The War of Hundreds of Large Models
百模大战

Qihong Ruan

May 7, 2024

Abstract

This paper develops a unified model to analyze the competitive dynamics of large language model (LLM) development, incorporating productivity growth, proprietary competition, open-sourcing initiatives, and community diversity. We introduce a new model that captures the key roles of computing power, algorithms, and data as scarce resources, and incorporates Nash bargaining between firms and resource providers, separating equilibria based on firm size, and the endogenous choice of open-sourcing by companies. We show that open-source models and communities can reduce duplicative investments and resource waste caused by arms races, promote diversity and inclusivity in the digital ecosystem, and foster long-tail economics. We discuss the competition between China and the US in LLM development and highlight the parallels between open-sourcing in LLMs and decentralized finance (DeFi). Our model provides insights into industry standards, international competition, and the concentration of power and resources in the LLM industry, offering guidance for policymakers, researchers, and practitioners navigating this rapidly evolving landscape.

1 Introduction

Large models, which means models beyond large language models, including multimodular models, have the potential to improve productivities in all aspects of human economic activities and all kinds of industries. The battle of large models has been hot across the globe. The development of large models is of relevance to national competeveness. Due to the unique features of large models, including externatilities (positive externatilities such as improving productivities, or negative externatiliteis such as labor replacement or sysmetic risks), non-rivalry, data-reinforcing effects, privacy concerns. We are facing unprecedented issues with the coming of large models. We need an economic framework to understand the battle of hundreds of large models to inform the optimal policies for the welfare of human beings.

The current landscape of large models consist of priprietary models such as chatgpt and also open-source models like Llama by Meta. Proprietary models can extract economic rent from the gap between the productivity their model and the most productive open-source model.

The scarce resources to make large models include computing power, algorithms made by talents, and data (open-source data and consumer data). The productivities of large models follow possion jump processes, with intensity proportional to computing power algo talents, data devoted to the model. Many firms of large models may have differnet starting points of the producitivity (can easily measured by their score in multiple tasks) of their large model. In a world without open-source models, they invest their resources to improve their own models, leading to arm races. And from the perspective of the whole society, the resources invested in the relative low states of productivities models are wasted, leading to suboptimal outcomes of resource allocation in the society. In contrast, in a world with open-source model with relatively high productivities, the firms with starting point lower than that open-source model will optimally choose to innovate based on that open-source models. The state-of-the-art open-source model serves a powerful aggregator of the decentralized computing power, talents, and data, so that these resources can be put together and innovate in a relatively promising direction given the current information environment (the state of producitivity of the open-source model), this leads to more efficient resource allocation in the society. Also, the choice of making the model open-source can be an endogenous choice of a firm, since make their model open-source can attract more efforts in the route and increase the probabitliy of making breakthourgh in their route of models (measure by the idensity of possion jump). This may help close the gap between the open-source model and the most productive proprietary model, and eventualy beneficial for the firm making their model open-source.

Based on the data from huggingface, a popular open-source community of large models, there have been about 700,000 open-source models. The downloads of models follow a Zipf's Law for Cities: An Explanation in Gabaix [1999], meaning that the log regression of downloads on the rank of model is close to -1. From this results, open-source communities of large models has led to a diverse and inclusive ecosystem just like different cities. We have large models like metropolitan New York City and also boutique models like small town Ithaca with Cornell Unviersity. It is worth emphasizing the long-tail of models are non-trivial just like Ithaca is non-trivial with Cornell Unversity. The likes of model even have a fatter tail, indicating diverse preferences on the models. This results highlight that open-source models, enables communicty mebmers to build an divesiryt and icnlusive ecosystem of larege models, and at the same time of making resources allocaiton efficient. We may model this by making the productivity multidimentional, since large models can be useful for multiple tasks, meeting the diverse preferences of developers and consumers.

The development of large models, given its huge externatilities, is vital for the national competiveness. Among the countries that compete for large models, two most prominent ones are China and the US. This can be seen the Meta's releasing of Llama, the menu of countries for user to download llama list the US as the first, while China is not among the list. This is an indication of the competiton of China and US in the arena of large models and AI. The developement of large models are majorly powered by BigTech private firms such as Amazon, Google, and Meta, while Chinese government take more initiatives to power up the development of AI, including making AI development a national strategy, building computing power clusters in provinces, subsize home-made chips, boosting AI talent education, build data bureau to facilitate the circulation of data factor in production. This leads to the comparisions of the Socialism Large Models and Captalism Large Models. The two approachs may need a economic framework to tell their pros and cons and understand their competition dynamics.

Also, the open-source of large models draw parralles to the open-source of protocols in the DeFi space. The open-source of DeFi protocols fundamentally change the ecosystem of traditional financial system. It opens the composablity and interoperablity to developers and allows consumers to build finance logo based on their diverse needs. From the data of DeFiLLamma, the total value locked(TVL) of DeFi protocols are distributed with a long-tailed, highlighting the diverse and inclusive nature of the DeFi system. The topic is of especial interests as the era is standing on the intersection of Centrealized Financen and Decentralized Finance. The notions of Open-source, self-custody, composiblity, interopoerablity, are crucial for the next-generation financial system. The parallels between open-source Large models and DeFi system are especially meaningful.

The competition of AI fueled by the recent developments of large models is relevant for the national competitiveness. It is crucial for regulators to understand the policy implications. It is of policy relevance given the unique features of large models including non-rivalry and externalities. We need a framework to understand the battle of large models and inform optimal policies. The rapid development of large language models (LLMs) has led to significant advancements in natural language processing and generated substantial interest from both industry and academia [Brown et al., 2020, Chowdhery et al., 2022, Hoffmann et al., 2022]. However, the competitive landscape of LLM development has raised concerns about potential arms races, duplication of efforts, and inefficient resource allocation [Strubell et al., 2019, Ahmed et al., 2022].

This paper presents a unified model that captures the key aspects of LLM development, including productivity growth, proprietary competition, open-sourcing initiatives, and the role of community diversity. By incorporating these factors into a coherent framework, we shed light on the dynamics of LLM development and draw parallels to other domains, such as decentralized finance (DeFi).

Open source models aggregate and integrates the decentralized talent and computing power and lead them to innovate based on the state-of-the-art (SOTA) quality of open-source model, this leads to a more efficient outcome from the perspective of the society.

The companies of proprietary models enjoy economic rents from the differences between the productivity of proprietary models and the most leading open-source models.

Since open-source can aggregate decentralized computing power and talents and may help reduce the gap between open-source and the most leading proprietary models, open source can be an endogenous and strategic choice. The companies may strategically choose the timing and quality of the release of open-source model.

In the context of the AI competition, my model sheds light on the competition between socialism large models and capitalism AI.

