firm without collaboration or access to its resources. Even if a firm is established in one part of the supply chain and another in a subsequent part, vertical leapfrogging is unlikely due to model development's interdependence and continuous nature.

Input foreclosure is when a merged company stops providing its competitors with the materials, resources, services, etc., needed to develop their products (5). Assuming market dominance by economies-of-scale for AI firms through specializing their models is analogous to a merger, both displaying vertical integration.

There is the worry that a closed-source model core hurts smaller firms building specialized models on their core, distorting competition and driving the large firms' market power.
"'

3 The Model's Journey to Specialization

Through this iterative process of pruning, sparsity induction, and fine-tuning, the foundation model evolves into a specialized expert.

First, pruning is then employed to remove unnecessary parameters, reducing the model's size and complexity. This pruning induces sparsity within the model by setting numerous parameters to zero. Sparsity can make the model more efficient and effective in specialized tasks or domains. A highly-specialized model is, therefore, likely highly sparse. The optimal sparsity of a model also increases with the amount of data used for training. So in expectation, foundation models by design have a high amount of baseline training data and need to be fine-tuned to high sparsity in specialization. After pruning, fine-tuning is conducted using task-specific data to adapt the model to the target task or domain. Fine-tuning helps the model regain any lost performance during pruning and refines its parameters for optimal task performance.

4 Scaling is for Generalists

Foundational models are overparametrized as a result of the breadth and magnitude of data they initially train on, and this is largely why they perform well (8). However, overparametrized models are also costly to train and unspecialized. (16)More mechanistically, overparametrization typically occurs when a model has more learnable parameters than is required for a task—as a general, but fine-tunable model would be.(17) Sparse model training can make model training cheaper and is necessary for fine-tuning. (**?**)

We suggest that FM providers will adopt a Chinchilla training regime and increase the volume of training data to enhance performance.

Chinchilla Scaling Law

Chinchilla scaling laws claim that data precedes model size for model performance. It is denoted as follows:

$$\hat{L} = \left(L^{\frac{1}{N}} \cdot \frac{D}{E} \cdot \frac{A}{B} \right)^{(2)}$$

As such, large models (foundation models) are currently not maxxing out the amount of data they could use for their models. By providing more training tokens, they could improve performance.(8)

A key finding from this line of research on neural scaling suggests that scaling surpasses intricate, expert-designed systems.

Optimally-trained foundation models that follow the Chinchilla training regime do demonstrate superior performance (Chinchilla vs. Gopher) across various tasks, including Language Modelling, Reading Comprehension, Fact Checking, Question Answering, Common Sense, MMLU (Multimodal Language Understanding), and BIG-bench (26). However, domain experts engage in reasoning-based tasks where scaling laws are not yet applicable or at least explored.

Notably, new work suggests that the confidence intervals for the optimal scaling policies based on Hoffmann et al.'s (2022) estimates are very tight.

$$L(N, D) = 1.69 + \frac{406.4}{N^{0.34}} + \frac{410.7}{D^{0.28}}$$

However, a revised model by Besiroglu et al. (2024) still aligns with the scaling recommendations in Hoffman et al.'s work.

$$L(N, D) = 1.82 + \frac{514.0}{N^{0.35}} + \frac{2115.2}{D^{0.37}}$$

Nevertheless, per the Chinchilla training regime, it is also assumed that foundation model scaling will involve sparse model training to address the expenses associated with training their over-parametrized models. Research indicates that optimal sparsity rises with the quantity of data utilized for training while preserving a fixed number of non-zero parameters.(4)

But this scarcity also comes at a cost, with (very) high sparsity affecting robustness and accuracy, which, independently considered, hurts the case for scaling FMs for specialization. (16)

Additionally, sparsity also slows the training process, and it can be difficult to sparsify all the necessary parts of the model. However, inaccuracy can supposedly be mitigated with longer training periods and less pruning, but at a financial and specialization cost.

