Artificial Intelligence, Data Corruption, and Labor Displacement

Zhifeng Cai *

February 11, 2024

Abstract

The rise of generative artificial intelligence (AI) has fueled the conern that AI will ultimately replace human jobs. This paper introduces a data corruption channel, arguing that the growing use of AI could eventually reduce its own productivity, due to a surge in AI-generated data being used in training datasets. This diminished productivity implies that AI's impact on displacing human labor may not be as significant in the long run. Through a model of social learning with both human and AI production, it is shown that the quantitative relevance of this channel depends on the relative quality of data generated by AI versus human activities. Calibrating the model based on recent experimental findings from the AI literature, it is found that the data corruption channel could mitigate AI's initial labor displacement effect by 30% over the long run. Taxes on AI adoption need to be implemented as private agents fail to internalize the impact of their AI adoption on aggregate data quality.

*First version: July 2023. Cai: Department of Economics, Rutgers University, 75 Hamilton St, New Brunswick, NJ 08901, USA., (email: zhifeng.cai@rutgers.edu). I thank Diego Perez, Laura Veldkamp, and Yucheng Yang for comments. ChatGPT provides text-editing services. All errors remain my own.

1 Introduction

The era of artificial intelligence (AI) has dawned upon us. What was once relegated to the realm of science fiction has now permeated our everyday existence, making its presence felt across numerous sectors and industries. With AI's development advancing at an unparalleled rate, it brings forth a unique set of challenges, among which the effects on the labor market stand out as a pressing issue. Several recent works have focused on exploring the displacement effects of AI on labor, particularly in terms of suppressing labor demand and intensifying income inequality (Acemoglu and Restrepo, 2018, 2019; etc). The advent of sophisticated large language models, such as ChatGPT, amplifies worries about AI displacing human labor in the foreseeable future. Recent empirical studies, including those by Yilmaz et al. (2023) and Hui et al. (2023), shed light on AI's immediate detrimental effects on the labor market, manifesting through reduced wages and decreased employment opportunities over the short run.

This paper examines the *long run* impacts of AI on the labor market. To this end, it moves beyond pure empirical analyses to integrate economic theories with cutting-edge insights from AI computer science research. The novel perspective this paper brings to the debate regarding the tension between AI and labor is the notion of "data corruption", which posits that the widespread adoption of AI may degrade the quality of data used to train subsequent AI generations, thereby adversely affecting the quality of AI over time.

The idea of data corruption is illustrated in figure 1. Generative AI models, like ChatGPT, rely on vast datasets for effective training (indicated by the black arrow). But where do these data come from? Before the inception of the age of AI, these datasets were primarily composed of data generated through human activities (blue arrow). These "real data" are then fed into AI models for training. Now, with the increased popularity of generative AI models, those training datasets are corrupted with "synthetic data" which are generated not through authentic human activities, but through AI models. For instance, AI-powered tools like Midjourney and DallE2 allow users to produce digital artwork that, once shared online, may become part of the training data for future AI models. This cycle of using synthetic data in training could significantly affect AI productivity over the long run.

The issue of data corruption has recently attracted significant attention within the AI and computer science community. Table 1 surveys recent articles in the computer science literature that explore the negative implications of training AI models with iterations of synthetic data. A com-

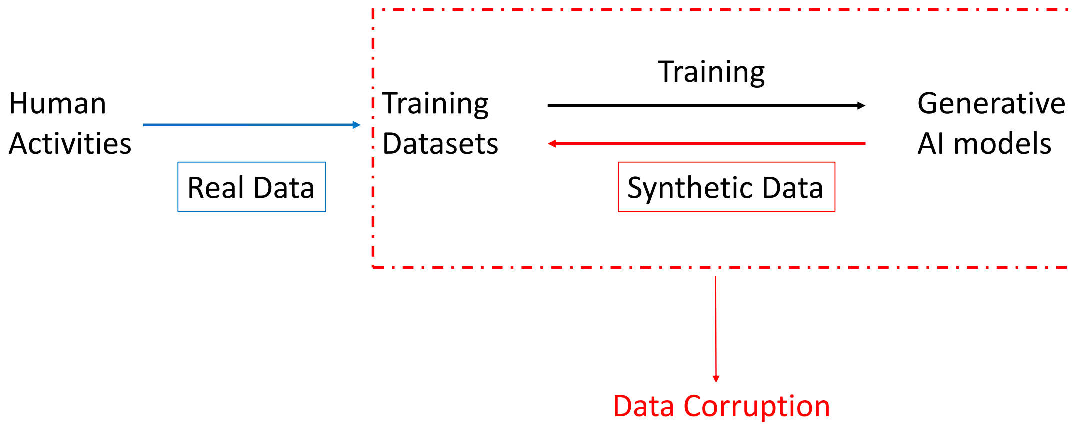

Figure 1: Data Corruption with Generative AI

mon theme in these studies is that relying on synthetic data for AI training can erode the AI's capacity to perform its designated tasks effectively. We leverage these insights and experimental results to calibrate and inform some deep parameters in our economic model, to which we now turn.

The primary goal of this paper is to introduce an economic framework that connects findings from the AI domain to ongoing economic discussions about the competition between human labor and AI technology. Previous theoretic frameworks, such as those in Acemoglu and Restrepo (2018), do not suit our analysis as they do not account for the pivotal role of data quality in determining AI productivity.

