The Macroeconomic Impact of AI-Generated Data

Zhifeng Cai *

September 9, 2024

Abstract

With the rise of generative AI, there has been an increase in AI-generated synthetic data. Some argue that this type of data offers certain advantages over human-generated data, while others caution against the risk of "data corruption," where AI-generated data could degrade future datasets and eventually reduce AI's productivity. To assess the macroeconomic impact of AI-generated data, this paper proposes a social learning model of endogenous AI adoption, where both AI and labor can produce and generate data of differential quality. Drawing on recent experimental findings from AI research, the model suggests that the data corruption risk could have significant quantitative implications, potentially reducing AI-driven labor displacement by 30% over the long term. The equilibrium is inefficient because individuals do not consider the impact of their actions on overall data quality. Therefore, regulations on AI adoption and limits on AI's access to data may enhance overall welfare.

*First version: July 2023. This paper was circulated under the title "Artificial Intelligence, Data Corruption, and Labor Displacement". Cai: Department of Economics, Rutgers University, 75 Hamilton St, New Brunswick, NJ 08901, USA., (email: zhifeng.cai@rutgers.edu). I thank Todd Keister, Diego Perez, Laura Veldkamp, Venky Venkateswaran, Liyan Yang, and Yucheng Yang for comments. ChatGPT and Claude provide text-editing services. All errors are my own.

1 Introduction

With the rapid adoption of generative AI, there has been a significant increase in AI-generated synthetic data. This data is now being widely used across various fields, from research and commercial forecasting to gaming and training new AI models. For example, tools like Midjourney and DALL-E2 enable users to create digital artwork, which, once shared online, may become part of future AI training datasets. Companies like Amazon use synthetic data to train their Alexa language assistant, while Apple has trained its Mm1 model using pre-trained data from GPT-4 and LLaMA. A report by the European Union Law Enforcement Agency predicts that by 2026, up to 90% of online content may be AI-generated. As the use of synthetic data expands, businesses specializing in its generation have emerged. The synthetic data market is expected to grow from 381.3 million in 2022 to 2.1 billion by 2028, with an annual growth rate of 33.1%.

Many attribute the rise of AI-generated data to its advantages over human-generated data. Synthetic data offers benefits such as anonymity, enhanced privacy, and compliance with legal requirements. It is also used in situations where human-generated data is scarce, boosting data availability and improving outcomes in machine learning and AI training contexts.

More recently, concerns have emerged within the AI research community that AI-generated content may crowd out human-generated data, potentially leading to undesirable outcomes when these datasets are used to train future AI models. This issue of "data corruption" gained significant attention since 2023. For example, Shumailov et al. (2024) demonstrate that recursively training AI models on AI-generated data can result in "model collapse," a phenomenon also observed by many other studies surveyed in Table 1, where the quality of AI-generated content deteriorates significantly over time.

The concept of data corruption is illustrated in Figure 1. Generative AI models, such as ChatGPT, rely on large-scale datasets for effective training (represented by the black arrow). But where do these datasets come from? Before the widespread use of AI, these datasets primarily consists of data generated by human activities (depicted by the blue arrow). These "real data" were then fed into AI models for training. However, with the growing popularity of generative AI models, training datasets are increasingly mixed with "synthetic data"—produced not through authentic human activities, but by AI models themselves. For example, AI tools like Midjourney and DALL-E2 allow users to create digital artwork that, once shared online, can become part of the training data for future AI models. Such incorporation of synthetic data into training could have significant

Table 1: Recent AI Papers on Data Corruption

Title	*Authors*	*Date*
The Curse of Recursion: Training on Generated Data Makes Models Forget	Shumailov et al. (Oxford)	May 2023
Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet	Martínez et al. (Madrid)	June 2023
Self Consuming Generative Model Go Mad	Alemohammad et al. (Stanford)	July 2023
Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow	Rio-Chanona et al. (Harvard)	July 2023
Will Large-scale Generative Models Corrupt Future Datasets?	Hataya et al. (Kyoto)	Aug 2023
Generative artificial intelligence enhances individual creativity but reduces the collective diversity of novel content	Doshi et al. (UCL)	Aug 2023
The Curious Decline of Linguistic Diversity: Training Language Models on Synthetic Text	Guo et al. (École Polytechnique)	Nov 2023
Large Language Models Suffer From Their Own Output: An Analysis of the Self-Consuming Training Loop	Briesch et al. (Johannes Gutenberg)	Nov 2023
AI-Generated Images Introduce Invisible Relevance Bias to Text-Image Retrieval	Xu et al. (Chinese Academy of Sciences)	Nov 2023
Nepotistically Trained Generative-AI Models Collapse	Bohacek and Farid (Stanford)	Nov 2023
Under the Surface: Tracking the Artifactuality of LLM-Generated Data	Das et al. (U of Minnesota)	Jan 2024
Model Collapse Demystified: The Case of Regression	Dohmatob et al. (New York University)	Apr 2024
Self-Consuming Generative Models with Curated Data Provably Optimize Human Preferences	Ferbach et al. (Montréal)	June 2024
Towards Theoretical Understandings of Self-Consuming Generative Models	Fu et al. (USTC)	June 2024
Regurgitative Training: The Value of Real Data in Training Large Language Models	Zhang et al. (Tsinghua University)	July 2024

Note: Papers are listed in chronological order. For the sake of space, only the first academic affiliation is listed. The list does not meant to be exhaustive.

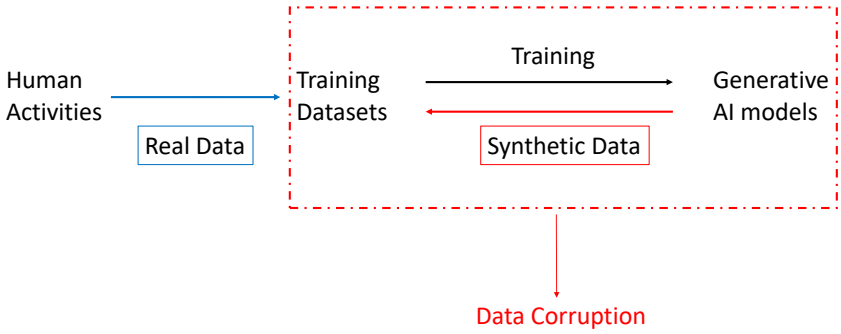


Figure 1: Data Corruption with Generative AI

effects on AI productivity over the long run.

Given the growing importance of AI-generated data and the ongoing debate surrounding it, this paper aims to explore the macroeconomic implications of the rise of AI-generated data, particularly in the context of AI-human competition and potential regulatory measures. The key contribution of this work, compared to previous studies, is to provide a micro-founded model that examines the relationship between AI and data quality. To achieve this, the paper develops a social-learning model with endogenous AI adoption, where both AI and labor can produce, and their production behavior generates useful data—potentially of varying quality—for training future AI models. Data is valuable in this model because it can help predict the value of certain economic fundamentals (Farboodi and Veldkamp, 2022). A key advantage of AI over human labor is that AI has access to rich historical data generated by past economic activities, whether conducted by humans or AI. Since AI-generated data may differ in quality from human-generated data, the *composition* of economic activities carried out by AI versus humans could affect overall data quality and, consequently, AI productivity. In turn, AI productivity affects the attractiveness of adopting AI over human labor, which determines the equilibrium share of AI-driven production and the resulting composition between real and synthetic data. As a result, the model features an equilibrium relationship between AI adoption and data quality.

One of the key theoretical insights from the model is that the quantitative impact of AI-generated data—specifically how AI adoption affects data quality—depends on AI's capacity to produce high-quality new information rather than relying solely on historical public data. Consider two

extreme scenarios. In the first, AI functions like traditional machine learning, drawing entirely from historical data without generating new information. In this case, AI activities and the resulting data add no new value beyond what is already in the dataset. If AI crowds out human activities, the quality of the training datasets would degrade over time, leading to severe data corruption. In contrast, imagine AI is highly effective at producing high-quality new data, similar to how AlphaGo generates valuable insights in the game of Go without relying on historical data. In this scenario, the model predicts no data corruption, and increased AI adoption would actually improve data quality overall.

The questions is, is reality closer to the first possibility or the second? This paper attempts to answer that by combining the model with recent computer science research on AI-generated data (see Table 1). This body of research evaluates how foundational AI models, such as large language models (LLMs) and image generation tools, are affected by the influx of AI-generated data. The findings generally show that recursively training AI models with AI-generated data leads to a decline in quality in later generations. This evidence helps the model quantify the relative quality of AI-generated versus human-generated data. The resulting calibration suggests that AI generates new data (or ideas) at only 8% of the quality of human-generated data. This significant gap is driven by the sharp decline in AI quality as synthetic data is incorporated into the training process (Chen et al. 2023; Alemohammad et al. 2023; Martínez et al. 2023). This indicates that the data corruption channel could be quantitatively important across a wide range of generative tasks.[1]

Given the presence of data corruption, what are the model's implications for the labor market, particularly regarding AI-human competition? How should we regulate AI adoption, and should we limit AI's access to data? These are crucial questions that have stimulated significant academic and policy discussions in recent years. The model developed in this paper provides a framework to address these issues.

The model predicts that labor displacement is less concerning over the long run due to the rise of AI-generate data. To illustrate this, consider the model's dynamics after a one-time permanent reduction in AI technology costs. The size of this shock is calibrated to cause an immediate spike in AI adoption to 100%, representing a scenario where AI completely displaces labor in the short run—a severe case of labor displacement. Over time, however, as more AI-generated data is

[1]Note that this study is quantified with experimental findings on generative tasks, most notably content generation tasks such as image generation and text processing. It does not rule out the possibility that AI might generate highly useful data in certain contexts, such as gaming. Exploring other applications remains an interesting avenue for further research.

produced, the quality of this data, being lower than human-generated data, gradually reduces the quality of the training dataset and hence AI productivity. As a result, the share of firms using AI decreases, and the equilibrium stabilizes with AI making up about 70% of total production. Thus, the model suggests that the data corruption channel can reverse approximately 30% of the initial labor displacement over the long run.[2]

The fact that labor displacement is less concerning over the long run does not allieviate the need for government intervention. Intervention is necessary due to a data externality: individuals do not account for the impact of their private actions on the quality of aggregate datasets. We consider in the model two margins of private actions, one on AI adoption and the other on AI's access to historical data. Both are central to current debates on AI regulation. For example, the EU Artificial Intelligence Act, the first comprehensive law on AI, outlines various areas where AI usage should be regulated, including prohibited practices and data governance. The artificial intelligence and Data Act (AIDA) of Canada also outlines similar requirements for AI adoption and data usage. The model shows that, with data corruption, taxing AI adoption—even in the long run—can correct the negative externality where increased AI usage lowers overall data quality.

We also explore whether limiting AI's access to datasets improves welfare. A common justification for such regulations is to protect privacy and ensure safety and fairness in AI systems. Beyond these ethical concerns, this paper highlights an additional efficiency benefit of constraining AI's access to data: it "forces" AI to prioritize innovation and generate its own new information, instead of relying on public data, which helps improve the quality of AI-generated data. The model shows that, as data corruption intensifies, the cost of restricting AI's access to data decreases over time, strengthening the case for imposing limits on AI's access to data.

The paper connects to the literature that investigates the influence of data on the macroeconomy, as discussed in the works of Farboodi et al. (2019); Jones and Tonetti (2020); Abis and Veldkamp (2021); Farboodi and Veldkamp (2022). The paper's novelty lies in proposing a model that jointly determines data quality and AI adoption. It further provides insights into policy implications and the regulation of AI to manage the quality of data efficiently.

The paper also relates to recent research on the impact of AI on labor, as explored by Acemoglu and Restrepo (2018, 2019, 2022); Acemoglu et al. (2022); Moll et al. (2022) and Alonso et al. (2022). In this context, the paper considers the endogeneity of AI productivity with respect to data quality.

[2]Consistent with this paper's findings, Rio-Chanona et al. (2023) observe that the introduction of large language models like ChatGPT leads to a reduction in human-generated content on platforms such as Stack Overflow.

It examines the implications of this relationship for labor displacement effects across various tasks characterized by different levels of uncertainty. It is the first paper that introduces experimental evidence from the computer science literature to discipline a model of AI.

Moreover, the paper is linked to the literature on information economics and social learning, as explored by Morris and Shin (2002); Veldkamp (2005); Amador and Weill (2010); Ordonez (2013) and Fajgelbaum et al. (2017). Its key contribution is the application of a social learning framework to analyze the issue of AI adoption. It builds on the insights in this literature that more economic activities lead to more data being generated and hence in general better data quality. This paper further argue that the *composition* of data from different sources, i.e. how much activities are conducted by AI v.s. human, matters for aggregate data quality.