2 Model Setup

2.1 Scarce Resources

Consider an economy with three scarce resources:

1. Computing power: Provided by a monopoly, denoted by M.
2. Algorithms: Developed by firms using existing models and algorithmic talent, denoted by A.
3. Data: Consists of existing open-source datasets and user-generated data, denoted by D.

There are N firms, indexed by $i \in \{1, 2, \ldots, N\}$, that develop large language models (LLMs) using these resources.

2.2 Model Productivity and User Choice

The productivity of firm i's LLM is given by:

$$q_i = f(C_i, A_i, D_i) \tag{1}$$

where C_i, A_i, and D_i represent the computing power, algorithmic capabilities, and data used by firm i, respectively. The function $f(\cdot)$ is increasing in each argument and exhibits diminishing returns [Brynjolfsson and Hitt, 2000].

The quality of firm i's LLM evolves according to a Poisson jump process:

$$dq_i(t) = \lambda_i dN_i(t) \tag{2}$$

where λ_i is the average jump size and $N_i(t)$ is a Poisson process with intensity μ_i. The intensity of jumps is proportional to the firm's computing power, algorithmic capabilities, and data:

$$\mu_i = \alpha C_i + \beta A_i + \gamma D_i \tag{3}$$

where α, β, and γ are positive constants.

Users choose to contribute data to LLMs based on their productivity. The share of users contributing data to firm i's LLM is given by:

$$s_i = \frac{q_i}{\sum_{j=1}^{N} q_j} \tag{4}$$

This specification captures the idea that users are more likely to contribute data to more productive LLMs [Rochet and Tirole, 2003].

2.3 Nash Bargaining

Firms engage in Nash bargaining with the monopoly computing power provider and algorithmic talent [Nash Jr, 1950].

2.3.1 Computing Power

The surplus generated by firm i and the monopoly M is given by:

$$S_{iM} = R_i(C_i) - P_M(C_i) - \bar{R}_i - \bar{P}_M \tag{5}$$

where $R_i(C_i)$ is the revenue generated by firm i using computing power C_i, $P_M(C_i)$ is the cost of providing C_i for the monopoly, and $\bar{R}_i$ and $\bar{P}_M$ are the disagreement payoffs for firm i and the monopoly, respectively.

The Nash bargaining solution is given by:

$$\max_{C_i} (R_i(C_i) - \bar{R}_i)^{\beta} (P_M(C_i) - \bar{P}_M)^{1-\beta} \tag{6}$$

where $\beta \in (0, 1)$ represents the bargaining power of firm i relative to the monopoly.

2.3.2 Algorithmic Talent

Similarly, the surplus generated by firm i and algorithmic talent is given by:

$$S_{iA} = R_i(A_i) - W_A(A_i) - \bar{R}_i - \bar{W}_A \tag{7}$$

where $R_i(A_i)$ is the revenue generated by firm i using algorithmic capabilities A_i, $W_A(A_i)$ is the cost of providing A_i for the talent, and $\bar{R}_i$ and $\bar{W}_A$ are the disagreement payoffs for firm i and the talent, respectively.

The Nash bargaining solution is given by:

$$\max_{A_i} (R_i(A_i) - \bar{R}_i)^{\gamma} (W_A(A_i) - \bar{W}_A)^{1-\gamma} \tag{8}$$

where $\gamma \in (0, 1)$ represents the bargaining power of firm i relative to the algorithmic talent.

3 Proprietary Models versus Open-Source Models

Open-source models have the potential to aggregate decentralized talent and computing power, leading to more efficient outcomes for society Von Hippel and Von Krogh [2003], Lerner and Tirole [2002]. Proprietary model companies enjoy economic rents from the productivity gap between their models and the leading open-source models Greenwald and Stiglitz [2014]. As open-source models can reduce this gap, companies may strategically choose the timing and quality of their open-source releases Casadesus-Masanell and Ghemawat [2012], Lerner and Tirole [2002], Jiang et al. [2018].

4 Model Setup

Consider an economy with a continuum of firms indexed by $i \in [0, 1]$. Each firm can develop proprietary models and contribute to open-source models. The productivity of firm i's proprietary model is denoted by A_i^P, while the productivity of the open-source model is A^O.

4.1 Open-Source Model Productivity

The productivity of the open-source model is determined by the aggregation of decentralized talent and computing power contributions, as well as the quality of open-source releases:

$$A^O = f\left(\int_0^1 t_i di, \int_0^1 c_i di, \int_0^1 q_i di\right) \tag{9}$$

where t_i, c_i, and q_i represent the talent contribution, computing power contribution, and the quality of open-source release by firm i, respectively, and $f(\cdot)$ is an increasing and concave function Von Hippel and Von Krogh [2003], Jiang et al. [2018].

4.2 Proprietary Model Productivity

The productivity of firm i's proprietary model is given by:

$$A_i^P = g(t_i, c_i, A^O, q_i) \tag{10}$$

where $g(\cdot)$ is increasing in all arguments, reflecting the positive spillovers from the open-source model to proprietary models and the impact of the firm's own open-source release quality Lerner and Tirole [2002], Jiang et al. [2018].

4.3 Firm Optimization

Each firm aims to maximize its discounted sum of profits over an infinite time horizon, considering the economic rents from the productivity gap between its proprietary model and the open-source model, as well as the costs of talent, computing power, and open-source release quality:

$$\max_{t_{i,\tau}, c_{i,\tau}, q_{i,\tau} \tau=0^\infty} \quad \sum \tau = 0^\infty \delta^\tau \left[(A_{i,\tau}^P - A_\tau^O) - C(t_{i,\tau}, c_{i,\tau}) - K(q_{i,\tau}) \right] \tag{11}$$

subject to:

$$A_{i,\tau}^P = g(t_{i,\tau}, c_{i,\tau}, A_\tau^O, q_{i,\tau}) \tag{12}$$

$$A_\tau^O = f\left(\int_0^1 t_{i,\tau} di, \int_0^1 c_{i,\tau} di, \int_0^1 q_{i,\tau} di\right) \tag{13}$$

where $\delta \in (0, 1)$ is the discount factor, $C(\cdot)$ is the cost function for talent and computing power investments, and $K(\cdot)$ is the cost function for the quality of open-source releases Casadesus-Masanell and Ghemawat [2012], Jiang et al. [2018].

5 Equilibrium Analysis

The equilibrium of the model is characterized by the following conditions:

1. Each firm chooses its talent investment $t_{i,\tau}|_{\tau=0}^{\infty}$, computing power investment $ci, \tau|_{\tau=0}^{\infty}$, and open-source release quality $q_{i,\tau}^{*}|_{\tau=0}^{\infty}$ to maximize its discounted sum of profits, given the strategies of other firms.

2. The open-source model productivity A_{τ}^{O*} is determined by the aggregation of decentralized contributions and open-source release qualities at each time τ.