To this end, I develop a model where production occurs through either human labor or AI technology. A novel aspect of this model is the assumption that AI productivity is directly influenced by the quality of the data it is trained on. The concept of data in the model is consistent with the social learning literature, that data emerges from economic activities (Veldkamp 2005; Ordonez 2013; Fajgelbaum et al. 2017). Both AI and human labor can generate new information and also act upon it, creating a feedback loop where the data generated by their activities informs the training of future generations of AI. The key advantage of AI relative to human labor is that AI has access to all historical data from economic activities, regardless of whether it was produced by humans or previous versions of AI. The relative composition between economic activities carried

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 et al. (Stanford)	Nov 2023
Under the Surface: Tracking the Artifactuality of LLM-Generated Data	Das et al. (U of Minnesota)	Jan 2024

Note: Papers are listed in chronological order. For the sake of space, only the first academic affiliation is listed.

out by AI versus humans thus could potentially impact the overall data quality.

To illustrate, consider a scenario in the healthcare sector where diagnosing a patient is the task at hand. There's inherent uncertainty about the best method for diagnosis, which is a key variable. Decision-makers must estimate this variable to choose the most effective treatment path. If a human professional, like a doctor, is tasked with the diagnosis, there are associated labor costs. The human directly engages with the patient—observing symptoms and gathering data—which then informs the diagnosis. This direct human interaction provides a unique, firsthand data signal on the health condition. On the other hand, choosing AI for the task incurs technology deployment costs. The AI's prediction capabilities and thus its productivity hinge on the historical data it has been trained with. Continuing with the healthcare example, an AI would utilize data records from previous patients to diagnose the current one. It's also posited that AI can generate its own data signals about the condition, akin to creating simulated patient data through generative models, to improve prediction accuracy for the task at hand. Note that all past data – diagnoses of previous patients – could be coming from diagnoses from either human doctors or AI technologies.

I first establish an information equilibrium in which AI adoption and data quality are jointly determined – the quality of data determines the attractiveness of AI technology relative to labor, and the relative proportion of AI and labor determines the data quality. Through this framework, the paper explores the model's positive and normative implications: How does introducing endogenous data quality change our understanding of the AI versus labor debate, particularly regarding AI's potential to replace jobs over the long run? And, with data quality playing a role in AI adoption, what new considerations emerge for AI regulation?

The paper's main finding is that the long-term impact of AI on job displacement may not be as severe as previously thought. The data corruption channel plays a crucial role here in reverting the labor displacement of AI. Consider a situation where AI becomes more affordable following a significant reduction in expenses (possibly due to advancements in big data technology leading to tools like ChatGPT). The immediate effect would likely be a rapid expansion in AI adoption, leading to a decrease in labor demand. Nevertheless, because AI-driven activities generally produce information of lower quality compared to human activities, a surge in AI utilization could gradually degrade the overall quality of data. This deterioration in data quality could, in turn, diminish the effectiveness of AI, mitigating the initial rush towards AI adoption. Consequently, the market share dynamics for AI might exhibit a bell-shaped curve, with an initial increase in the short term followed by a decrease in the long term.

I conduct numerical simulations to evaluate the quantitative relevance of this data corruption channel. I first show that a key statistic that determines the magnitude of this channel is the relative quality of data produced by human labor activities and that generated by AI. If the information quality from AI were on par with that from human labor, an increase in AI's market share would not result in much of a decline in overall data quality, thus avoiding a significant long-term reversal in labor displacement trends. Conversely, if the quality of information from AI were inferior to that produced by humans, the data corruption channel would emerge as significantly impactful.

To identify the relative quality of data from AI versus human sources, I explore the latest research from the forefront of AI and computer science literature. As shown in table 1, a growing body of works investigate how training AI models with synthetic (AI-generated) data can degrade these models' productivity. Specifically, several studies (Chen et al. 2023; Alemohammad et al. 2023; Martínez et al. 2023) have assessed the decline in data quality by analyzing changes in the Fréchet Inception Distance (FID) as synthetic data is incorporated into AI training processes. I calculate the corresponding metrics within my model and calibrate the relative data quality of AI v.s. human to replicate this key piece of evidence.

The resulting calibration suggests that AI provides much less information compared to human labor: in the benchmark calibration data produced by AI is found to be only 8% as informative as that from human labor. Given such a large disparity, the data corruption channel is quantitatively important. When analyzing the dynamics of AI adoption following a one-time, permanent reduction in AI technology costs, the model predicts an immediate spike in AI adoption to 100%, which later stabilizes to around 70% in the long term. Therefore, the data corruption channel can reverse approximately 30% of the initial labor displacement over the long run.[1]

The data corruption channel also leads to an information externality, which calls for government intervention: individual agents don't consider their AI adoption decisions' impact on equilibrium aggregate data quality. Depending on the relative quality of AI and human-generated data. Given that AI tends to generate less information comparing to human activities, this externality is negative, indicating the need for a tax on AI. Unlike classic results in capital taxation that capital tax should be zero over the long run, with the presence of this information externality government intervention in AI technology remains necessary even in the long run.