2 Model

2.1 Primitives, Technology, and Friction

Time is assumed to be infinite, spanning from $-\infty$ to $+\infty$. In each period, there exist $\bar{N}$ number of entrepreneurs. Each entrepreneur has a lifespan of 1 period and is replaced by a newborn entrepreneur at the start of the next period. These entrepreneurs have the ability to produce a good of quality A_t^i, which they can sell to a representative household at a price of pA_t^i. The representative household is assumed with a utility function that is linear in the quality A_t^i, and for convenience, we normalize the slope of this linear relationship to 1. Therefore, we have $p = 1$.[3]

The quality of the production A_t^i is endogenous, and is influenced by the entrepreneur's action a_t^i. Specifically, I adopt the following functional form as in Farboodi and Veldkamp (2022):[4]

$$A_t^i = \bar{A} - \left(\theta_t - a_t^i\right)^2 \tag{1}$$

Where $\bar{A}$ is the maximum quality that can be achieved by some optimal action θ_t. It is assumed

[3]By assuming short-lived entrepreneurs, we bypass the need to address the issue of data hoarding, as highlighted by Jones and Tonetti (2020), which demonstrates the substantial social costs resulting from the nonrivalrous nature of data. By abstracting from this aspect, we can direct our attention towards the innovative mechanism presented in this paper.

[4]Farboodi and Veldkamp (2022) also explore a general setting under which the quality is given by:

$$A_t^i = g\left[\left(\theta_t - a_t^i\right)^2\right],$$

where g is a differentiable and strictly decreasing function. All of this paper's result goes through with this generalization.

that θ_t evolves according to

$$\theta_t = \rho\theta_{t-1} + \eta_t \tag{2}$$

where η_t is a noise with mean 0 and precision γ_η.

There is an information friction in the model: θ_t is not directly observable to any economic participants. This information frictions is the only friction in the model. Absent this friction, enterprenuers would always take actions alligned with the fundamental and the equilirbium achieves the first best social optima.

2.2 Timing, Information, and the Firm's Problem

Each entrepreneur faces a choice between using labor, incurring an idiosyncratic labor cost of φ_t^i, or employing AI technology, incurring a capital cost of R_t. The key difference between labor and AI is th way they predict the fundamental and thus take the associated actions.

At the start of period t, entrepreneurs are born with their individual labor cost φ_t^i, which is drawn from a distribution $F(\varphi)$. Then the fundamental θ_t is realized, together with a public signal that is observable by all individuals:

$$S_t = \theta_t + \varepsilon_t^S$$

Here, ε_t^S represents noise in the public signal with a mean of 0 and precision γ_S. The parameter γ_S captures the level of uncertainty associated with a particular task. For instance, tasks like go-playing might have low uncertainty, while more complex real-life tasks such as diagnosing patients can exhibit higher task uncertainty.

Labor Production: The entrepreneur then make a decision regarding whether to utilize their own labor or adopt AI technology for the task. If they choose to employ their own labor, it generates a private signal given by:

$$s_t^{Li} = \theta_t + \varepsilon_t^{Li}. \tag{3}$$

The precision of this private signal is denoted as γ_L. Based on this signal, the entrepreneur can produce the good using human labor at a cost of φ_t^i. The cost encompasses both the expenses associated with information acquisition and the production cost with labor. In this case, the entrepreneur's optimal action is determined by:

$$W^L(S_t, s_t^{Li}) = \max_{a_t^i} E(A_t^i(a_t^i)|S_t, s_t^{Li}) \tag{4}$$

The ex-ante payoff from labor production can be expressed as:

$$V^L(S_t, \varphi_t^i) = E(W^L(S_t, s_t^{Li})|S_t) - \varphi_t^i$$

Here, the superscript l indicates that the entrepreneur is utilizing their own labor for production.

AI production: Let's now explore the considerations of entrepreneurs who choose to adopt AI in their production processes. The adoption of AI incurs a cost, denoted as R_t, which entrepreneurs in period t need to pay for the AI product. R_t is treated as an exogenous required rate of return, consistent with the perspective of Acemoglu and Restrepo (2018) and others, who view AI technology as a capital good. Once the entrepreneur purchases the AI product by paying R_t, the AI carries out an action a_t^i on behalf of the entrepreneur.

How does AI pick the action? The AI has access to what is referred to as "big data," which encompasses an extensive dataset containing all past public signals and past actions taken by entrepreneurs. Mathematically, this can be represented as the following information set:

$$\Omega_{t-1} = \{S_{t-j}, D_{t-j}^i\}_{j=1,2,3,...,-\infty}^{i=1,2,3,...\bar{N}}$$

Here, S_{t-j} denotes past public signals, and D_{t-j}^i represents data points regarding past actions taken by previous entrepreneurs. The process of how this data is generated will be discussed later.

In addition to historical data, the AI is also capable of generating its own data, which can be expressed as:

$$s_t^{Ai} = \theta_t + \varepsilon_t^{Ai}, \varepsilon_t^{Ai} \sim N(0, 1/\gamma_A)$$

The utilization of these two sources of information by the AI can be interpreted as follows. The use of large datasets incorporating past data (Ω_{t-1}) to predict future outcomes is a common practice in many applications of machine learning. For example, retail firms may utilize historical data to estimate future market demand, and hospitals might employ data from past patient records to make diagnoses. In addition to historical data, AI can also create its own, new data points and act upon it. This captures AI's "creativity" or its ability to "innovate". The quality of such new data is given by γ_A. If $\gamma_A \to 0$, AI in the model converges to a pure machine learning tool while if $\gamma_A \to \infty$, AI's action would only depend on its newly generated signal and not on past data. In this case, therefore, AI would be very creative.[5]

It is important to note that the emphasis on prediction versus new data generation may vary across different industries. In certain domains like automated driving or Go-playing, where the

[5]See section X for an analysis of AI's optimal action.

reward system is relatively straightforward and deterministic, or in research settings where AI can explore novel drug combinations in high-dimensional spaces, the generation of new data could be highly valuable. Conversely, in a market context where AI is used to estimate demand, prediction based on past data might play a more significant role. The model accommodates both dimensions. In the numerical section we attempt to use experimental evidence from Large language and image generation models to assess the importance of both dimensions.

Based on both sources of information, the AI picks an action to maxmizes its value to the entrepreneur:

$$W^A(S_t, s_t^{Ai}, \Omega_{t-1}) = \max_{a_t^i} E(A_t^i(a_t^i)|S_t, s_t^{Ai}, \Omega_{t-1}) \tag{5}$$

The payoff from AI adoption is then given by:

$$V^A(S_t) = E(W^A(S_t, s_t^{Ai}, \Omega_{t-1})|S_t)$$

Here, the superscript A indicates that the entrepreneur has adopted AI for production.

Given the value functions, the agent selects the option that results in the highest ex-ante value:

$$\max\{V^l(S_t, \varphi_t^i), V^A(S_t)\}$$

Due to the monotonicity of $V^l(S_t, \varphi_t^i)$ with respect to φ_t^i, there exists a cutoff value denoted as $\bar{\varphi}$ such that agents adopt AI only if their individual φ_t^i exceeds this cutoff. This determines the number of labor-adopting entrepreneurs, denoted by N_t. The number of AI-adopting entrepreneurs is thus given by $\bar{N} - N_t$.

2.3 Data Generation

Let us now discuss the data generating process and how data points D_t^i are related to economic fundamental. When an entrepreneur takes an action, a data point is recorded as a noisy signal about that action:

$$D_t^i = a_t^i + \varepsilon_t^{Di}. \tag{6}$$

Here, the noise $\varepsilon_t^{Di} \sim N(0, \bar{N}/\gamma_D)$ can be interpreted as the information loss in the data collection process. Following Fajgelbaum et al. (2017), we assume that the precision of these data points D_t^i are inversely proportional to $\bar{N}$. This prevents the signals from being fully revealing as $\bar{N}$ tends to infinity. This assumption captures the notion that noise persists even with big data due to the increasing complexity of large economies and the challenges of handling larger datasets with constrained computational power.

In our model, we make the assumption that actions are directly observable, while private information is not. This assumption is commonly found in the social learning literature (see, for example, Fajgelbaum et al. 2017), where companies consider their data as valuable assets and refrain from sharing it with other firms. For instance, Facebook's private data is not directly visible to external parties, but its various business activities are, allowing outsiders to make partial inferences about the information held by Facebook based on its actions.

Given this setup, the idiosyncratic data points in a given period can be summarized into the aggregate means that reflect information generated by human labor and AI, respectively:

$$X_t^l = \frac{1}{N_t} \sum_{i \leq N_t} D_t^i \tag{7}$$

$$X_t^A = \frac{1}{\bar{N} - N_t} \sum_{i > N_t} D_t^i \tag{8}$$

Here, N_t represents the number of entrepreneurs adopting labor, and hence $\bar{N} - N_t$ is the number of entrepreneurs adopting AI. Thus, the overall information set available for AI at time t is:

$$\Omega_{t-1} = \{S_{t-j}, X_{t-j}^l, X_{t-j}^A\}_{j=1,2,3,\ldots,-\infty} \tag{9}$$

2.4 Equilibrium

Definition 2.1 *A competitive equilibrium, given sequences of exogenous shocks and parameters, consists of a sequence of individual decision rules* $\{\mathbb{1}_{At}^i, a_t^i\}_{t=-\infty,\ldots,+\infty}^{i=1,2,\ldots\bar{N}}$, *a sequence of aggregate labor-adopting entrepreneurs* $\{N_t\}_{t=-\infty,\ldots,+\infty}$, *a sequence of data points* $\{X_t^l, X_t^A\}_{t=-\infty,\ldots,+\infty}$ *and a sequence of beliefs* $\{\Omega_t\}_{t=-\infty,\ldots,+\infty}$ *such that*

1. $\{\mathbb{1}_{At}^i, a_t^i\}_{t=-\infty,\ldots,+\infty}^{i=1,2,\ldots\bar{N}}$ *solves individual entrepreneur's problem given the sequence of beliefs.*
2. $\{N_t\}_{t=-\infty,\ldots,+\infty}$ *is given by* $N_t = \sum_{i=1}^{\bar{N}} \mathbb{1}_{At}^i, \forall t$.
3. $\{X_t^l, X_t^A\}_{t=-\infty,\ldots,+\infty}$ *are given by:*

$$X_t^l = \frac{1}{N_t} \sum_{i \leq N_t} \left(a_t^i + \varepsilon_t^{Di}\right)$$

$$X_t^A = \frac{1}{\bar{N} - N_t} \sum_{i > N_t} \left(a_t^i + \varepsilon_t^{Di}\right)$$

4. *Beliefs evolve according to:*

$$\Omega_t = \{S_{t-j}, X_{t-j}^l, X_{t-j}^A\}_{j=0,1,2,3,\ldots,-\infty}, \forall t$$

3 Characterization

Our focus will be on the limiting case as the number of entrepreneurs in the economy, $\bar{N}$, tends to infinity. In this scenario, the number of labor-adopting agents versus AI-adopting agents becomes deterministic. Therefore, we can analyze the share of labor-adopting entrepreneurs:

$$n_t = \frac{N_t}{\bar{N}} \to F(\bar{\varphi}_t)$$

Here, $\bar{\varphi}_t$ represents the threshold labor cost above which an entrepreneur would choose to adopt AI. The share of labor-adopting entrepreneurs converges to the function $F(\bar{\varphi}_t)$.

$$n_t = \frac{N_t}{\bar{N}} \to F(\bar{\varphi}_t)$$

3.1 A Recursive Structure in Data Quality

To characterize the information equilibrium, we begin by identifying a recursive structure in the evolution of aggregate data accumulation. Following Farboodi and Veldkamp (2022), we measure the quality of data by the conditional precision of fundamental given the dataset Ω_{t-1}:

$$\gamma_t = \frac{1}{Var\left(\theta_t | \Omega_{t-1}\right)}$$

which is the precision of belief given all past information Ω_{t-1}. It is worth noting that due to the linear quadratic form of the agents' objective functions, the only state variable that needs to be tracked is the precision, rather than the mean.

We now derive a law of motion for γ_t. To do so, note that the information sets defined in equation 9 admit the following recursion:

$$\Omega_t = \Omega_{t-1} \cup \{S_t, X_t^l, X_t^A\}. \tag{10}$$

Thus to derive the law of motion for conditional beliefs we just need to see how beliefs evolve through S_t, X_t^l, and X_t^A. We turn now to this task.

3.2 Optimal Actions Human v.s. AI

We now derive the optimal action rules for human labor and for AI. Taking the first order condition with respect to equation 4 and 5, and exploit the quadratic structure of the objective function (equation 1), we obtain the following lemma for the optimal decision rules:

Lemma 3.1 *The optimal action rule for human labor depends on the public signal and the private signal generated through labor:*

$$a_t^i = E(\theta_t|S_t, s_t^{Li}) = \frac{\gamma_S}{\gamma_S + \gamma_L} S_t + \frac{\gamma_L}{\gamma_S + \gamma_L} s_t^{Li}. \tag{11}$$

The optimal decision rule for AI depends on the public signal, past history of data, and private signal generated by AI:

$$a_t^i = E(\theta_t|\Omega_{t-1}, S_t, s_t^{Ai}) = \frac{\gamma_S}{\gamma_S + \gamma_A + \gamma_t} S_t + \frac{\gamma_t}{\gamma_S + \gamma_A + \gamma_t} \mu_t + \frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t} s_t^{Ai}. \tag{12}$$

where γ_t is given by equation 3.1 and $\mu_t = E(\theta_t|\Omega_{t-1})$ is the conditional mean of θ_t given past information.