3. The proprietary model productivity of each firm $A_{i,\tau}^{P*}$ is determined by its investments $(t_{i,\tau}, c_{i,\tau})$, the open-source model productivity A_{τ}^{O*}, and its open-source release quality $q_{i,\tau}^{*}$ at each time τ.

The equilibrium can be solved using dynamic programming techniques, such as the Bellman equation, to obtain the optimal strategies of firms Stokey and Lucas Jr [1989], Ljungqvist and Sargent [2012].

6 Implications and Discussion

The extended model incorporates the strategic choice of timing and quality of open-source releases by firms. Companies balance the benefits of contributing to the open-source model, which can improve their proprietary models, with the costs of investing in talent, computing power, and the quality of open-source releases Jiang et al. [2018]. The timing and quality of open-source releases become dynamic decisions for firms, as they seek to maximize their long-term profits while considering the actions of other firms and the evolution of the open-source model. This strategic behavior can lead to delays in open-source contributions or the release of lower-quality models compared to the socially optimal outcome Lerner and Tirole [2002], Jiang et al. [2018]. The model highlights the importance of incentives and strategic considerations in shaping the development of open-source models and their impact on proprietary model productivity. The efficiency of the open-source model and its ability to close the productivity gap with proprietary models depend on the endogenous choices of firms regarding their contributions and release strategies Casadesus-Masanell and Ghemawat [2012], Jiang et al. [2018]. Future research could extend the model by incorporating heterogeneity among firms, such as differences in their capabilities or market positions, and exploring the implications for industry dynamics and welfare. Empirical studies could also test the predictions of the model using data on open-source contributions, release strategies, and firm performance in the AI industry.

6.1 Open-Source Models and Social Welfare

In the absence of open-source models, each firm's LLM quality starts from a low initial level and improves independently through Poisson jumps. However, when an open-source model with quality q_O is introduced, firms can build upon this model, effectively starting from a higher initial quality level.

Let $W(q_1, \ldots, q_N)$ be a measure of social welfare that depends on the quality levels of all LLMs in the market. We assume that W is increasing in each q_i, reflecting the positive impact of LLM quality on society.

In the absence of open-source models, the expected social welfare at time t is given by:

$$\mathbb{E}[W(q_1(t), \ldots, q_N(t))] = \int_0^\infty W(q_1, \ldots, q_N) \prod_{i=1}^N p_i(q_i, t) dq_1 \ldots dq_N \quad (14)$$

where $p_i(q_i, t)$ is the probability density function of firm i's LLM quality at time t, which evolves according to the Poisson jump process.

When an open-source model with quality q_O is introduced, the expected social welfare at time t becomes:

$$\mathbb{E}[W(q_1(t), \ldots, q_N(t))] = \int_{q_O}^\infty W(q_1, \ldots, q_N) \prod_{i=1}^N p_i(q_i, t) dq_1 \ldots dq_N \quad (15)$$

The lower bound of the integral changes from 0 to q_O, reflecting the fact that firms can now start building their LLMs from the open-source quality level.

Comparing the expected social welfare with and without open-source models, we can show that the introduction of an open-source model improves social welfare by:

$$\Delta W = \int_0^{q_O} (W(q_O, \ldots, q_O) - W(q_1, \ldots, q_N)) \prod_{i=1}^N p_i(q_i, t) dq_1 \ldots dq_N > 0 \quad (16)$$

This positive welfare gain arises because open-source models allow firms to start from a higher quality level, reducing duplicative efforts and accelerating the development of high-quality LLMs.

In summary, the incorporation of the Poisson jump process for LLM quality and the analysis of open-source models' impact on social welfare provide new insights into the dynamics of LLM development. The model suggests that open-source initiatives can play a crucial role in mitigating the inefficiencies associated with proprietary competition and improving overall welfare in the LLM ecosystem.

7 Open-Source Models and Community Contributions

7.1 Release of Open-Source Models

Consider a scenario where a company releases an open-source LLM with productivity q_O. Although q_O is lower than the productivity of the most advanced proprietary model, it is still relatively high. The release of this open-source model allows other firms and community members to build upon it, reducing the need for repetitive investments and facilitating the development of LLMs for diverse use cases [Von Hippel and Von Krogh, 2003, Lakhani and Von Hippel, 2003].

7.2 Endogenous Choice of Open-Sourcing

The decision to release an open-source model can be an endogenous choice for a company seeking to maximize the efforts of algorithmic talent on its specific technological trajectory. By attracting more talent to work on its open-source model, the company increases the likelihood of breakthrough innovations in its chosen direction [Lerner and Tirole, 2002, West and O'mahony, 2008]. The company's objective function can be written as:

$$\max_{q_O} P(B|q_O) \cdot V(B) - C(q_O) \tag{17}$$

where $P(B|q_O)$ is the probability of a breakthrough innovation B given the open-source model quality q_O, $V(B)$ is the value of the breakthrough, and $C(q_O)$ is the cost of developing and maintaining the open-source model.

7.3 Community Diversity and Long-Tailed Distribution

The LLM development community consists of M members, indexed by $j = 1, 2, \ldots, M$, with heterogeneous preferences and capabilities. Each community member j's contribution to LLM development is denoted by $y_j(t)$, which follows a power-law distribution [Clauset et al., 2009, Alstott et al., 2014, Broido and Claffy, 2019]:

$$P(y_j(t) > y) \propto y^{-\alpha} \tag{18}$$

where $\alpha > 1$ is the scaling parameter. This long-tailed distribution captures the diversity of community contributions, with a few members making significant contributions while the majority make smaller but collectively important contributions. The choice of a power-law distribution is supported by empirical evidence from open-source software development [Cosentino et al., 2017, Zhang et al., 2019] and online community participation [Muchnik et al., 2013, Yan et al., 2016].

The long-tailed distribution of community contributions can be rationalized by the preferential attachment mechanism [Barab'asi and Albert, 1999a, Merton, 1968]. In the context of LLM development, this implies that community members are more likely to contribute to and build upon models that have already received significant attention and contributions from others. This rich-get-richer phenomenon leads to a skewed distribution of downloads, likes, and other measures of popularity across LLMs, reflecting the diversity of preferences and capabilities within the community.

The long-tailed distribution of LLM popularity is non-trivial and corresponds to the notion of diversity and inclusion. It suggests that a few dominant models are likely to emerge and capture a disproportionate share of the community's attention and contributions. However, the presence of a long tail also highlights the importance of niche and specialized models that cater to specific use cases and user preferences [Anderson, 2004, Brynjolfsson et al., 2006]. This allows for a diverse ecosystem of LLMs, where both dominant and niche models coexist and contribute to the overall progress of the field, promoting diversity and inclusivity.