[1]Consistent with the channel of the paper, Rio-Chanona et al. (2023) find that the introduction of large language models like ChatGPT reduces human-generated content and posts on Stack Overflow.

The paper 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. 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.

Additionally, the paper connects to the existing body of 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.

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 and Technology

Time is assumed to be infinite, spanning from $-\infty$ to $+\infty$. In each period, there exist $\bar{N}$ 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 $p = 1$.[2]

The quality of the production is endogenous, and is influenced by the entrepreneurs action a_t^i.

[2]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.

Specifically, I adopt the following functional form as in Farboodi and Veldkamp (2022):[3]

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

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

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

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

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. Given the setting of incomplete information, the primary objective for each entrepreneur is to predict the true value of θ_t, potentially aided by AI technology.

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

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

Here, ε_t^S represents noise 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.

The entrepreneur must 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:

$$s_t^{li} = \theta_t + \varepsilon_t^{li}. \tag{2}$$

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 en-

[3]In a recent version of Farboodi and Veldkamp (2022), they 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 strictly decreasing function. All of this paper's result goes through with this generalization.

trepreneur's optimal action is determined by:

$$W^l(S_t, s_t^{li}) = \max_{a_t^i} E(A_t^i(a_t^i)|S_t, s_t^{li}) \tag{3}$$

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.

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.

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:

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

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 past information (Ω_{t-1}) by AI to predict future outcomes is a common practice in many applications. 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. Additionally, AI can generate its own data, similar to generative models creating synthetic data to train the AI model.

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 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.

Based on both sources of information, the AI then selects the action on behalf of the entrepreneur:

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

The payoff from AI adoption is then given by:

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

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 generated as a noisy signal about that action:

$$D_t^i = a_t^i + \varepsilon_t^{Di}. \tag{4}$$

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 observable, while private signals are not directly observable. 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 two aggregate statistics that reflect information generated by human labor and AI, respectively:

$$X_t^l = \frac{1}{N_t} \sum_{i \leq N_t} \left(a_t^i + \varepsilon_t^{Di} \right) \tag{5}$$

$$X_t^A = \frac{1}{\bar{N} - N_t} \sum_{i > N_t} \left(a_t^i + \varepsilon_t^{Di} \right) \tag{6}$$

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:

$$\Omega_{t-1} = \{S_{t-j}, X_{t-j}^l, X_{t-j}^A\}_{j=1,2,3,\ldots,-\infty} \tag{7}$$

2.4 Equilibrium

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

1. $\left\{\mathbb{1}_{At}^i, a_t^i\right\}_{t=-\infty,\ldots,+\infty}^{i=1,2,\ldots\bar{N}}$ *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. $\left\{X_t^l, X_t^A\right\}_{t=-\infty,\ldots,+\infty}$ *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 now 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 introduce the concept of "stock of knowledge":

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

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 7 admit the following recursion:

$$\Omega_t = \Omega_{t-1} \cup \{S_t, X_t^l, X_t^A\}. \tag{8}$$

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 Human-Generated Data

We first characterize the information generated by human labor X_t^l (equation 5). Solving the optimal action rule for those who produce with labor (equation 3), we have:

$$a_t^i = E(\theta_t|S_t, s_t^{li}) = \frac{\gamma_S}{\gamma_S + \gamma_l} S_t + \frac{\gamma_l}{\gamma_S + \gamma_l} s_t^{li}$$

This action generates a data point:

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

Since S_t is publically known, observing D_t^i is equivalent to observing the idiosyncratic part $\frac{\gamma_l}{\gamma_S+\gamma_l} s_t^{li} + \varepsilon_t^{Di}$, which is equivalent to the following signal of the fundamental θ_t:[4]

$$\theta_t + \left(\varepsilon_t^{li} + \frac{\gamma_S + \gamma_l}{\gamma_l} \varepsilon_t^{Di} \right).$$

Hence we obtain an expression for the summary statistic of human data (equation 5):

$$\begin{aligned} X_t^l &= \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\theta_t + \varepsilon_t^{li} + \frac{\gamma_S + \gamma_l}{\gamma_l} \varepsilon_t^{Di} \right) \\ &\sim N\left(\theta_t, \frac{1}{N_t} \left(\frac{1}{\gamma_l} + \left(\frac{\gamma_S + \gamma_l}{\gamma_l} \right)^2 \frac{\bar{N}}{\gamma_D} \right) \right) \end{aligned}$$

where we plug in the precision of ε_t^{li} and ε_t^{Di}.

Note that as the number of entrepreneurs $\bar{N} \to \infty$, this signal becomes:

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

where n_t is the *share* of labor-adopting firms. Thus we obtain the following:

Lemma 3.1 *The information contents in human production activities can be summarized in X_t^l 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 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 idea that agents' actions give weight to the public signal, and part of the variation in actions does not contribute additional information beyond what is already observed in the public information. Importantly, this term is less than 1, indicating that the actions of agents do not fully enhance the information content of the public signal.

3.3 Information Content in AI activities

If the agent chooses to adopt AI tech, his optimal action is:

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

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.