The optimal actions are conditional means given available data. For labor, it is a weighted average of public data S_t and labor-generated data s_t^{Li}. The higher the quality of labor-generated data, the more sensitive the action is with respect to it. The same holds for AI as well: the higher the quality of historical data (measured by γ_t), the more weight AI puts on it and the less weight AI would put on other sources of information, including its own privately-generated data S_t^{Ai}.

Also note that the action's loading on privately-generated data is generally lower for AI:

$$\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t} < \frac{\gamma_L}{\gamma_S + \gamma_L} \tag{13}$$

if the values of γ_A and γ_L are similar, due to the presence of historical datasets (γ_t). The fact that AI can have access to historical datasets serves as a double-edged sword: on the one hand, it can significantly enhance AI's ability to take the "better" action, but on the other hand, it means that AI's action is more of a reflection of what has been done in the past (and hence recorded in the dataset) and less of what AI truly adds in terms of new information.[6]

Corollary 3.1 *AI's action is less informative (sensitive) of its privately generated data compared to human, given that the quality of its own signal is not too much greater than those generated by human.*

3.3 Data Generation

We now turn to data generation, holding fixed the relative share of AI and labor adoption.

The labor signal X_t^L: given the optimal action rule derived in lemma 3.1, the human action generates a data point:

$$D_t^i = a_t^i + \varepsilon_t^{Di} = \frac{\gamma_S}{\gamma_S + \gamma_l} S_t + \frac{\gamma_l}{\gamma_S + \gamma_l} s_t^{Li} + \varepsilon_t^{Di}$$

[6]Smart readers would have noticed that equation 13 doesn't holds when AI's private signal is of very high quality $\gamma_A \to \infty$. In this case, the weight on private signal would approach 100%, meaning that the AI would no longer rely on public datasets to be trained. One could think of this extreme case as in the fied of gaming, especially when Alphago came out, which are trained solely through playing against itself. The model predict that in this case, the AI-generated data would be of much higher quality than human-generated data.

Since S_t is publically known, observing D_t^i is equivalent to observing the idiosyncratic part $\frac{\gamma_I}{\gamma_S+\gamma_I}s_t^{Li} + \varepsilon_t^{Di}$, which is equivalent to the following signal of the fundamental θ_t:[7]

$$\theta_t + \varepsilon_t^{Ii} + \frac{\gamma_S+\gamma_L}{\gamma_L}\varepsilon_t^{Di}.$$

Summing up all the data points for human-generated data (equation 7):

$$X_t^I = \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\theta_t + \varepsilon_t^{Li} + \frac{\gamma_S+\gamma_L}{\gamma_L}\varepsilon_t^{Di}\right)$$

which follows a normal distribution and has a mean equal to the fundamental θ_t, as all other noise terms have means equal to zero.

Taking the number of entrepreneurs $\bar{N}$ to ∞, we obtain the following result:

Lemma 3.2 *The information contents in human production activities can be summarized in X_t^I which is an unbiased signal of the fundamental θ_t with precision $n_t\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D$.*

The quality (precision) of the signal is influenced by the share n_t because when more agents adopt labor, it leads to more information content being generated by labor. The term $\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2$ captures the fact that agents' actions give certain weight to the public signal and privately-generated signal, and, because the public signal is already observed by all, only the part of the variation in actions that comes from the private signal (measured by $\frac{\gamma_L}{\gamma_S+\gamma_L}$, see equation 11) contribute additional information to the overall datasets.

The AI signal X_t^A: given the optimal action rule derived in lemma 3.1, the AI action generates the following data point:

$$D_t^i = a_t^i + \varepsilon_t^{Di} = \frac{\gamma_S}{\gamma_S+\gamma_A+\gamma_t}S_t + \frac{\gamma_t}{\gamma_S+\gamma_A+\gamma_t}\mu_t + \frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}s_t^{Ai} + \varepsilon_t^{Di},$$

which is a weighted average of the public signal S_t, the information in the big dataset (summarized by the conditional mean μ_t), and the AI-generated data s_t^{Ai}. Given that the public signal and datasets are observable by all, only the part of the variations in actions that comes from the AI-generated private signal contributes and enhances the overall quality of the dataset. Hence we have:

Lemma 3.3 *The information contents in AI production activities can be summarized in X_t^A which is a signal of θ_t with precision $(1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\gamma_D$.*

The precision of the signal depends on the AI adoption share $1-n_t$ because more adoption generates more AI-related data. The middle term $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2$ captures the fact that agents' ac-

[7] To obtain this, multiply the idiosyncratic part $\frac{\gamma_I}{\gamma_S+\gamma_I}s_t^{Ii} + \varepsilon_t^{Di}$ by $\frac{\gamma_S+\gamma_I}{\gamma_I}$, and we obtain $s_t^{Ii} + \frac{\gamma_S+\gamma_I}{\gamma_I}\varepsilon_t^{Di}$. The signal expression can be obtain by noticing that $s_t^{Ii} = \theta_t + \varepsilon_t^{Ii}$ (equation 3).

tions place certain weight on the public signal and prior history, and that part of the variation in actions does not contribute additional information beyond what is already incorporated into aggregate information. Hence, the valuable source of variation comes from the private signal part, and the AI optimally assigns a weight of $\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}$ to it.

Comparing Lemma 3.2 and Lemma 3.3, one can see that the relative quality of human and AI generated information depend on two aspects. First, it depends on the relative share of labor n_t and AI adoption $1-n_t$. Second, it depends on the weight each action assigns to the privately-generated data. For labor this weight is $\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2$ and for AI it is $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2$. One can easily see that, given a positive γ_t, the AI's weight is generally lower than the labor's weight if the quality of the privately-generated data is similar. This reflects that AI can rely on big datasets on past actions to make predictions, this dilutes the impact of the private signal on AI's actions.

Corollary 3.2 *The AI-generated data carries less information compared to human-generated data, given that the quality of AI's privately generated data is not too much greater than those generated by human.*

3.4 The Evolution of Information and the Data Corruption Channel

We now derive the law of motion of the stock of knowledge γ_t. Utilizing the recursion (equation 10), we have:

$$Var\left(\theta_t|\Omega_t\right) \quad = \quad Var\left(\theta_t|\Omega_{t-1}, S_t, X_t^l, X_t^A\right)$$

From lemma 3.2 and 3.3, this conditional variance is given by:

$$Var\left(\theta_t|\Omega_t\right) = \frac{1}{\underbrace{\gamma_t}_{\text{prec. of } \Omega_{t-1}} + \underbrace{\gamma_s}_{\text{prec. of } S_t} + \underbrace{n_t\left(\frac{\gamma_l}{\gamma_S+\gamma_l}\right)^2\gamma_D}_{\text{prec. of } X_t^l} + \underbrace{(1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\gamma_D}_{\text{prec. of } X_t^A}}$$

where the four components in the denominator captures the information content in Ω_{t-1}, S_t, X_t^l and X_t^A respectively. Given this expression, we obtain the law of motion for γ_t:

Lemma 3.4 *Given γ_t and n_t, γ_{t+1} is given by:*

$$\begin{aligned}\frac{1}{\gamma_{t+1}} &= Var\left(\theta_{t+1}|\Omega_t\right) \\ &= \rho^2\frac{1}{\gamma_t+\gamma_s+n_t\left(\frac{\gamma_l}{\gamma_S+\gamma_l}\right)^2\gamma_D+(1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\gamma_D}+\frac{1}{\gamma_\eta}\end{aligned} \tag{14}$$

Note that this is a partial equilibrium result in the sense that the share of labor-adopting entrepreneurs n_t is holding fixed. It is nonetheless useful for us to understand how n_t could affect

knowledge accumulation. We now analyze what happens to aggregate information quality γ_{t+1} when we change the relative composition of AI and human (i.e. varying the value of n_t):

Theorem 3.1 *Holding fixed current data quality γ_t, an increase in AI adoption (lower n_t) leads to lower data quality in the future (lower γ_{t+1}) if and only if the relatively quality of information generated by AI is not too much higher than those generated by labor:*

$$\frac{\gamma_A}{\gamma_L} < \frac{\gamma_S + \gamma_t}{\gamma_S}. \tag{15}$$

proof: It can be observed from equation 14 that perturbations of n_t affect γ_{t+1} through the relative magnitude of $\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D$ and $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2 \gamma_D$. If the former is greater than the latter, a decrease in n_t leads to a decrease in γ_{t+1}. Simplifying this condition yields equation 15.

This theorem is the key result of the paper: with greater AI adoption, the aggregate quality of data can be hurt, because more AI-generated data corrupts the quality of the dataset. This data corruption channel hinges on the key condition that the relative quality of AI versus human generated data must be sufficiently low. We will quantify this aspect of the model in the section 4.1.

A perhaps unexpected implication can be drawn from the theorem: if the precision of signals produced by both AI and humans is identical, i.e., $\gamma_L = \gamma_A$, then increased AI usage could detriment future knowledge generation, as condition 15 is satisfied. Why could this occur, considering they both generate equal-equality data?

This result can be understood as follows. In a context of social learning, the information obtained by agents is not directly perceived but rather inferred from their actions. These actions are noisy indicators of their private signals, as they also respond to additional data sources. For humans, their actions are influenced by the public signal S_t. For the AI, its actions are predominantly affected by historical data Ω_{t-1}, apart from the public signal S_t. Therefore, given that it learns from more information sources, AI efficiently assigns less weight to its own private signal than humans do. As a result, less private data is incorporated into the collective knowledge pool in comparison to human activities.

Another interesting implication of this theorem is that the lower the public uncertainty (higher γ_S), the lower the value on the right hand side of equation 15, hence the more likely that AI can increase the quality of aggregate data.[8]

[8]The game of Go serves as a good example. The reward system in Go is relatively deterministic, with no uncertainty. According to the prediction of this theorem, employing AI would significantly enhance knowledge creation,

From condition 15, one can easily see that the higher the quality of data, (i.e., higher γ_t), the larger the value on the right-hand side of equation 15, suggesting that the adoption of AI is more likely to hinder data quality improvement. This is based on the same reasoning that higher γ_t values prompt AI to depend more on past data, assigning less weight to its own signal generation, which leads to less information being integrated into the collective knowledge. Note also that in situations where prior knowledge is extremely precise, AI implementation yields the most significant advantages. In such cases, we can expect a surge in AI adoption which could suppress long-term knowledge growth.

The remaining task for characterizing the equilibrium is to endogenize n_t. We now close the model by determining the equilibrium n_t via the individual optimality condition.

3.5 Closing the Model: Endogenizing n_t

From the quadratic objective function, we know that optimal action is simply the conditional expectation of the fundamental parameter θ_t. $a_t^i = E(\theta_t|\text{information set})$, where agents use all the available information to forecast the fundamental, and take actions accordingly. Specifically, we now derive the ex-post payoff of the entrepreneur who adopts labor to produce (equation 4):

$$W^l\left(S_t, s_t^{li}\right) = \bar{A} - \max_{a_t^i} E\left(\left(\theta_t - a_t^i\right)^2 |S_t, s_t^{li}\right)$$

Take first order condition with respect to a_t^i, we obtain:

$$a_t^i = E\left(\theta_t|S_t, s_t^{l,i}\right)$$

Plug it back into $W^l\left(S_t, s_t^{li}\right)$ which becomes:

$$\begin{aligned} W^l\left(S_t, s_t^{li}\right) &= \bar{A} - Var\left(\theta_t|S_t, s_t^{li}\right) \\ &= \bar{A} - \frac{1}{\gamma_S + \gamma_l} \end{aligned}$$

where the last equality follows because the precision of the public signal and private signal are given by γ_S and γ_L respectively, and that these two signals are uncorrelated. Given that $W^l\left(S_t, s_t^{li}\right)$ does not depend on the specific realization of S_t, but only the second moments, we can derive the ex-ante payoff for labor adoption (equation 2.2):

$$\begin{aligned} V^l\left(\varphi_t^i\right) &= E\left(W^l\left(S_t, s_t^{li}\right)|S_t\right) - \varphi_t^i \\ &= \bar{A} - \frac{1}{\gamma_S + \gamma_l} - \varphi_t^i \end{aligned} \tag{16}$$

which aligns perfectly with the observation of the AlphaGo system defeating world champion Ke Jie and continuing to improve its gameplay, becoming a subject of study for Go players globally.

Similarly we can derive the payoff of adopting AI technology

$$\begin{aligned} V^A(\gamma_t) &= \bar{A} - Var\left(\theta_t | S_t, \Omega_{t-1}, s_t^{Ai}\right) \\ &= \bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - R_t \end{aligned} \tag{17}$$

Hence the threshold that determines the share of entrepreneur using each technology is pinned down by equalizing the payoffs in those two cases:

$$V^l(\bar{\varphi}_t) = V^A(\gamma_t)$$

Plugging in the expressions for the two payoff functions:

$$\bar{A} - \frac{1}{\gamma_S + \gamma_l} - \bar{\varphi}_t = \bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - R_t$$

This gives the following threshold as a function of current stock of knowledge:

$$\bar{\varphi}_t(\gamma_t) = R_t + \frac{1}{\gamma_S + \gamma_t + \gamma_A} - \frac{1}{\gamma_S + \gamma_l}$$

This equation, together with the equation characterizing the dynamic evolution of information, pins down the information equilibrium in this model:

Theorem 3.2 *The dynamics of the model is fully characterized by the following system:*

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t(\gamma_t)\left(\frac{\gamma_l}{\gamma_S+\gamma_l}\right)^2 \gamma_D + (1 - n_t(\gamma_t))\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{18}$$

where the share of labor-adopting entrepreneurs is given by:

$$n_t(\gamma_t) = F(\bar{\varphi}_t(\gamma_t)) = F\left(R_t + \frac{1}{\gamma_S + \gamma_t + \gamma_A} - \frac{1}{\gamma_S + \gamma_l}\right),$$

where $F(.)$ is the cummulative distribution function for the labor cost and R_t is the cost of AI in period t.