The emergence of a long-tailed distribution in LLM downloads and likes can be fundamentally attributed to the diversity and heterogeneity within the LLM development community. A key factor driving this phenomenon is the heterogeneous preferences and goals of different community members, leading to a demand for specialized LLMs tailored to specific use cases and requirements. As Falkinger [2020] demonstrates in the context of product markets with quality uncertainty, heterogeneous consumer preferences can result in a long-tailed distribution of product demand, with a few dominant products coexisting with a diverse range of niche offerings catering to specialized preferences. Analogously, in the LLM ecosystem, the diverse preferences and needs of developers, researchers, and end-users contribute to the emergence of a long tail of specialized and niche LLMs alongside a few dominant models.

Moreover, the development and application of LLMs require a diverse set of knowledge and expertise, spanning domains such as natural language processing, machine learning, domain-specific knowledge, and user experience design. As Akcigit et al. [2022] highlight, collaborativity and knowledge diversity play a crucial role in fostering innovation and productivity. In the LLM context, this diversity manifests through the collaboration and complementary expertise of diverse community members, facilitating the development of a diverse range of LLMs and contributing to the long-tailed distribution of downloads and likes. The collective efforts of community members with diverse backgrounds, skills, and perspectives enable the creation of LLMs that cater to a wide array of use cases and requirements, thereby shaping the long-tailed distribution.

Let $d_i(t)$ denote the number of downloads or likes of LLM i at time t. We can model the evolution of $d_i(t)$ using a stochastic process that incorporates the preferential attachment mechanism [Mandelbrot, 1953, Kong et al., 2008]:

$$dd_i(t) = \eta \frac{d_i(t)}{\sum_{k=1}^{N} d_k(t)} dt + \sigma_i dW_i(t) \tag{19}$$

where $\eta > 0$ is a parameter capturing the strength of the preferential attachment effect, σ_i is the volatility of the process, and $W_i(t)$ are independent Brownian motions. The first term on the right-hand side of the equation represents the preferential attachment mechanism, with the growth rate of downloads or likes being proportional to the current level of popularity relative to the total popularity of all LLMs. The second term captures the stochastic nature of the process, allowing for random fluctuations in popularity. This stochastic process can generate a long-tailed, power-law distribution of downloads or likes across LLMs in equilibrium [Gabaix, 1999, Newman, 2005]. The skewed distribution arises from the self-reinforcing nature of the preferential attachment mechanism, where popular models attract more attention and contributions, further cementing their dominance.

Open-source initiatives can play a crucial role in shaping the long-tailed distribution of LLM popularity. By providing access to high-quality models and encouraging community participation, open-source projects can facilitate the emergence of a diverse ecosystem of LLMs, where both dominant and niche models coexist and contribute to the overall progress of the field.

8 Consumer Preferences and Long-Tailed Distribution

To capture the diverse preferences of consumers and generate results that fit the long-tail pattern observed in the LLM ecosystem, we introduce Constant Elasticity of Substitution (CES) preferences.

8.1 CES Utility Function

Consider a continuum of consumers indexed by $i \in [0, 1]$, each with CES preferences over a variety of LLM products indexed by $j \in [0, N]$. The utility function of consumer i is given by [Dixit and Stiglitz, 1977, Anderson, 1987]:

$$U_i = \left(\int_0^N (\beta_{ij} c_{ij})^\rho dj \right)^{1/\rho} \tag{20}$$

where c_{ij} is consumer i's consumption of LLM product j, $\beta_{ij} > 0$ is a product-specific preference parameter capturing the idiosyncratic tastes of consumer i for product j, and $\rho \in (0, 1)$ is a parameter related to the elasticity of substitution between LLM products, with a lower value of ρ indicating higher substitutability.

8.2 Consumer Optimization

Each consumer i maximizes their utility U_i subject to a budget constraint:

$$\max_{c_{ij}} U_i = \left(\int_0^N (\beta_{ij} c_{ij})^\rho dj \right)^{1/\rho} \quad \text{s.t.} \quad \int_0^N p_j c_{ij} dj \leq y_i \tag{21}$$

where p_j is the price of LLM product j and y_i is consumer i's income. The first-order conditions for optimality yield the following demand function for consumer i and product j:

$$c_{ij} = (\beta_{ij}/p_j)^{1/(1-\rho)} (y_i/P_i) \tag{22}$$

where $P_i = \left(\int_0^N (\beta_{ij}^{1/(1-\rho)} p_j^{-\rho/(1-\rho)})^{1-\rho} dj \right)^{1/(1-\rho)}$ is the consumer-specific price index.

8.3 Aggregate Demand and Pareto Distribution of Preferences

The aggregate demand for LLM product j is given by:

$$C_j = \int_0^1 c_{ij} di = (p_j^{-1/(1-\rho)}/P^\rho) \int_0^1 \beta_{ij}^{1/(1-\rho)} y_i di \tag{23}$$

where $P = \left(\int_0^N p_j^{1-\rho} dj \right)^{1/(1-\rho)}$ is the aggregate price index. Assuming that the preference parameters β_{ij} follow a Pareto distribution with shape parameter $\alpha > 1$ and scale parameter $\beta_{\min} > 0$ [Reed and Jorgensen, 2001, Gabaix, 2009]:

$$P(\beta_{ij} > x) = (\beta_{\min}/x)^{\alpha} \quad \text{for} \quad x \geq \beta_{\min} \tag{24}$$

the aggregate demand C_j follows a power law with a tail index of $(\alpha(1-\rho)-1)/(1-\rho)$, provided that $\alpha > 1/(1-\rho)$. This condition ensures that the preference heterogeneity is sufficiently high relative to the substitutability between products, generating a long-tail pattern in product demand. The choice of the Pareto distribution for modeling consumer preferences is justified by empirical evidence and theoretical arguments from the economics and marketing literature, which have found that consumer preferences and product popularity often follow power-law or Pareto-like distributions in various contexts [Brynjolfsson et al., 2003, Goel et al., 2010].

9 US-China Competition and Policy Implications

9.1 Introduction

To analyze the competition between socialism large models and capitalism large models, we can develop a model that captures the unique aspects of the two systems and draws upon the insights from top economics and finance papers. The model will aim to inform the pros and cons of the two systems and provide a clear mathematical framework in the standard of top economics and finance journal papers.

9.2 Setup

Consider an economy with two types of large models: socialism large models (SLM) and capitalism large models (CLM). The SLM are developed and maintained by a centralized entity, such as a government or a public institution, while the CLM are developed by private firms driven by profit maximization incentives.