[4] To obtain this, multiply the idiosyncratic part $\frac{\gamma_l}{\gamma_S+\gamma_l} s_t^{li} + \varepsilon_t^{Di}$ by $\frac{\gamma_S+\gamma_l}{\gamma_l}$, and we obtain $s_t^{li} + \frac{\gamma_S+\gamma_l}{\gamma_l} \varepsilon_t^{Di}$. The signal expression can be obtain by noticing that $s_t^{li} = \theta_t + \varepsilon_t^{li}$ (equation 2).

As in the case with labor production, this action generates a data point:

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

where both S_t and μ_t are prior knowledge. Hence the only valuable part of this data point is the idiosyncratic component $\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}s_t^{Ai} + \varepsilon_t^{Di}$, which can be arranged into the following signal about fundamental θ_t :

$$\theta_t + \left(\varepsilon_t^{li} + \frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\varepsilon_t^{Di}\right).$$

Hence we obtain an expression for the summary statistic of AI-generated data (defined in equation 6):

$$\begin{aligned} X_t^A &= \frac{1}{\bar{N}-N_t}\sum_{i>N_t}\left(\theta_t + \varepsilon_t^{Ai} + \frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\varepsilon_t^{Di}\right) \\ &\to N(\theta_t, \frac{1}{1-n_t}\left(\frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\right)^2\frac{1}{\gamma_D}) \end{aligned}$$

where the limit in the last equation is taken with respect to $\bar{N} \to \infty$. $1-n_t$ is the share of AI-adopting firms.

Lemma 3.2 *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 share $1-n_t$ because more AI-related activities generates more information content for the AI signal. The middle term $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2$ captures the fact that agents' actions 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.

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 8), we have:

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

From lemma 3.1 and 3.2, 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:

Theorem 3.1 *Given γ_t and n_t, γ_{t+1} is given by:*

$$
\begin{aligned}
\frac{1}{\gamma_{t+1}} &= Var\left(\theta_{t+1}|\Omega_t\right) \\
&= \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t\left(\frac{\gamma_l}{\gamma_S+\gamma_l}\right)^2 \gamma_D + (1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}
\end{aligned} \tag{9}
$$

Note that this is a partial equilibrium result in the sense that the share of labor-adopting entrepreneurs n_t is given. 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 3.2 *Holding fixed a 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 in AI is sufficiently low compared to those generated through labor:*

$$
\frac{\gamma_A}{\gamma_l} < \frac{\gamma_S + \gamma_t}{\gamma_S}. \tag{10}
$$

proof: It can be observed from equation 9 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 10.

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 10 is satisfied. Why could this occur, considering they both provide equal-equality information?

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 10, hence the more likely that AI can increase the quality of aggregate data.[5]

From condition 10, one can easily see that the larger the existing body of knowledge (i.e., higher γ_t), the larger the value on the right-hand side of equation 10, suggesting that the adoption of AI is more likely to hinder knowledge growth. 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 3):

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

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

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

[5] 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, 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.

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):

$$\begin{aligned} V^l\left(\varphi_t^i\right) &= E\left(W^l\left(S_t, s_t^{li}\right)|S_t\right) - \varphi_t^i \\ &= \bar{A} - \frac{1}{\gamma_S + \gamma_l} - \varphi_t^i \end{aligned} \tag{11}$$

Similarly we can derive the payoff of adopting AI technology

$$\begin{aligned} V^A\left(\gamma_t\right) &= \bar{A} - Var\left(\theta_t | S_t, \Omega_{t-1}, s_t^{Ai}\right) \\ &= \bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - R_t \end{aligned} \tag{12}$$

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

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

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:

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

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

Theorem 3.3 *The dynamics of the model is fully characterized by the following system:*

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t\left(\gamma_t\right)\left(\frac{\gamma_l}{\gamma_S + \gamma_l}\right)^2 \gamma_D + \left(1 - n_t\left(\gamma_t\right)\right)\left(\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{13}$$

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

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

Note that the knowledge stock parameter γ_t appears in three distinct locations within the law of motion equation 13. Firstly, it embodies information passed down from previous periods.

Figure 2: Law of Motion and Policy Functions

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 provides a visualization of the equilibrium functions.[6] 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, demonstrating the standard information transmission effect. The decreasing part indicates that an abundance of information spurs greater AI adoption (as illustrated in the right panel). Increased AI adoption can, in turn, impede knowledge growth when γ_l and γ_A are equal.

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 $\bar{A} - \frac{1}{\gamma_S + \gamma_l}$. Hence, the relative (expected) productivity of labor versus AI,

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

is a constant and independent of γ_t.[7]

[6] 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.

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

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 knowledge base (higher γ_t) results in AI being relatively more productive than labor, leading to increased AI adoption.

An important consideration is that even though AI can be *more* productive, it may contribute *less* information to aggregate knowledge 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: $\frac{\gamma_A}{\gamma_I}$, as shown in equation 10. 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 distri-

task.

Figure 3: Full Synthetic Loops

butions 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.

