Generative AI and Data Quality:

Implications for Productivity, Labor Displacement, and Policy

March 12, 2026

Abstract

Generative AI has been increasingly consuming and producing huge amounts of data. We propose a social learning model of AI, emphasizing a data-AI feedback loop: data quality affects AI productivity, which influences AI adoption and, consequently, the composition (AI vs. human-generated) and quality of future data. Calibrated to evidence on synthetic training loops, the model predicts hump-shaped labor dynamics—short-term displacement that partially reverses as data quality deteriorates. In a competitive market, AI should be taxed to correct the data-quality externality; a concentrated AI industry over-corrects, making a subsidy optimal.

1 Introduction

This paper explores the economic implications of the rise of generative artificial intelligence (AI). Modern AI has achieved remarkable success by learning from vast collections of human-created data. Large language models derive their understanding of language structure and meaning from billions of pages of human-written texts—from books and articles to online content. AI image generators have learned to recognize objects, textures, and scenes by training on enormous databases of photographs. This foundation of high-quality human-created data has been crucial for AI to generate relevant and accurate outputs for real-world applications (Liu et al., 2024).

This situation is changing. The rapid adoption of generative AI has led to an unprecedented surge in AI-generated content. The European Union Law Enforcement Agency projects that AI-generated content could constitute up to 90% of online content in the long term.[1] Rio-Chanona et al. (2023) shows that the introduction of large language models like ChatGPT has reduced human-generated content on platforms like Stack Overflow, while Brooks et al. (2024) documents an increasing presence of AI-generated content on Wikipedia, often of inferior quality.

This trend suggests that future AI will likely be trained on a mixture of AI-created and human-created content—if human-created content remains available at all.[2] This has raised concerns among both AI researchers and the broader community. Shumailov et al. (2024) and many other computer science studies (summarized in Table 1[3]) have documented "model collapse"—a phenomenon where the quality of AI-generated content progressively deteriorates when models are recursively trained on AI-generated data.[4] These academic endeavors have attracted attention from major media outlets such as Forbes[5], The New York Times[6], and the Harvard Business Review[7], all of which have reported extensively on the risks of relying too heavily on AI-generated

[1] The report is available here.

[2] This shift is already underway: Amazon uses synthetic, AI-generated data to train its Alexa language assistant, while Apple has trained its MM1 model using data generated from GPT-4 and LLaMA. The synthetic data industry is experiencing rapid growth, with market projections showing an expansion from 381.3 million in 2022 to 2.1 billion by 2028, representing an annual growth rate of 33.1%.

[3] The table only covers works between May 2023 and July 2024. However, the literature is experiencing rapid growth, and newer publications beyond this period are not included. Interested readers are encouraged to explore the subsequent developments independently.

[4] These academic findings are consistent with industry observations, such as the 2023 GitClear report that identifies a "downward pressure" on the quality of AI-generated code used with GitHub Copilot. The report is available here.

[5] https://www.forbes.com/sites/bernardmarr/2024/08/19/why-ai-models-are-collapsing-and-what-it-means-for-the-future-of-technology/

[6] https://www.nytimes.com/2024/07/19/technology/ai-data-restrictions.html

[7] https://hbr.org/2023/11/has-generative-ai-peaked

data.

This paper studies how generative AI changes the joint evolution of productivity, labor demand, and data quality. The key departure from the existing automation literature is that AI is not only a production technology; it also changes the data environment on which future AI relies (Aghion et al., 2019; Korinek and Suh, 2024; Acemoglu, 2024). We build a social-learning model with endogenous AI adoption in which firms choose between AI and human labor, and those production choices generate data of different informational quality. Here, quality refers to how much *new* information a data point adds for predicting future fundamentals. AI has a productive advantage because it can draw on the full historical dataset, but that same reliance on historical data can make its actions less informative at the margin. The composition of production between AI and humans therefore affects future data quality, future AI productivity, and future adoption. This is the AI-data feedback loop at the core of the paper.

The central theoretical insight is that access to big data can make AI *more productive today* while making its incremental data contribution *less valuable for future learning*. Big data is therefore a double-edged sword. It improves current performance, but it also makes AI's actions more likely to reproduce existing patterns rather than reveal new information. Humans face the opposite tradeoff: they are often less productive because they cannot consult the full historical record, but they rely more on their own signals and experimentation, which can add more novel information to the aggregate dataset.

Because current adoption changes the future data environment, the model features multiplicity, stability, and cyclicality. Which regime emerges depends on the relative quality of AI- and human-generated data. When AI data is much better than human data, adoption can be self-reinforcing and multiple steady states arise. When AI data is much worse, the feedback loop works in the opposite direction: high data quality raises adoption today, but the resulting AI-generated data lowers data quality tomorrow, generating cycles. When the two sources are similar, the model converges to a unique stable steady state.

To evaluate how the rise of AI can affect the economy, we model a one-time permanent reduction in AI costs in two scenarios. Our first experiment, the "AlphaGo experiment," explores what happens when AI-generated data is far superior in quality to human-generated data, allowing AI models to improve through "self-play" (learning from their own generated data). In this case, even

a small reduction in AI costs can trigger a dramatic shift from zero to complete AI adoption. This is because of the self-reinforcement loop that more AI adoption leads to higher data quality, which leads to even more AI adoption. This transition features a rapid increase in AI-generated data and massive improvements in AI productivity, consistent with the real-world evolution from the original AlphaGo, which relied on human data, to AlphaGo Zero, which learned entirely through self-play (Silver et al., 2017; Holcomb et al., 2018). The key point of this experiment is that high-quality data generated by AI can lead to explosive growth of AI and economy, even if the model abstracts away from forces such as exogenous productivity growth or capital accumulation.

Our second experiment, the "ChatGPT experiment," examines the current rise of generative AI. To quantify the relative quality of AI versus human-generated data, we draw on recent computer science research (summarized in Table 1). These studies evaluate how foundational AI models—including large language models and image generation tools—perform when trained on AI-generated data. They consistently find quality deterioration when AI models are recursively trained on AI-generated data. Based on this research, our calibration suggests that AI generates new data (or ideas) at only 8% of the quality of human-generated data. This substantial gap is identified with a sharp decline in AI quality when synthetic, AI-generated data enters the training process of AI (Chen et al. 2023; Alemohammad et al. 2023; Martínez et al. 2023). Our estimation indicates that the rise of AI-generated data imposes significant risks of corrupting future datasets in those tasks undertaken by generative AI.

Given this risk of data corruption, our model suggests that concerns about labor displacement by AI may not be as severe as commonly feared. To illustrate this, we simulate model dynamics following a one-time stylized cost reduction shock which causes AI adoption rate to spike to 100%, completely displacing human labor in the short term. However, as more AI-generated data accumulates, its lower quality gradually degrades the training dataset and thus AI productivity. Over the long run, labor displacement is partially reversed due to the declined AI productivity and the economy stabilizes with AI handling about 70% of total production. This suggests that data quality concerns can reverse roughly 30% of the short-run labor displacement over time.

The fact that labor displacement is less severe in the long run does not remove the case for policy intervention. The core inefficiency is a data externality: private adopters do not internalize how their production choices affect the common data pool used by future AI. Our analysis

therefore focuses on two policy margins, AI adoption and AI's access to historical data, both of which are central to current policy debates. The European Union Artificial Intelligence Act and Canada's Artificial Intelligence and Data Act (AIDA), for example, both address deployment and data governance. In the competitive benchmark, our model shows that when AI-generated data lowers overall data quality, taxing AI adoption can help correct the negative externality.[8]

In addition to AI adoption, the planner can also consider restricting AI's access to historical datasets. We analyze this data-access policy and show that limiting data access can improve welfare by forcing AI to generate more novel information, with the case for such restrictions strengthening as data corruption becomes more severe.

That competitive benchmark is only part of the story, because the AI industry is highly concentrated in practice. We therefore endogenize the AI price by modeling competitive, oligopolistic, and monopolistic market structures. A forward-looking monopolist internalizes the effect of current pricing on future data quality, creating a novel interaction between market power and the data externality. Quantitatively, the monopolist's markup over-corrects the externality by restricting AI adoption below the socially efficient level. This reverses the direction of optimal policy: relative to the competitive equilibrium, AI should be taxed; relative to the monopoly equilibrium, it should be subsidized. We characterize the optimal subsidy as a general-equilibrium fixed-point problem, show that it is front-loaded during the transition, and extend the welfare analysis to a dynamic social planner.

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); Cong et al. (2021); Abis and Veldkamp (2021); Cong et al. (2022); Farboodi and Veldkamp (2022). The paper proposes a novel modeling approach 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 automation on labor, as explored by Acemoglu and Restrepo (2018, 2019, 2022); Acemoglu et al. (2022); Moll et al. (2022); Alonso et al. (2022); Jones (2023); Korinek and Suh (2024) and Acemoglu (2024). In this context, the paper considers the endogeneity of AI productivity with respect to data quality. The novel AI-data

[8] If greater AI adoption improves overall data quality, our model suggests a subsidy to AI adoption.

feedback loop allows the model to deliver new implications for labor displacement effects across various tasks with different levels of data quality. It is also the first paper that introduces direct experimental evidence from the computer science literature to discipline an economic model of AI.

More broadly, two features distinguish generative AI from traditional learning models or human experts and make the data-quality feedback loop particularly relevant. First, foundation models are trained on internet-scale datasets—orders of magnitude more historical data than any human expert or conventional model would access.[9] Because generative AI draws on such a large pool of historical data, its actions largely replicate existing patterns rather than contributing novel information, activating the data corruption channel. Second, generative AI produces content—text, images, code—at unprecedented scale, and this output enters future training datasets, closing the feedback loop.

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 better data quality. This paper further argues that the *composition* of data from different sources, i.e. the different types of activities conducted by AI vs. human, matters for aggregate data quality.

The rest of the paper proceeds as follows. Section 2 presents the model and Section 3 characterizes the AI-data feedback loop. Section 4 disciplines the quantitative mechanism using evidence from the computer science literature on recursive training. Section 5 endogenizes AI supply, Section 6 studies the planner's allocation and policy implementation, and Section 7 extends the baseline to human-AI collaboration.

[9]This large information gap is central to our mechanism: as we show formally in Appendix B, when entrepreneurs also observe a large fraction of historical data, the heterogeneity that drives the model's dynamics vanishes.

2 Model

2.1 Primitives, Technology, and Friction

Time is assumed to be infinite and runs from $-\infty$ to $+\infty$. The economy is populated by $\bar{N}$ entrepreneurs and a representative household. Each entrepreneur has a lifespan of 1 period and is replaced by a newborn entrepreneur at the start of the next period.[10] These entrepreneurs have the ability to produce a good of quality A_t^i, which they can sell to the 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 $p = 1$.

The quality of production A_t^i is endogenous, and is influenced by entrepreneurs' action a_t^i. Specifically, we adopt the following functional form as in Morris and Shin (2002) and Farboodi and Veldkamp (2022):[11]

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

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

The variable θ_t represents the underlying economic fundamental of this economy. It is assumed 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 participants. This information friction is the only friction in the model. Absent this friction, entrepreneurs

[10] The assumption of short-lived entrepreneurs is made to maintain analytical tractability and to keep the model focused on the aggregate data-quality feedback loop, which is the mechanism that generates the policy-relevant externality. Importantly, even with long-lived entrepreneurs, as long as each entrepreneur is atomistic (measure zero), no individual action has a perceptible effect on aggregate data quality γ_t. Each entrepreneur therefore treats γ_t as exogenous regardless of lifespan, and the data-quality externality persists. Extending the framework to long-lived entrepreneurs would introduce idiosyncratic information heterogeneity as agents accumulate different private histories, but it would not eliminate the externality that is the paper's central concern. In Section 6.3, we introduce a discount factor β that captures the degree to which longer-lived agents would internalize the future consequences of current AI adoption, without these complications.

[11] 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.

would always take actions aligned with the fundamental and the equilibrium achieves the first best allocation. θ_t can be interpreted as reflecting general economic conditions, market demand, or some optimal approach to research and development that is not easily available to individuals and requires efforts to obtain valuable information. The task of the AI in the model would be to predict the value of θ_t, given available data, and take an action accordingly. Because θ_t is time-varying, even if AI can have access to potentially unlimited amount of data from history, old data becomes obsolete and less useful over time. This constrains AI's ability to predict θ_t accurately.

2.2 Timing, Information, and the Firm's Problem

Each entrepreneur upon birth can choose to produce with their own labor or with an AI technology. Labor production would involve an idiosyncratic labor cost of φ_t^i while production with AI incurs a capital cost of R_t. Besides the cost differences, labor and AI production also differ in the way they predict the economic fundamental and, hence, take the actions.

Specifically, 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 available to 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 aggregate uncertainty.

Labor Production: If the entrepreneurs choose to use their own labor, a private signal is generated:

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

The precision of this private signal is denoted as γ_L. The signal can be thought of as generated by information acquisition of the entrepreneur or obtained through "learning-by-doing" when the entrepreneur is conducting the task with their own labor. Based on this signal, the entrepreneur can produce the good using human labor at the cost of φ_t^i. One can interpret the cost as encompassing both the expenses associated with information acquisition and the production cost of labor. The

entrepreneurs' optimal actions are 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 payoff from labor production, given the cost, 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 entrepreneurs are utilizing their own labor for production.

AI production: We now explore the case of entrepreneurs who adopt AI in their production processes. The adoption of AI incurs a cost, denoted as R_t. 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 choose the action? The AI has access to "big data," which is an extensive dataset containing all past public signals and past actions taken by entrepreneurs. Mathematically, this can be represented as the following information set:

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

Here, S_{t-j} denotes past public signals in period $t-j$, and D_{t-j}^i represents data points regarding past actions taken by entrepreneur i in period $t-j$. The process of how this data is generated will be discussed in the following section.[12]

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

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

The use of large datasets integrating 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, novel information and act upon it. This captures AI's "creativity" or its ability to "innovate", in addition to

[12] In the benchmark model we assume that human labor cannot observe any of the historical dataset. In appendix B, we consider an extension where we allow labor to observe a fraction of the past historical dataset and we show that our key result is robust to this extension as long as AI possess an information advantage (i.e. observing more in the big dataset) relative to human.

replicating what people did in the past. The quality of such new information 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 information and not on historical past data.

It is important to note that the emphasis on prediction versus new information generation may vary across different tasks. In certain areas such as chess and Go-playing, where the reward objective and the environment are relatively straightforward and deterministic, or in research settings where AI can explore novel drug combinations in high-dimensional spaces, the generation of new information by AI could be highly valuable. For instance, the AlphaGo software is able to significantly improve upon itself through self-play (Silver et al., 2017). Conversely, in a market context where AI is used to estimate demand of a product in some local markets, 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 models and image generation AI tools to assess the relative importance of both dimensions in those tasks.

Based on both sources of data, the AI picks an action to maximize the value of 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:

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

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 $\bar{N} - N_t$. After production, the entrepreneurs exit the economy and are replaced by a new group of entrepreneurs in the following period. The timeline is summarized as in Figure 1. Note that the only inter-period linkage is informational: that past actions become the data that feed into future versions of AI.

Remark on AI-Labor complementarity/substitutability: The relationship between labor and AI (or more generally, automation) has been an important topic in recent economic research, with

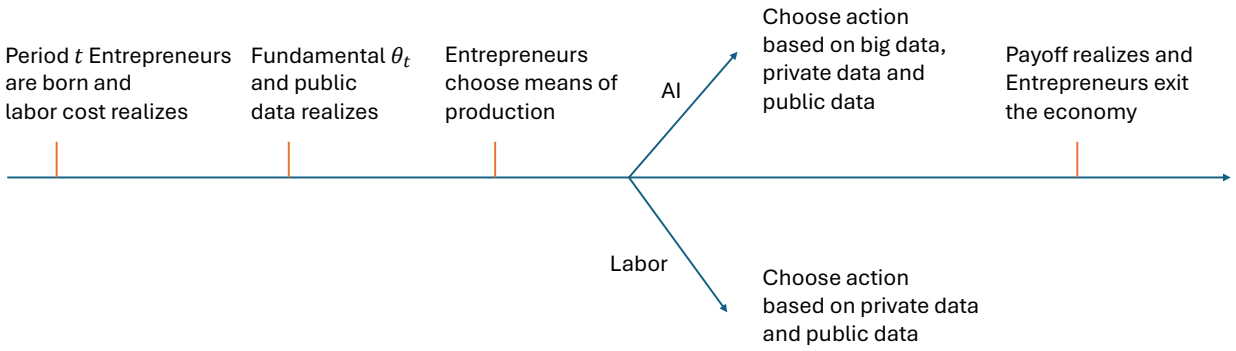

Figure 1: Timeline within a Period

studies suggesting that AI can act as either a complement or a substitute for labor, depending on the nature of tasks. AI often complements non-routine cognitive tasks and creative work, while substituting for routine and repetitive tasks that are easier to automate. To focus on the AI-data tradeoff, this paper adopts an agnostic approach on this issue, assuming that an individual's payoff function does not depend on the AI adoption decisions of others. As a result, the model abstracts from *action* complementarity or substitutability, to focus on the interactions from *information* linkages across agents. Extending the model to integrate rich strategic responses to AI adoption would be a fruitful avenue for future research.