Note that the knowledge stock parameter γ_t appears in three different locations within the law of motion equation 18. Firstly, it embodies information passed down from previous periods. Secondly, it influences the equilibrium level of AI adoption represented by $1 - n_t$. Lastly, it also impacts the information generated through AI-associated activities because it alters the relative weight AI places on its privately-produced signal.

Figure 2 plots the equilibrium functions, including the law of motion for γ_t and the AI adoption share as a function of γ_t.[9] The left panel displays the law of motion for γ_t, which exhibits a non-linear trend, starting with an increase and then a decrease. The initial increase is due to more information from the past implying more information for the future, which is the standard

[9]Parameters used: $\rho = 0.95, \gamma_S = 0.05, \gamma_D = 0.1, \gamma_L = 0.1, \gamma_A = 0.1, \gamma_\eta = 0.5, R_t = 3.68, \forall t$. The function F follows a log normal distribution with a mean of -2 and a standard deviation of 2.

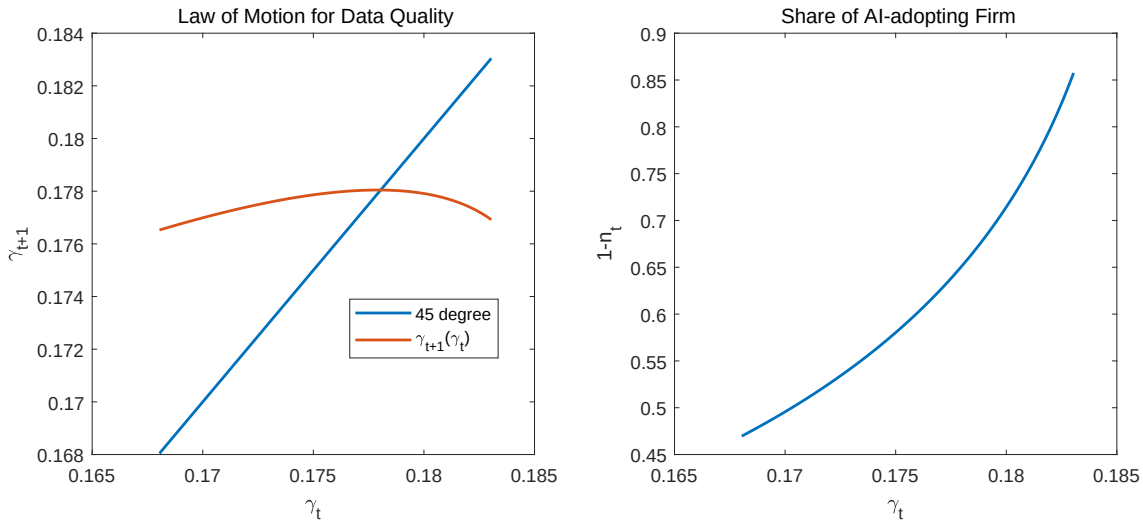


Figure 2: Law of Motion and Policy Functions

information transmission effect. The decreasing part indicates that an abundance of information encourages greater AI adoption (as illustrated in the right panel). Increased AI adoption can, in turn, reduce the quality of aggregate data.[10]

When the precision of the AI-generated signal γ_A approaches infinity, the model simplifies to one where the relative productivity of AI and labor remains constant, similar to Acemoglu and Restrepo (2018). Under this circumstance, AI can predict θ_t flawlessly using only the private signal, making AI's productivity consistently equal to the upper limit $\bar{A}$. For labor-adopting entrepreneurs, their expected productivity equals $\bar{A} - \frac{1}{\gamma_S+\gamma_I}$. Hence, the relative (expected) productivity of labor versus AI, using quality of the product produced as a measure of producitivty, is given by

$$\frac{\bar{A} - \frac{1}{\gamma_S+\gamma_I}}{\bar{A}}$$

which is a constant and independent of γ_t.[11]

More generally, this relative productivity is given by

$$\frac{\bar{A} - \frac{1}{\gamma_S+\gamma_I}}{\bar{A} - \frac{1}{\gamma_S+\gamma_A+\gamma_t}}$$

from which it is evident that a higher data quality (higher γ_t) results in AI being relatively more

[10] It is true in this numerical example where γ_L=γ_A. This satisfies the condition in theorem 3.1 and hence more AI adoption hurts data quality.

[11] In Acemoglu and Restrepo (2018, page 1495, assumption 1), relative productivity is assumed to be fixed within a task.

productive than labor, leading to more AI adoption.

An important implication is that even though AI can be *more* productive, it may contribute *less* to aggregate data accumulation than humans. This is because AI can draw upon rich historical data, potentially increasing its productivity. However, learning from the past and taking corresponding actions do not necessarily contribute additional information to the collective knowledge pool. In fact, the more accurate the historical data, the less AI relies on its privately-generated signal, which can make its actions even less informative compared to those of humans.

4 Quantification of the Data Corruption Channel

We now quantify the model. The key aspect of the calibration is to discipline the relative quality of information generated by AI v.s. human: $\frac{\gamma_A}{\gamma_L}$, as shown in equation 15. We will now discuss the evidence to inform this aspect of the model.

4.1 FID Score and Full Synthetic Loop

We consult the computer science literature for how to measure the quality of information produced by AI. A crucial metric in this assessment is the Fréchet Inception Distance (FID), a popular measure that gauges the similarity between outputs from a generative model and the data used to train it. Mathematically the FID score calculates the distance between two probability distributions. In practise, it effectively compares the empirical distribution of real data (e.g. images) with the synthetic distribution produced by AI. A lower FID score indicates a closer resemblance between these two distributions, and higher quality of AI.

We first calculate the FID score between the true value of fundamental θ_t and its conditional distribution given the information available to AI, Ω_{t-1}. Given that both distributions are normally distributed, we apply a specific formula for the FID score adapted to (single dimensional) normal distributions (Dowson and Landau, 1982), where the distance between two normal distributions with means μ_x, μ_y and standard deviations σ_x, σ_y is calculated as:

$$(\mu_x - \mu_y)^2 + (\sigma_x - \sigma_y)^2$$

We determine the average FID score in the model by averaging across a large panel of simulations. The reason for running the simulation is that the mean of the probability distributions could change due to exogenous shocks. By averaging across a large panel of simulations we can wash out the noises and obtain a robust measure of AI performance.

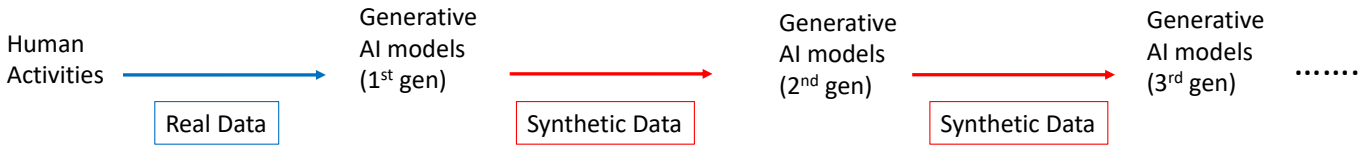


Figure 3: Full Synthetic Loops

As discussed in the introduction, the emerging computer science literature on data corruption consistently finds that increasing reliance on synthetic data generated by AI can degrade data quality. In this literature the general methodology involves training generative AI software with real-world data and then using the synthetic data it produces to train subsequent AI generations, creating "full synthetic loops." This process, illustrated in figure 3, repeats over several iterations, with each AI generation being trained exclusively on synthetic data produced by the preceding generation (Alemohammad et al., 2023).

The key finding from those studies is that the FID score tends to increase with each iteration of full synthetic loops, suggesting a decline in AI-generated data quality. For instance, Alemohammad et al. (2023, see their figure 8) observed over 400% increase in FID scores during full synthetic loops. Martinez et al. (2023, see their figure 3(c)) notes a rise of over 700%, and Bohacek and Farid (2023, figure 2) find an approximately 300% increase throughout full synthetic loops in a most recent study.

I argue that this pattern of increasing FID scores across synthetic loops can help identify a key parameter in our information model: the relative quality of information produced by AI compared to humans. Figure 4 demonstrates this by showing the relationship between the quality of AI-generated information and the FID score's trajectory, with a steeper increase in FID score implying faster detorioration of data quality. One can easily see that the lower the quality of information provided by AI, the more the FID score increases during full synthetic loops. Thus, faster growth of the FID score indicates lower quality of information provided by AI. In the calibration I thus match the most conservative estimate of the FID score increase provided by Bohacek and Farid (2023), which is 300%. This provides an upper bound for the relative information quality and

even in this case the calibration indicates that the relative quality of information provided by AI is only 8% of that by human.

Our calibration results imply that AI contributes minimally to the generation of new information or knowledge beyond what is already encapsulated within the aggregate dataset. This finding aligns with the perspective that AI fundamentally operates as a machine learning tool, which extrapolates predictions and insights based on pre-existing data and historical information. Consequently, while AI may surpass human capabilities in terms of productivity, this advantage does not translate into significant enhancements in the overall quality of the aggregated data. This suggests that the value added by AI in the context of generating novel insights or improving the informational quality of the dataset is limited. It underscores the notion that AI's primary function remains within optimizing and analyzing existing information rather than contributing original knowledge or fundamentally novel insights to the dataset.

4.2 Calibration of Other Parameters

I now describe how the remaining parameters are calibrated, beginning with the calibration of the fundamental process as defined in equation 2. The calibration is based on the autocorrelation and standard deviation of the logarithm of output, resulting in standard parameter values: $\rho = 0.99$ and $\gamma_\eta = 1/(0.055)^2$.[12] We set the standard deviation of investment cost to 0.0155, following Fajgelbaum et al. (2017). The mean of the cost distribution is not separately identified from the AI adoption cost R, and is therefore normalized to 0.

Next, we address the calibration of information parameters: $\gamma_A, \gamma_L, \gamma_D$, and γ_S. According to equation 15, it's important to note that the absolute values of γ_A and γ_L are not as critical as their ratio, which determines the quantitative strength of the data corruption channel. Hence, we normalize the precision of the labor signal to $\gamma_L = 100$, making the precision of the AI signal γ_A a direct indicator of relative information quality. As discussed in section 4.1 and based on the experimental findings by Bohacek and Farid (2023) showing that the FID score increases by 300% over a full synthetic loop, we set $\gamma_A = 8$.

For γ_D, which measures the information loss during the data collection process (indicating that data in our model merely serves as a noisy mirror of past economic activities as seen in equation 6), we calibrate it to match the initial data quality observed in the full synthetic loop 3. The initial

[12] This implies that the standard deviation of the innovation η_t is 0.055, consistent with those used in existing literature.

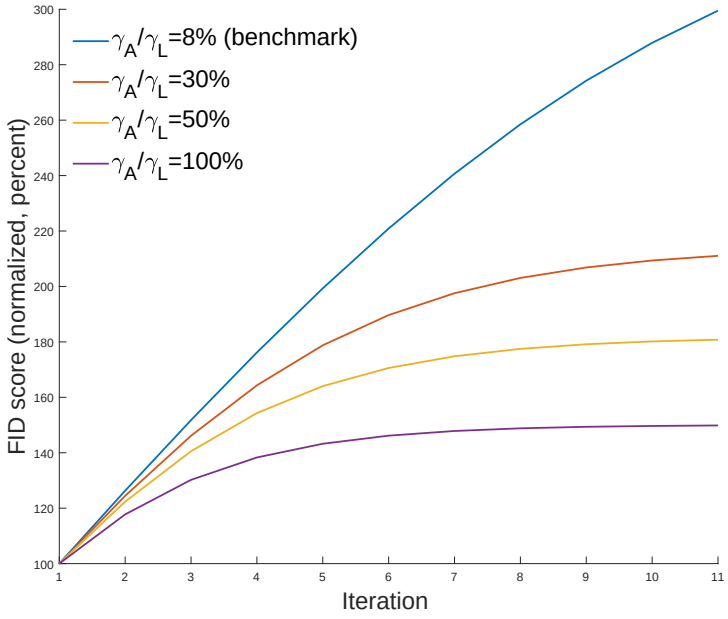


Figure 4: FID score in the model for full synthetic loops

quality of the dataset, trained by humans and assessed using the FID score, is established at 90 by Bohacek and Farid (2023). Aiming to match this values leads us to set γ_D=35.

Lastly, we assign $\gamma_S = 0$ due to the absence of additional targets for measuring the precision of public signals. However, in section 4.5, we conduct a sensitivity analysis to evaluate the impact of a higher γ_S value on our findings, to better understand how variations in this parameter could potentially influence the outcomes of our model. We find that our results doe not change with a changing precision of public signal, as long as the model is recalibrated to match the evolution of FID score during full synthetic loops as shown in section 4.1.