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 1. The calibration is based on the autocorrelation and standard deviation of the logarithm of output, resulting in standard parameter values: $\rho = 0.99$ and $\gamma_\eta = 1/(0.055)^2$.[8] 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: $\gamma_A, \gamma_I, \gamma_D$, and γ_S. According to equation 10, 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

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

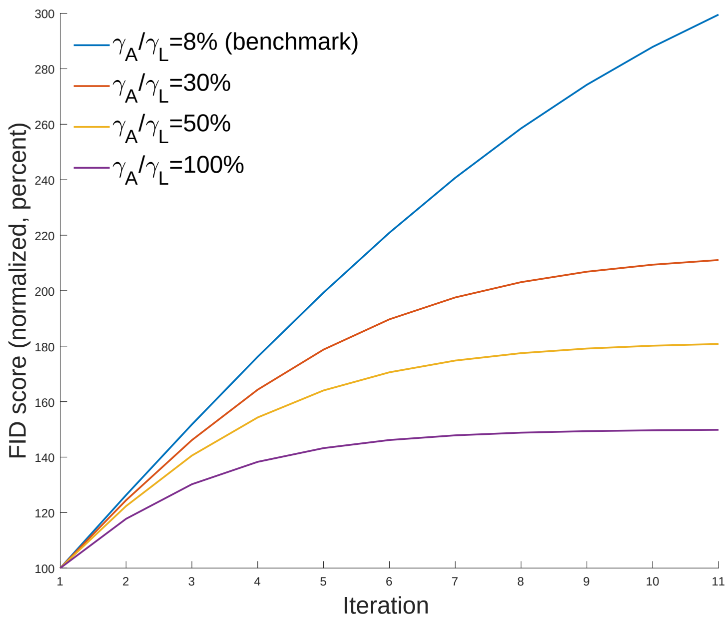

Figure 4: FID score in the model for full synthetic loops

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 4), we calibrate it to match the initial data quality observed in the full synthetic loop 3. The initial 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.

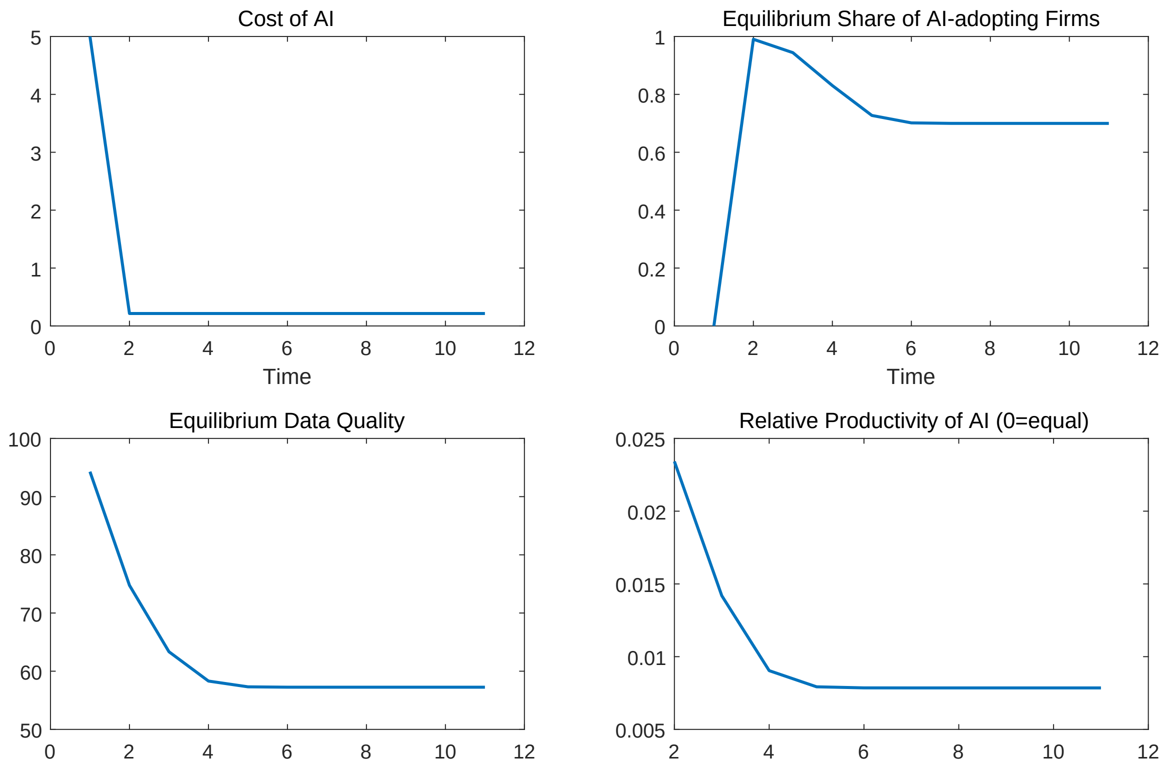

Figure 5: Model Dynamics – the ChatGPT Experiment

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 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 13. 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