2.3 Data Generation Process

We 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.[13]

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) and is consistent with real-world observations where companies consider their data as valuable assets and refrain from sharing it with other entities. 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 private data held by Facebook.[14]

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

We now define an equilibrium. The information equilibrium is competitive in the sense that each agent is atomic and takes the evolution of aggregate data as given. The data then evolves consistent with each agent's choices and actions.

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*

[13] 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.

[14] In the model we do not allow agents to trade their private information or sell their data in a marketplace. Since data quality is heterogeneous and perhaps hard to observe in reality, a data market would likely face severe adverse selection problems, potentially leading to market failure as in Akerlof (1970). We therefore assume away such a market in the model. Exploring frictional data markets could be a promising direction for future research, see Farboodi and Veldkamp (2022) for an example of frictionless data market.

$\{\Omega_t\}_{t=-\infty,\ldots,+\infty}$ *such that*

1. $\left\{\mathbb{1}^i_{At}, a^i_t\right\}^{i=1,2,\ldots\bar{N}}_{t=-\infty,\ldots,+\infty}$ *solves individual entrepreneur's problem given the sequence of beliefs.*

2. $\{N_t\}_{t=-\infty,\ldots,+\infty}$ *is given by* $N_t = \sum_{i=1}^{\bar{N}} \mathbb{1}^i_{At}, \forall t$.

3. $\left\{X^l_t, X^A_t\right\}_{t=-\infty,\ldots,+\infty}$ *are given by:*

$$X^l_t = \frac{1}{N_t} \sum_{i \leq N_t} \left(a^i_t + \varepsilon^{Di}_t\right)$$

$$X^A_t = \frac{1}{\bar{N} - N_t} \sum_{i > N_t} \left(a^i_t + \varepsilon^{Di}_t\right)$$

4. *Information evolves according to:*

$$\Omega_t = \{S_{t-j}, X^l_{t-j}, X^A_{t-j}\}_{j=0,1,2,3,\ldots,-\infty}, \forall t$$

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. The share of labor-adopting entrepreneurs converges to the function $F(\bar{\varphi}_t)$.

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

Here, $\bar{\varphi}_t$ represents the threshold labor cost above which an entrepreneur would choose to adopt AI.

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}:

$$\gamma_t = \frac{1}{Var\left(\theta_t | \Omega_{t-1}\right)} \tag{10}$$

It is worth noting that due to the linear quadratic form of the entrepreneurs' objective functions, the only state variable that needs to be tracked is the precision γ_t, 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:

$$\Omega_t = \Omega_{t-1} \cup \{S_t, X_t^l, X_t^A\}. \tag{11}$$

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 (Optimal Actions and Endogenous Weight Decay) *The optimal action for human labor depends on the public data and the private data:*

$$a_t^i = E(\theta_t | S_t, s_t^{Li}) = \frac{\gamma_S}{\gamma_S + \gamma_L} S_t + \underbrace{\frac{\gamma_L}{\gamma_S + \gamma_L} s_t^{Li}}_{\text{Information content in labor action}}. \tag{12}$$

The optimal decision for AI depends on the public data, historical data, and private data generated by AI:

$$a_t^i = E(\theta_t | \Omega_{t-1}, S_t, s_t^{Ai}) = \frac{\gamma_t}{\gamma_S + \gamma_A + \gamma_t} \mu_t + \frac{\gamma_S}{\gamma_S + \gamma_A + \gamma_t} S_t + \underbrace{\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t} s_t^{Ai}}_{\text{Information content in AI action}}. \tag{13}$$

where γ_t is given by equation 10 and $\mu_t = E(\theta_t | \Omega_{t-1})$ is the conditional mean of θ_t given historical dataset.

Moreover, at a steady state ($\gamma_t = \gamma^*, n_t = n^*$), the AI's optimal action (equation 13) can be expanded as:

$$a_t^{Ai} = \underbrace{\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma^*} s_t^{Ai}}_{\text{private signal}} + \underbrace{\frac{\gamma_S}{\gamma_S + \gamma_A + \gamma^*} S_t}_{\text{current public signal}} + \underbrace{\frac{\rho\, \gamma^*}{\gamma_S + \gamma_A + \gamma^*} \sum_{\tau=1}^{\infty} \lambda^{\tau-1} \left[\omega_S S_{t-\tau} + \omega_l X_{t-\tau}^l + \omega_A X_{t-\tau}^A\right]}_{\text{historical data with exponential decay}}, \tag{14}$$

where $\omega_S, \omega_l, \omega_A$ are the steady-state Kalman innovation weights on the public signal, labor data, and AI data, respectively, and the weight on each historical signal decays at the geometric rate $\lambda = \rho\alpha \in (0,1)$ per period. Here ρ is the persistence of the fundamental and α is the Kalman smoothing weight (the fraction of information carried forward from the prior rather than obtained from new data).

The optimal actions are a weighted average of those information perceived by human and AI, where the weights are functions of the different qualities (precision) of those information. For

labor, this means taking a weighted average of public data S_t and labor-generated data s_t^{Li}. As the quality of labor-generated data increases, the action becomes more sensitive to it. For AI, its optimal action is a weighted average of public information, conditional mean from the historical data, and some private information. The higher the quality of historical data (denoted by γ_t), the more weight AI assigns to it, and consequently, less weight is placed on other sources of data, including AI's own privately generated data S_t^{Ai}. Note that, because both the public signal and historical data are already part of the big dataset, the only valuable source of variation for both human an AI comes from its private information (highlighted in red).

Because of its access to historical data (γ_t), AI's action generally depends less on its privately generated information than labor's action does:

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

as long as γ_A is not significantly greater than γ_L. While having access to historical data strengthens AI's ability to make "better" decisions, it also means that AI's actions are more reflective of past events (already recorded in the datasets) and less influenced by the newly-generated information AI could contribute to aggregate data quality.

An important feature of the model is that the AR(1) structure of fundamentals (equation 2) endogenously generates exponential weight decay on older information—mirroring the weight-decay and lookback-window techniques widely used in machine learning practice.[15] Equation (14) in Lemma 3.1 formalizes this: the weight on each historical signal decays at the geometric rate $\lambda = \rho\alpha \in (0, 1)$ per period.

The decay rate $\lambda = \rho\alpha$ has an intuitive decomposition: ρ captures the mean-reversion of fundamentals (older states are less relevant for predicting θ_t), while $\alpha < 1$ reflects the fact that each period's new signals partially supersede the information in the prior, further reducing the weight on older data.

The first term in equation (14), $\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma^*} s_t^{Ai}$, captures genuinely new information that the AI generates independently of historical data. This component is analogous to the novel insights produced through processes such as reinforcement learning or fine-tuning on proprietary data, where

[15] In machine learning, weight decay refers to downweighting or discarding older training data so that the model adapts to recent patterns (Krogh and Hertz, 1991; Goodfellow et al., 2016). Common implementations include exponentially decaying learning rates and finite lookback windows; see Loshchilov and Hutter (2019) for a widely-adopted modern variant.

AI discovers patterns not previously recorded in existing datasets. Crucially, this is the only term in the AI's action that contributes new information to the aggregate dataset (Lemma 3.3). Equation (14) thus embeds an endogenous exploration–exploitation tradeoff: the private-signal term represents *exploration* (novel information), while the historical-data term represents *exploitation* (leveraging accumulated knowledge). Data quality γ_t governs the balance—when γ_t is high, AI shifts weight toward the historical dataset (exploitation); when γ_t is low, AI relies more on its private signal (exploration).

Note that, in contrast, human labor's action (equation 12) depends only on the current-period signals S_t and s_t^{Li}—it involves no historical data component and hence no lookback window. The exponential decay structure in equation (14) is therefore specific to the AI's decision rule, arising precisely because AI, but not human labor, has access to the historical dataset Ω_{t-1}.

This distinction has two implications. First, the exponential decay is precisely the condition that preserves the Markov structure of the model: because past signals in the AI's information set decay geometrically, the entire history of beliefs can be summarized by the single state variable γ_t. This provides a tight conceptual bridge between the weight-decay heuristic in machine learning and the Bayesian information structure in our model. Second, when further restrictions are imposed AI such as a finite lookback window of T periods, the law of motion for γ_t acquires an additional correction term that subtracts the decayed precision of signals dropping out of the window; this correction vanishes as $T \to \infty$ because $\rho < 1$, yielding the tractable recursive form in Lemma 3.4. We present the full finite-lookback derivation in Appendix C.

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:

$$D_t^i = a_t^i + \varepsilon_t^{Di} = \underbrace{\frac{\gamma_S}{\gamma_S + \gamma_L} S_t}_{\text{Public Info.}} + \underbrace{\frac{\gamma_L}{\gamma_S + \gamma_L} s_t^{Li} + \varepsilon_t^{Di}}_{\text{Idiosyncratic Info.}}$$

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:[16]

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

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

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

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^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 quality (or precision) of the signal is affected by the share n_t, as a larger share of labor increases the amount of information generated by labor. The term $\left(\frac{\gamma_L}{\gamma_S+\gamma_L} \right)^2$ reflects how agents' actions are weighted between the public data and the privately generated data. Since the public signal is observed by everyone, only the portion of the variation in actions that stems from the private signal (captured by $\frac{\gamma_L}{\gamma_S+\gamma_L}$, see equation 12) contributes additional information to the overall dataset.

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

$$D_t^i = a_t^i + \varepsilon_t^{Di} = \underbrace{\frac{\gamma_S}{\gamma_S + \gamma_A + \gamma_t} S_t}_{\text{Public Info.}} + \underbrace{\frac{\gamma_t}{\gamma_S + \gamma_A + \gamma_t} \mu_t}_{\text{Dataset Info.}} + \underbrace{\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t} s_t^{Ai} + \varepsilon_t^{Di}}_{\text{Idiosyncratic Info.}},$$

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 private information generated by AI s_t^{Ai} (and observed by others with a noise ε_t^{Di}). 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 biased 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 is influenced by the AI adoption share $1 - n_t$, as greater adoption leads to more AI-generated data. The middle term $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t} \right)^2$ reflects how agents' actions put

[16] 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 3).

certain weight on the public data and historical data, and those variations in actions do not add new information beyond what is already integrated in the aggregate data. Hence, the only valuable source of information one can extract from AI' action comes from the variation due to its privately-generated data, to which the AI's action optimally assigns a weight of $\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}$.

Comparing Lemma 3.2 and Lemma 3.3, it is clear that the relative quality of human- and AI-generated data depend on two aspects. First, it depends on the relative shares of labor n_t and AI adoption $1-n_t$. Second, it depends on how much weight each type assigns to privately generated data. For labor, this weight is $\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2$, while for AI, it is $\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2$. Given a positive γ_t, it is clear that AI generally assigns a lower weight compared to labor if the quality of the privately generated data is similar. This reflects the fact that AI can leverage large datasets from past actions, which dilutes the impact of the private signal on AI's actions.

Corollary 3.1 *AI-generated data contains less valuable information than human-generated data, provided the quality of AI's privately generated data is not significantly greater than that of human-generated data.*

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 11), 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, we obtain the law of motion for γ_t:

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

$$\begin{aligned}\frac{1}{\gamma_{t+1}} &= Var\left(\theta_{t+1}|\Omega_t\right) \\ &= \rho^2 \frac{1}{\underbrace{\gamma_t}_{\text{prec. of } \Omega_{t-1}} + \underbrace{\gamma_S}_{\text{prec. of } S_t} + \underbrace{n_t\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D}_{\text{prec. of } X_t^l} + \underbrace{(1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\gamma_D}_{\text{prec. of } X_t^A}} + \frac{1}{\gamma_\eta} \qquad (16)\end{aligned}$$

where the four components in the denominator capture the information content in Ω_{t-1}, S_t, X_t^l and X_t^A respectively. 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 relative quality of information generated by AI is not too much higher than that generated by labor:*

$$\frac{\gamma_A}{\gamma_L} < \frac{\gamma_S + \gamma_t}{\gamma_S}. \tag{17}$$

proof: It can be observed from equation 16 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 17.

This theorem presents the key result of the paper: with increased AI adoption, the overall quality of the data can decline, as a larger share of AI-generated data corrupts the dataset. This data corruption channel is present when the quality of AI-generated data, relative to human-generated data, is sufficiently low.

A perhaps unexpected implication of the theorem arises when the quality of private information produced by both AI and humans is identical, i.e., $\gamma_L = \gamma_A$. In this case, greater AI usage could harm future data quality, as Condition 17 holds. But why would this occur if AI and humans generate private information of equal quality? In the context of social learning, agents infer information from others' actions, which are noisy reflections of their privately-generated information, as they also respond to other data sources. Human actions are influenced by the public signal S_t, while AI actions are predominantly influenced by historical data Ω_{t-1} in addition to the public signal S_t. Since AI processes more diverse data sources, it places less emphasis on its own privately-generated data than humans do. Consequently, AI activities contribute less valuable information to the aggregate data quality compared 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 17, hence the more likely that AI can increase the quality of aggregate data.

One can also show that higher quality historical data may impede further data quality improvement. From Condition 17, it is clear that the higher the quality of historical data, (i.e., higher γ_t), the larger the value on the right-hand side of equation 17, suggesting that the adoption of AI

is more likely to impede data quality improvement. This occurs because higher γ_t values lead AI to rely more on historical data, placing less weight on its own signal generation, thus contributing less new information to aggregate data quality. In cases where prior data quality is very high, AI adoption offers great returns. In such a situation, we can expect a surge in AI adoption, which may suppress long-term data quality.

We now close the model by solving for the equilibrium value of n_t based on individual optimality conditions.

3.5 Closing the Model: Endogenizing n_t

Since each agent's optimal action equals the conditional expectation of θ_t, the expected payoffs under labor and AI production reduce to:

$$V^l\left(\varphi_t^i\right) = \bar{A} - \frac{1}{\gamma_S + \gamma_L} - \varphi_t^i, \tag{18}$$

$$V^A\left(\gamma_t\right) = \bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - R_t. \tag{19}$$

The payoff difference reflects the informational advantage of AI (access to Ω_{t-1}) against its cost R_t. Equalizing the two payoffs yields the adoption threshold:

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

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

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

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t\left(\gamma_t\right)\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + \left(1 - n_t\left(\gamma_t\right)\right)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{20}$$

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

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

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

Note that the data quality parameter γ_t appears in three different locations within the law of motion equation 20. Firstly, it embodies information passed down from previous periods. Sec-

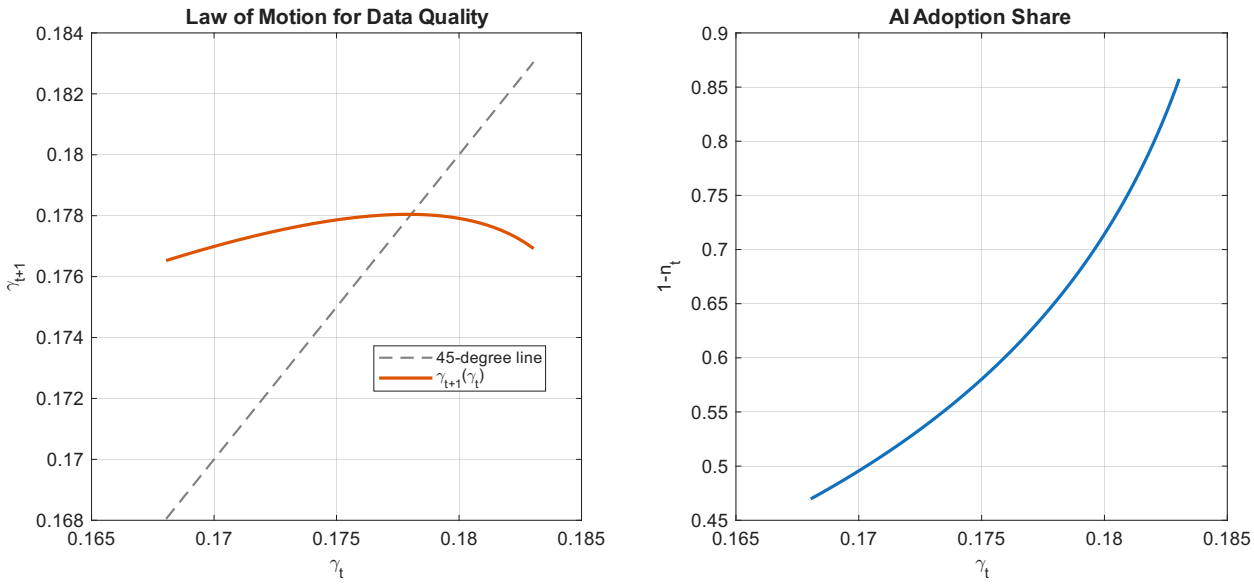

Figure 2: **Law of Motion and Policy Functions.** Left panel: the law of motion $\gamma_{t+1}(\gamma_t)$ for data quality under baseline parameters. Right panel: the AI adoption share $1 - n_t$ as a function of current data quality γ_t. Parameters: $\rho = 0.95$, $\gamma_S = 0.05$, $\gamma_D = 0.1$, $\gamma_L = 0.1$, $\gamma_A = 0.1$, $\gamma_\eta = 0.5$, $R_t = 3.68$. F follows a log-normal distribution.

ondly, 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.[17] 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 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.[18]

When $\gamma_A \to \infty$, the model simplifies to one where the relative productivity of AI and labor is constant and independent of γ_t, similar to Acemoglu and Restrepo (2018).[19]

[17] 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.