9.3 Socialism Large Models (SLM)

The productivity of SLM, denoted by A_{SLM}, is a function of the resources allocated by the centralized entity, R_{SLM}, and the coordination efficiency, $\theta \in [0, 1]$:

$$A_{SLM} = \theta f(R_{SLM}) \tag{25}$$

where $f(\cdot)$ is an increasing and concave function, reflecting the diminishing returns to resource allocation Acemoglu et al. [2012]. The parameter θ captures the ability of the centralized entity to coordinate resources and internalize externalities Holmstrom [1999]. The centralized entity aims to maximize social welfare, W, which is a function of the productivity of SLM and the resources allocated:

$$\max_{R_{SLM}} W(A_{SLM}, R_{SLM}) = A_{SLM} - c(R_{SLM}) \tag{26}$$

subject to a resource constraint:

$$R_{SLM} \leq \bar{R} \tag{27}$$

where $c(\cdot)$ is the cost function, which is increasing and convex in R_{SLM}, and $\bar{R}$ is the total available resources for AI development in the economy.

9.4 Capitalism Large Models (CLM)

The productivity of CLM, denoted by A_{CLM}, is determined by the resources allocated by private firms, R_{CLM}, and the market size, M:

$$A_{CLM} = g(R_{CLM}, M) \tag{28}$$

where $g(\cdot)$ is increasing in both arguments. The market size M is endogenously determined by the demand for AI services, which is a function of the productivity of CLM and the price, p:

$$M = h(A_{CLM}, p) \tag{29}$$

where $h(\cdot)$ is increasing in A_{CLM} and decreasing in p, capturing the positive network effects and the self-reinforcing nature of AI adoption in a market economy Goldfarb and Trefler [2018]. Private firms maximize their profits, π, by choosing the optimal level of resource allocation and price:

$$\max_{R_{CLM}, p} \pi(A_{CLM}, R_{CLM}, p) = pM - c(R_{CLM}) \tag{30}$$

where $c(\cdot)$ is the cost function, which is increasing and convex in R_{CLM}.

9.5 Equilibrium and Comparative Analysis

The equilibrium outcomes of the two systems can be determined by solving the optimization problems of the centralized entity and the private firms. The equilibrium productivity levels, A_{SLM} and A_{CLM}, the corresponding resource allocations, R_{SLM} and R_{CLM}, and the price p^* will depend on the functional forms and parameters of the model. To compare the performance of the two systems, we can analyze the relative magnitudes of A_{SLM} and A_{CLM}, as well as the social welfare implications. The model can also be extended to incorporate dynamic aspects, such as the evolution of technology and the strategic interactions between the two systems Aghion et al. [2005]. The key trade-off between the two systems lies in the coordination efficiency (θ) of the centralized entity in the SLM and the market size effects ($h(\cdot)$) and the profit incentives in the CLM. If θ is high, the SLM may achieve higher productivity and social welfare compared to the CLM by effectively internalizing externalities and avoiding duplication of efforts Bloom et al. [2013]. However, if the market size effects and the profit incentives are strong enough, the CLM may outperform the SLM in terms of productivity and innovation by harnessing the power of decentralized decision-making and market competition Hayek [1945]. Furthermore, the model can be used to study the potential for cooperation or competition between the two systems. If the SLM and CLM can share resources or knowledge, there may be potential for synergies and improved outcomes for both systems Akcigit et al. [2016]. Alternatively, if the two systems engage in a zero-sum game or an arms race, it may lead to inefficiencies and suboptimal outcomes Acemoglu and Restrepo [2018].

9.6 Conclusion

This model provides a stylized framework for analyzing the competition between socialism large models and capitalism large models, drawing upon the insights from top economics and finance papers. By capturing the unique aspects of the two systems, such as the coordination efficiency in the SLM and the market size effects and profit incentives in the CLM, the model generates insights into the potential outcomes and trade-offs of this competition. The model highlights the key role of institutional factors, such as the ability of the centralized entity to internalize externalities and the strength of market forces, in determining the relative performance of the two systems. It also suggests potential avenues for cooperation and the risks of inefficient competition between the SLM and CLM. Future research can build upon this model by incorporating more realistic assumptions, empirical evidence, and policy implications, thereby contributing to the understanding of the economic and social consequences of large model development in different institutional contexts.

9.7 Computing Power and Energy Consumption

The monopoly provider of computing power is based in the US, while China is developing its own computing power companies and infrastructure. The competition between the two countries in this domain has led to initiatives to build computing power clusters in China [Zhang, 2021]. The arms race in computing power investment can lead to negative externalities in terms of energy consumption. Policymakers should consider measures to internalize these externalities, such as carbon taxes or energy efficiency standards [Acemoglu et al., 2012].

9.8 Algorithmic Capabilities and Open-Source Models

Open-source LLMs can help bridge the gap in algorithmic capabilities between the US and China. The availability of high-quality open-source models can enable Chinese firms and researchers to build upon existing knowledge and contribute to the global LLM community [Sun et al., 2021]. Policymakers should support the development and adoption of open-source models to promote innovation and collaboration across borders. This can be achieved through funding for open-source projects, tax incentives for companies contributing to open-source development, and the establishment of international standards and protocols [Athey and Luca, 2017].

9.9 Algorithmic Talent and Human Capital

The quality and quantity of algorithmic talent in the US and China play a crucial role in the development of LLMs. The competition between the two countries extends to the attraction and retention of top talent in the field [Chen and Zhang, 2021]. Policymakers should invest in education and training programs to develop domestic talent pools and create incentives for international talent to contribute to their respective LLM ecosystems. This can include scholarships, research grants, and visa programs for high-skilled workers [Hanson, 2011].

9.10 Players, Capabilities, and Policies

Consider two countries: the United States (US) and China (CN). Each country $i \in \mathrm{US}, \mathrm{CN}$ has a set of N_i companies engaged in LLM development, with each company j's productivity denoted by $q_{ij}(t)$. The overall technological capability of each country i is given by:

$$Q_i(t) = \sum_j q_{ij}(t) \tag{31}$$

Let $x_i(t)$ denote the investment of country i in LLM development at time t, which can include direct funding, tax incentives, or other support measures, and define $s_i(t) \in [0, 1]$ as the degree of openness or collaboration of country i at time t, with higher values indicating more open-source sharing and international collaboration. Each country i's policy vector is given by:

$$p_i(t) = (x_i(t), s_i(t)) \tag{32}$$

9.11 Impact of Policies on LLM Development

The productivity growth of company j in country i is influenced by the country's policies and the global open-source knowledge pool:

$$dq_{ij}(t) = (\alpha_i x_i(t) + \beta_i Q(t))dt + \sigma_{ij} dW_{ij}(t) \tag{33}$$

where α_i and β_i are country-specific coefficients capturing the effectiveness of domestic investment and open-source knowledge absorption, respectively, $Q(t) = s_{\mathrm{US}}(t)Q_{\mathrm{US}}(t) + s_{\mathrm{CN}}(t)Q_{\mathrm{CN}}(t)$ is the global open-source knowledge pool, weighted by each country's degree of openness, and $W_{ij}(t)$ are independent Brownian motions representing company-specific productivity shocks.