Let:

- **Accuracy**: Represents the performance accuracy of the trained model.
- **Model Sparsity**: Denotes the degree of sparsity achieved in the model through pruning or other techniques.
- k: A scaling factor determining the steepness of the accuracy curve relative to model sparsity changes. Higher k implies more significant accuracy changes with small sparsity changes.
- m: Midpoint parameter influencing the accuracy curve transition from low to high values. Higher m shifts the curve right, requiring higher sparsity for high accuracy.
- e: Base of natural logarithm, used to exponentiate $-k \cdot (\text{Model Sparsity} - m)$ in the denominator. Smoothens the sigmoid curve mapping sparsity to accuracy. Potentially do experiments to test this

$$\text{Accuracy} = \frac{1}{1 + e^{-k \cdot (\text{Model Sparsity} - m)}}$$

Villalobos and Atkinson provide a model for this. If x is the fraction of weights pruned at each iteration, and n is the number of iterations, then compute increases by a factor of $\frac{(1-(1-x)^n)}{x}$. (28)

> **Robustness:** Model robustness refers to a model's capacity to effectively perform on unseen data beyond its training set, showcasing resilience against diverse forms of noise, variations, and uncertainties in the data. A robust model demonstrates the ability to generalize and perform at a similar level across various scenarios and domains, including science.

4.0.1 Key Takeaway

Currently have specialized models like GnOME,(22), which excels in material discovery, and AlphaFold, (24), which excels in 3D protein structure prediction.

Foundation models are generative agents that can do reasonably well at many things, but these two examples have unique deep learning architecture.

Research shows that scaling laws may hold across modalities and deep learning models, but even then, scaling multi-modal models is not where the large firms we focus on, at this time, have established their market dominance. (11)

Arguably, no machine learning model is currently very competent in a given, specific domain and could complete another unique, highly specialized task to the same degree.

5 Why Specialization Does Not Scale

If scaling a model could allow FM providers to easily monopolize product markets, they would. However, scaling FMs is not the way to develop specialized models and expert-designed systems. The Chinchilla Scaling law suggests that specialized models are not needed or larger model sizes. Models as of today are under-trained and could be scaled largely with data.

However, to the extent to which the Chinchilla Scaling Law is true, the scope of "specialized" tasks does not include the ones held by a domain expert in terms of complex reasoning. Regardless, large AI firms, the providers of FMs, will find ways to expand their product offerings, so it is a question of what types of products and problems they will develop unique algorithms/architectures for.

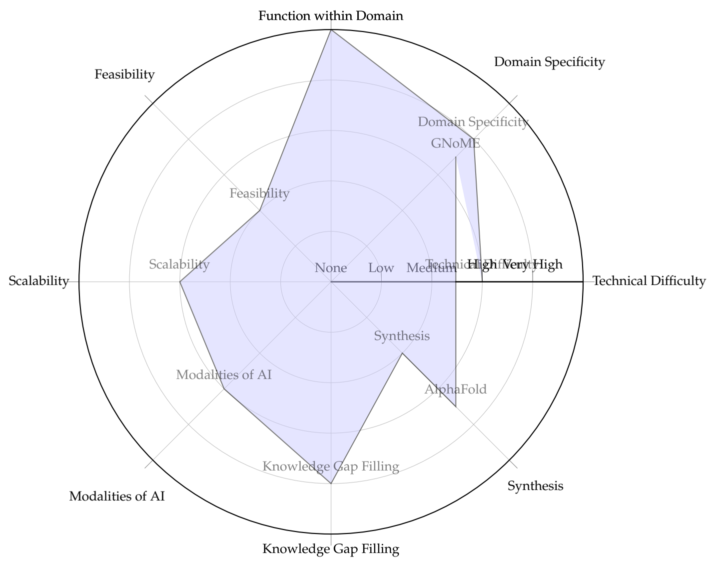

Figure 1: Radar Chart for Specialization of AI Models within the Innovation Pipeline

5.1 The Costs of Specializing and Deciding Where to Spend

We propose a model for the cost of developing a specialized model architecture for AI applications. The cost function, denoted as C, is designed to capture the resources (capital) and time required for development, taking into account various characteristics outlined in the table below.