[18] 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.

[19] In this limit, AI perfectly predicts θ_t using only its private signal, achieving productivity $\bar{A}$. Labor productivity equals $\bar{A} - \frac{1}{\gamma_S+\gamma_L}$, so the relative productivity ratio $(\bar{A} - \frac{1}{\gamma_S+\gamma_L})/\bar{A}$ is a constant.

Theorem 3.3 *When historical datasets are useful: $\gamma_t > 0$ and the relative quality of AI v.s. human signal satisfies:*

$$\frac{\gamma_L - \gamma_t}{\gamma_L} < \frac{\gamma_A}{\gamma_L} < \frac{\gamma_S + \gamma_t}{\gamma_S}, \tag{21}$$

AI is more productive than humans, but its action generates data of lower quality.

The interval in Condition 21 is non-empty: the lower bound is less than 1 (and negative when $\gamma_t > \gamma_L$), while the upper bound exceeds 1. The lower bound captures that AI's access to the historical dataset Ω_{t-1} compensates for a weaker private signal, so AI can be more productive than humans even when $\gamma_A < \gamma_L$. The upper bound ensures that AI-generated data is less informative than human-generated data: because AI relies heavily on the historical dataset, its actions reveal less new information.[20] Above this upper bound, AI is both more productive and generates higher-quality data.

As the historical dataset improves (higher γ_t), the interval in Condition 21 widens: the lower bound falls and the upper bound rises. Intuitively, a richer dataset lets AI "free-ride" on accumulated knowledge—it can make accurate predictions without generating much new information of its own. This is because AI's action primarily reflects what it extracts from the historical dataset Ω_{t-1}, so observers of AI's action learn little beyond what the dataset already contains. The better the dataset, the stronger this effect: AI becomes more productive, but its actions become less informative. This tension—between AI's productivity gain and its data quality loss—is the central mechanism driving the model's dynamics.

3.6 Model Dynamics

Due to the interaction between AI adoption and data quality, the model displays rich dynamics over time. Depending on the relative quality of AI v.s. human data, we divide the analysis into the following three cases:

Case 1: high quality AI-generated data

In this case multiple steady states with different levels of aggregate data quality can arise. When data quality is low, AI is not very productive relative to humans, hence AI adoption is low

[20] It is important to distinguish between the AI signal γ_A, which measures the quality of information privately generated by AI, and AI-generated data, which is the observable information inferred from AI actions. The latter depends on how much weight AI places on its private signal versus the historical dataset.

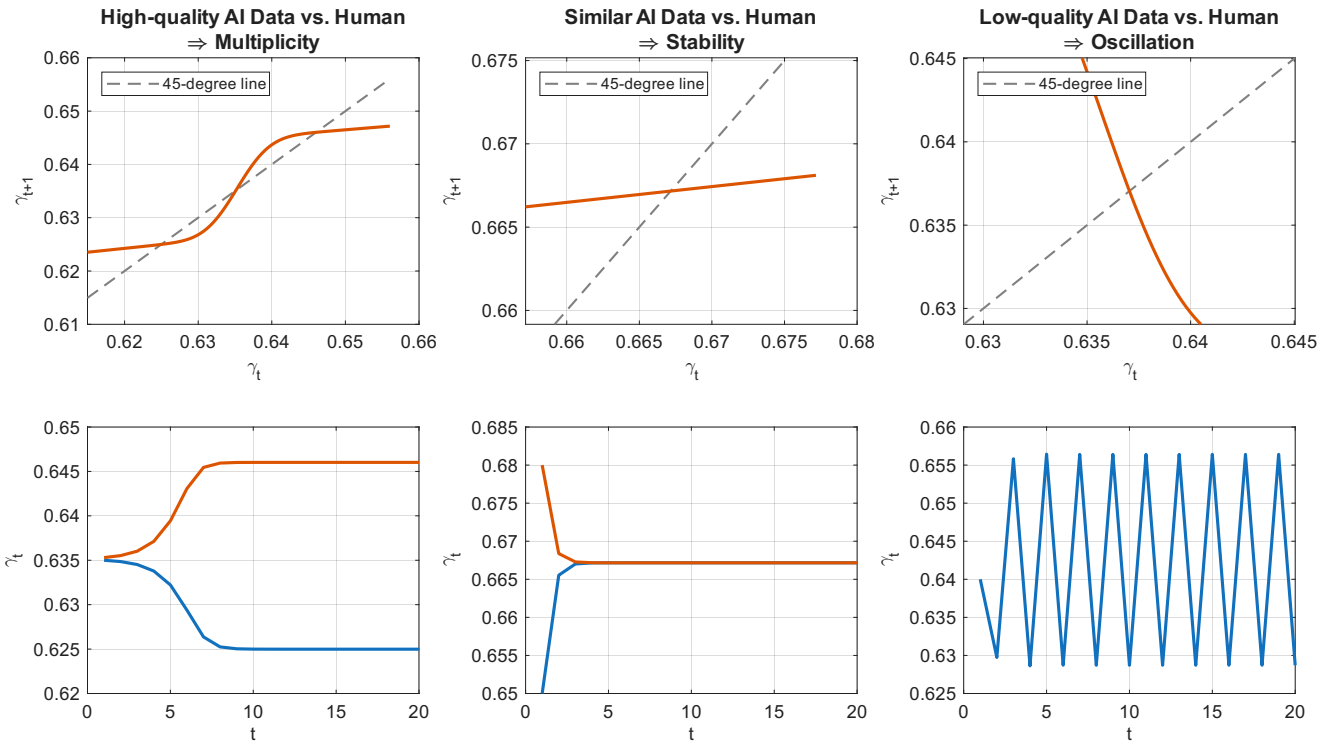

Figure 3: **Model Dynamics: Multiplicity, Stability, and Cyclicality.** Top row: law of motion $\gamma_{t+1}(\gamma_t)$ under three cases—high-quality AI data (left, $\gamma_A \gg \gamma_L$), equal quality ($\gamma_A = \gamma_L$, middle), and low-quality AI data ($\gamma_A = 0.1\gamma_L$, right). Bottom row: corresponding transition dynamics of data quality γ_t over time. Multiple steady states arise when AI data quality is high; a unique stable steady state obtains when qualities are similar; endogenous cycles emerge when AI data quality is low.

and therefore almost all new data are generated by humans and thus, by assumption, of lower quality. This confirms the low aggregate data quality in this steady state. When the initial data quality is high, on the other hand, AI productivity is high, leading to many agents to switch to AI. This creates high-quality AI-generated data, which confirms the initial high data quality in equilibrium. The law of motion for data quality is depicted in the top left panel of Figure 3.

Theorem 3.4 *Given that $\gamma_S > 0$ and $\gamma_A \to \infty$, the dynamic system in Theorem 3.2 admits at least two steady states for a sufficiently compressed labor cost distribution $F(.)$. In those two steady states, the steady state with high data quality is associated with a high AI adoption rate.*

In this case the equilibrium feedback loop between AI adoption and data quality is *strongly positive*: higher AI adoption leads to higher overall data quality, which leads to even greater share of AI adoption. This means that the derivative $\gamma_{t+1}(\gamma_t)$ has a steep and positive slope around the middle steady state (see top left panel) and hence multiple steady states can arise.

The multiplicity suggests certain fragility when AI data is of much higher quality compared to human-generated data. This is illustrated in the bottom left panel of Figure 3: small variations in the initial condition of the model can create vastly different outcome over the long run, as they may converge to different steady states.

An AlphaGo Experiment: How Explosive Growth Happens

A natural application of this situation would be where AI-generated data is far superior in quality compared to human-generated data. The game of Go would be a natural candidate as the quality of AI can improve through self-play (using data generated by playing against itself). We start by parameterizing the model. We perceive relatively little fundamental uncertainty, as the reward system and objective function of those games are relatively deterministic compared to other real-world tasks which often involve greater uncertainty. We thus set $\rho = 0.999$, and $\gamma_\eta = 10$, indicating that the fundamental (optimal solution) hardly moves over time. Further increasing the value of ρ does not affect the simulation results. We then set the quality of the public information to be relatively low to capture that little is known (a priori) about the optimal solution: $\gamma_S = 0.01$. We also assume that AI-generated data can be highly useful in this setting, so that the quality of AI-generated data is 100 times more valuable than those generated by labor: $\gamma_A = 1$ and $\gamma_L = 0.01$. Finally, we set $\gamma_D = 10$ so that it is relatively easy to record data into the

database for the training of AI.

The resulting law of motion for data quality is displayed as the red solid line in the top left panel of Figure 4. One can see that the model displays multiple steady states: a low steady state where no AI is adopted, and a high steady state with 100 percent AI adoption.

We now simulate the model assuming that the world starts at the low steady state with zero AI adoption, which captures the long-held belief before AlphaGo emerged that a computer can never defeat a human player in the game of go. We consider a shock that reduces the cost of AI (R_t) by 5% in period 20. Such a reduction in cost shifts the law of motion for data quality up from the red solid line to the yellow dashed line (see top left panel of Figure 4), and the low steady state disappears as a result. The equilibrium thus starts to transit to the high steady state. During this transition, the share of AI-generated data increases dramatically from 0 percent to 100 percent. This is consistent with the transition from AlphaGo (which still uses human inputs for training) to AlphaGo Zero, where all data input is generated by itself. As a result, aggregate data quality improves and AI productivity also increases dramatically within a relatively short time frame, which is consistent with the AlphaGo project, which was formed in 2014 and improved massively in a relatively short time, beating Lee Sedol in 2016 and world champion Ke Jie in 2017. Although framed in the context of Go-playing, the numerical example aims to capture something more general: rapid economic growth and productivity improvement where AI can "self-improve" through generating and consuming high-quality data. The explosive growth is similar to some types of the transformative technological progress such as those discussed in Trammell and Korinek (2023).

Case 2: Similar quality AI-generated data compared to human-generated data

In this case it is assumed that $\gamma_A = \gamma_L$. This implies that shifts in the share of AI adoption would not have a significant impact on overall data quality because the loadings on new information are not so much different across humans and AI.[21] In such a case the equilibrium feedback loop between AI adoption and data quality is weak, and does not significantly affect the aggregate law of motion for γ_t. Hence the model displays stability with global convergence to a unique steady state, as in a standard model of Bayesian learning. Results are displayed in the middle panel of Figure 3.

[21] The loadings of actions on private information won't be exactly the same between AI and human, but they are not so much different in this case.

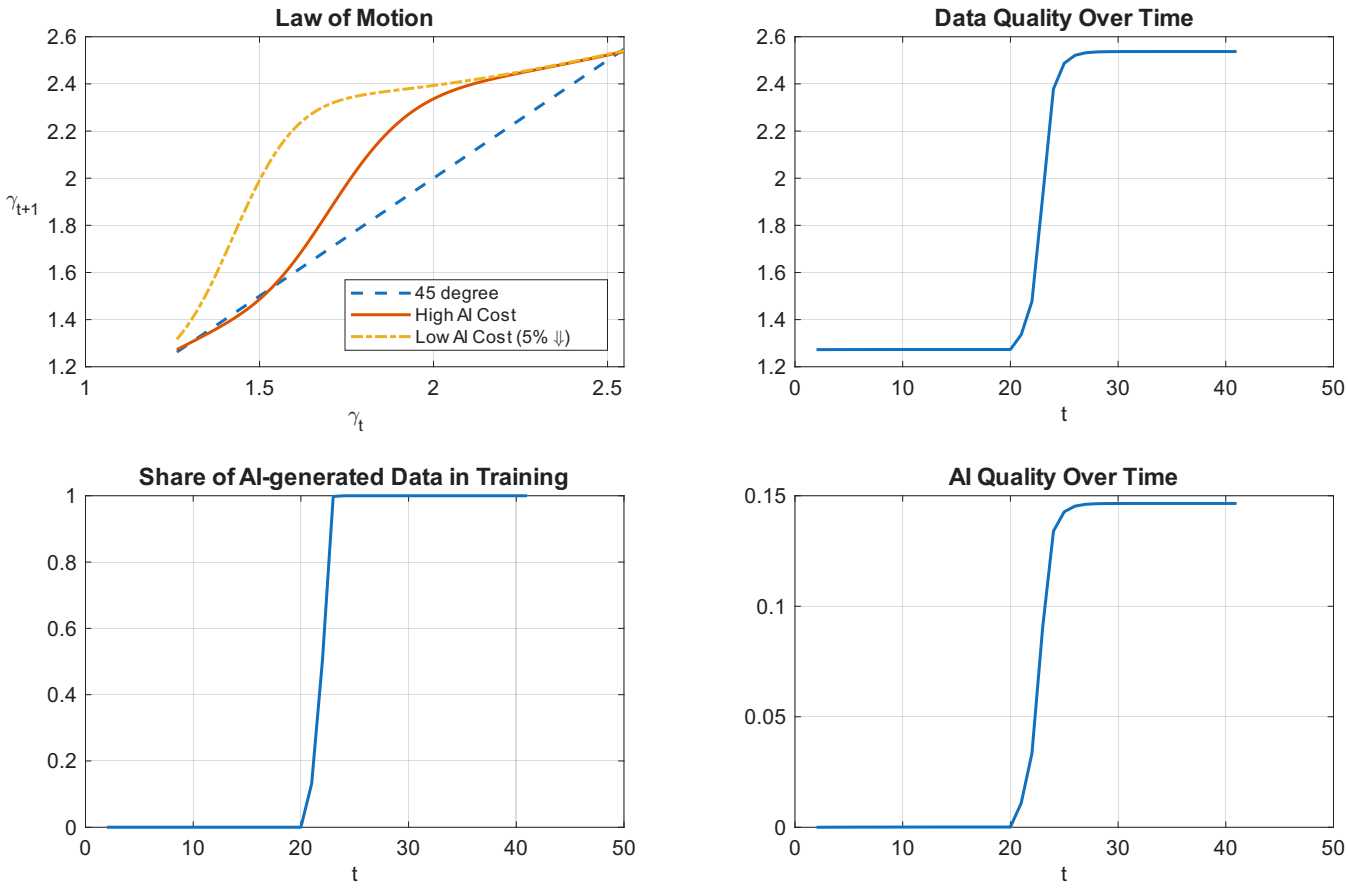

Figure 4: **Model Dynamics: The AlphaGo Experiment.** A 5% reduction in AI cost at $t = 20$ triggers a transition from zero to full AI adoption. Top-left: the law of motion shifts from the solid red line (pre-shock) to the dashed yellow line (post-shock), eliminating the low steady state. Top-right: AI adoption jumps from 0% to 100%. Bottom panels: data quality and AI productivity increase sharply. Parameters: $\rho = 0.999$, $\gamma_S = 0.01$, $\gamma_A = 1$, $\gamma_L = 0.01$, $\gamma_D = 10$, $\gamma_\eta = 10$.