9.12 Objectives, Payoffs, and Equilibrium Analysis

Each country i aims to maximize its expected long-term technological capability:

$$\max_{p_i(t)} \mathbb{E}\left[\int_0^\infty e^{-r_i t} Q_i(t) dt\right] \tag{34}$$

where r_i is the discount rate of country i. The payoff of country i depends on its own technological capability $Q_i(t)$ and the relative capability of the other country $Q_j(t)$:

$$\pi_i(t) = Q_i(t) - \gamma_i(Q_j(t) - Q_i(t)) \tag{35}$$

where $\gamma_i \in [0, 1]$ captures the degree of strategic interdependence or rivalry between the two countries. The equilibrium strategies $p_i^{(}t) = (x_i^{(}t), s_i^*(t))$ for each country i can be derived using dynamic game theory techniques, such as Markov perfect equilibria or open-loop equilibria [Fudenberg and Tirole, 1991, Basak and Hu, 2016]. Comparative statics analysis can be conducted to examine the impact of changes in key parameters, such as the effectiveness of domestic investment, absorptive capacity for open-source knowledge, or degree of strategic rivalry, on the equilibrium strategies and outcomes.

10 Parallels to Other Innovation Domains

The dynamics observed in the LLM ecosystem, such as the role of open-source models and the competition between the US and China, have parallels in other innovation domains, such as drug discovery and patent protection [Sampat, 2018]. In the pharmaceutical industry, the debate around patent protection and generic drugs mirrors the discussion on proprietary and open-source LLMs. Policymakers can learn from the experiences in these domains to design effective intellectual property regimes and innovation policies for the LLM ecosystem [Cockburn et al., 2016].

11 DeFi and the Introduction of Open-Source in Finance

Decentralized Finance (DeFi) has emerged as a transformative force in the financial industry, introducing open-source principles and technologies to a traditionally closed and proprietary system [Chen et al., 2020, Sch"ar, 2021]. DeFi protocols, built on blockchain networks such as Ethereum, enable the creation of transparent, permissionless, and interoperable financial applications, challenging the dominance of traditional financial intermediaries [Zetzsche et al., 2020, Werner et al., 2021]. The open-source nature of DeFi protocols has several implications for the financial system. First, it allows for greater transparency and auditability, as the underlying code is publicly available and can be scrutinized by anyone [Gudgeon et al., 2020]. Second, open-source DeFi protocols enable permissionless innovation, as developers can build upon existing infrastructure and create new financial products and services without the need for centralized authorities [Amler et al., 2021]. Third, the composability and interoperability of DeFi protocols allow for the creation of complex and interconnected financial systems, where different components can be combined and leveraged to unlock new use cases and efficiencies [Owen, 2020].

11.1 Long-Tailed Distribution of TVL in DeFi Protocols

One interesting observation in the DeFi ecosystem is that the distribution of Total Value Locked (TVL) across different protocols follows a long-tailed pattern [Pulse, 2021, Llama, 2021]. TVL is a measure of the total value of assets that are locked or staked within a particular DeFi protocol, and it serves as a proxy for the adoption and popularity of that protocol [Liu and Tsyvinski, 2020]. Empirical data from DeFi analytics platforms such as DeFi Pulse and DeFi Llama reveal that a small number of protocols account for a disproportionately large share of the total TVL in the ecosystem, while a long tail of smaller protocols collectively holds a significant portion of the remaining value [Angeris and Chitra, 2020, Singh and Singh, 2020]. This distribution is reminiscent of the power-law or Pareto distribution observed in various other contexts, such as city sizes, income distribution, and website traffic [Gabaix, 2016, Clauset et al., 2009]. The long-tailed distribution of TVL in DeFi protocols can be explained by several factors, including network effects, first-mover advantages, composability, and the open-source nature of DeFi, which enables the rapid proliferation of new protocols and forks, leading to a highly diverse and fragmented ecosystem [Gudgeon et al., 2020, Young, 2020, Zhang et al., 2021].

11.2 A Model for DeFi, Open-Source, and Long-Tailed TVL Distribution

To capture the dynamics of the DeFi ecosystem, we propose a model that incorporates open-source protocols, stochastic growth, and preferential attachment. Consider a DeFi ecosystem with N open-source protocols, indexed by $i = 1, 2, \ldots, N$. Each protocol i has a certain level of total value locked (TVL), denoted by V_i, which evolves according to a geometric Brownian motion (GBM) [Hull, 2003]:

$$dV_i(t) = \mu_i V_i(t)dt + \sigma_i V_i(t)dW_i(t) \tag{36}$$

where $V_i(t)$ is the TVL of protocol i at time t, μ_i is the expected growth rate, σ_i is the volatility, and $W_i(t)$ is a standard Brownian motion. The growth rate μ_i can be decomposed into an organic growth rate r_i due to the protocol's inherent utility and attractiveness, and a network effect and composability benefit λ_i derived from the open-source ecosystem [Cong et al., 2021]:

$$\mu_i = r_i + \lambda_i \tag{37}$$

The network effects and composability benefits λ_i can be modeled using a preferential attachment mechanism [Barab'asi and Albert, 1999b]. The probability Π_i that a new user or developer chooses to interact with protocol i is proportional to its TVL:

$$\Pi_i = \frac{V_i}{\sum_{j=1}^{N} V_j} \tag{38}$$

This preferential attachment mechanism reinforces the growth of popular protocols, leading to a rich-get-richer phenomenon [Liu et al., 2021]. The combination of the open-source ecosystem, stochastic growth, and preferential attachment mechanism can generate a long-tailed distribution of TVL across DeFi protocols. The distribution can be approximated by a power law [Clauset et al., 2009]:

$$P(V > v) \propto v^{-\alpha} \tag{39}$$

where $P(V > v)$ is the probability that a protocol has a TVL greater than v, and $\alpha > 0$ is the scaling exponent. The smaller the value of α, the heavier the tail of the distribution. The long-tailed distribution emerges due to the heterogeneity in protocol growth rates and the reinforcing effects of preferential attachment [Mitzenmacher, 2004]. A few protocols with high TVL and strong network effects dominate the ecosystem, while a large number of smaller protocols coexist, serving niche markets and providing a diverse range of services [Wang et al., 2021].