4.3 Quantitative Results: the ChatGPT Experiment

We will now describe the experiment designed to evaluate the long-run effects of AI adoption, termed the "ChatGPT" experiment. This experiment begins in a hypothetical world at time $t = 1$, where AI technology is either nonexistent or the costs associated with its adoption are prohibitively high (e.g. $R_0 \to \infty$). We then examine the model dynamics following a negative shock in period $t = 2$, which reduces the cost of AI to R_1. The cost stays constant after this period. This scenario could be interpreted as a significant technological breakthrough, such as the introduction of ChatGPT, which substantially lowers the barriers to AI accessibility and affordability.

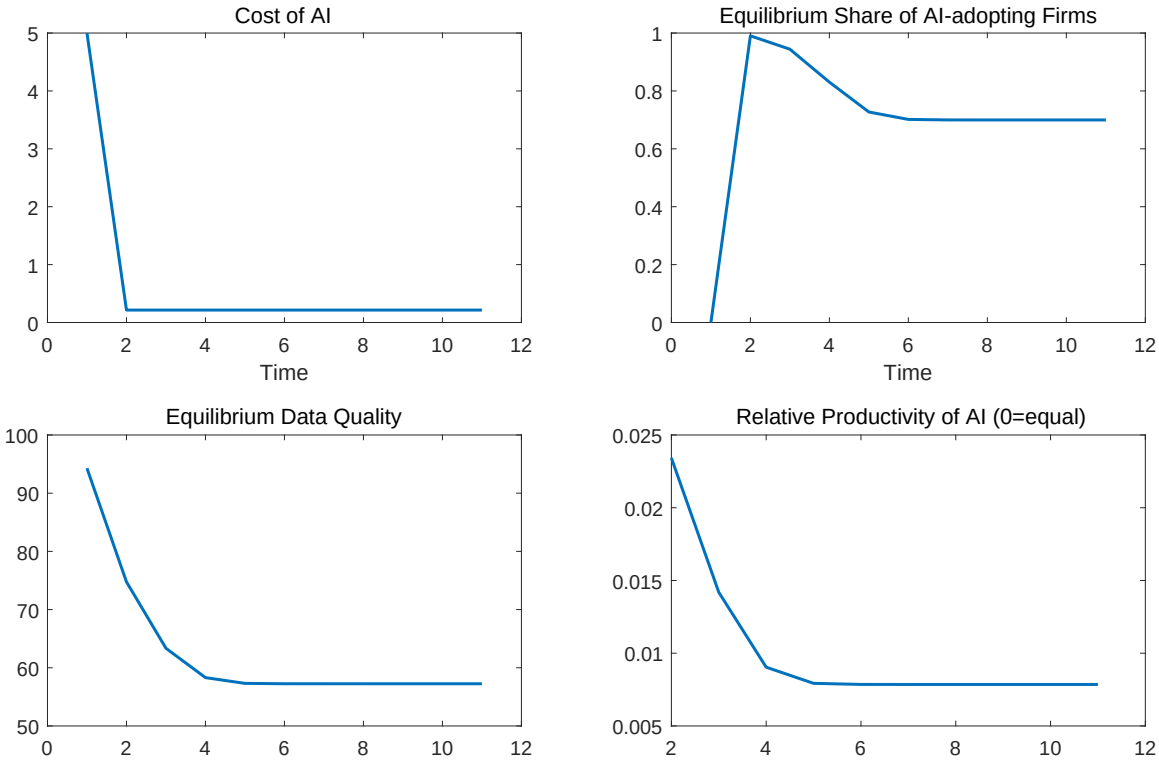


Figure 5: Model Dynamics – the ChatGPT Experiment

Figure 5 displays the model dynamics. The top left panel plots the one-time permanent reduction in the cost of AI. Following this exogenous cost reduction, we analyze the model dynamics according to equation 18. The top right panel shows a rapid increase in the proportion of firms adopting AI, nearly reaching 100% before settling back to approximately 75%. This initial surge and subsequent decline highlight the impact of the data corruption channel: the reduced costs encourage widespread AI adoption, which in turn hurts the aggregate data quality, mitigating the advantages of adopting AI technology. The bottom left panel tracks the decline in aggregate data quality over time (γ_t), a direct consequence of increased reliance on AI, which, as illustrated in the bottom right panel, leads to a reduction in AI's relative productivity and, consequently, its attractiveness.[13] This experiment suggests that the labor displacement effects observed with the introduction of AI may partially reverse over time. This reversion effect could be significant, as demonstrated in this numerical example.

Figure 6 compares equilibrium functions under varying AI costs. The information law of motion when AI costs are high (represented by the yellow line) is located above that when the costs are low (indicated by the red line). Hence, if we begin with a steady state characterized by high costs of AI, following exogenous shocks that reduces the AI costs, the equilibrium information quality would gradually diminish over time and converge to the low steady state associated with lower costs of AI.

4.4 Equal Information Quality: A Counterfactual

To understand the importance of the data corruption channel,we conduct a counterfactual experiment where AI and human activities generate information of identical quality, setting $\gamma_A = \gamma_L$ to 100 for both, while other parameters remain as they were in the baseline calibration. Results are displayed in figure 7, comparing the baseline (blue solid line) against the scenario of equal information quality (red dashed line).

The top left panel shows that, in this counterfactual, the cost reduction needed to achieve 100% AI adoption in the short term is almost the same as in the baseline. However, the adoption dynam-

[13]The relative productivity of AI is calculated as the percentage differences in expected ex-post payoffs between AI adoption and labor adoption. Specifically it is equal to:

$$\frac{V^A(R_t) - E(V^l(\phi^i))}{E(V^l(\phi^i)}$$

where $V^A(R_t)$ and $V^l(\phi^i)$ are ex-post payoffs and are given by equation 17 and 16. The expectation is taken over idiosyncratic labor costs ϕ^i.

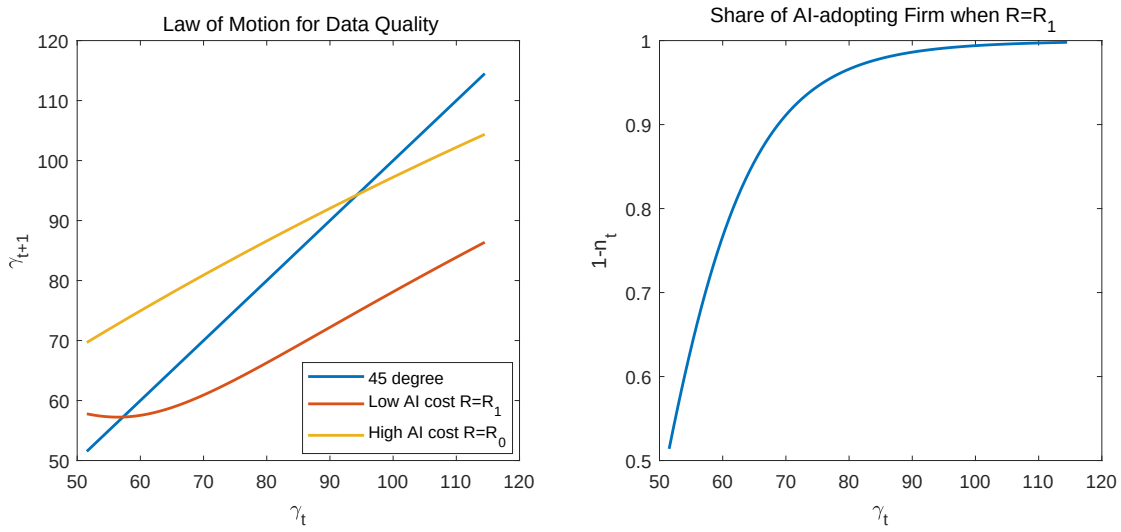


Figure 6: Equilibrium Functions when $\gamma_A = \gamma_L$

ics differ markedly; with equal information quality, there is no significant reversal in AI adoption rates, which remain close to 100% in the long run, in stark contrast to the 30% reversal observed in the baseline. This difference arises because data quality does not deteriorate as significantly when AI generates information of comparable quality to humans. Interestingly, even with equal information quality, we observe a gradual decline in data quality over time, attributed to AI's lesser reliance on private information compared to humans, and hence AI' action reveals less private information and contribute less to aggregate data quality compared to humans. Equal information quality implies a reduced data corruption effect, leading to a slower decline in AI's relative productivity and sustaining its appeal as an efficient choice for firms in the long term.

This counterfactual experiment concludes that the data corruption channel has a significant long-term impact on the labor market by diminishing AI's productivity. In our baseline scenario, this channel is responsible for reversing 30% of AI's short-term labor displacement effect. This reversal is entirely absent when information quality is equalized between humans and AI.

4.5 The Role of Public Signal

In our baseline calibration, we set the precision of the public signal, γ_S, to zero, which naturally raises the question of how the model's dynamics might be affected by lower aggregate uncertainty, indicated by a higher value of γ_S. This section explores changes in the model's behavior with a more precise public signal.

To start, I set the precision of the public signal γ_S to 1. With this change, other parameters need

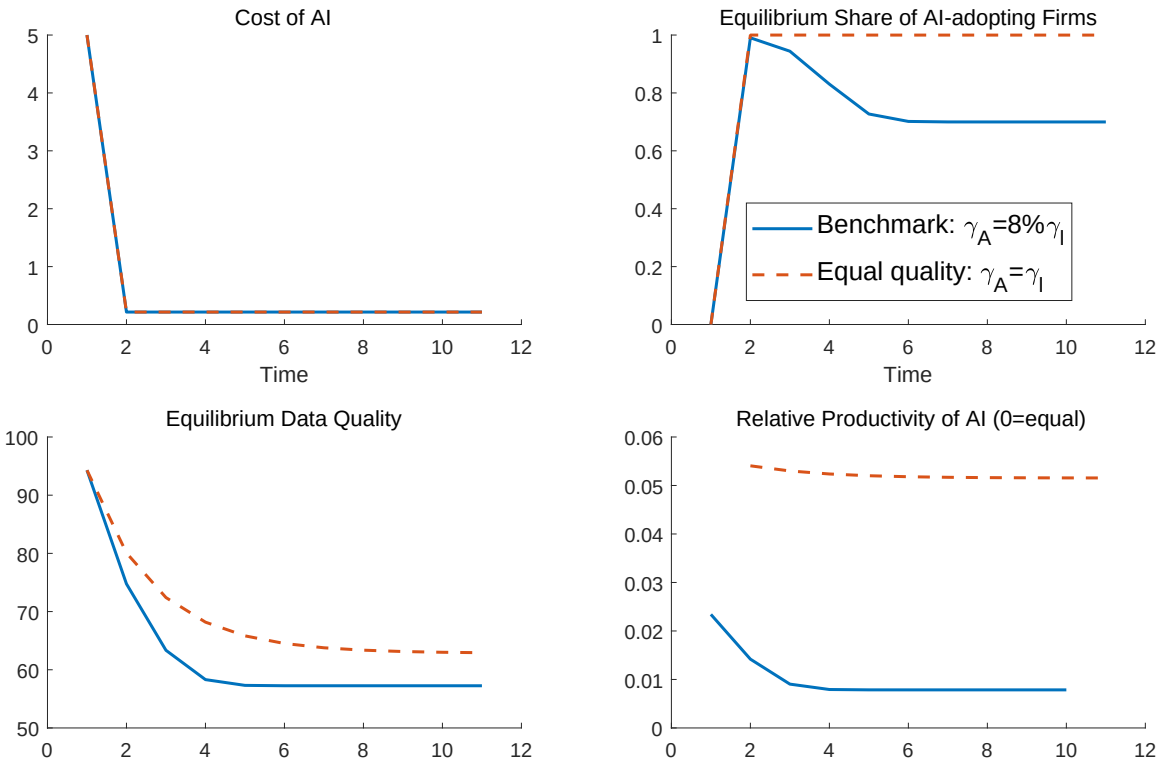


Figure 7: The ChatGPT Experiment with Equal Information Quality

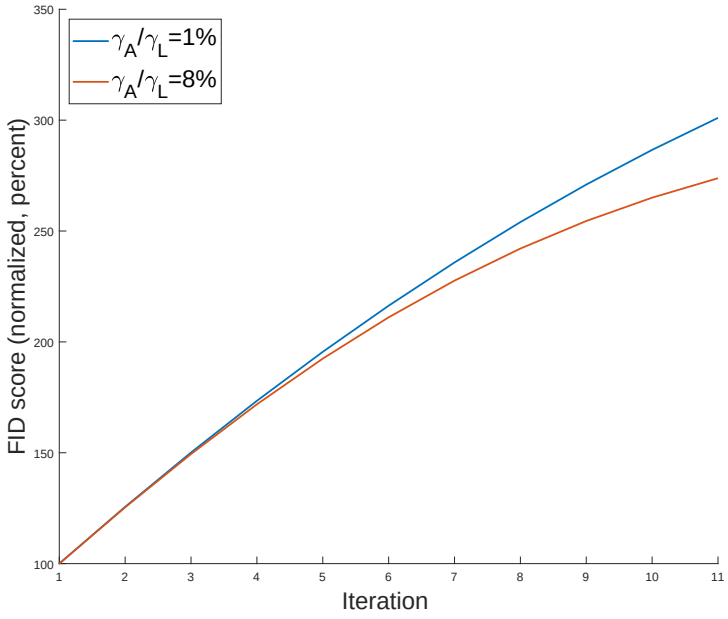


Figure 8: FID score with public signal $\gamma_S = 1$

to be recalibrated, particularly the relative quality of information generated by AI versus human. Figure 8 illustrates why a recalibration is necessary by comparing two circumstances. The red line represents the case where the relative information quality remains at the benchmark level of 8%. However, with a more accurate public signal, this 8% difference is insufficient to achieve a 300% increase in the FID score. The lower uncertainty environment means that the contributions of human and AI to data quality are closer, as both are following the precise public signal, diminishing the data corruption effect and lessening the deterioration in data quality during full synthetic loops.