Case 3: Low quality AI-generated data

In this case we set γ_A to be 10 percent of γ_L, suggesting that AI-generated data is of lower quality than labor. In this case, endogenous cycles can arise: higher data quality today leads to higher AI productivity and hence more AI adoption. This creates more low-quality AI-generated data and leads to lower overall data quality in the future. Thus the aggregate law of motion for γ_t can take a negative slope (top right panel of Figure 3).

One can formally see the negative slope by examining the dynamic equation in Theorem 3.1 in the special case where $\gamma_A \to 0$. In this case, AI is a pure machine learning tool and thus AI-generated data does not contribute to improvements in overall data quality. The law of motion for data becomes:

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t \uparrow + \gamma_s + n_t \downarrow (\gamma_t) \left(\frac{\gamma_L}{\gamma_S + \gamma_L}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}$$

An increase in γ_t has a direct positive impact on γ_{t+1} but a negative impact through n_t because there is less fresh data generated through human production. The negative impact overpowers the positive direct effect if the derivative of $n'(.)$ is sufficiently large. This is guaranteed by a sufficiently compressed distribution of labor cost.

Theorem 3.5 *Given that $\gamma_A \to 0$, the law of motion for γ_t is downward slope at the steady state for a sufficiently compressed labor cost distribution.*

The implication of the downward-sloping law of motion is that it leads to endogenous cycles in both data quality and AI adoption. This is shown in the bottom right panel of Figure 3, where the economy bounces around situations with high data quality and low data quality. The AI adoption rate would also exhibit similar oscillation as the AI adoption rate is a monotonic function of overall data quality γ_t.

The model can deliver rich dynamics depending on the quality of AI-generated data compared to human data.[22] We now attempt to quantify this critical aspect of the model using experimental evidence from the computer science literature.

[22]The sensitivity of adoption to data quality is amplified if production costs depend on task complexity. Suppose $C_{it}^H = \varphi_{it} + \kappa_H \mathbb{E}[(\theta_t - a_{it}^H)^2]$ and $C_t^A = R_t + \kappa_A \mathbb{E}[(\theta_t - a_{it}^A)^2]$, where $\kappa_H, \kappa_A \geq 0$. The adoption cutoff becomes $\bar{\varphi}_t = R_t + \frac{1+\kappa_A}{\gamma_S+\gamma_t+\gamma_A} - \frac{1+\kappa_H}{\gamma_S+\gamma_L}$. When $\kappa_A > \kappa_H$—reflecting that AI excels at information-intensive tasks—the derivative $\partial\bar{\varphi}_t/\partial\gamma_t$ is larger in magnitude, so $n_t(\gamma_t)$ responds more strongly to changes in data quality. This amplifies both the boom phase (more aggressive AI adoption when γ_t is high) and the correction phase (sharper pullback when γ_t falls), but preserves the qualitative structure of the model.

4 Quantification of Data Quality in Generative Tasks

We now quantify the model in the context of tasks conducted by the current version of generative AI. Those tasks include producing a variety of novel content, such as images, video, music, speech, text, software code and product designs. The key aspect of the calibration is to discipline the relative quality of information generated by AI vs. human: $\frac{\gamma_A}{\gamma_L}$ in those tasks, which will play a crucial role in determining model dynamics. We will now discuss the evidence to inform this aspect of the model.

4.1 FID Score and Fully 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 practice, 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. The FID score can be thought of as a loss function—where a higher score reflects lower AI quality.

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.

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

Figure 5: **Fully Synthetic Loops.** Illustration of the iterative training process: an initial AI model is trained on real (human-generated) data, and each subsequent generation is trained exclusively on synthetic data produced by the preceding generation. This methodology is used in the computer science literature to measure data quality degradation.

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 "fully synthetic loops." This process, illustrated in Figure 5, 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)) note a rise of over 700%, and Bohacek and Farid (2023, figure 2) find an approximately 300% increase throughout full synthetic loops in the most recent study.

We 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 6 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 deterioration 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 we 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 of AI and even at this upper bound 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 the current version of generative 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

We 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 $\gamma_\eta = 1/(0.055)^2$.[23] 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 17, 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 $\gamma_A = 8$.

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

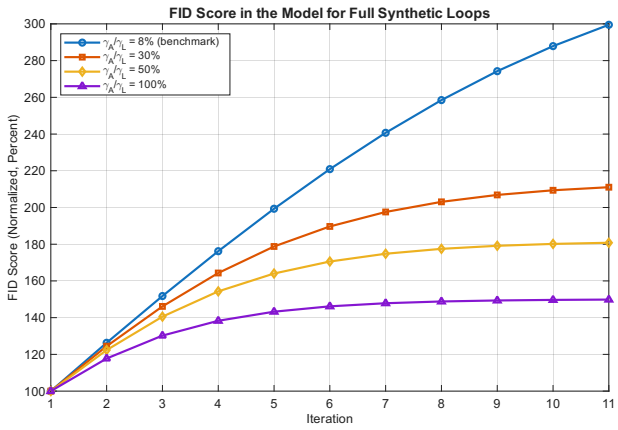

Figure 6: **FID Score in the Model for Full Synthetic Loops.** Each line corresponds to a different value of AI signal quality γ_A, showing how the FID score increases across synthetic loop iterations. Lower γ_A produces steeper FID growth, indicating faster quality degradation. The calibration matches the 300% FID increase documented by Bohacek and Farid (2023), yielding $\gamma_A = 8$ (approximately 8% of $\gamma_L = 100$).

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 5. 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 value leads us to set $\gamma_D = 35$.

Lastly, we assign $\gamma_S = 0$ due to the absence of additional targets for measuring the precision of public signals. However, in Appendix E.2, we conduct a sensitivity analysis to evaluate the impact of a higher γ_S value on our findings. We find that our results do 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

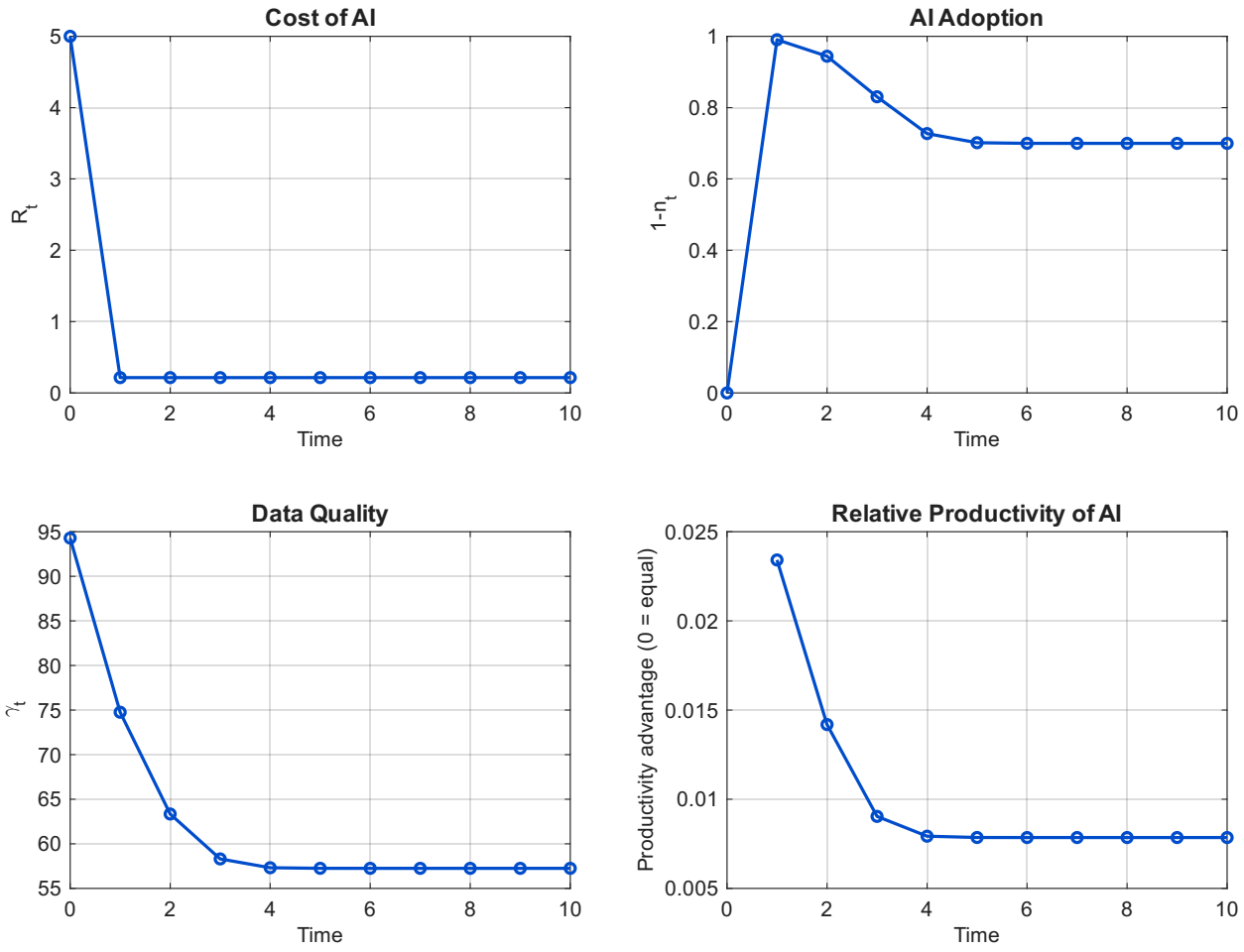

Figure 7: **Model Dynamics: The ChatGPT Experiment.** Top-left: a one-time permanent reduction in AI cost R_t. Top-right: the share of AI-adopting firms rises to nearly 100% before partially reversing to approximately 75%, reflecting data corruption. Bottom-left: aggregate data quality γ_t declines over time as AI-generated data accumulates. Bottom-right: AI's relative productivity advantage erodes. Parameters calibrated to the ChatGPT era using FID scores from the computer science literature (see Section 4).

of ChatGPT, which substantially lowers the barriers to AI accessibility and affordability.

Figure 7 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 20. 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

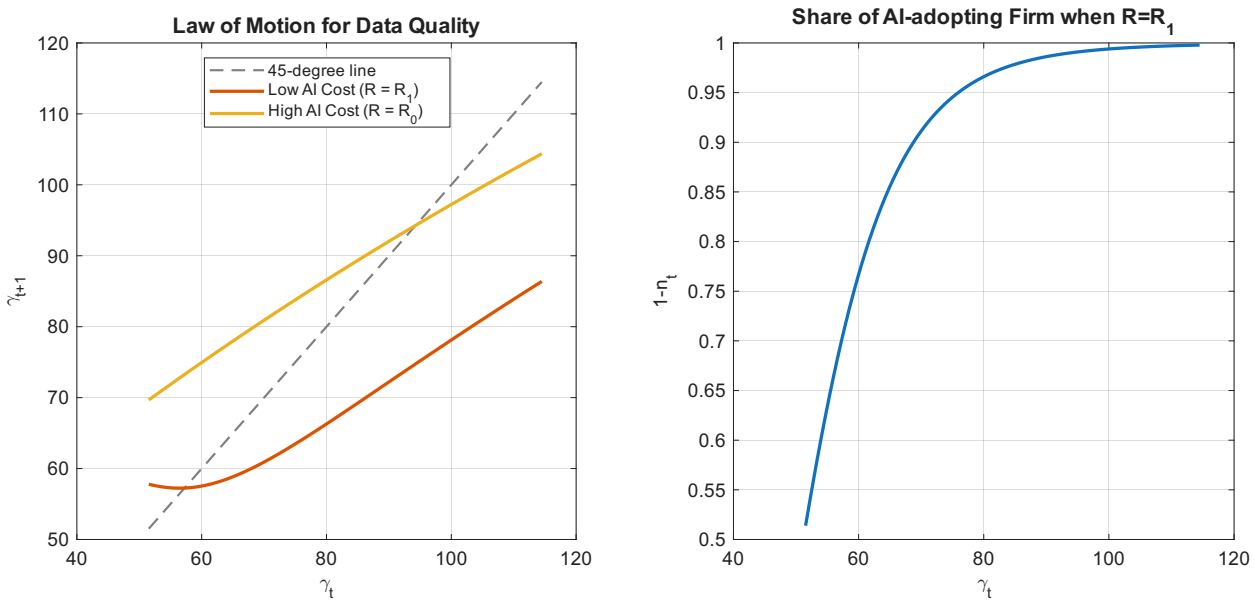

Figure 8: **Equilibrium Functions Under Varying AI Costs.** Left: the law of motion $\gamma_{t+1}(\gamma_t)$ under high AI costs (yellow) and low AI costs (red). The downward shift shows that lower AI costs reduce long-run data quality. Right: the corresponding AI adoption share. When $\gamma_A = \gamma_L$, the law of motion shifts moderately.

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.[24] 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 8 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 reduce 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.

[24]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 19 and 18. The expectation is taken over idiosyncratic labor costs ϕ^i.

We conduct two robustness checks in Appendix E.1 and E.2. First, a counterfactual experiment with equal information quality ($\gamma_A = \gamma_L$) confirms that the 30% reversal in labor displacement is entirely attributable to the data corruption channel. Second, varying the precision of the public signal γ_S does not fundamentally change the data corruption channel's magnitude once the model is recalibrated to match FID score evidence.

4.4 Review of Evidence from Computer Science

The model's key prediction—that increased AI adoption degrades data quality and reduces future AI productivity—finds substantial support in recent computer science research (see Table 1 for a summary). Multiple studies document "model collapse": when AI models are recursively trained on AI-generated data, output quality progressively deteriorates. Alemohammad et al. (2023) finds that generative image models trained on synthetic data amplify artifacts, leading to declining precision and diversity. Shumailov et al. (2024) shows that recursive training on model-generated content produces "irreversible defects," a finding that extends beyond large language models to other AI architectures. Additional studies confirm these patterns across different contexts and AI systems (Guo et al., 2023; Doshi and Hauser, 2024; Briesch et al., 2023).

There is also evidence that generative AI crowds out human-generated content. Rio-Chanona et al. (2023) finds that the launch of ChatGPT is associated with a 16% decrease in weekly contributions to Stack Overflow, while Brooks et al. (2024) documents increasing AI-generated content on Wikipedia. This crowding-out effect reinforces the data corruption channel: as AI displaces human content creators, the composition of training data shifts further toward lower-quality AI-generated content.

A caveat is that our calibration is specific to generative AI tasks, and the ratio γ_A/γ_L is likely to vary substantially across domains:

- **Image generation**: This is the domain with the strongest and most direct evidence, which our calibration directly targets. Bohacek and Farid (2023) report approximately 300% FID increases in synthetic training loops, Alemohammad et al. (2023) find over 400%, and similar patterns are confirmed in Briesch et al. (2023). Our calibration targets the most conservative of these estimates.

- **Text and natural language processing**: Shumailov et al. (2024) document "model collapse" in language models iteratively trained on their own output, with progressive loss of distributional tails and linguistic diversity. The rate of degradation varies with corpus diversity and task specificity, but the qualitative pattern of quality loss is robust across architectures (Alemohammad et al., 2023).

- **Code generation**: The theoretical mechanism of model collapse applies in principle (Shumailov et al., 2024; Gerstgrasser et al., 2024), and industry evidence already points in this direction: Harding (2024) documents that code churn has doubled and refactoring rates have halved since the widespread adoption of AI coding assistants. The degradation may be slower than in image generation due to the more structured nature of programming languages, but the feedback loop is qualitatively similar.

- **Scientific discovery and other domains**: In settings where AI explores combinatorial spaces that humans cannot efficiently search—such as protein structure prediction (Jumper et al., 2021) and materials discovery (Merchant et al., 2023)—AI may contribute genuinely novel information, suggesting γ_A / γ_L could be substantially higher. Our model accommodates this through the condition in Theorem 3.1: when γ_A is sufficiently high, AI adoption improves rather than degrades data quality, and the data corruption concern does not apply.

Our results should therefore be interpreted as highlighting an interdisciplinary approach for combining economic models with computer science evidence in the context of generative content production, rather than as universally applicable across all AI applications. Moreover, recent advances in reasoning-capable AI systems—such as OpenAI's o-series models and DeepSeek-R1, which employ chain-of-thought reasoning and reinforcement learning—suggest that γ_A may be rising over time as AI develops the capacity to generate genuinely novel insights rather than merely recombining patterns from historical data. If this trend continues, the data corruption concern may eventually diminish for tasks where reasoning models are deployed. A quantitative model that jointly captures cross-domain heterogeneity in γ_A / γ_L, endogenous technological progress, and the long-run accumulation of AI-generated data is beyond the scope of this paper, but our framework provides the building blocks for such an analysis, and identifies key theoretic conditions for model dynamics.