The cost function is defined as follows:

$$C = R \times T \times \prod_{i=1}^{n} f_i$$

Where:

- C represents the overall cost of development.
- R denotes the amount of capital (resources) invested in the project.
- T signifies the time required for development.
- f_i represents the factor associated with each characteristic, where $i = 1, 2, ..., n$.

In this model, each characteristic from the table corresponds to an indexed factor f_i, indicating its influence on the overall cost. For example, let's say we have the following factors:

$$f_1, f_2, f_3, \ldots, f_n$$

These factors represent characteristics such as the audience's size, expertise, absorbency level for technology, broadness of the area, use-case characteristics, assistive vs. agentic nature, time horizon of tasks, and length of the task chain.

We assign numerical values to these factors based on their impact. For instance, favorable characteristics may have $f_i = 1$, while unfavorable ones may have $f_i > 1$, indicating an additional cost.

With this model, we can estimate the total cost of developing a specialized model architecture by considering the resources invested, time required, and the influence of various characteristics. Fine-tuning the factors allows us to analyze different scenarios and optimize the development process accordingly.

6 Considerations

7 Open-Source, Specialization and Market Entrants

Although fine-tuned models are less precisely engineered for specific use cases, their capabilities should not be underestimated, and open-source could help unlock some of their specialized potential (18).

They can lead to developing sophisticated, specific use cases without training a new foundation model architecture from scratch, significantly reducing the incentive for producers to build their own models.

However, despite lower costs associated with fine-tuning compared to developing and training new models, it is a non-negotiable expense that can be prohibitive to smaller firms (18).

Research finds that the learning rate must be re-increased to improve compute efficiency when retraining on a new dataset, performing continual pre-training on a model (18). With continual pre-training, model performance can be improved at or above the level of the non-rewarmed model, but the costs of increasing the learning rate are significant.

Today's largest foundation models are typically closed-source, limiting developers' flexibility. However, Meta's release of LLaMA3 (6), a large open-source model, offers a path forward for those who can bear the costs of continual pre-training and do not want to develop their own model core.

Table 1: Characteristics of Specialized Models

Characteristics	Description
Audience	
Size of Audience	Large (e.g., mainstream applications like language translation)
	Niche (e.g., specialized scientific research)
Expertise of Audience	Unsophisticated (e.g., everyday users)
	Specialized (e.g., domain experts in specific fields like medicine or finance)
Absorbency Level for Technology	High (e.g., pragmatic problems with clear applications)
	Low (e.g., abstract or theoretical problems with many unknowns)
Broadness of the Area	Wide (e.g., general scientific research)
	Narrow (e.g., specific subfields like astrophysics)
Use-Case Characteristics	
Productivity-Enhancing Tools	Tools that streamline workflows or automate repetitive tasks.
Synthesis	Models that generate new insights or ideas by combining existing information.
Knowledge Retrieval	Models that efficiently access and retrieve relevant information from large datasets.
Predictive	Models that forecast future outcomes based on historical data.
Classification	Models that categorize data into distinct groups or classes.
Assistive vs. Agentic	
Assistive	Models that provide support and assistance to human users in decision-making or problem-solving tasks.
Agentic	Models that autonomously perform tasks without direct human intervention.
Time-Horizon of Tasks	
Long	Long time horizons, often involving chained tasks with complex dependencies (e.g., drug discovery).
Medium	Medium time horizons, typically chained with many tasks but with less complex dependencies (e.g., software development).
Short	Short time horizons, single tasks with immediate responses (e.g., chatbots or question-answering systems).
Length of Task Chain	
	Tasks often involve a long, medium, or short chain depending on the complexity and nature of the problem being addressed.