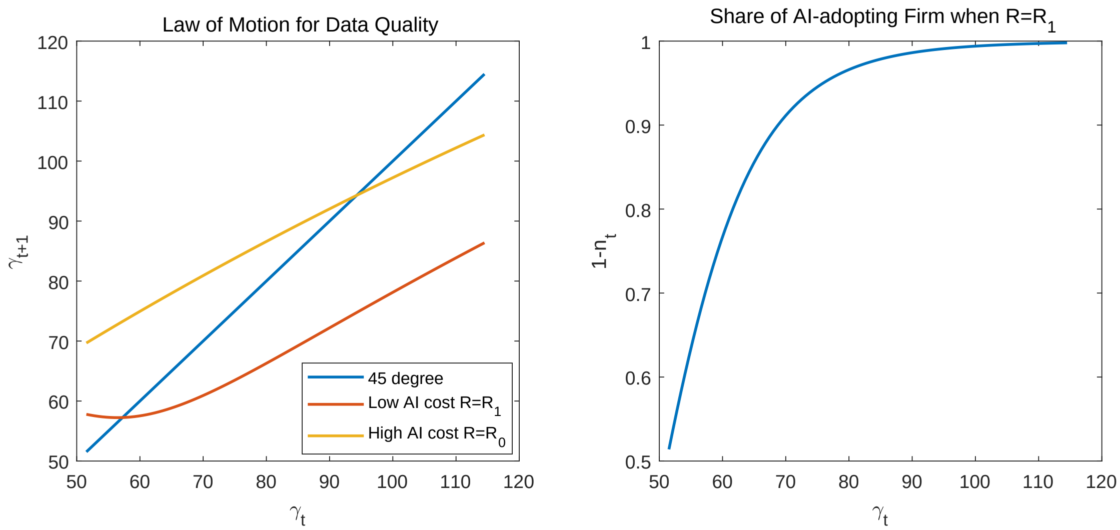

Figure 6: Equilibrium Functions when $\gamma_A = \gamma_l$

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.[9] 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.

[9] 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 12 and 11. The expectation is taken over idiosyncratic labor costs ϕ^i.

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 $\gamma_A = \gamma_I$ 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 dynamics 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 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%.