To replicate the observed 300% increase in the FID score, we find that the gap in private information between AI and humans must be widened. A recalibration indicates that this gap corresponds to AI's relative information quality being at 1% (as shown by the blue line in Figure 8). All other parameters remain the same as the benchmark calibration.

Figure 9 presents the numerical results when we conduct the ChatGPT experiment under this scenario of low uncertainty. The blue line is the benchmark result for comparison purposes. The red dashed line is the case with low uncertainty. Following the shock that reduces AI adoption costs, the immediate spike in the share of AI-adopting firms mirrors the original experiment, almost reaching 100%. Intriguingly, the reversal effect observed is quite similar to that of the

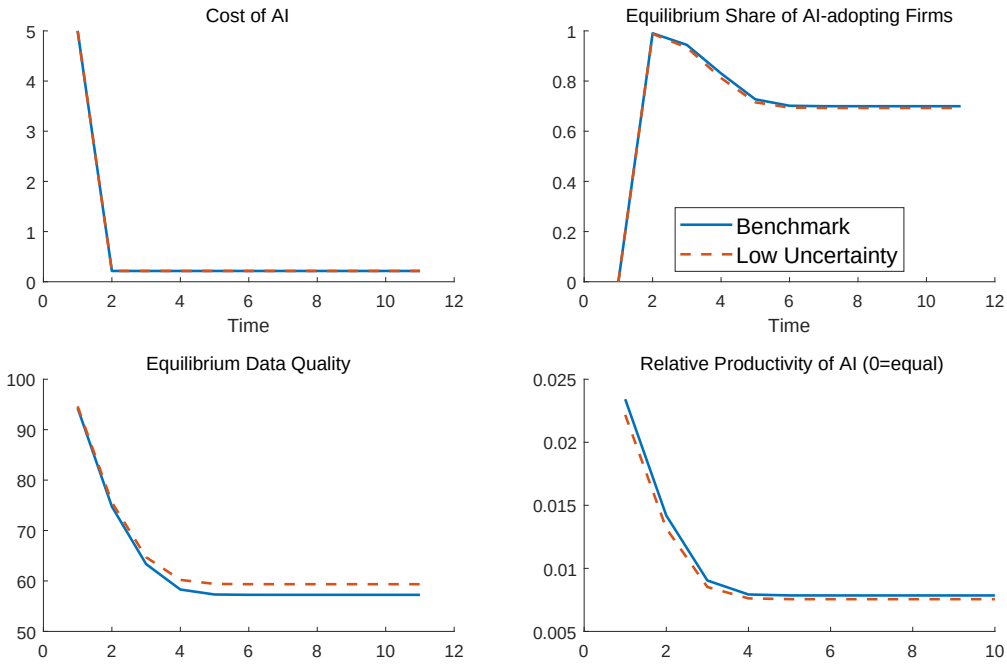


Figure 9: Model Dynamics with Low Uncertainty

benchmark scenario (as depicted in the top right panel of Figure 9), with the equilibrium share of AI-adopting firms stabilizing around 75%. This suggests that even under conditions of low uncertainty, the data corruption channel remains significant enough to counteract 30% of the short-term labor displacement effect.

We conclude that varying the precision of the public signal does not fundamentally change the impact of the data corruption channel. More critically, it underscores that when the model is calibrated to replicate the increase in FID score observed during synthetic loops as in the AI literature, the quantitative magnitude of the data corruption channel is effectively pinned down, rendering it robust to changes in other model parameters.

4.6 Review of Evidence from Computer Science

This section examines evidence in support of the model's key mechanism. The primary proposition of the model, which results in less labor displacement over the long run, is that increased AI adoption tends to produce less useful data points, subsequently reducing productivity for future AI. There is substantial evidence in computer science to support it. A recent paper reveals that the effectiveness of ChatGPT has decreased over time (Chen et al. 2023). The authors evaluate

the performance of the March version of ChatGPT against its June version across a range of tasks such as prime-finding and code generation. They conclude that in most of these tasks, ChatGPT's capability to generate correct responses has lessened.

What causes this decline? A major factor of concern is that since the introduction of ChatGPT, and large language models (LLMs) in general, they have been broadly adopted to perform a variety of tasks including text and image generation, app development, and more, thereby generating vast amounts of content posted on the internet. This AI-generated content is then utilized to train successive generations of AI models, which might have unforeseen consequences (Martínez et al. 2023). This aligns precisely with the feedback mechanism proposed in the paper.

Concurrent with this paper, contemporary scientific research in computer science has expressed similar worries, suggesting that the use of model-generated data to train an LLM could result in potentially problematic outcomes, such as quality degradation, reinforced biases, and loss of novelty. Alemohammad et al. (2023) examines generative image models trained on model-generated data and finds that training these models on synthetic data, as opposed to human-generated data, tends to progressively amplify artifacts (refer to their Figure 1). Consequently, "...models are destined to see a progressive decline in their quality (precision) or diversity (recall)." Shumailov et al. (2024) employs large language models like ChatGPT and shows that the use of model-generated content leads to "irreversible defects in the resulting models." They show that this phenomenon can occur in a variety of AI models, not limited to large language models. Research across various contexts and employing different AI software consistently shows that training AI models with synthetic content tends to degrade the quality of outputs, particularly affecting their diversity, according to studies such as Guo et al. (2023) and Doshi and Hauser (2024). The extent of this decline in data quality is influenced by the mix of real versus synthetic data in the training set, as highlighted by Briesch et al. (2023).

Additionally, there is evidence of a displacement effect where generative AI crowds out the production of human-generated content. Rio-Chanona et al. (2023) utilizes a difference-in-differences approach to compare the United States with Russia and China, where access to ChatGPT is more restricted. Their findings suggest that the launch of ChatGPT is associated with a 16% decrease in weekly contributions to Stack Overflow, indicating a significant crowding-out effect of generative AI on human contributions.

5 Efficiency

The regulation of AI has been a focal point in the concurrent debate surrounding AI. The first law on AI, the EU AI regulation for example, has set restrictions on AI adoptions and data regulations in varioius areas. In this section, we investigate the efficiency of the competitive equilibrium, focusing particularly on the efficiency of the AI adoption decisions and AI's usage of data.

We seek to answer the following questions: 1) Is the private sector's decision to adopt AI socially efficient? If given the choice, would a benevelent social planner do differently in terms of how much shares of firms to adopt AI? 2) If the benevelent planner were to be given the chance to constraint AI's acces to (certain) data, would the planner find it optimal to do so?

5.1 Formulation of Planner's Problem

To explore these questions, we establish a constrained social planner's problem. The planner is constrained in the sense that he is subject to the same information friction as private individuals. Thus, the planner cannot directly interfere with the data accumulation process. It is assumed that the planner can pick the labor-cost threshold $\bar{\varphi}$ which determines the share of AI adoption. It can also decide what data is available to the AI:

$$\Omega_t^A \subseteq \Omega_t, \forall t$$

where Ω_t^A is the subset of historical data available to AI, and is picked by the social planner:

$$W^A(S_t, s_t^{Ai}, \Omega_{t-1}^A) = \max_{a_t^i} E(A_t^i(a_t^i)|S_t, s_t^{Ai}, \Omega_{t-1}^A) \tag{19}$$

The following lemma illustrates the potential societal tradeoff of constraining AI's access to historical datasets:

Lemma 5.1 *Given any $\Omega_{t-1}^A \subseteq \Omega_{t-1}$, the optimal action of AI is given by*

$$a_t^i = E(\theta_t|\Omega_{t-1}^A, S_t, s_t^{Ai}) = \frac{\gamma_S}{\gamma_S + \gamma_A + \kappa\gamma_t} S_t + \frac{\kappa\gamma_t}{\gamma_S + \gamma_A + \kappa\gamma_t}\mu_t + \frac{\gamma_A}{\gamma_S + \gamma_A + \kappa\gamma_t} s_t^{Ai}. \tag{20}$$

And the value of AI production is given by:

$$V^A = \bar{A} - \frac{1}{\gamma_S + \gamma_A + \kappa\gamma_t} - R_t \tag{21}$$

where κ is the ratio of conditional precision given Ω_t and Ω_t^A:

$$\kappa = \frac{1/Var\{\theta_t|\Omega_{t-1}^A\}}{1/Var\{\theta_t|\Omega_{t-1}\}} \leq 1$$

where the last inequality follows because $\Omega_{t-1}^A \subseteq \Omega_{t-1}$.

The direct cost of constraining AI's access to data is that it reduces AI productivity. This re-

duces the value of AI production, as can be seen in equation 21 where the introduction of κ reduces the value of V^A. It is also clear that all that matters is the quality of the dataset available to AI, instead of the specific data points contained in it. Thus, it is sufficient to consider a social planner picking the value of κ.

From individual perspective there would be no benefit from constraining AI's access to data, as the value of V^A is unambiguously increasing in κ. However, there is a social benefit – more stringent restriction of public data access (lower κ) makes AI's action put more weight on its privately-generated data. This can be seen from equation 20 where the weight on the private data s_t^{Ai}, $\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}$, increases with lower value of κ. Constraining AI's acces to public data "forces" it to put more weight on its private one, and this can enhance overall data quality, particularly in an economy populated by AI:

Lemma 5.2 *Given current data quality γ_t and some share of labor adoption n_t, the future data quality γ_{t+1} is given by:*

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n * \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n) * \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}$$

Thus, constraining AI's access to data (lower value of κ) leads to higher quality of data in the future (higher value of γ_{t+1}).

We now formulate a steady-state version of the constrained social planner's problem. The planner can choose the relative share of AI adoption, $\bar{\varphi}$, as well as the quality of data available to AI, κ:

$$\max_{\bar{\varphi},\kappa} \int^{\bar{\varphi}} V^l(\varphi)\, d\varphi + \int_{\bar{\varphi}} V^A d\varphi$$

$$s.t.$$

$$\frac{1}{\gamma} = \rho^2 \frac{1}{\gamma + \gamma_s + n * \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n) * \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}$$

$$n = F(\bar{\varphi})$$

$$V^l(\varphi) = \bar{A} - \frac{1}{\gamma_s + \gamma_L} - \varphi$$

$$V^A = \bar{A} - \frac{1}{\gamma_S + \gamma_A + \kappa\gamma} - R$$

$$\kappa \leq 1$$

The objective of the social planner is to maximize the ex-ante utility of the entrepreneurs (before the φ shock realizes). The planner can choose the threshold $\bar{\varphi}$, which pins down the share of

labor vs. AI adopting entrepreneurs (n), which in turn affects the steady state data quality γ from the first constraint. From Lemma 5.2 one also knows that lower value of κ can also enhance data quality. Thus, those conditions jointly pin down $\gamma(\bar{\varphi}, \kappa)$ as a function of $\bar{\varphi}$ and κ.

Plug in $\gamma(\bar{\varphi}, \kappa)$ and the value function for labor adoption $V^l(\varphi)$ and AI adoption $V^A(\gamma)$, the social planner's problem can be written compactly as:

$$\max_{\bar{\varphi}, \kappa \leq 1} \int^{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_l} - \varphi \right] d\varphi + \int_{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \kappa \gamma(\bar{\varphi}, \kappa)} - R \right] d\varphi \tag{22}$$

The key difference between the social planner's problem and private agents' problem is that the social planner understands that information quality is affected by AI adoption and data access (the expression of $\gamma(.)$ in red) while private agents takes γ as fixed. This presents an information externality that the planner needs to correct: individuals when they take their actions do not take into account the impact of their actions on the overall data quality.

5.2 AI Adoption

We now focus on the AI adoption margin of the planner's problem. The following first order condition pins down the socially efficient threshold $\bar{\varphi}^{sp}$:

$$\underbrace{\frac{1}{\gamma_S + \gamma_A + \kappa \gamma(\bar{\varphi}^{sp})} + R - \frac{1}{\gamma_S + \gamma_l} - \bar{\varphi}^{sp}}_{\text{Private Benefits}} + \underbrace{\frac{(1 - n(\bar{\varphi}^{sp}))}{(\gamma_S + \gamma_A + \kappa \gamma(\bar{\varphi}^{sp}))^2} \gamma'(\bar{\varphi}^{sp})}_{\text{Externality}} = 0$$

The optimality condition consists of two parts: the "private benefits" and the "externality". The "private benefits" term is identical to the private optimality condition. The "externality" term is what the social planner considers but the private agent does not. Note that the externality term depends on how information γ varies with the adoption threshold $\bar{\varphi}$.