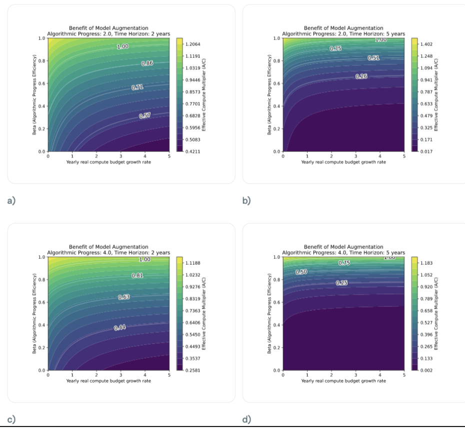

Figure 2: Model augmentation simulations with varied parameter configurations

Enhancing older models with new algorithms may offer no advantage if deficiencies lead to future performance degradation (13).

Therefore, it's often more beneficial to train the model from scratch.

Barnett shares a model for why this is true, where the parameter β specifies the portion of new algorithmic advancements that we can utilize in each iteration, i. We establish the model based on the following equations.

Their findings are as follows: In most scenarios, the benefits of model augmentation seem modest or counterproductive unless the re-use efficiency closely approaches 1 or the number of iterations is kept very small.

- They maintain the assumption of exogenous growth in computational resources.
- The initial adjusted computational capacity matches the training compute at the outset.
- Now, incorporating a parameter governing algorithmic efficiency, denoted by P, which escalates at a rate g.
- Each iteration's adjusted compute is computed as the product of the real compute for that iteration, algorithmic efficiency, and a factor β, added to the adjusted compute from the previous iteration.
- Consequently, if this adjusted compute lags behind the effective compute, it often indicates a preference for starting afresh, especially if $\beta > 1$.

The simulations in the figure below show that unless the re-use efficiency is very close to 1 or we limit the number of iterations significantly, the advantages of augmentation are small or even harmful.

7.1

Still, smaller firms could flourish by targeting niche markets where even slight quirks in capability are highly valued. Highly specialized data often leads to the development of highly specialized models. Leveraging open-source models, smaller firms can harness their talent and limited yet specialized data to capitalize on large open-source models.

Small firms that already have access to (private) specialized data and cater to niche, specialized markets (27).

- Niche, expert consumers play a significant role by providing valuable data to smaller training pools, where each additional training token holds higher value (27).
- Even without access to open model cores, quantization could mean that smaller language models can be just as good as private APIs in quality and use less memory, resulting in cost savings on compute.

may benefit from their hyper-specialization because they can fine-tune their models to use cases that large firms are not prioritizing

8 Models

Through this iterative process of pruning, sparsity induction, and fine-tuning, the foundation model evolves into a specialized expert.

First, pruning is then employed to remove unnecessary parameters, reducing the model's size and complexity. This pruning induces sparsity within the model by setting numerous parameters to zero. Sparsity can make the model more efficient and effective in specialized tasks or domains. A highly-specialized model is, therefore, likely highly sparse. The optimal sparsity of a model also increases with the amount of data used for training. So in expectation, foundation models by design have a high amount of baseline training data and need to be fine-tuned to high sparsity in specialization. After pruning, fine-tuning is conducted using task-specific data to adapt the model to the target task or domain. Fine-tuning helps the model regain any lost performance during pruning and refines its parameters for optimal task performance.

9 Conclusion

9.1 Discussion

Large firms may prioritize certain functions over others based on their potential rewards, leading to trade-offs against vertical integration. High R&D intensity is characteristic of foundation models, with firms needing to decide between product and process innovation. Technical sophistication and the continuous nature of model development discourage vertical integration, favouring research and development (R&D) and vertical disintegration (VD). And scaling foundation models may not lead to monopolizing product markets, as specialization requires more than just larger model sizes. The function of scientific expertise is not captured in the generalist functions that a scaled foundation model would do well at.