5 The Supply Side of AI

The preceding analysis treats the cost of AI adoption R_t as an exogenous parameter. In practice, the AI industry is highly concentrated: a small number of firms control the development and deployment of frontier AI models.[25] If those providers set prices strategically, market power affects not only current adoption but also the future evolution of data quality. This section therefore endogenizes R_t and studies how competition, Cournot pricing, and forward-looking monopoly interact with the AI-data feedback loop.

5.1 Demand for AI Services

The adoption decision characterized in the previous sections implies a downward-sloping demand curve for AI services. Recall that entrepreneur i adopts AI if and only if her idiosyncratic labor cost φ_t^i exceeds the threshold $\bar{\varphi}_t$ derived in Section 3.5. The aggregate demand for AI services—the fraction of entrepreneurs who adopt AI—is therefore

$$D(R_t, \gamma_t) = 1 - n_t = 1 - F\left(R_t + \frac{1}{\gamma_S + \gamma_A + \gamma_t} - \frac{1}{\gamma_S + \gamma_L}\right). \tag{22}$$

Since F is increasing, demand is strictly decreasing in R_t: a higher AI price reduces adoption. Moreover, demand is increasing in data quality γ_t: higher data quality raises AI productivity, making AI more attractive relative to labor.

5.2 Competitive Benchmark

Under perfect competition among AI providers, the equilibrium price equals marginal cost:

$$R_t = c_t. \tag{23}$$

This is the benchmark implicitly assumed in the previous sections, where R_t was treated as exogenous. Under perfect competition, AI providers earn zero profit and all surplus accrues to entrepreneurs. The data quality dynamics are governed by the law of motion in Theorem 3.2 with the adoption share determined by $n_t = F\left(c_t + \frac{1}{\gamma_S+\gamma_A+\gamma_t} - \frac{1}{\gamma_S+\gamma_L}\right)$.

[25] As of 2024, the frontier AI market is dominated by a handful of firms—including OpenAI, Google DeepMind, Anthropic, and Meta AI—with significant barriers to entry arising from the enormous computational and data requirements of training large models.

5.3 Static Imperfect Competition

We now consider quantity (Cournot) competition among $N \geq 1$ symmetric AI providers, each producing at constant marginal cost c_t. Each firm takes the quantities of its competitors as given and chooses its own quantity to maximize current-period profit. In the symmetric Nash equilibrium, the first-order condition yields the standard Cournot markup (see Appendix A.1 for a derivation):

$$R_t - c_t = \frac{1 - F(\bar{\varphi}_t)}{N \cdot f(\bar{\varphi}_t)}, \tag{24}$$

where f denotes the density of the labor cost distribution F. The markup is the ratio of demand to the slope of demand, scaled by the number of competitors N. As $N \to \infty$, the markup vanishes and we recover the competitive benchmark $R_t = c_t$.

When $N = 1$, the market is served by a single monopolist with markup

$$R_t - c_t = \frac{1 - F(\bar{\varphi}_t)}{f(\bar{\varphi}_t)}. \tag{25}$$

The per-period profit of each firm is $\pi_t = (R_t - c_t) \cdot [1 - F(\bar{\varphi}_t)]/N$, which in the monopoly case ($N = 1$) becomes $\pi_t = (R_t - c_t) \cdot [1 - F(\bar{\varphi}_t)]$. Note that the Cournot firms in this subsection are *myopic*: they maximize current-period profit without considering how their pricing affects future data quality. We relax this assumption next.

5.4 Dynamic Monopoly

A key feature of our model is that AI adoption today affects data quality tomorrow, which in turn determines future AI productivity and demand. A forward-looking monopolist—one who recognizes this intertemporal linkage—has an incentive to manage data quality as a strategic asset. We now characterize the pricing behavior of such a monopolist.

The monopolist chooses a sequence of AI prices $\{R_t\}_{t=0}^{\infty}$ to maximize the present value of profits, taking into account that current pricing affects future data quality and hence future demand. Given data quality γ_t, the adoption threshold is

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

so that $n_t = F(\bar{\varphi}_t)$ firms use human labor and $1 - n_t$ adopt AI. Per-period monopoly profit π_t is as defined in Section 5.3. Data quality evolves according to the law of motion (20). The dynamic

monopolist's problem is:

$$M(\gamma_t) = \max_{R_t} \left\{ \pi_t + \beta\, M\big(\gamma_{t+1}\big) \right\} \tag{26}$$

subject to the law of motion for γ_{t+1} (equation 20) with $n_t = F(\bar{\varphi}_t)$, where $\beta \in [0, 1)$ is the monopolist's discount factor.

The first-order condition is:

$$\underbrace{\frac{\partial \pi_t}{\partial R_t}}_{\text{static profit}} + \beta \underbrace{M'(\gamma_{t+1}) \cdot \frac{\partial \gamma_{t+1}}{\partial R_t}}_{\text{dynamic data-quality effect}} = 0. \tag{27}$$

Expanding the static profit term:

$$\big[1 - F(\bar{\varphi}_t)\big] - (R_t - c_t)\, f(\bar{\varphi}_t) \;+\; \beta\, M'(\gamma_{t+1}) \cdot \frac{\partial \gamma_{t+1}}{\partial n_t} \cdot \big[-f(\bar{\varphi}_t)\big] = 0, \tag{28}$$

where $\frac{\partial \gamma_{t+1}}{\partial n_t}$ captures the marginal effect of shifting one firm from AI to human labor on next-period data quality. Under data corruption, this derivative is positive: more human labor improves data quality.

The first term in (27) is the standard monopoly tradeoff between higher margins and lower volume. The second term reflects the monopolist's forward-looking incentive: raising R_t reduces AI adoption, which—under data corruption—improves future data quality ($\partial \gamma_{t+1} / \partial R_t > 0$) and hence future demand and profits ($M'(\gamma_{t+1}) > 0$). Thus, the dynamic monopolist charges a *higher* price than the myopic (static) monopolist, further restricting AI adoption to protect the value of future data. This adds an aggregate-level dimension to the exploration–exploitation tradeoff identified in Section 2: while Lemma 3.1 describes how each AI agent balances its private signal (exploration) against historical data (exploitation), the dynamic monopolist manages this balance at the market level by restraining adoption today to preserve data quality—the informational commons—for the future.[26]

5.5 Comparing Market Equilibria

Given data quality γ_t and AI production cost c_t, the three market structures deliver a clear ordering. Competitive pricing ($R_t = c_t$) yields the highest AI adoption and, under data corruption, the lowest long-run data quality. The static monopolist restricts AI adoption through its markup (equation 25), raising data quality relative to the competitive benchmark. The dynamic monopo-

[26] When $\beta = 0$, the dynamic monopolist reduces to the static monopolist of Section 5.3 (with $N = 1$), and the first-order condition reduces to $\partial \pi_t / \partial R_t = 0$, which yields the static markup in equation (25).

list restricts adoption even further, reflecting the forward-looking data-quality effect captured in the second term of equation (27).

This ordering—competitive, static monopoly, dynamic monopoly—illustrates how market power interacts with the data-quality feedback loop. The monopolist's markup, while set for profit-maximizing reasons, has the side effect of shifting production toward human labor, which generates data of higher informational content. The dynamic monopolist amplifies this effect by explicitly investing in future data quality. Whether this restriction is socially excessive or insufficient depends on the social planner's optimum, which we characterize next.

6 Efficiency and Regulation of AI

Section 5 showed that market structure changes AI adoption and may partially internalize the data externality, but not necessarily in the right amount. We now turn to welfare. The relevant question is not only whether the competitive benchmark over-adopts AI, but also how that benchmark compares to a constrained planner once the supply side is endogenous. The policy discussion is already central in current debates: for example, the EU's AI Act places restrictions on both AI deployment and data use. Beyond adoption, the high concentration of the AI industry raises questions about whether market power distorts the pace and direction of AI deployment. The dominance of a few frontier AI providers may restrict access and inflate costs, but it may also slow adoption in ways that partially mitigate data externalities. Whether open-source alternatives or competitive entry would improve welfare depends on how market structure interacts with the data-quality feedback loop. In this section, we study a constrained planner who faces the same information frictions as private agents but internalizes how current adoption shapes future data quality.

We focus on two main questions:

1. Is the private sector's decision to adopt AI socially efficient? Would a benevolent social planner, if given the chance, make different decisions about how many companies should adopt AI?

2. How should government respond to different market structures in the AI sector?

6.1 Formulation of Planner's Problem

To explore these questions, we set up a constrained social planner's problem. The planner is constrained in the sense that he faces the same information friction as private individuals, meaning that he cannot observe θ_t and also cannot directly interfere with the data accumulation process. The planner chooses the labor threshold $\bar{\varphi}$, which determines the share of AI adoption.[27]

We formulate the social planner's problem. In each period, total surplus consists of the aggregate surplus of labor-using entrepreneurs, the aggregate surplus of AI-using entrepreneurs, and the AI provider's profit:

$$U_t = \underbrace{\int_0^{\bar{\varphi}_t} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_L} - \varphi\right] dF(\varphi)}_{\text{labor-using entrepreneurs}} + \underbrace{\left[1 - F(\bar{\varphi}_t)\right] \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - R_t\right]}_{\text{AI-using entrepreneurs}} + \underbrace{(R_t - c_t)\left[1 - F(\bar{\varphi}_t)\right]}_{\text{monopoly profit } \pi_t}. \tag{29}$$

The AI price R_t is a transfer between entrepreneurs and the AI provider: it appears as a cost in the second term and as revenue in the third. Combining these two terms, R_t cancels, and total surplus simplifies to

$$U_t = \int_0^{\bar{\varphi}_t} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_L} - \varphi\right] dF(\varphi) + \left[1 - F(\bar{\varphi}_t)\right] \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} - c_t\right]. \tag{30}$$

The planner's welfare depends on the real resource cost c_t, not the market price R_t. This is why the planner's allocation is independent of market structure: regardless of whether the AI market is competitive or monopolistic, the efficient allocation is the same. We begin with the steady-state analysis. In Section 6.3, we extend the analysis to a fully dynamic setting with discounting.

The planner chooses the adoption threshold $\bar{\varphi}$ to maximize steady-state total surplus. The steady-state law of motion and $n = F(\bar{\varphi})$ jointly define $\gamma(\bar{\varphi})$ as a function of the adoption threshold. The planner's steady-state problem is:

$$\max_{\bar{\varphi}} \int^{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_L} - \varphi\right] dF(\varphi) + \int_{\bar{\varphi}} \left[\bar{A} - \frac{1}{\gamma_S + \gamma_A + \gamma\left(\bar{\varphi}\right)} - c\right] dF(\varphi) \tag{31}$$

The key distinction between the social planner's problem and the decisions made by private agents is that the social planner understands how AI adoption affects information quality, while private agents treat γ as fixed. This creates an information externality that the planner needs to correct: private individuals, when making decisions, do not consider the broader impact of their

[27] In Appendix D, we also analyze a second policy instrument: restricting AI's access to historical data through a parameter $\kappa \leq 1$. We show that limiting data access can improve welfare by forcing AI to generate more novel information.

actions on the overall data quality.

6.2 AI Adoption

The first-order condition that determines the socially efficient threshold $\bar{\varphi}^{sp}$ is given by:

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

This optimality condition consists of two parts: the "private benefits" and the "externality." The "private benefits" term mirrors the condition for individual optimization, while the "externality" term represents what the social planner takes into account, but private agents do not. The externality term depends on how the data quality γ changes with the adoption threshold $\bar{\varphi}$.

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

Theorem 6.1 *1. Increasing AI adoption hurts data quality ($\gamma'\left(\bar{\varphi}^{sp}\right)>0$) if and only if the relative quality of information generated by AI is not too much higher than that generated by labor:*

$$\frac{\gamma_A}{\gamma_L}<\frac{\gamma_S+\gamma}{\gamma_S} \tag{32}$$

2. Given the presence of data corruption, AI adoption hurts data quality, hence it should be taxed in the competitive equilibrium.[28]

Condition 32 is identical to Condition 17, where the externality is tied to the relative quality of information generated by humans and AI. When human-generated information is of higher quality compared to AI, the externality of AI adoption is negative, meaning that more AI adoption lowers the overall data quality. This negative externality justifies the taxation of AI, even in the long run. Thus, based on our benchmark calibration, a tax on AI adoption is necessary to achieve socially optimal data quality.

The theorem suggests that government regulation should vary by task, depending on the level of uncertainty and quality of existing historical data. For example, in tasks with minimal uncertainty, like Go-playing, the case for taxation is weaker. However, in areas with high uncertainty,

[28] In Section 6.4, we show that the direction of optimal policy reverses under monopoly.

such as forecasting market demand or developing new drugs, the theorem implies that more uncertainty calls for greater regulation and taxation. Thus, government policy should be tailored to specific tasks, with regulation determined on a case-by-case basis.

6.3 Dynamic Social Planner

We now extend the planner's problem to a dynamic setting by introducing a discount factor $\beta \in (0,1)$ that captures the weight placed on future welfare. The discount factor β can be interpreted as the planner's patience in weighting successive generations of short-lived entrepreneurs.

Using the per-period total surplus U_t from equation (30), the planner's dynamic problem is:

$$S(\gamma_t) = \max_{\bar{\varphi}_t} \left\{ U_t(\bar{\varphi}_t, \gamma_t) + \beta\, S\big(\gamma_{t+1}\big) \right\} \tag{33}$$

subject to the law of motion for data quality (equation 20) with $n_t = F(\bar{\varphi}_t)$.

The first-order condition is:

$$\underbrace{\frac{\partial U_t}{\partial \bar{\varphi}_t}}_{\text{static welfare effect}} + \beta \underbrace{S'(\gamma_{t+1}) \cdot \frac{\partial \gamma_{t+1}}{\partial \bar{\varphi}_t}}_{\text{dynamic data-quality effect}} = 0. \tag{34}$$

The static welfare effect captures the within-period impact of shifting the adoption margin:

$$\frac{\partial U_t}{\partial \bar{\varphi}_t} = \left[\frac{1}{\gamma_S + \gamma_A + \gamma_t} - \frac{1}{\gamma_S + \gamma_L} + c_t - \bar{\varphi}_t \right] \cdot f(\bar{\varphi}_t).$$

In the competitive equilibrium (where $R_t = c_t$), the threshold is $\bar{\varphi}_t^{ce} = c_t + \frac{1}{\gamma_S+\gamma_A+\gamma_t} - \frac{1}{\gamma_S+\gamma_L}$, which sets this term exactly to zero: the competitive equilibrium maximizes static welfare. Hence, any inefficiency in the competitive equilibrium is entirely *dynamic*—it arises from the failure of private agents to internalize the effect of current adoption on future data quality.[29]

The dynamic data-quality effect captures the intertemporal linkage. Under data corruption, shifting the adoption threshold upward (more labor, less AI) improves future data quality: $\partial \gamma_{t+1}/\partial \bar{\varphi}_t > 0$, because more human production generates data of higher informational content (by Theorem 6.1). Since better future data quality raises future welfare, $S'(\gamma_{t+1}) > 0$. The product of these two terms is strictly positive, which leads to the following result:

Theorem 6.2 *Under data corruption (Condition 32 holds) and $\beta > 0$, the competitive equilibrium features over-adoption of AI along the transition path. That is, the planner's optimal threshold satisfies $\bar{\varphi}_t^{sp} > \bar{\varphi}_t^{ce}$*

[29] When $\beta = 0$, the dynamic correction vanishes and the planner's first-order condition reduces to $\partial U_t/\partial \bar{\varphi}_t = 0$, which coincides with the competitive equilibrium threshold for any given γ_t.

for all t, implying that the planner would always choose strictly less AI adoption than arises in the perfect competitive equilibrium.