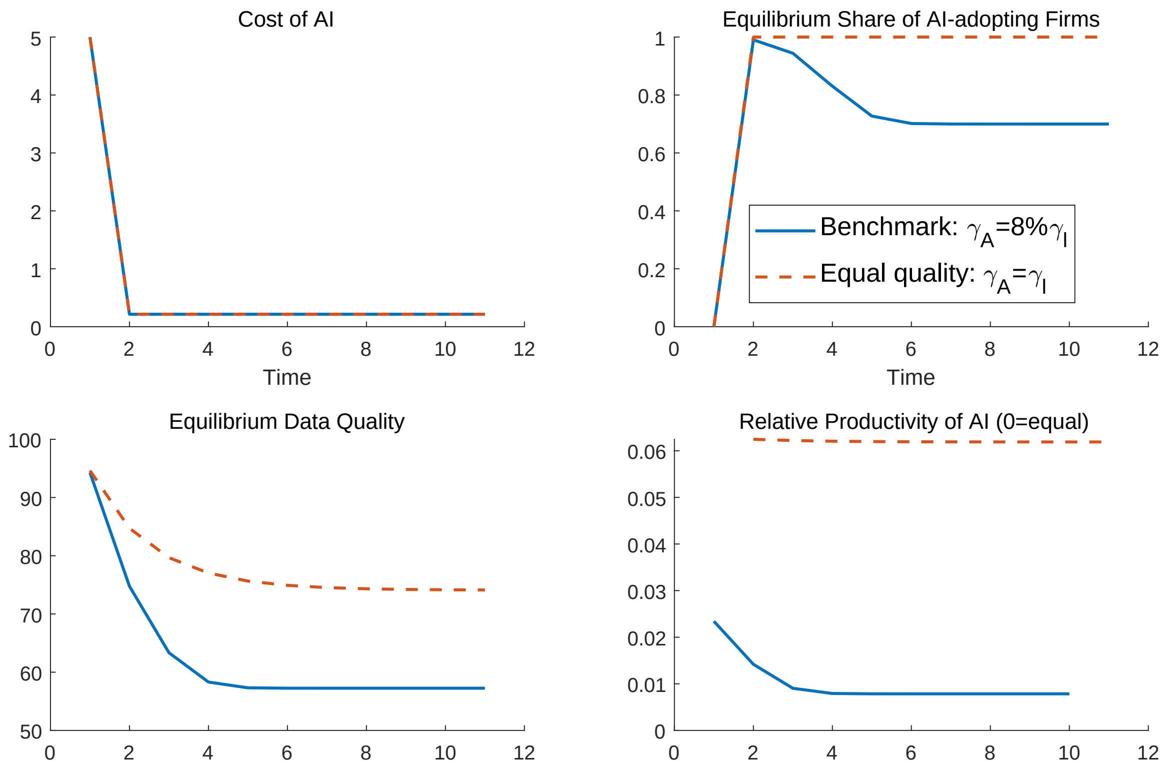

Figure 7: The ChatGPT Experiment with Equal Information Quality

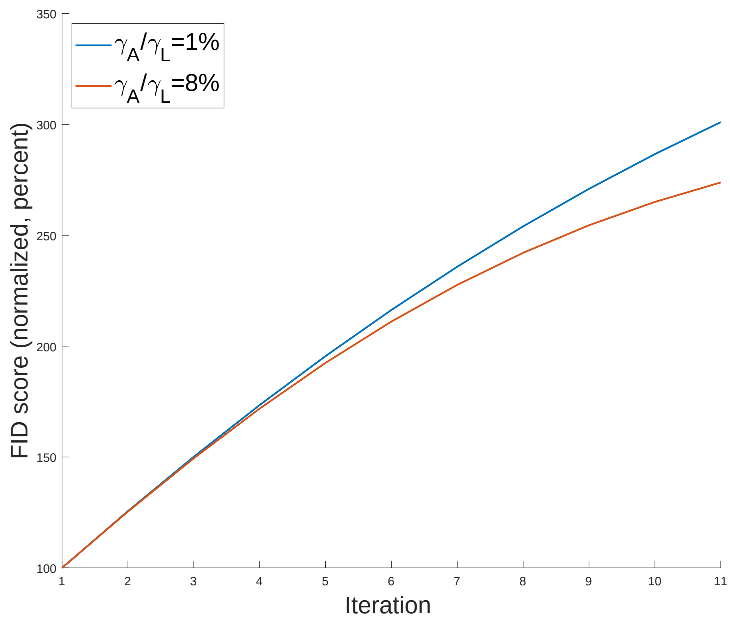

Figure 8: FID score with public signal $\gamma_S = 1$

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 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

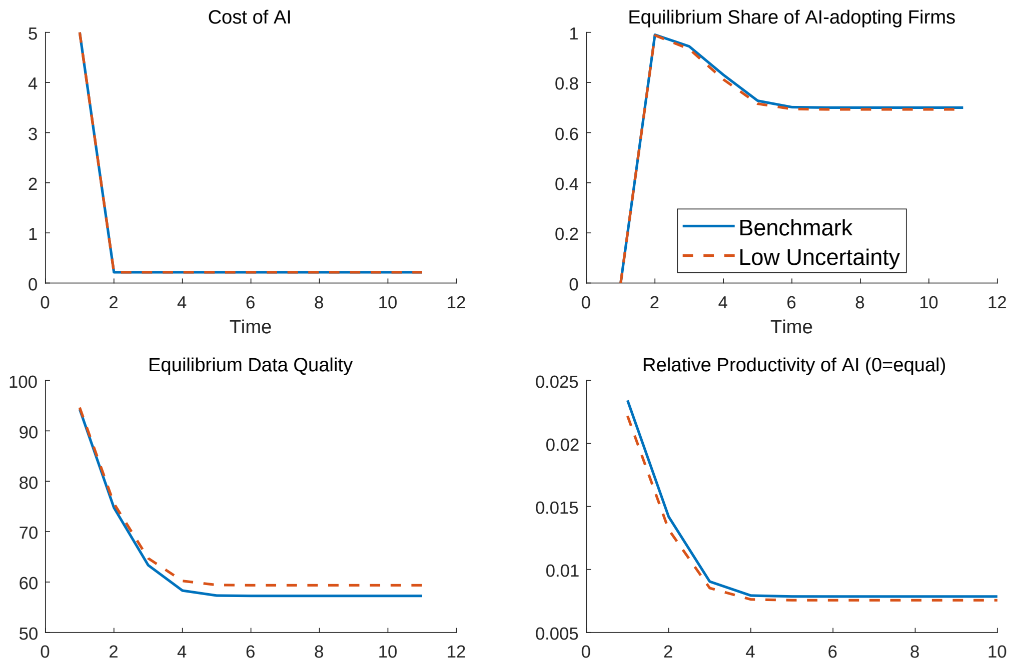

Figure 9: Model Dynamics with Low Uncertainty

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. (2023) 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

In this section, we investigate the efficiency of the information equilibrium, focusing particularly on the efficiency of the AI adoption decision. We seek to answer questions such as: Is the private sector's decision to adopt AI socially efficient? If given the choice, would a planner change the mass of AI adoption?

To explore these questions, we establish a benevolent social planner's problem. To focus on the main tradeoff, we focus on the steady-state version of the planner's problem. The planner can choose which private agents adopt AI technology, but otherwise cannot affect the knowledge accumulation procedure.

$$V^{sp} = \max_{\gamma, \bar{\varphi}} \int^{\bar{\varphi}} \left[V^{l}(\varphi) \right] d\varphi + \int_{\bar{\varphi}} \left[V^{A}(\gamma) \right] d\varphi$$

$$s.t.$$

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

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

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 determines the steady state information quality γ from the first constraint. The two constraints jointly pin down $\gamma(\bar{\varphi})$ as a function of $\bar{\varphi}$.

Plug in $\gamma(\bar{\varphi})$ and the value function for labor adoption $V^{l}(\varphi)$ and AI adoption $V^{A}(\gamma)$ (equation 11 and 12):

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

One can see from this expression that 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 (the expression $\gamma(\bar{\varphi})$ in red) while private agents takes γ as fixed.

The following first order condition pins down the socially efficient threshold $\bar{\varphi}^{sp}$:

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

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. $\gamma'(\bar{\varphi}^{sp}) > 0$ if and only if the adjusted quality of human data is greater than the AI data:*

$$\frac{\gamma_I}{\gamma_A} > \frac{\gamma_S}{\gamma_S + \gamma(\bar{\varphi}^{sp})} \tag{14}$$

2. If the government were to levy a tax on AI adoption, the tax rate is positive if and only if $\gamma'(\bar{\varphi}^{sp}) > 0$

Condition 14 takes a similar force to condition 10, 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 aggregate information content. 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_I 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.