9.2 Implications

In conclusion, we may witness the emergence of an adapted or modified multi-sided platform (MSP) model. Here, a foundation model provider might offer diverse "products," appealing to distinct "customer" groups, yet this phenomenon stems more from incidental outcomes than deliberate value creation strategies.

In contrast to traditional multi-sided markets, where products complement each other and network effects contribute to market dominance, achieving critical mass for a foundation model provider is not the primary factor.(19) However, having a high-quality pool of training data proves advantageous. While large model providers leverage their consumers to enhance model capabilities, the effects lack recursion—existing models don't exponentially evolve based on user data. Instead, firms gain an edge through horizontally integrated product development. The allocation of resources becomes pivotal, particularly for specialized use cases that require new models to be built from scratch. Understanding the heuristics guiding resource allocation in such cases becomes imperative.

References

[1] 13.5. Training on Multiple GPUs — Dive into Deep Learning 1.0.3 documentation.

[2] [1901.09451] Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting.

[3] [2304.14997] Towards Automated Circuit Discovery for Mechanistic Interpretability.

[4] 2.4: Scaling Laws — AI Safety, Ethics, and Society Textbook.

[5] Input Foreclosure as Theory of Harm in in Vertical and Conglomerable Mergers by Amanda Athayde :: SSRN.

[6] Introducing Meta Llama 3: The most capable openly available LLM to date.

[7] Market concentration implications of foundation models: The Invisible Hand of ChatGPT.

[8] Papers with Code - Training Compute-Optimal Large Language Models.

[9] Acemoglu, D., Autor, D., and Patterson, C. Bottlenecks: Sectoral imbalances and the us productivity slowdown, Apr 2023.

[10] Acemoglu, D., Griffith, R., Aghion, P., and Zilibotti, F. Vertical Integration and Technology: Theory and Evidence.

[11] Aghajanyan, A., Yu, L., Conneau, A., Hsu, W.-N., Hambardzumyan, K., Zhang, S., Roller, S., Goyal, N., Levy, O., and Zettlemoyer, L. Scaling Laws for Generative Mixed-Modal Language Models.

[12] Aghion, P., Jones, B. F., and Jones, C. I. Artificial Intelligence and Economic Growth.

[13] Barnett, M. The limited benefit of recycling foundation models, 2023. Accessed: 2024-04-21.

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

[15] Filippini, L. Leapfrogging in a Vertical Product Differentiation Model. 245–256.

[16] Frantar, E., Riquelme, C., Houlsby, N., Alistarh, D., and Evci, U. Scaling Laws for Sparsely-Connected Foundation Models.

[17] Gormely, I. Machine Learning Robustness: New Challenges and Approaches.

[18] Gupta, K., Thérien, B., Ibrahim, A., Richter, M. L., Anthony, Q., Belilovsky, E., Rish, I., and Lesort, T. Continual Pre-Training of Large Language Models: How to (re)warm your model?

[19] HAGIU, A., AND WRIGHT, J. Multi-sided platforms. *International Journal of Industrial Organization 43* (2015), 162–174.

[20] IDE, E., AND TALAMAS, E. Artificial intelligence in the knowledge economy, 2024.

[21] JOSKOW, P. L. VERTICAL INTEGRATION.

[22] JUMPER, J., EVANS, R., PRITZEL, A., AND ET AL. Highly accurate protein structure prediction with alphafold. *Nature 596* (2021), 583–589.

[23] LAMBERTINI, L., AND ROSSINI, G. IS VERTICAL DISINTEGRATION PREFERABLE TO INTEGRATION WHEN THERE IS PROCESS R&D? 401–416.

[24] MERCHANT, A., AND CUBUK, E. D. Millions of new materials discovered with deep learning, Nov 2023.

[25] MISHKIN, P. Economic impacts research at openai.