11.3 Parallel with the LLM Ecosystem

The emergence of open-source principles in the financial system through DeFi shares several parallels with the development of the Large Language Model (LLM) ecosystem. Both domains have witnessed the transformative impact of open-source technologies, enabling permissionless innovation, collaborative development, and the democratization of access to powerful tools and resources [Nguyen et al., 2022, Gao et al., 2022]. Moreover, the long-tailed distribution of TVL in DeFi protocols mirrors the distribution of contributions and adoption in open-source LLM communities, where a few dominant models and platforms coexist with a diverse array of smaller and more specialized initiatives [Huang et al., 2022, Qiu et al., 2022]. This similarity suggests that the underlying dynamics of network effects, composability, and user preferences shape the structure and evolution of both ecosystems. Drawing insights from the DeFi experience, the LLM community can leverage the power of open-source collaboration while also being mindful of the risks and challenges associated with concentration and fragmentation. Fostering a healthy and sustainable ecosystem that balances innovation, inclusivity, and stability will be crucial for realizing the full potential of LLMs in various domains, from natural language processing and content generation to knowledge management and decision support [Bommasani et al., 2021, Zeng et al., 2022].

12 Conclusion

This paper presents a comprehensive model of the LLM ecosystem that captures the key roles of computing power, algorithms, and data as scarce resources. The model incorporates Nash bargaining between firms and resource providers, separating equilibria based on firm size, and the endogenous choice of open-sourcing by companies. The analysis highlights the potential of open-source LLMs to reduce arms races, enable diverse applications, and bridge the algorithmic capability gap between the US and China. We show that open-source models and communities can reduce duplicative investments and resource waste caused by arms races, promote diversity and inclusivity in the digital ecosystem, and foster long-tail economics. Policymakers should consider measures to internalize the negative externalities of computing power investment, support open-source initiatives, and invest in human capital development. The parallels drawn between the LLM ecosystem and other innovation domains, such as drug discovery and patent protection, provide valuable insights for the design of effective policies and incentive mechanisms. Furthermore, we explore the similarities between the LLM ecosystem and the emergence of open-source principles in the financial system through DeFi, highlighting the common underlying dynamics shaping the structure and evolution of both ecosystems. As LLMs and DeFi continue to evolve and mature, drawing insights and lessons from each other's experiences can help navigate the challenges and opportunities ahead. By embracing open-source principles while addressing the risks and limitations, these ecosystems can chart a path towards a more inclusive, resilient, and impactful future, redefining the boundaries of artificial intelligence and finance.

References

Daron Acemoglu and Pascual Restrepo. Artificial intelligence, automation, and work. *NBER Working Paper*, (24196), 2018.

Daron Acemoglu, Philippe Aghion, Leonardo Bursztyn, and David Hemous.

The environment and directed technical change. *American Economic Review*, 102(1):131–66, 2012.

Philippe Aghion, Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728, 2005.

Rizwan Ahmed, Alexandre Masson, and Emma Strubell. Measuring the carbon intensity of ai in cloud instances. *arXiv preprint arXiv:2206.05229*, 2022.

Ufuk Akcigit, Murat Alp Celik, and Jeremy Greenwood. Buy, keep, or sell: Economic growth and the market for ideas. *Econometrica*, 84(3):943–984, 2016.

Ufuk Akcigit, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi. Collaborativity and knowledge diversity. *American Economic Review*, 112(8):2603–2640, 2022.

Jeff Alstott, Ed Bullmore, and Dietmar Plenz. Powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777, 2014.

Hannes Amler, Peter Sardon, Lukas Eckey, Christian Badertscher, and Peter Egger. Defi: On the prospects of decentralized financial market infrastructure and stability-enhancing regulation. *arXiv preprint arXiv:2112.09135*, 2021.

Chris Anderson. *The long tail*. Random House Business, 2004.

Lindray J Anderson. A demand curve analysis for tourist arrivals. *Tourism Management*, 8(4):283–290, 1987.

Guillermo Angeris and Tarun Chitra. Improved price oracles: Constant function market makers. *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies*, pages 80–91, 2020.

Susan Athey and Michael Luca. The adoption and impact of open source software by us local governments. *NBER Working Paper*, (23929), 2017.

Albert-L'aszl'o Barab'asi and R'eka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999a.

Albert-L'aszl'o Barab'asi and R'eka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999b.

Gopal K Basak and Minyi Hu. Stochastic differential games: A control theoretic approach. *Springer*, 2016.

Nicholas Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4): 1347–1393, 2013.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Anna Broido and KC Claffy. Scale-free and efficient behavior of the internet router network. *IEEE/ACM Transactions on Networking*, 27(2):844–857, 2019.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

Erik Brynjolfsson and Lorin M Hitt. Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4):23–48, 2000.

Erik Brynjolfsson, Yu Jeffrey Hu, and Michael D Smith. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11):1580–1596, 2003.

Erik Brynjolfsson, Feng Zhu, Amy Austin Renshaw, Michael Greenstone, and Jeffrey Watson. Niches and networks: Prospects for expanding the use of information technology. *Frontiers of Broadband, Electronic and Mobile Commerce*, pages 65–104, 2006.

Ramon Casadesus-Masanell and Pankaj Ghemawat. Competitive strategy in the presence of open source software. *Management Science*, 58(6):1056–1073, 2012.

Wenhu Chen, Zhi Xu, Shuyu Shi, Yang Zhao, and Jun Zhao. Blockchain and cryptocurrencies: Model, techniques, and applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(9):3211–3221, 2020.

Xiangyu Chen and Rong Zhang. Global race for ai talent: The rapidly evolving landscape. *arXiv preprint arXiv:2104.06625*, 2021.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*, 2022.

Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

Iain M Cockburn, Jean O Lanjouw, and Mark Schankerman. Patents and the global diffusion of new drugs. *American Economic Review*, 106(1):136–64, 2016.

Lin William Cong, Zhiguo He, and Jiasun Wang. Decentralized finance: On the interaction of blockchain and financial intermediation. *Available at SSRN 3889499*, 2021.

Valerio Cosentino, Javier Luis, and Jordi Cabot. Github towards open collaboration and release: A case study using github in free/open-source projects courses. *Journal of Computer Science Education*, 27(2):81–97, 2017.

Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308, 1977.

Josef Falkinger. Long tails and specialization in markets with quality uncertainty. *Journal of Economic Theory*, 190:105116, 2020.

Drew Fudenberg and Jean Tirole. *Game theory*. MIT press, 1991.

Xavier Gabaix. Zipf's law for cities: an explanation. *The Quarterly Journal of Economics*, 114(3):739–767, 1999.

Xavier Gabaix. Power laws in economics and finance. *Annu. Rev. Econ.*, 1 (1):255–294, 2009.

Xavier Gabaix. Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206, 2016.

Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. Scaling language models: Methods, analysis insights from training gopher. *arXiv preprint arXiv:2112.11446*, 2022.

Sharad Goel, Andrei Broder, Evgeniy Gabrilovich, and Bo Pang. The anatomy of the long tail: Ordinary people with extraordinary tastes. *Wsdm. acm. org*, pages 105–114, 2010.