The following theorem shows the direction of this derivative and hence the optimal regulation by government:

Theorem 5.1 *1. Increasing AI adoption hurts data quality ($\gamma'(\bar{\varphi}^{sp}) > 0$) if and only if the relatively quality of information generated by AI is not too much higher than those generated by labor:*

$$\frac{\gamma_A}{\gamma_L} < \frac{\gamma_S + \gamma}{\gamma_S} \tag{23}$$

2. Given the presence of data corruption, AI adoption hurts data quality, hence it should be taxed.

Condition 23 is similar to condition 15, where the nature of externality depends on the relative quality of information generated by human and AI. When human activities can create more information content compared to AI, the information externality of AI adoption is negative, in the

sense that more AI adoption reduces the quality of data. The negative externality thus calls for a taxation of AI, even over the long run. Note that, if AI generates more precise data, but γ_L is not much less than γ_A, a tax should still be imposed on AI adoption. The reason is that AI are less efficient at incorporating new information into aggregate knowledge, and a tax on AI reflects this relative inefficiency. Thus, according to our benchmark calibration, a tax on AI adoption should be implemented to achieve a socially efficient level of data quality.

A caveat of the theorem is that it focuses on the steady state, which means it primarily considers optimal government regulation in the long run. Over the shorter term, the case for regulation could be less compelling due to the dynamic nature of information externalities: today's AI adoption decisions impact tomorrow's knowledge stock, thereby influencing the welfare of future generations. Consequently, the extent and nature of regulation would depend on the time frame the government considers, as well as the weight assigned to the welfare of each generation. For instance, if the government's concern is limited to the welfare of the current generation of entrepreneurs, regulation may not be deemed necessary.

This theorem also suggests that the government should regulate different tasks with potentially different levels of fundamental uncertainty and stock of knowledge differently. For example, the task of Go-playing has minimal uncertainty, which weakens the case for taxation. However, tasks that feature a lot of uncertainty, such as forecasting future market demand or developing a new drug, the theorem suggests that greater uncertainty implies more regulation and taxation. Therefore, government policy should be task-based, determined on a case-by-case basis.

5.3 Data Access

We now focus on the data access margin of the planner's problem: the optimal choice of κ. Note from equation 22 the choice of κ affects overall welfare only through the value of AI. From individual AI's perspective, it is always optimal to use 100% of all the data in hand:

$$V^A \Uparrow= \bar{A} - \frac{1}{\gamma_S + \gamma_A + \kappa \Uparrow \gamma} - R.$$

The planner acknowledges that the choice of κ affects aggregate data quality:

$$V^A = \bar{A} - \frac{1}{\gamma_S + \gamma_A + \kappa \Uparrow \gamma(\kappa) \Downarrow} - R$$

$\gamma'(\kappa) < 0$ because constraining public data usage of AI forces it to act more aggressively upon its own privately generated data, raising overall data quality and hence the value of γ (see lemma 5.2).

Thus optimal κ solves:

$$\max_{\kappa} \kappa\gamma(\kappa)$$

and $\kappa < 100\%$ can be potentially welfare improving. It is straightforward take the first order condition:

$$\underbrace{\gamma}_{\text{Cost of Limiting Data Access}} + \underbrace{\kappa\frac{\partial\gamma}{\partial\kappa}}_{\text{Benefit of Limiting Data Access}} = 0$$

where the first term γ is the cost of restricting data access. The higher the quality of the dataset the larger the cost of restricting access to it. The second term is the societal benefit of constraining AI's access to data. This could raise overall data quality and hence improve productivity of future AI.

Theorem 5.2 *With data corruption, the long-run data quality is lower. Given the lower data quality, the cost of constraining AI's data access becomes smaller, and hence this is more likely to be socially desirable.*

With data corruption, the quality of data deteriorates (the value of γ goes down) over the long run (see the bottom left panel of figure 5). Hence the cost of limiting data access of AI becomes smaller. The benefit, on the other hand, can be shown to vary less proportional to changes in the data quality. Thus the net benefit increases when the value of γ goes down, and it is therefore more likely that constraining AI's access to data can be welfare-improving.

6 Extensions and Robustness

This section offers two extensions to the base model to examine its robustness. We first relax the assumption that entrepreneurs, upon inception, cannot observe any prior history. Instead, we posit that they are born with the ability to observe a certain fraction of past information. We also relax the assumption that the signal generated by AI is unrelated to past history, positing instead that the quality of the signal may improve with the quality of historical data observed. In this expanded model, we outline the model equilibrium and discuss its implications.

In the baseline model, it's assumed that entrepreneurs are incapable of observing past data history, while AI can observe the entire history. This assumption is mainly for clarity and to underscore the informational differences between humans and machines, emphasizing that machines (AI) are more efficient at collecting and processing extensive historical big data. In this

section, we relax this assumption and demonstrate that the primary findings of the paper remain unchanged.

We propose that each entrepreneur, upon their inception, is able to observe a fraction of past data, denoted by $\kappa < 1$. This can be interpreted as entrepreneurs being able to observe information within their local communities and networks, which only constitute a fraction of the total amount of information available in the entire economic system. For instance, an entrepreneur from Michigan may be very familiar with business practices, culture, and customer preferences in Michigan, but less familiar with those in Florida or New York. Conversely, AI can collect and process data from all states in the US and even from the rest of the world.

Specifically, assume that the information set an entrepreneur born with (denoted by superscript e) is given by:

$$\Omega_{t-1}^{e} = \{S_{t-j}, D_{t-j}^{i}\}_{j=1,2,3,\ldots,-\infty}^{i=1,2,3,\ldots\kappa\bar{N}}$$

Note that the superscript for the number of noises now spans from 1 to $\kappa\bar{N} < \bar{N}$. As we will take $\bar{N} \to \infty$, there will be no integer problem associated with this notation.

This information set admits a similar recursion as shown in Y:

$$\Omega_{t}^{e} = \Omega_{t-1}^{e} \cup \{S_t, X_t^{le}, X_t^{Ae}\}$$

where

$$X_t^{le} = \frac{1}{\kappa N_t} \sum_{i \leq \kappa N_t} \left(a_t^i + \varepsilon_t^{Di}\right)$$

$$X_t^{Ae} = \frac{1}{\kappa\left(\bar{N} - N_t\right)} \sum_{i > \kappa N_t}^{\kappa\bar{N}} \left(a_t^i + \varepsilon_t^{Di}\right)$$

Denote the summary statistics for human and AI in the evolution of information for entrepreneurs. Denote the prior information content to be $\gamma_t^e = \frac{1}{Var\left(\theta_t | \Omega_{t-1}^e\right)}$.

We also relax the assumption that the precision of the signal generated by AI is independent of the quality of the historical data the AI has been trained on. We refine this by postulating that the precision of the AI-generated signal follows the following functional form:

$$\gamma_A\left(\gamma_t\right) = a + b\gamma_t \tag{24}$$

where a and $b \geq 0$ are some constants. When $b = 0$, we go back to our baseline assumption.

Now, let's derive the model dynamics. We'll begin by determining the optimal actions for entrepreneurs choosing different modes of production.

If the agent chooses to use labor, his optimal action is:

$$a_t^i = E(\theta_t|\Omega_{t-1}^e, S_t, s_t^{li}) = \frac{\gamma_S}{\gamma_S+\gamma_l+\gamma_t^e}S_t + \frac{\gamma_t^e}{\gamma_S+\gamma_l+\gamma_t^e}\mu_t^e + \frac{\gamma_l}{\gamma_S+\gamma_l+\gamma_t^e}s_t^{li}$$

where γ_t is given by equation 3.1 and $\mu_t^e = E\left(\theta_t|\Omega_{t-1}^e\right)$ is the conditional mean of θ_t given past information, and the only informative part in this action is $\frac{\gamma_l}{\gamma_S+\gamma_l+\gamma_t^e}s_t^{Ai}$. Thus, it can be shown that the summary statistics for labor is (taking $\bar{N} \to \infty$):

$$X_t^{le} \sim N\left(\theta_t, \frac{1}{\kappa n_t}\left(\frac{\gamma_S+\gamma_l+\gamma_t^e}{\gamma_l}\right)^2\frac{1}{\gamma_D}\right)$$

We now derive the signal associated with AI activities. The optimal action of AI is given by:

$$a_t^i = E(\theta_t|\Omega_{t-1}, S_t, s_t^{Ai}) = \frac{\gamma_S}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}S_t + \frac{\gamma_t}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}\mu_t + \frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}s_t^{Ai}$$

where $\gamma_A(\gamma_t)$ is given by equation 24. Note that the only information content here is $\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}s_t^{Ai}$.

Hence, the summary statistic for AI activity is (taking $\bar{N} \to \infty$):

$$X_t^{Ae} \to N(\theta_t, \frac{1}{\kappa(1-n_t)}\left(\frac{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}{\gamma_A(\gamma_t)}\right)^2\frac{1}{\gamma_D})$$

With these two summary signals, we can derive the law of motion for γ_t and γ_t^e in a joint manner. Note that we need to keep track of both measures of information as they enter into the law of motion for both variables. Specifically, γ_{t+1} is given by the following function of γ_t and γ_t^e:

$$\frac{1}{\gamma_{t+1}} = \rho^2\frac{1}{\gamma_t+\gamma_s+n_t\left(\frac{\gamma_l}{\gamma_S+\gamma_l+\gamma_t^e}\right)^2\gamma_D+(1-n_t)\left(\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}\right)^2\gamma_D} + \frac{1}{\gamma_\eta} \tag{25}$$

And γ_{t+1}^e is given by the following function of γ_t and γ_t^e:

$$\frac{1}{\gamma_{t+1}^e} = \rho^2\frac{1}{\gamma_t^e+\gamma_s+\kappa n_t\left(\frac{\gamma_l}{\gamma_S+\gamma_l+\gamma_t^e}\right)^2\gamma_D+\kappa(1-n_t)\left(\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}\right)^2\gamma_D} + \frac{1}{\gamma_\eta} \tag{26}$$

From the two law of motion, one can examine how changes in n_t affects knowledge accumulation by examining whether

$$\left(\frac{\gamma_l}{\gamma_S+\gamma_l+\gamma_t^e}\right)^2 - \left(\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t}\right)^2 > 0$$

which boils down to:

$$\frac{\gamma_l}{\gamma_A(\gamma_t)} > \frac{\gamma_S+\gamma_t^e}{\gamma_S+\gamma_t}$$

when this condition is satisfied, reducing n_t (or more AI adoption) reduces knowledge accumulation for both AI and entrepreneurs.

The insights obtained from the baseline model remain. First, given that AI has access to more data points than humans, $\gamma_t > \gamma_t^e$, and therefore,

$$\frac{\gamma_S + \gamma_t^e}{\gamma_S + \gamma_t} < 1$$

Hence, even if human and AI generate signals of equal precision, we would still expect AI adoption to hinder knowledge accumulation. However, now an offsetting force is introduced: γ_A is increasing in γ_t. So when γ_t is very large, γ_A would also be large, which would make the inequality harder to satisfy. This is due to the fact that the signal generated by AI improves in quality in relation to the overall data quality. However, this offsetting force is likely not strong enough to overturn the main result of the paper, as we will see in the full model dynamics.

To close the model, we now derive the endogenous share of labor-adopting entrepreneurs. In this case the value of adopting labor is:

$$V^l = \bar{A} - \frac{1}{\gamma_S + \gamma_l + \gamma_t^e} - \varphi_t^i$$

while the value of adopting AI is

$$V^A = \bar{A} - \frac{1}{\gamma_S + \gamma_A(\gamma_t) + \gamma_t} - R_t$$

Hence the threshold cost of labor is pinned down by

$$\bar{\varphi}(\gamma_t, \gamma_t^e) = R_t + \frac{1}{\gamma_S + \gamma_A(\gamma_t) + \gamma_t} - \frac{1}{\gamma_S + \gamma_l + \gamma_t^e} \tag{27}$$

Thus model dynamics is fully characterized by equation 25 and 26, where the share of labor-adopting entrepreneurs is given by

$$n_t(\gamma_t, \gamma_t^e) = F(\bar{\varphi}(\gamma_t, \gamma_t^e))$$

and $\bar{\varphi}(\gamma_t, \gamma_t^e)$ is given by equation 27.

In contrast to our benchmark model, this system is characterized by two endogenous state variables representing the information possessed by entrepreneurs and AI, respectively. Figure 10 simulates the economy and carries out the ChatGPT experiment similar to section 4. It is observed that the labor displacement effect is also partially reversed in this case (top right panel). In this numerical example, I set $\kappa = 0.01$, indicating that the information possessed by humans is only 1% of that by AI. The difference in the overall level of information between γ_t and γ_t^e is much smaller, roughly equal to 60 percent (0.27 vs. 0.16) because of the existence of a public signal.[14]

Upon a reduction in the cost of AI, the share of AI-adopting firms increases. This displacement effect is partially reversed as in the baseline experiment. The logic for the reversal is basically

[14]Parameters used: $\rho = 0.95, \gamma_S = 0.05, \gamma_D = 2, \gamma_L = 0.1, \gamma_A = 0.1 + 0.1 * \gamma_t, \gamma_\eta = 0.5, \kappa = 0.01, R_t = 0.95, \forall t$. The function F follows a log normal distribution with a mean of -2 and a standard deviation of 2.