[26] RAE, J. W., BORGEAUD, S., CAI, T., MILLICAN, K., HOFFMANN, J., SONG, F., ASLANIDES, J., HENDERSON, S., RING, R., YOUNG, S., RUTHERFORD, E., HENNIGAN, T., MENICK, J., CASSIRER, A., POWELL, R., FAMILY=DRIESSCHE, GIVEN=GEORGE, P. D. U., HENDRICKS, L. A., RAUH, M., HUANG, P.-S., GLAESE, A., WELBL, J., DATHATHRI, S., HUANG, S., UESATO, J., MELLOR, J., HIGGINS, I., CRESWELL, A., MCALEESE, N., WU, A., ELSEN, E., JAYAKUMAR, S., BUCHATSKAYA, E., BUDDEN, D., SUTHERLAND, E., SIMONYAN, K., PAGANINI, M., SIFRE, L., MARTENS, L., LI, X. L., KUNCORO, A., NEMATZADEH, A., GRIBOVSKAYA, E., DONATO, D., LAZARIDOU, A., MENSCH, A., LESPIAU, J.-B., TSIMPOUKELLI, M., GRIGOREV, N., FRITZ, D., SOTTIAUX, T., PAJARSKAS, M., POHLEN, T., GONG, Z., TOYAMA, D., FAMILY=AUTUME, GIVEN=CYPRIEN DE MASSON, P. U., LI, Y., TERZI, T., MIKULIK, V., BABUSCHKIN, I., CLARK, A., CASAS, D. D. L., GUY, A., JONES, C., BRADBURY, J., JOHNSON, M., HECHTMAN, B., WEIDINGER, L., GABRIEL, I., ISAAC, W., LOCKHART, E., OSINDERO, S., RIMELL, L., DYER, C., VINYALS, O., AYOUB, K., STANWAY, J., BENNETT, L., HASSABIS, D., KAVUKCUOGLU, K., AND IRVING, G. Scaling Language Models: Methods, Analysis & Insights from Training Gopher.

[27] SANTESTEBAN, C., AND LONGPRE, S. How Big Data Confers Market Power to Big Tech: Leveraging the Perspective of Data Science.

[28] VILLALOBOS, P., AND ATKINSON, D. Trading off compute in training and inference, 2023. Accessed: 2024-04-21.

[29] WANG, H., FU, T., DU, Y., GAO, W., HUANG, K., LIU, Z., CHANDAK, P., LIU, S., VAN KATWYK, P., DEAC, A., ANANDKUMAR, A., BERGEN, K., GOMES, C. P., HO, S., KOHLI, P., LASENBY, J., LESKOVEC, J., LIU, T.-Y., MANRAI, A., MARKS, D., RAMSUNDAR, B., SONG, L., SUN, J., TANG, J., VELIČKOVIĆ, P., WELLING, M., ZHANG, L., COLEY, C. W., BENGIO, Y., AND ZITNIK, M. Scientific discovery in the age of artificial intelligence. 47–60.

[30] XU, Y., LIU, X., CAO, X., HUANG, C., LIU, E., QIAN, S., LIU, X., WU, Y., DONG, F., QIU, C.-W., QIU, J., HUA, K., SU, W., WU, J., XU, H., HAN, Y., FU, C., YIN, Z., LIU, M., ROEPMAN, R., DIETMANN, S., VIRTA, M., KENGARA, F., ZHANG, Z., ZHANG, L., ZHAO, T., DAI, J., YANG, J., LAN, L., LUO, M., LIU, Z., AN, T., ZHANG, B., HE, X., CONG, S., LIU, X., ZHANG, W., LEWIS, J. P., TIEDJE, J. M., WANG, Q., AN, Z., WANG, F., ZHANG, L., HUANG, T., LU, C., CAI, Z., WANG, F., AND ZHANG, J. Artificial intelligence: A powerful paradigm for scientific research. *The Innovation 2*, 4 (2021), 100179.