Proof. At the competitive equilibrium threshold $\bar{\varphi}_t^{ce}$, the static welfare effect $\partial U_t / \partial \bar{\varphi}_t$ equals zero. The dynamic term $\beta S'(\gamma_{t+1}) \cdot \partial \gamma_{t+1} / \partial \bar{\varphi}_t$ is strictly positive when $\beta > 0$, since $S'(\gamma_{t+1}) > 0$ (higher data quality raises future surplus) and $\partial \gamma_{t+1} / \partial \bar{\varphi}_t > 0$ (more labor adoption improves future data quality under data corruption). Hence the planner's first-order condition (34) is not satisfied at $\bar{\varphi}_t^{ce}$; the planner raises $\bar{\varphi}_t$ above $\bar{\varphi}_t^{ce}$, reducing AI adoption. □

Choice of β. In our baseline quantitative exercises, we set $\beta = 0.66$, interpreting each model period as spanning approximately 10 years. The 10-year generational length follows Guerreiro et al. (2022), who model each cohort of workers as active for approximately one decade when studying optimal taxation of automation.[30] With an annual discount factor of $\delta = 0.96$, a 10-year generational cycle gives $\beta = 0.96^{10} \approx 0.66$. For comparison, we also examine $\beta = 0.54$, which corresponds to a 15-year generational cycle ($0.96^{15} \approx 0.54$). A longer generational cycle implies a lower β because one must wait longer between successive cohorts. We examine the sensitivity of all results to β over the range $[0.45, 0.75]$ in Section 6.6.

6.4 Comparing Planner, Monopoly, and Competitive Equilibrium

Theorem 6.2 establishes that the competitive equilibrium features too much AI adoption from a social standpoint. Even the static (Cournot) monopolist restricts AI adoption through its markup (equation 24), and the dynamic monopolist analyzed in Section 5.4 restricts it further. A natural question is whether the monopolist's market structure brings the equilibrium closer to the planner's optimum, or whether it over-corrects.

The answer depends on the relative magnitudes of two forces. On one hand, the monopoly markup reduces AI adoption, which partially corrects the data externality. On the other hand, the monopolist sets the markup to maximize *profit*, not welfare. The monopolist's dynamic first-order

[30] Guerreiro et al. (2022) use a 10-year working-life cohort in their analysis of robot taxation. In our setting, this horizon also aligns with the deeper cycle of data accumulation, market-wide adoption shifts, and their feedback into AI training, which occurs over a longer horizon than individual model updates.

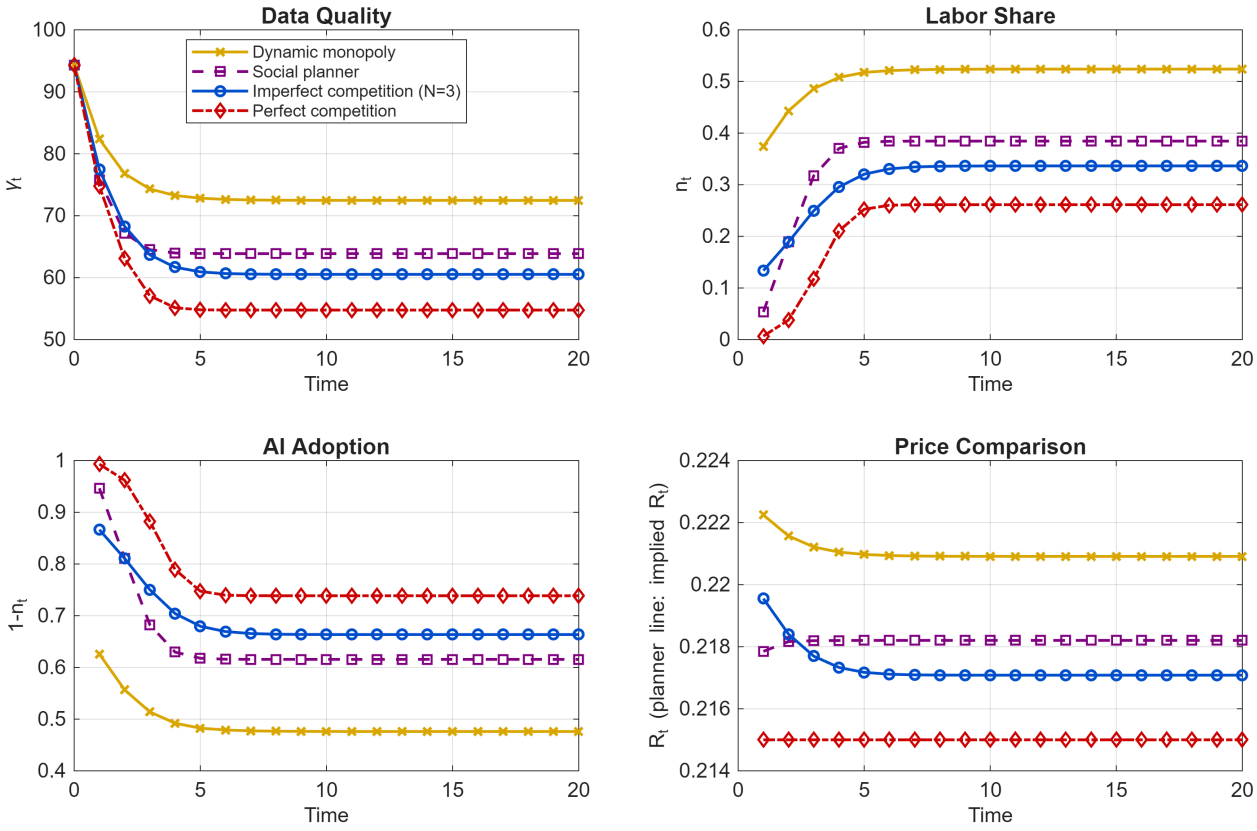

Figure 9: **Transition Dynamics Across Market Structures** ($\beta = 0.66$). The four panels display the transition paths of data quality (γ_t), labor share (n_t), AI adoption ($1 - n_t$), and AI price (R_t) following a cost reduction shock. "Dynamic monopoly": forward-looking monopoly; "Social planner": the planner's allocation; "Imperfect competition (N=3)": myopic Cournot oligopoly equilibrium; "Perfect competition": myopic competitive equilibrium with R=c. The planner's allocation lies between the myopic competitive equilibrium and the dynamic monopoly in both AI adoption and long-run data quality.

condition (27) involves the private return to data quality $M'(\gamma_{t+1})$, which reflects future *profits*, while the planner's condition (34) involves the social return $S'(\gamma_{t+1})$, which reflects future total *welfare*. Moreover, the monopolist's markup simultaneously restricts AI adoption *and* transfers surplus from entrepreneurs to the AI provider, generating a welfare loss that has no counterpart in the planner's problem.

To determine which force dominates, we turn to the quantitative analysis. Using the baseline calibration from Section 4 with $\beta = 0.66$, Figure 9 plots the transition dynamics under all four regimes following a cost reduction shock.

Several patterns emerge from the figure. First, the perfect competitive equilibrium exhibits the highest AI adoption and the lowest long-run data quality, consistent with the over-adoption result

of Theorem 6.2. Second, both the imperfect competitive (Cournot) equilibrium and the dynamic monopoly restrict AI adoption through their markups, with the dynamic monopolist restricting further due to the forward-looking motive to maintain high data quality. Third—and most importantly—the planner's allocation lies *between* the competitive and monopolistic extremes: the planner allows more AI adoption than the dynamic monopolist but less than the competitive market. This indicates that the monopolist's markup over-corrects the data externality.

This leads to a central insight for AI regulation: optimal policy is *market-structure dependent*. In a competitive AI market, the data externality is unambiguously negative and a tax on AI adoption is warranted (Theorem 6.1). Lacking a forward-looking motive, even in the presence of myopic oligopolists there is a need to tax AI adoption. In a dynamic monopolistic AI market, the monopolist's markup could (over-)correct the externality, and the planner would prefer to *subsidize* AI adoption relative to the monopoly outcome. We formalize the implementation of such a subsidy next.

6.5 Implementing the Planner's Allocation

We now consider how a social planner can implement the efficient allocation through the monopolistic AI market, given the currently concentrated AI sector. The policy instrument is a proportional subsidy τ_t on the AI price, financed by lump-sum taxes. This is a general equilibrium problem: the subsidy changes the monopolist's demand curve, so the monopolist re-optimizes its pricing, which in turn affects adoption, data quality, and future demand. We must therefore solve for the subsidy jointly with the monopolist's optimal response.

The monopolist's problem. Given a subsidy sequence $\{\tau_t\}_{t=0}^{\infty}$, the monopolist posts a price R_t but entrepreneurs face an effective price $(1-\tau_t)R_t$. The adoption threshold becomes

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

and the monopolist's per-period profit remains $(R_t - c_t)[1 - F(\bar{\varphi}_t)]$, since the subsidy is paid to entrepreneurs, not to the monopolist.[31] The dynamic monopolist solves:

$$M(\gamma_t; \{\tau_s\}) = \max_{R_t} \left\{ (R_t - c_t)\left[1 - F\left(\bar{\varphi}_t(\tau_t, R_t)\right)\right] + \beta\, M\left(\gamma_{t+1}; \{\tau_s\}\right) \right\} \tag{35}$$

subject to the law of motion for data quality with $n_t = F(\bar{\varphi}_t(\tau_t, R_t))$. The subsidy enters the monopolist's problem through the demand curve: a higher τ_t shifts the adoption threshold down, increasing demand at any posted price, and the forward-looking monopolist adjusts R_t accordingly. In particular, the monopolist may *partially absorb* the subsidy by raising the posted price, so the effective price to entrepreneurs does not fall one-for-one with the subsidy.

Fixed-point characterization. The optimal subsidy sequence $\{\tau_t^*\}$ must satisfy: given $\{\tau_t^*\}$, the monopolist's optimal pricing $\{R_t^*\}$ from (35) generates adoption and data quality paths that coincide with the social planner's allocation from (33). This is a fixed-point problem in the sequence $\{\tau_t\}$: the planner's target adoption requires

$$(1 - \tau_t^*) R_t^*(\{\tau_s^*\}) = R_t^{sp}, \tag{36}$$

where $R_t^{sp} \equiv \bar{\varphi}_t^{sp} - \frac{1}{\gamma_S + \gamma_A + \gamma_t} + \frac{1}{\gamma_S + \gamma_L}$ is the effective AI price that implements the planner's optimal adoption threshold $\bar{\varphi}_t^{sp}$, and $R_t^*(\{\tau_s^*\})$ is the monopolist's optimal posted price given the subsidy sequence. Because R_t^* depends on the entire future path of subsidies (through the monopolist's forward-looking Bellman equation), and the planner's target depends on the equilibrium data quality path (which is itself determined by adoption), the problem must be solved jointly. We compute this fixed point numerically by iterating between the monopolist's finite-horizon dynamic programming problem and the implied subsidy update.[32]

Figure 10 plots the equilibrium subsidy sequence for $\beta = 0.66$. Three features are notable.

First, the subsidy is *strictly positive* throughout the transition, confirming that the monopolist persistently restricts AI adoption below the socially efficient level. Even after the monopolist re-optimizes given the subsidy, the posted price R_t^* is higher than the planner's implied effective

[31] The monopolist collects the full posted price R_t. The government pays $\tau_t R_t$ per unit to adopting entrepreneurs, funded by lump-sum taxes on all entrepreneurs. From the monopolist's perspective, the subsidy shifts the demand curve outward: at any posted price R_t, more firms adopt because the effective user cost is lower.

[32] The subsidy must be financed. Total expenditure in period t is $G_t = \tau_t R_t^* [1 - F(\bar{\varphi}_t)]$. The government's balanced-budget constraint is

$$T_t^M + T_t^E = G_t, \tag{37}$$

where T_t^M is a lump-sum tax on the monopolist and T_t^E is a lump-sum tax on entrepreneurs. Because both taxes are lump-sum, they do not distort any margins and the split between T_t^M and T_t^E is irrelevant for equilibrium allocations.

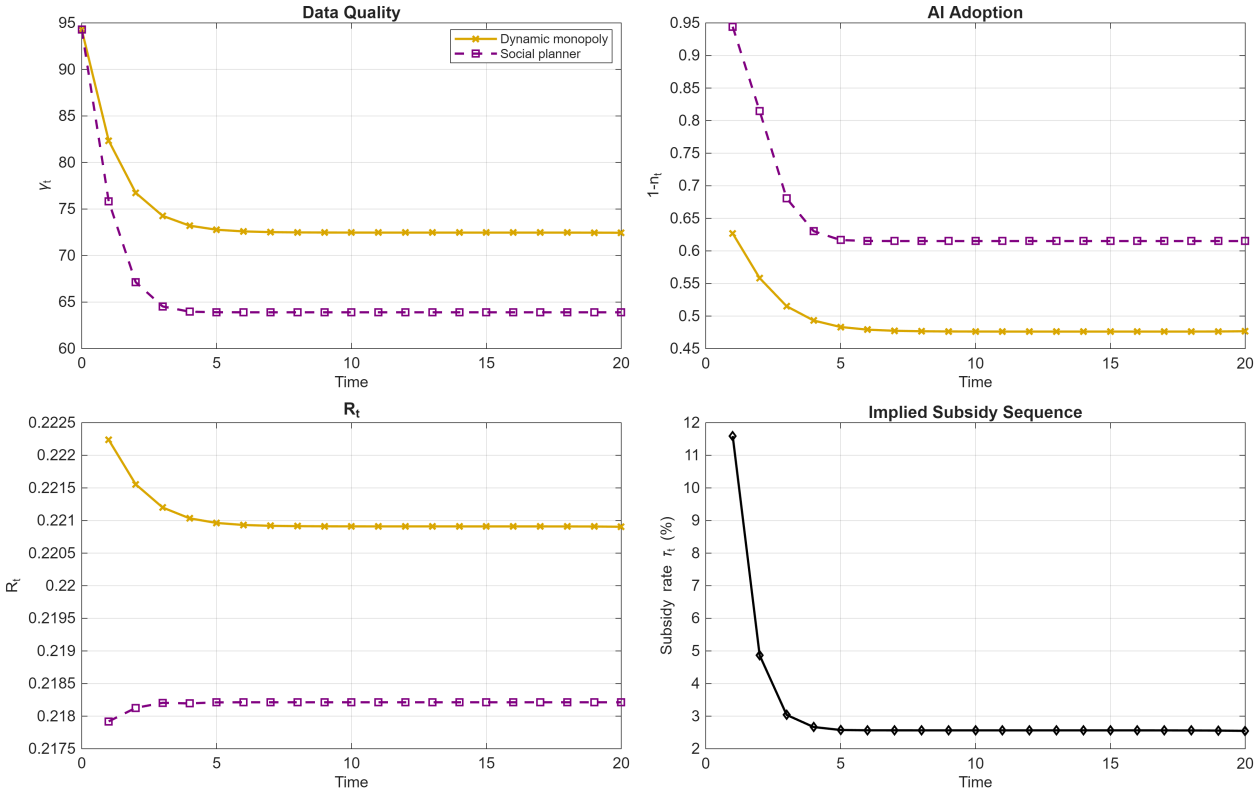

Figure 10: **Transition Dynamics and Implied AI Subsidies** ($\beta = 0.66$). The panels compare the dynamic monopoly (yellow) to the social planner (purple). Top panels: data quality and AI adoption paths. Bottom-left: R_t, showing the monopolist's posted price versus the planner's implied effective price. Bottom-right: the equilibrium subsidy rate τ_t^* that implements the planner's allocation through the monopolist's re-optimized pricing. The subsidy is front-loaded, starting at approximately 11% and declining to approximately 3% at the steady state.

price, so a positive wedge is always needed.

Second, the subsidy is *front-loaded*: it peaks at approximately 11% at the onset of the cost reduction and declines toward approximately 3% at the steady state. The front-loading reflects the general equilibrium dynamics: in the early transition, data quality is declining rapidly and the monopolist's markup is particularly distortive. The subsidy must be large enough not only to lower the effective price but also to offset the monopolist's strategic response—which includes raising the posted price in anticipation of subsidized demand. As the economy converges to a steady state, both the data quality gap and the monopolist's strategic response stabilize, and the required subsidy settles at a modest level.

Third, even at the steady state, a positive subsidy is required. This indicates that the monopolist's long-run pricing permanently restricts AI adoption below the socially efficient level, and general equilibrium forces do not eliminate this wedge.

6.6 The Role of the Planning Horizon

The discount factor β governs the weight placed on future welfare and profits. As discussed in Section 6.3, β can be interpreted as reflecting the planner's patience, the monopolist's discount rate, or—in reduced form—the effective lifespan of economic agents. We now examine how β affects the allocations of the social planner and dynamic monopolist.

Figure 11 displays steady-state allocations as a function of β. As β increases, both the planner and the monopolist restrict AI adoption: a longer planning horizon places more weight on preserving data quality for the future. Correspondingly, steady-state data quality increases with β for both regimes (bottom-right panel).

Importantly, the gap between the planner's implied price and the monopolist's posted price *narrows* as β increases (bottom-left panel). This occurs because the dynamic monopolist's incentive to preserve data quality grows with β: future profits become more valuable, and maintaining data quality is an investment in future demand. As the monopolist becomes more forward-looking, its behavior moves closer to the planner's optimum, reducing the need for corrective subsidies.