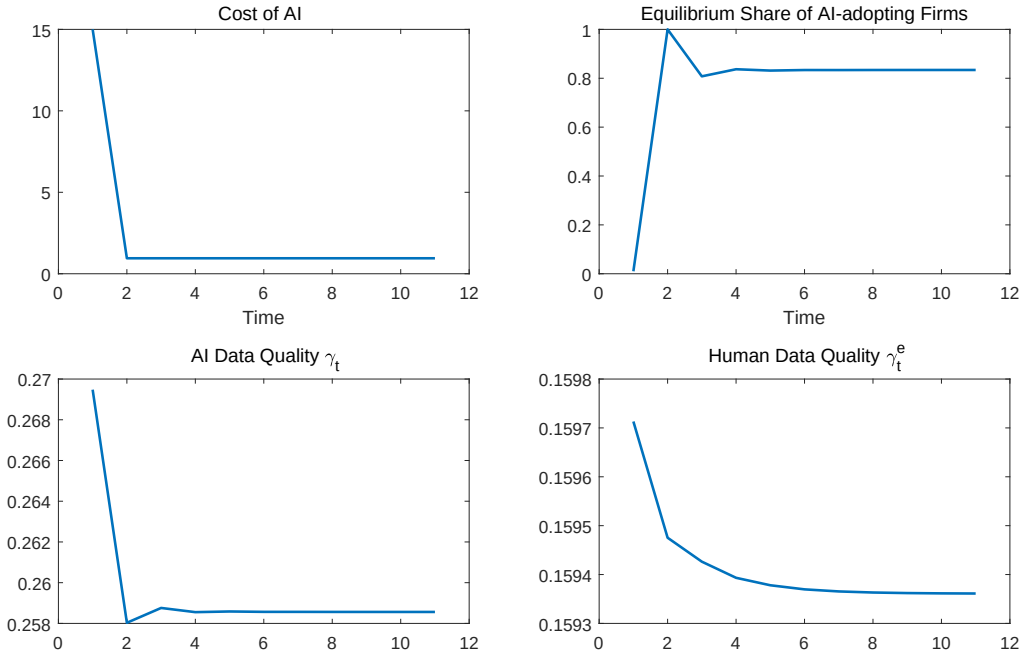


Figure 10: Model Dynamics with Low Uncertainty

the same as in the baseline experiment: the data quality of AI γ_t decreases (bottom left panel), reducing the efficacy and hence the productivity of AI products. However, there is an additional force here: the relative attractiveness of AI also depends on the efficiency of using labor, which in this case is endogenous to past information quality up to the coefficient κ. The bottom right panel indicates that the change in labor data quality γ_t^e decreases to a much lesser extent, indicating that there is virtually no change in the productivity of labor. This is because $\kappa < 1$, and therefore the productivity of labor is less affected by the endogenous evolution of data quality.

7 Conclusion

Data is the lifeblood of AI. Quaity of data determines quality of AI. This paper proposes a data corruption channel and explores its implications for the long-term effects of labor displacement by artificial intelligence (AI). It presents a model where the adoption of AI and the quality of data are jointly determined within an information equilibrium. The central focus is a key statistic: the relative quality of data produced by humans versus AI. When this statistic is calibrated with recent experimental findings from computer science literature, it is found that information produced

by AI is much inferior to those produced by human, and hence, the data corruption channel is quantitatively important, and can reverse 30% of the labor displacement effect in the long run. Nonetheless, AI should still be regulated, possibly through taxation, in the long run to achieve the socially efficient level of data quality.

The model is presented in a purposefully simple manner to highlight the key force at work. It can be extended in various ways to explore many interesting questions. For instance, while this paper assumes short-lived entrepreneurs for simplicity, future research could explore social learning in the context of long-lived firms. This would introduce additional complexities, such as data privacy issues highlighted in works like Jones and Tonetti (2020). A future exciting area of research would be to create a comprehensive quantitative model that considers various forces shaping the co-evolution of data and AI, allowing for a more detailed and practical discussion of government policies.

References

Abis, Simona and Laura Veldkamp, "The Changing Economics of Knowledge Production," *Working Paper*, 2021.

Acemoglu, Daron and Pascual Restrepo, "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, June 2018, *108* (6), 1488–1542.

_ **and** _ , "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, May 2019, *33* (2), 3–30.

_ **and** _ , "Tasks, Automation, and the Rise in U.S. Wage Inequality," *Econometrica*, 2022, *90* (5), 1973–2016.

_ **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Alemohammad, Sina, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G. Baraniuk, "Self-Consuming Generative Models Go MAD," 2023.

Alonso, Cristian, Andrew Berg, Siddharth Kothari, Chris Papageorgiou, and Sidra Rehman, "Will the AI revolution cause a great divergence?," *Journal of Monetary Economics*, 2022, *127*, 18–37.

Amador, Manuel and Pierre-Olivier Weill, "Learning from Prices: Public Communication and Welfare," *Journal of Political Economy*, 2010, *118* (5), 866–907.

Bohacek, Matyas and Hany Farid, "Nepotistically Trained Generative-AI Models Collapse," 2023.

Briesch, Martin, Dominik Sobania, and Franz Rothlauf, "Large Language Models Suffer From Their Own Output: An Analysis of the Self-Consuming Training Loop," 2023.

Chen, Lingjiao, Matei Zaharia, and James Zou, "How is ChatGPT's behavior changing over time?," 2023.

Doshi, Anil R. and Oliver P. Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," 2024.

Dowson, D.C and B.V Landau, "The FrÃ©chet distance between multivariate normal distributions," *Journal of Multivariate Analysis*, 1982, *12* (3), 450–455.

Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel, "Uncertainty Traps*," *The Quarterly Journal of Economics*, 05 2017, *132* (4), 1641–1692.

Farboodi, Maryam and Laura Veldkamp, "A Model of the Data Economy," *Working Paper*, 2022.

_ **, Roxana Mihet, Thomas Philippon, and Laura Veldkamp**, "Big Data and Firm Dynamics," *AEA Papers and Proceedings*, May 2019, *109*, 38–42.

Guo, Yanzhu, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel, "The Curious Decline of Linguistic Diversity: Training Language Models on Synthetic Text," 2023.

Jones, Charles I. and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, September 2020, *110* (9), 2819–58.

Martínez, Gonzalo, Lauren Watson, Pedro Reviriego, José Alberto Hernández, Marc Juarez, and Rik Sarkar, "Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet," 2023.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo, "Uneven Growth: Automation's Impact on Income and Wealth Inequality," *Econometrica*, 2022, *90* (6), 2645–2683.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *American Economic Review*, December 2002, *92* (5), 1521–1534.

Ordonez, Guillermo, "The Asymmetric Effects of Financial Frictions," *Journal of Political Economy*, 2013, *121* (5), 844 – 895.

Rio-Chanona, Maria Del, Nadzeya Laurentsyeva, and Johannes Wachs, "Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow," 2023.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson, "AI models collapse when trained on recursively generated data," *Nature*, 2024, *631*, 755–759.

Veldkamp, Laura, "Slow boom, sudden crash," *Journal of Economic Theory*, 2005, *124* (2), 230–257.

Appendix

A Proof

Proof of theorem 5.1

Plug in the conditional value functions into the social planner's problem:

$$V^{sp} = \max_{\gamma, \bar{\varphi}} \int^{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_I} - \varphi \right] d\varphi + \int_{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma} - R \right] d\varphi$$

We could arrange the constraint as $\gamma(\bar{\varphi})$, hence the value function becomes:

$$V^{sp} = \max_{\bar{\varphi}} \int^{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_I} - \varphi \right] d\varphi + \int_{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] d\varphi$$

Now consider perturbing $\bar{\varphi}$. The first order condition becomes:

$$\left[\bar{A} - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi} \right] - \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] + \int_{\bar{\varphi}} \left[\frac{\gamma'(\bar{\varphi})}{(\gamma_S + \gamma_A + \gamma(\bar{\varphi}))^2} \right] d\varphi = 0$$

Thus the socially efficient threshold $\bar{\varphi}^{sp}$ is pinned down by:

$$\underbrace{\frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi}^{sp})} + R - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi}^{sp}}_{\text{Private Benefit Term}} + \underbrace{\frac{(1 - n(\bar{\varphi}^{sp}))}{(\gamma_S + \gamma_A + \gamma(\bar{\varphi}^{sp}))^2} \gamma'(\bar{\varphi}^{sp})}_{\text{Externality Term}} = 0$$

Compared to the private equilibrium where the threshold is pinned down by equating the private gains and benefits:

$$\frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} + R - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi} = 0$$

We have an extra term $\int_{\bar{\varphi}} \left[\frac{\gamma'(\bar{\varphi})}{(\gamma_S + \gamma(\bar{\varphi}))^2} \right] d\varphi$ which captures the positive externality: the fact that one's economic behavior could generate data, improve the data quality and the quality of AI, and therefore the overall welfare. Because of this extra term, the government needs to tax the AI in the steady state. Let's say that the government needs to impose a wedge (tax) τ on the return of AI, and rebate all proceeds in a lump-sum fashion to all the households in a uniform way. Then the private CE becomes:

$$\left[\bar{A} - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi} \right] - \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] (1 - \tau) = 0$$

$$\left[\bar{A} - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi} \right] - \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] + \tau \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] = 0$$

$$\frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} + R - \frac{1}{\gamma_S + \gamma_I} - \bar{\varphi} + \tau \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma(\bar{\varphi})} - R \right] = 0$$

Hence we need the tax rate to be such that:

$$\tau\left[\bar{A}-\frac{1}{\gamma_S+\gamma_A+\gamma\left(\bar{\varphi}\right)}-R\right]=\int_{\bar{\varphi}}\left[\frac{\gamma'\left(\bar{\varphi}\right)}{\left(\gamma_S+\gamma_A+\gamma\left(\bar{\varphi}\right)\right)^2}\right]d\varphi$$

$$\tau=\frac{\left(1-F\left(\bar{\varphi}\right)\right)\left[\frac{\gamma'(\bar{\varphi})}{(\gamma_S+\gamma_A+\gamma(\bar{\varphi}))^2}\right]}{\bar{A}-\frac{1}{\gamma_S+\gamma_A+\gamma(\bar{\varphi})}-R}$$

where we factor out $\frac{\gamma'(\bar{\varphi})}{(\gamma_S+\gamma_A+\gamma(\bar{\varphi}))^2}$. But note that it depends on every term in this expression is positive except $\gamma'\left(\bar{\varphi}\right)$. Let us determine the sign of it. Write the constraint of the social planner as:

$$H\left(\gamma,\bar{\varphi}\right)=\frac{1}{\gamma}-\rho^2\frac{1}{\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D}-\frac{1}{\gamma_\eta}=0$$

$$\frac{\partial H}{\partial\bar{\varphi}}=\rho^2\frac{\left[\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D-\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right]f\left(\bar{\varphi}\right)}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}$$

$$\frac{\partial H}{\partial\gamma}=-\frac{1}{\gamma^2}-\rho^2\frac{-\left(1-2\left(1-F\left(\bar{\varphi}\right)\right)\frac{\gamma_A}{(\gamma_S+\gamma_A+\gamma)^2}\gamma_D\right)}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}$$

Hence

$$\gamma'\left(\bar{\varphi}\right)=-\frac{\frac{\partial H}{\partial\bar{\varphi}}}{\frac{\partial H}{\partial\gamma}}$$

One can show that $\frac{\partial H}{\partial\gamma}<0$:

$$\begin{aligned}\frac{\partial H}{\partial\gamma} &= -\frac{1}{\gamma^2}-\rho^2\frac{-1}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}\\ &\quad -\rho^2\frac{2\left(1-F\left(\bar{\varphi}\right)\right)\frac{\gamma_A}{(\gamma_S+\gamma_A+\gamma)^2}\gamma_D^A}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}\\ &< -\rho^2\frac{2\left(1-F\left(\bar{\varphi}\right)\right)\frac{\gamma_A}{(\gamma_S+\gamma_A+\gamma)^2}\gamma_D}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}<0\end{aligned}$$

because

$$\frac{1}{\gamma^2}>\rho^2\frac{1}{\left(\gamma+\gamma_s+F\left(\bar{\varphi}\right)\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2\gamma_D+\left(1-F\left(\bar{\varphi}\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2\gamma_D\right)^2}$$

Hence the sign of $\gamma'(\bar{\varphi})$ depends on $\frac{\partial H}{\partial \bar{\varphi}}$, which in turn depends on

$$\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2 \gamma_D - \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2 \gamma_D$$

Hence if

$$\frac{\gamma_I}{\gamma_A} > \frac{\gamma_S}{\gamma_S+\gamma(\bar{\varphi})}$$

we can show that

$$\left(\frac{\gamma_I}{\gamma_S+\gamma_I}\right)^2 \gamma_D - \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma}\right)^2 \gamma_D > 0$$

, then

$$\gamma'(\bar{\varphi}) > 0$$

which means that

$$\tau > 0$$

On the other hand, if AI is more efficient in generating data than labor:

$$\frac{\gamma_I}{\gamma_A} < \frac{\gamma_S}{\gamma_S+\gamma(\bar{\varphi})}$$

Then one can show that

$$\gamma'(\bar{\varphi}) < 0$$

and hence

$$\tau < 0$$