Avi Goldfarb and Daniel Trefler. Ai and international trade. *NBER Working Paper*, (24254), 2018.

Bruce Greenwald and Joseph E Stiglitz. Competition and innovation: Did arrow hit the bull's eye? *The rate and direction of inventive activity revisited*, pages 361–404, 2014.

Lewis Gudgeon, Daniel Perez, Dominik Harz, Benjamin Livshits, and Arthur Gervais. The decentralized financial crisis. *IEEE Security Privacy Workshops*, pages 1–13, 2020.

Gordon H Hanson. Immigration, productivity, and competitiveness in american industry. *Competing for Talent: The United States and High-Skilled Immigration*, pages 1–27, 2011.

Friedrich A Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.

Bengt Holmstrom. The firm as a subeconomy. *Journal of Law, Economics, and Organization*, 15(1):74–102, 1999.

Shih-Ting Huang, Yi-Ping Fang, and Hung-Yu Chen. Challenges in open-source machine learning software: A survey. *arXiv preprint arXiv:2208.07604*, 2022.

John Hull. *Options futures and other derivatives*. Pearson Education India, 2003.

Qixing Jiang, Li Zhang, and Huayu Shen. Open source software contribution strategy. *Decision Support Systems*, 111:1–12, 2018.

Juan Kong, Songfa Li, Amber Whiteley, and Lei Zhang. Experience vs. description: Explaining a preference reversal. *Journal of Behavioral Decision Making*, 21(4):363–372, 2008.

Karim R Lakhani and Eric Von Hippel. How open source software works:"free" user-to-user assistance. *Research policy*, 32(6):923–943, 2003.

Josh Lerner and Jean Tirole. Some simple economics of open source. *The journal of industrial economics*, 50(2):197–234, 2002.

Luyao Liu, Yuanjun Feng, Quang Nguyen, Jianbo Zhao, and Zhengpeng Wang. The complex evolutionary dynamics of defi. *Available at SSRN 3892784*, 2021.

Yukun Liu and Aleh Tsyvinski. Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727, 2020.

Lars Ljungqvist and Thomas J Sargent. *Recursive macroeconomic theory.* MIT Press, 2012.

DeFi Llama. Defi llama. `https://defillama.com/`, 2021.

Benoit B Mandelbrot. An informational theory of the statistical structure of languages. *Communication theory*, 84:486–502, 1953.

Robert K Merton. *The Matthew effect in science*, volume 159. Science, 1968.

Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

Lev Muchnik, Sen Pei, Lucas C Maia Parra, Saulo DS Reis, Jádson S Andrade Jr, Shlomo Havlin, and Hernán A Makse. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific reports*, 3(1):1–7, 2013.

John F Nash Jr. The bargaining problem. *Econometrica*, pages 155–162, 1950.

Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

Giang Nguyen, Lauren Purita, Milo Lesniewski, and Nicolas Papernot. Community and self-governance in open source machine learning: A case study of hugging face. *arXiv preprint arXiv:2203.13325*, 2022.

Taylor Owen. Component-based development in decentralized financial markets. *Available at SSRN 3606072*, 2020.

DeFi Pulse. Defi pulse. `https://defipulse.com/`, 2021.

An Qiu, Yilun Cui, Ji Zhang, Cihui Peng, Xiaojiang Li, Josiah Poon, Philip Yates, Tony Zhou, and Mingming Ding. Elude: Generating coherent explanations for ml model predictions. *arXiv preprint arXiv:2212.09396*, 2022.

William Jack Reed and Michael Jorgensen. Pareto analyses of the short-course heroin distribution system. *Reproductive Health*, 58(1):81–92, 2001.

Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.

Bhaven N Sampat. A survey of empirical evidence on patents and innovation. *NBER Working Paper*, (25383), 2018.

Fabian Sch"ar. Decentralized finance: On blockchain-and smart contract-based financial markets. *FRB of St. Louis Review*, 2021.

Kushagra Singh and Rakesh Kumar Singh. Defi: Building smart banking apps using decentralized finance. *International Journal of Research in Engineering, Science and Management*, 3(6):22–24, 2020.

Nancy L Stokey and Robert E Lucas Jr. *Recursive methods in economic dynamics*. Harvard University Press, 1989.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*, 2019.

Zihan Sun, Haibin Zhu, and Guoqiang Zhao. Open-source machine learning frameworks: Trends and challenges. *arXiv preprint arXiv:2111.00358*, 2021.

Eric Von Hippel and Georg Von Krogh. Open source software and the "private-collective" innovation model: Issues for organization science. *Organization science*, 14(2):209–223, 2003.

Shangqi Wang, Yun-Feng Ho, Ding-Yong Tsai, and Pin-Chun Wang. Decentralized finance: Blockchain technology and the quest for an open financial system. *The Journal of Fintech*, 4(01):59–85, 2021.

Sebastian M Werner, Swati Gubili, Ferdinand Oberdorf, Nicholas Peters, and Christian Peukert. Decentralized finance–the possibilities and limitations of blockchain-based financial services. *Available at SSRN 3852173*, 2021.

Joel West and Siobhan O'mahony. *Creating open source software*. Harvard Business Press Boston, MA, 2008.

Chaoyi Yan, Benjamin Cheung, Clare Kelly, Stanley Colcombe, R Cameron Craddock, Adriana Di Martino, Qingyang Li, Xi-Nian Zuo, F Xavier Castellanos, and Michael P Milham. On the dynamic nature of inter-human brain-to-brain synchrony. *Frontiers in human neuroscience*, 10: 190, 2016.

Ernst LLC Young. Defi: On the prospects for decentralized, disintermediated financial market infrastructure and stability-enhancing regulation. *Available at SSRN 3711777*, 2020.

Wangjun Zeng, Ning Liu, Chuan Qin, Haotian Zhang, and Xiang Ren. Large language models can be easily distracted by irrelevant context. *arXiv preprint arXiv:2205.11824*, 2022.

Dirk A Zetzsche, Douglas W Arner, and Ross P Buckley. Decentralized finance (defi). *Available at SSRN 3539194*, 2020.

Anqi Zhang. Regulation of artificial intelligence in china: Challenges, dynamics, and prospects. *Policy and Society*, pages 1–17, 2021.

Lingxiao Zhang, Jian Wei, Lei Chen, Jianjun Chen, Huaxi Xu, and Xiaoying Xie. Degate: Bridging decentralized finance with centralized finance. *arXiv preprint arXiv:2106.02197*, 2021.

Lisha Zhang, Kevin Tian, Shanshan Wang, Haipeng Zheng, Brian Adams, and Fei Zhang. Github and the software factory: A quantitative analysis of software project hosting platforms. *IEEE Transactions on Software Engineering*, 45(12):1221–1239, 2019.