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:

$$\gamma_A\left(\gamma_t\right) = a + b\gamma_t \tag{15}$$

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:

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

where $\gamma_A\left(\gamma_t\right)$ is given by equation 15. 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 $\bar{N} \to \infty$):

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

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:

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t\left(\frac{\gamma_l}{\gamma_S + \gamma_l + \gamma_t^e}\right)^2 \gamma_D + \left(1 - n_t\right)\left(\frac{\gamma_A(\gamma_t)}{\gamma_S + \gamma_A(\gamma_t) + \gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{16}$$

And γ^e_{t+1} is given by the following function of γ_t and γ^e_t:

$$\frac{1}{\gamma^e_{t+1}} = \rho^2 \frac{1}{\gamma^e_t + \gamma_s + \kappa n_t \left(\frac{\gamma_l}{\gamma_S + \gamma_l + \gamma^e_t}\right)^2 \gamma_D + \kappa \left(1 - n_t\right) \left(\frac{\gamma_A(\gamma_t)}{\gamma_S + \gamma_A(\gamma_t) + \gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{17}$$

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

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

which boils down to:

$$\frac{\gamma_l}{\gamma_A\left(\gamma_t\right)} > \frac{\gamma_S + \gamma^e_t}{\gamma_S + \gamma_t}$$

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, $\gamma_t > \gamma^e_t$, and therefore,

$$\frac{\gamma_S + \gamma^e_t}{\gamma_S + \gamma_t} < 1$$

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:

$$V^l = \bar{A} - \frac{1}{\gamma_S + \gamma_l + \gamma^e_t} - \varphi^i_t$$

while the value of adopting AI is

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

Hence the threshold cost of labor is pinned down by

$$\bar{\varphi}\left(\gamma_t, \gamma^e_t\right) = R_t + \frac{1}{\gamma_S + \gamma_A\left(\gamma_t\right) + \gamma_t} - \frac{1}{\gamma_S + \gamma_l + \gamma^e_t} \tag{18}$$

Thus model dynamics is fully characterized by equation 16 and 17, where the share of labor-adopting entrepreneurs is given by

$$n_t\left(\gamma_t, \gamma^e_t\right) = F\left(\bar{\varphi}\left(\gamma_t, \gamma^e_t\right)\right)$$

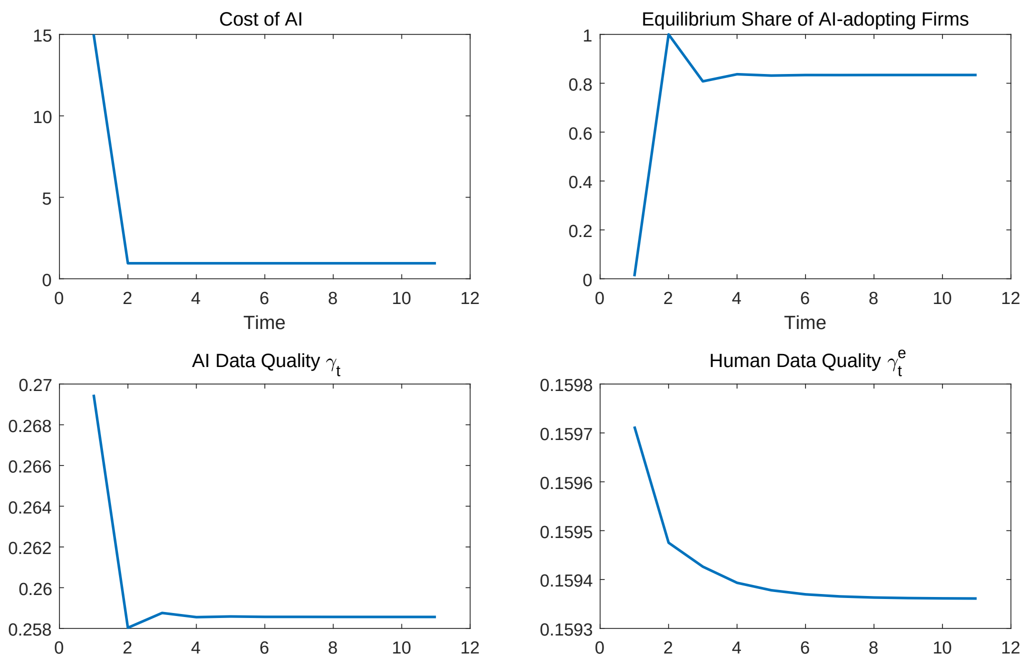

Figure 10: Model Dynamics with Low Uncertainty

and $\bar{\varphi}\left(\gamma_t, \gamma_t^e\right)$ is given by equation 18.

In contrast to our benchmark model, this system is characterized by two 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.[10]

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 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

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

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. 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.

Hui, Xiang, Oren Reshef, and Luofeng Zhou, "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," 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, "The Curse of Recursion: Training on Generated Data Makes Models Forget," 2023.

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

Yilmaz, Erdem Dogukan, Ivana Naumovska, and Vikas A. Aggarwal, "AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT," 2023.

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 $\gamma\left(\bar{\varphi}\right)$, hence the value function becomes:

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

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

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

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

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

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

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

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:

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

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

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

Hence we need the tax rate to be such that:

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

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

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

$$\tau < 0$$