Figure 12 directly plots the steady-state subsidy rate as a function of β. The required subsidy declines monotonically from approximately 2.85% at $\beta = 0.45$ to approximately 2.45% at $\beta = 0.75$.

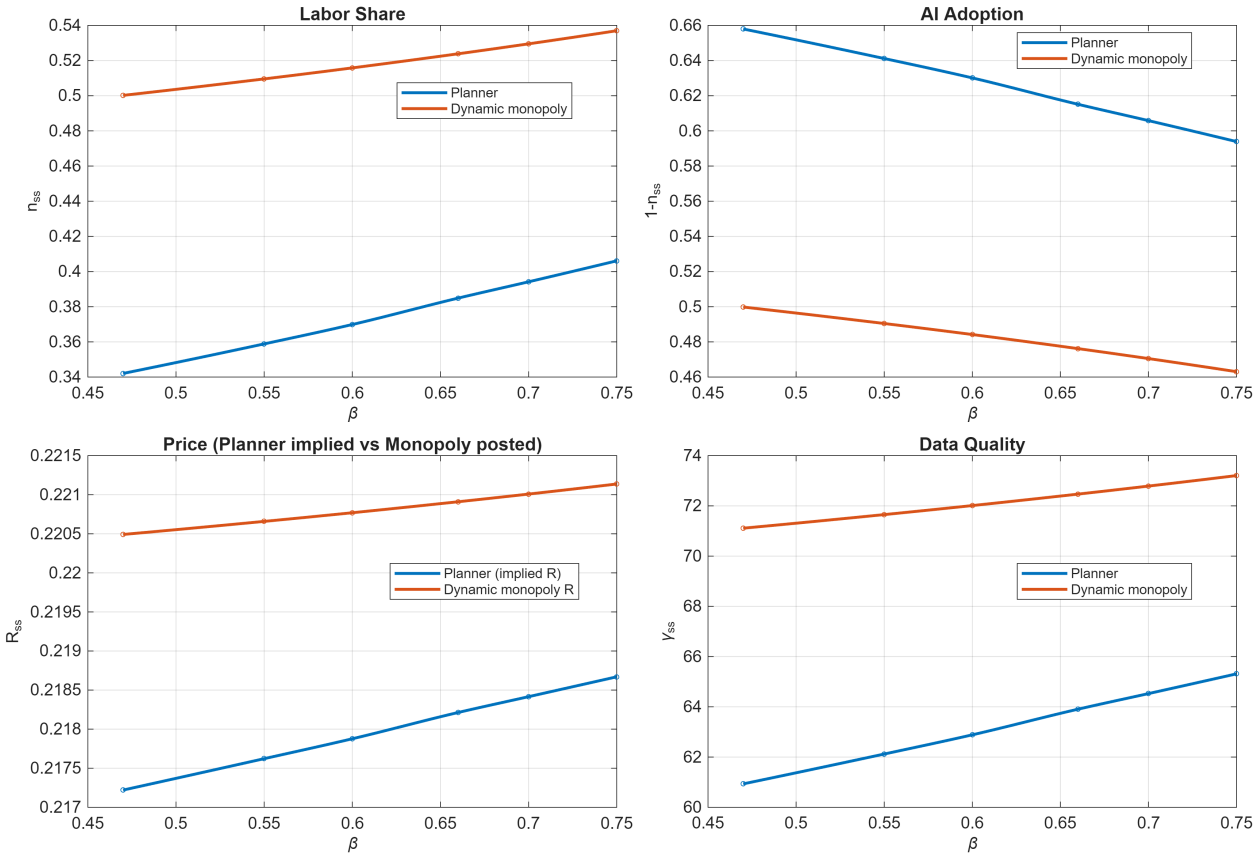

Figure 11: **Steady-State Allocations as a Function of β.** The panels display steady-state labor share (n_{ss}), AI adoption ($1 - n_{ss}$), AI price (R_{ss}), and data quality (γ_{ss}) for the social planner (purple, top row) and dynamic monopoly (yellow, bottom row) across values of β. As β increases, both the planner and monopolist choose less AI adoption and higher data quality. The gap between the planner's implied price and the monopolist's posted price narrows.

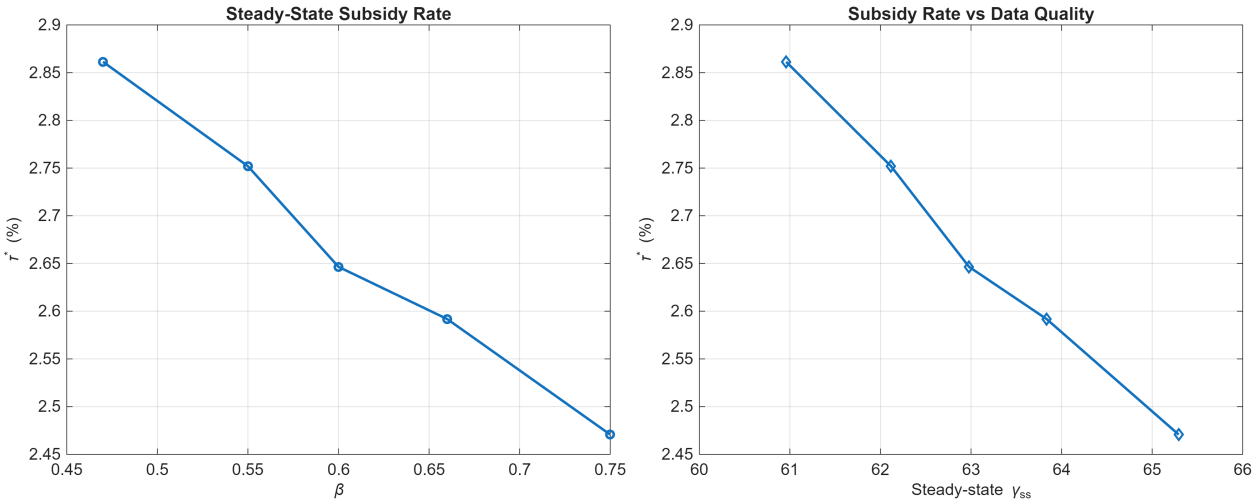

Figure 12: **Steady-State Subsidy as a Function of β.** Left panel: the steady-state subsidy rate τ^* that implements the planner's allocation given a dynamic monopolist, as a function of β. Right panel: the same subsidy rate plotted against the planner's steady-state data quality γ_{ss}^{pl}. The required subsidy declines as β increases, reflecting the improved alignment between monopoly behavior and social welfare when the monopolist is more forward-looking.

This confirms that a more patient monopolist is a better steward of data quality, reducing the welfare cost of market concentration.

Transition dynamics under different β. The effect of β is not limited to the steady state: it also shapes the *speed* and *pattern* of the transition. Figure 13 compares the transition dynamics under $\beta = 0.66$ and $\beta = 0.54$ for both the planner and the dynamic monopolist.

Three patterns are apparent. First, the higher β leads to uniformly lower AI adoption and higher data quality along the entire transition path, for both the planner and the monopolist. Second, the gap between the planner and monopolist—visible in the price panel—is smaller at $\beta = 0.66$ than at $\beta = 0.54$ throughout the transition, not only at the steady state. This confirms that the convergence between monopoly behavior and social welfare under a longer planning horizon is a robust feature of the transition dynamics, not an artifact of the steady-state analysis. Third, the optimal subsidy is strongly front-loaded under both β values: a large initial subsidy is needed to overcome the monopolist's under-adoption during the critical early transition when data quality is declining most rapidly. As the economy approaches its steady state, the required subsidy stabilizes at a small but positive level that corrects the residual monopoly markup.

These findings provide insights into the interaction between the planning horizon and market

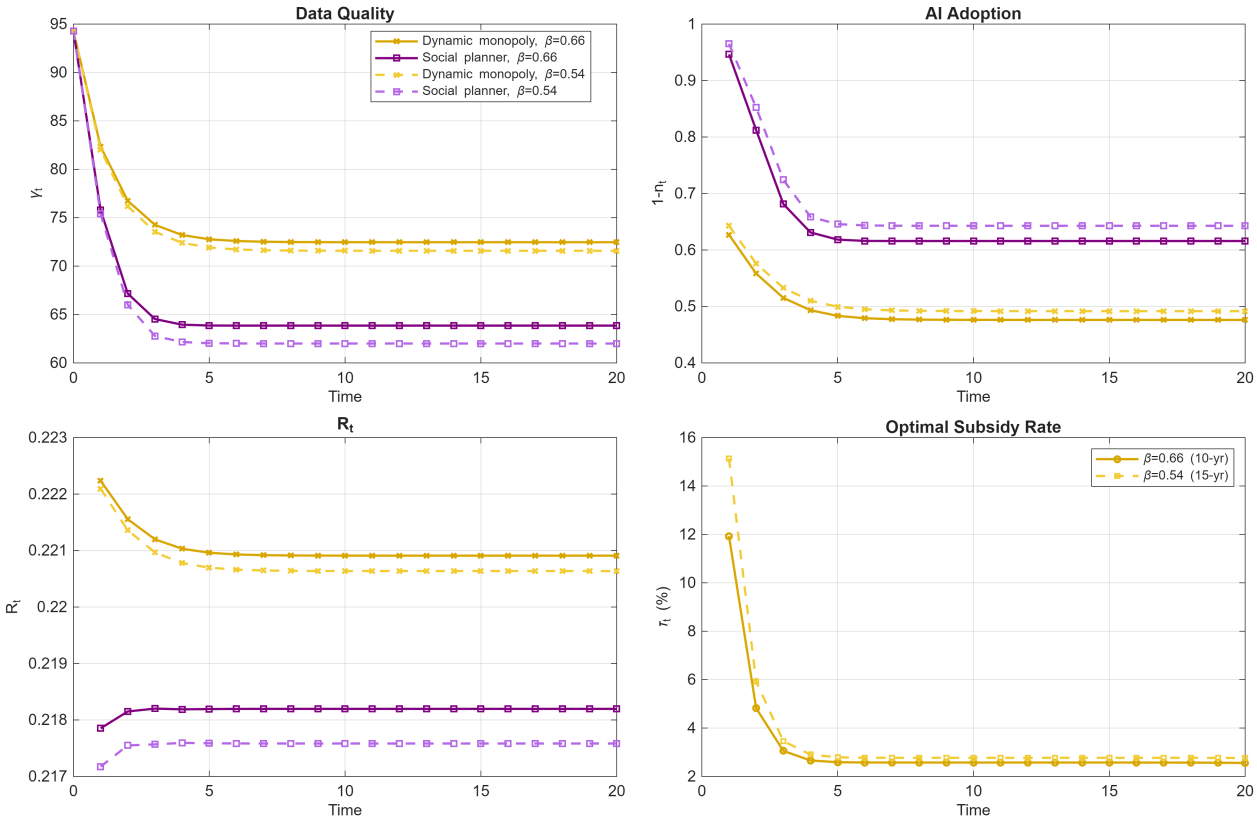

Figure 13: **Transition Dynamics: Social Planner vs. Dynamic Monopoly for $\beta = 0.66$ and $\beta = 0.54$.** Panels (a)–(c) display data quality (γ_t), AI adoption ($1 - n_t$), and R_t along the transition path. Solid lines: $\beta = 0.66$; dashed lines: $\beta = 0.54$. Purple: social planner; yellow: dynamic monopoly. Panel (d) shows the optimal subsidy rate τ_t that implements the planner's adoption path through a monopolistic AI provider. Both β values generate a front-loaded subsidy that starts high ($\approx 11.6\%$ for $\beta = 0.66$; $\approx 15.5\%$ for $\beta = 0.54$) and converges rapidly to a modest steady-state level (≈ 2.6–2.8%). The lower β requires a larger initial subsidy because the monopolist discounts future data quality more heavily.

structure. When economic agents or the policymaker are more patient, the dynamic value of data quality is better internalized by both the planner and the monopolist. A monopolist with a long planning horizon restricts AI adoption for forward-looking reasons that partially align with the planner's objective, reducing the need for regulatory intervention. Conversely, when agents are impatient or short-lived (β close to zero), the data externality is severe and market power provides little corrective benefit.

This analysis also sheds light on the role of *entrepreneur lifespan*. In the original model, entrepreneurs live for one period and make purely myopic adoption decisions. If entrepreneurs instead lived for multiple periods, they would directly internalize how their current AI adoption affects future data quality and hence their own future productivity—analogous to increasing β in our framework. Longer-lived agents would naturally adopt less AI (to preserve future data quality), reducing the severity of the data externality. In the context of the monopolistic AI market, a longer planning horizon further improves the alignment between the monopolist's pricing and social welfare, diminishing the need for subsidies.

7 AI-Human Collaboration

Sections 2–6 treat adoption as binary: each firm either uses human labor or AI. That simplification is useful in the baseline, but many real-world tasks combine the two inputs. This section therefore allows firms to choose a continuous collaboration intensity between human and AI. The extension is important because it tests whether the data-corruption mechanism relies on all-or-nothing substitution. It does not. Collaboration softens the mechanism by preserving some human-generated information inside each task, but the core feedback loop remains.

Each entrepreneur i at time t chooses a collaboration weight $m_{it} \in [0, 1]$, where $m_{it} = 0$ corresponds to fully human production and $m_{it} = 1$ to fully AI production. The implemented action is a weighted average:

$$a_{it} = (1 - m_{it})\, a_{it}^{L} + m_{it}\, a_{it}^{A},$$

where a_{it}^{L} and a_{it}^{A} are the optimal human and AI actions, respectively. Quality remains $A_{it} = \bar{A} - (\theta_t - a_{it})^2$.

Input costs are linear in the collaboration weight:

$$C_{it}(m) = (1 - m)\,\varphi_{it} + m\,R_t,$$

where φ_{it} is the idiosyncratic human labor cost and R_t is the AI price. The entrepreneur maximizes $V_{it}(m) = \mathbb{E}[A_{it}] - C_{it}(m)$. We will focus on the perfectly competitive case so $R_t = c_t$.[33] The first-order condition yields:

Theorem 7.1 (Optimal Collaboration Weight) *The optimal collaboration weight for firm i with labor cost φ_{it} is*

$$m_{it}^* = \underbrace{\frac{\gamma_A + \gamma_t}{\gamma_L + \gamma_A + \gamma_t}}_{\text{Bayesian updating}} + \underbrace{(\varphi_{it} - R_t) \cdot \frac{(\gamma_S + \gamma_L)(\gamma_S + \gamma_A + \gamma_t)}{2(\gamma_L + \gamma_A + \gamma_t)}}_{\text{cost differential}}, \quad m_{it}^* \in [0, 1]. \tag{38}$$

The theorem decomposes the optimal AI weight into two additive terms. The first term is the *Bayesian updating* component: it equals the Bayesian optimal weight on the AI forecast, allocating weight to each input in proportion to its informational precision ($\gamma_A + \gamma_t$ for AI versus γ_L for human labor), with no role for costs. This term increases with data quality γ_t, reflecting that better historical data makes AI's informational contribution more valuable. The second term is the *cost differential* component: when labor is more expensive than AI ($\varphi_{it} > R_t$), the firm tilts toward AI beyond the pure Bayesian benchmark; when labor is cheaper ($\varphi_{it} < R_t$), the firm tilts toward human input. Figure 14 illustrates this: the optimal AI weight $m^*(\varphi)$ increases smoothly in φ, with the Bayesian component determining the level at $\varphi = R$ and the cost differential governing the slope.

A key question is whether the collaboration extension preserves the positive feedback between data quality γ_t and AI usage m_t. Under the benchmark calibration, m_{it}^* is increasing in γ_t for all active firms: higher data quality raises the relative precision of AI forecasts, inducing each firm to increase its AI weight. The proof follows from differentiating the interior solution and verifying that the numerator is positive under our calibrated parameters. Figure 14 compares the individual decision rules under the binary and interior models at the steady-state data quality. In the binary model, the decision is a step function: firms with labor cost above the threshold $\bar{\varphi}$ adopt AI fully ($m = 1$), while those below do not ($m = 0$). In the interior model, the collaboration weight $m^*(\varphi)$ increases smoothly in φ, with firms near the margin mixing human and AI inputs rather than

[33] The objective $V_{it}(m)$ is strictly concave in m: $\frac{\partial^2 V_{it}}{\partial m^2} = -\frac{2(\gamma_L+\gamma_A+\gamma_t)}{(\gamma_S+\gamma_L)(\gamma_S+\gamma_A+\gamma_t)} < 0$.

making a discrete switch. The labor-cost density shows that most firms cluster near the threshold, so the intensive margin is quantitatively important.

Correlated signal innovations. In the baseline model, the private signals s_t^{Li} and s_t^{Ai} are independent of the public signal S_t. A natural extension is to allow signal errors to be correlated. If the human private signal is positively correlated with the public signal—for instance, because human judgment draws partly on publicly available narratives—then the effective new information content of human data is lower. This is formally equivalent to reducing the effective precision γ_L of novel information in human actions. In our collaboration framework, this shifts the optimal AI weight m^* upward, increasing equilibrium AI adoption: firms optimally rely more on AI when human signals are less informative at the margin. Conversely, if AI's private signal is less correlated with public noise—plausible when AI discovers patterns through computational methods distinct from conventional reasoning—AI provides additional diversification value, further raising m^*. The core data corruption mechanism remains robust: regardless of the correlation structure, what matters for future data quality is the share of genuinely new information in the aggregate dataset, which continues to depend on the composition of human versus AI activities.

Aggregation and data dynamics. Aggregate AI usage is $m_t = \int m_{it}^* \, dF(\varphi)$, with $1 - m_t$ denoting the aggregate human labor share. The law of motion remains

$$\frac{1}{\gamma_{t+1}} = \frac{\rho^2}{\gamma_t + \gamma_S + (1 - m_t)\left(\frac{\gamma_L}{\gamma_S + \gamma_L}\right)^2 \gamma_D^l + m_t \left(\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t}\right)^2 \gamma_D^A} + \frac{1}{\gamma_\eta}.$$

This is identical in form to the baseline law of motion (equation 20), with n_t replaced by $1 - m_t$. The data corruption channel operates through the same mechanism: higher AI usage (m_t) shifts the composition of data toward AI-generated records, which have lower informational content when $\gamma_A/(\gamma_S + \gamma_A + \gamma_t) < \gamma_L/(\gamma_S + \gamma_L)$.

Figure 15 compares the transition dynamics following the cost reduction shock under competitive pricing ($R_t = c_t$). The top panel shows that the interior model produces higher steady-state data quality than the binary model. This is because firms near the adoption margin use a mix of human and AI inputs rather than switching entirely to AI, preserving more human-generated data in the aggregate. The middle panel shows that aggregate AI intensity is lower in the interior

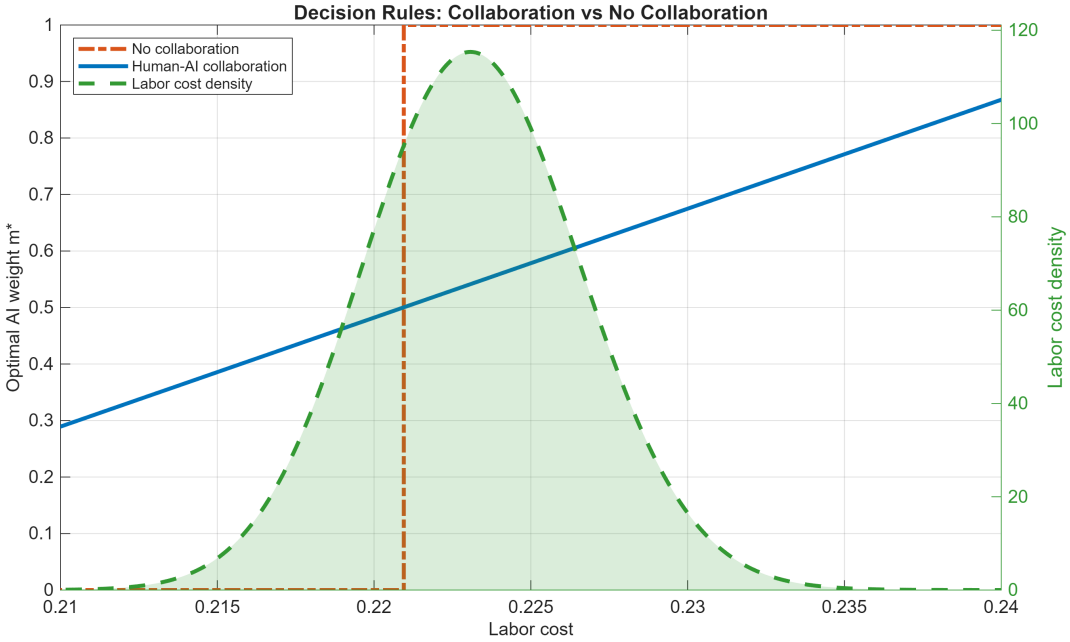

Figure 14: **Decision Rules: Collaboration vs. No Collaboration.** The red dashed line shows the no-collaboration benchmark: a step function at the adoption threshold where firms either fully use AI or fully use human labor. The blue solid line shows the collaboration model: the optimal AI weight increases smoothly in labor cost, with firms near the margin mixing human and AI inputs rather than making a discrete switch. The green dashed curve is the labor cost density, showing that most firms cluster near the threshold where the intensive margin matters most.

model: the smooth mixing means fewer firms are at full AI adoption. The bottom panel presents a recalibrated comparison: the interior model's AI cost is recalibrated so that initial AI adoption m_1 matches the binary model exactly. Despite identical impact adoption, the interior model's m_t path diverges over time, reflecting the endogenous feedback between data quality and the intensive margin of AI usage. Crucially, both models exhibit the same qualitative pattern—data quality declines following the cost shock and converges to a lower steady state—demonstrating that the data corruption mechanism is robust to allowing continuous human-AI collaboration.

8 Conclusion

Data is the lifeblood of generative AI, but AI also changes the data on which future models rely. This paper studies that feedback loop in an information-equilibrium model where AI adoption and data quality are jointly determined. The main message is that AI can be privately productive while socially degrading the information environment if it relies heavily on historical data and contributes little genuinely new information. Under the calibration disciplined by the recent computer science evidence on recursive synthetic training, this channel is quantitatively important: the initial labor displacement following an AI cost decline is partly reversed as AI-generated data lowers future AI productivity.

Once market structure is endogenous, the policy lesson becomes conditional on who sets AI prices. In a competitive AI market, firms over-adopt AI because they ignore the data externality, warranting a tax on adoption. In a monopolistic market, the forward-looking AI provider restricts adoption to protect future demand, but does so too aggressively relative to the social optimum, calling instead for a subsidy relative to the monopoly outcome. The optimal subsidy is front-loaded because the wedge is largest during the transition when data quality deteriorates most quickly. These findings underscore that AI regulation cannot be designed without taking market structure into account. The role of the planning horizon is also important: when agents or policymakers are more patient, the dynamic value of data quality is better internalized, reducing the need for corrective intervention.

The model is intentionally parsimonious so that the data-quality mechanism is transparent, but several extensions are natural. One is to study social learning with long-lived firms. Because

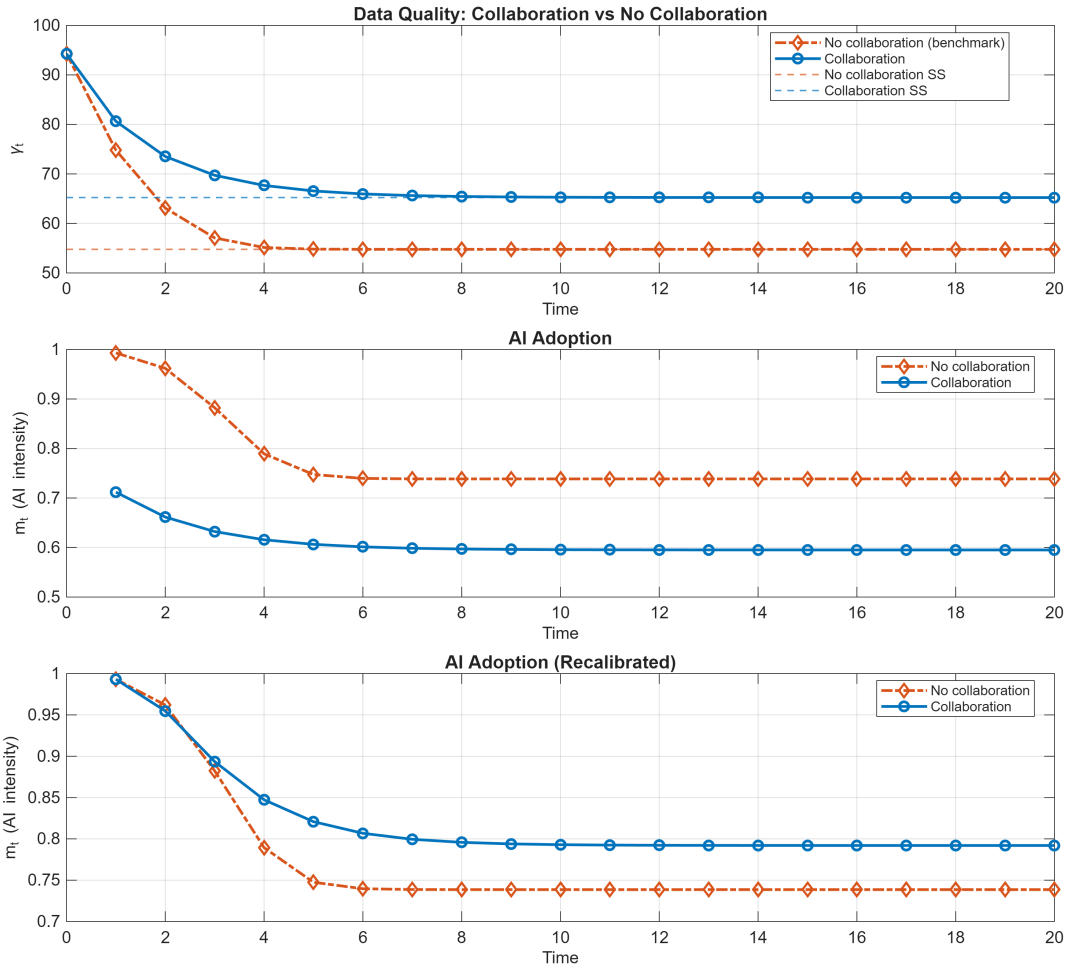

Figure 15: **Transition Dynamics: Collaboration vs. No Collaboration** ($R_t = c_t$). Top: data quality under the no-collaboration benchmark (red) and collaboration model (blue), with dashed lines indicating steady states. The collaboration model achieves higher long-run data quality. Middle: aggregate AI adoption under the same cost; the collaboration model produces lower aggregate AI usage because firms near the margin mix inputs rather than switching fully. Bottom: impact-matched comparison where the collaboration model's cost is recalibrated so that initial AI adoption equals the benchmark's impact adoption; despite identical initial AI intensity, the paths diverge over time.

each entrepreneur is atomistic, the data-quality externality would persist even with long-lived agents—no individual adoption decision has a perceptible effect on aggregate γ_t. However, long-lived agents would introduce richer heterogeneity in private histories and potential data-privacy issues (Jones and Tonetti, 2020). Another extension is anticipated technological progress. If AI is expected to become better at generating genuinely novel information, then preserving data quality today becomes even more valuable because better future models have more to gain from higher-quality data—creating a "wait and see" motive that, with long-lived firms, could endogenously slow early adoption. More broadly, the interdisciplinary approach developed here can be applied to study other tasks and technologies in which production choices shape the data available for future learning, particularly when combined with data from fields outside economics and finance.

References

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

Acemoglu, Daron, "The Simple Macroeconomics of AI," May 2024, (32487).

_ 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.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones, *Artificial Intelligence and Economic Growth*, University of Chicago Press, January

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

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.

Brooks, Creston, Samuel Eggert, and Denis Peskoff, "The Rise of AI-Generated Content in Wikipedia," 2024.

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

Cong, Lin William, Danxia Xie, and Longtian Zhang, "Knowledge Accumulation, Privacy, and Growth in a Data Economy," *Management Science*, 2021, *67* (10), 6480–6492.

_ , **Wenshi Wei, Danxia Xie, and Longtian Zhang**, "Endogenous growth under multiple uses of data," *Journal of Economic Dynamics and Control*, 2022, *141*, 104395. Markets and Economies with Information Frictions.

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.

Gerstgrasser, Matthias, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo, "Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data," *arXiv preprint arXiv:2404.01413*, 2024.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

Guerreiro, João, Sergio Rebelo, and Pedro Teles, "Should Robots Be Taxed?," *Review of Economic Studies*, 2022, *89* (1), 279–311.

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

Harding, Bill, "Coding on Copilot: 2023 Data Suggests Downward Pressure on Code Quality," White Paper, GitClear 2024.

Holcomb, Sean D., William K. Porter, Shaun V. Ault, Guifen Mao, and Jin Wang, "Overview on DeepMind and Its AlphaGo Zero AI," 2018, p. 67–71.

Jones, Charles I, "The A.I. Dilemma: Growth versus Existential Risk," Working Paper 31837, National Bureau of Economic Research November 2023.

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

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature*, 2021, *596* (7873), 583–589.

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," March 2024, (32255).

Krogh, Anders and John A. Hertz, "A Simple Weight Decay Can Improve Generalization," in "Advances in Neural Information Processing Systems," Vol. 4 1991, pp. 950–957.

Liu, Yang, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin, "Datasets for Large Language Models: A Comprehensive Survey," 2024.

Loshchilov, Ilya and Frank Hutter, "Decoupled Weight Decay Regularization," in "International Conference on Learning Representations" 2019.

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.

Merchant, Amil, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk, "Scaling Deep Learning for Materials Discovery," *Nature*, 2023, *624* (7990), 80–85.

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.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis, "Mastering the game of Go without human knowledge," *Nature*, October 2017, *550*, 354–.

Trammell, Philip and Anton Korinek, "Economic Growth under Transformative AI," October 2023, (31815).

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

Appendix

A Proof

Proof of Lemma 3.1, Part (ii): Endogenous Weight Decay

At steady state, denote the per-period precision of the labor and AI data signals by $\Delta^l = n^* \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D$ and $\Delta^A = (1-n^*)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma^*}\right)^2 \gamma_D$ (from Lemmas 3.2 and 3.3). Bayesian updating of the prior μ_{t-1} (precision γ^*) with the period-$(t-1)$ signals S_{t-1} (precision γ_S), X^l_{t-1} (precision Δ^l), and X^A_{t-1} (precision Δ^A) yields the posterior mean of θ_{t-1}:

$$E(\theta_{t-1}|\Omega_{t-1}) = \alpha\mu_{t-1} + \omega_S S_{t-1} + \omega_l X^l_{t-1} + \omega_A X^A_{t-1},$$

where the Kalman weights are the precision shares: $\alpha = \frac{\gamma^*}{\gamma^*+\gamma_S+\Delta^l+\Delta^A}$, $\omega_S = \frac{\gamma_S}{\gamma^*+\gamma_S+\Delta^l+\Delta^A}$, $\omega_l = \frac{\Delta^l}{\gamma^*+\gamma_S+\Delta^l+\Delta^A}$, $\omega_A = \frac{\Delta^A}{\gamma^*+\gamma_S+\Delta^l+\Delta^A}$, satisfying $\alpha+\omega_S+\omega_l+\omega_A = 1$. Since $\theta_t = \rho\theta_{t-1}+\eta_t$ and $E(\eta_t|\Omega_{t-1}) = 0$, we have $\mu_t = \rho\, E(\theta_{t-1}|\Omega_{t-1})$, giving the recursion:

$$\mu_t = \rho\left[\alpha\mu_{t-1} + \omega_S S_{t-1} + \omega_l X^l_{t-1} + \omega_A X^A_{t-1}\right].$$

Iterating and using $\rho\alpha < 1$:

$$\mu_t = \rho\sum_{\tau=1}^{\infty}(\rho\alpha)^{\tau-1}\left[\omega_S S_{t-\tau} + \omega_l X^l_{t-\tau} + \omega_A X^A_{t-\tau}\right].$$

Substituting into the AI's action rule $a^{Ai}_t = \frac{\gamma^*}{\gamma_S+\gamma_A+\gamma^*}\mu_t + \frac{\gamma_S}{\gamma_S+\gamma_A+\gamma^*}S_t + \frac{\gamma_A}{\gamma_S+\gamma_A+\gamma^*}s^{Ai}_t$ (equation 13) yields the stated expansion. □

Proof of Lemma 3.1

The optimal action a^i_t solves:

$$\max_{a^i_t} E\left[A^i_t\left(a^i_t\right)|S_t, s^{Li}_t\right] = \max_{a^i_t} E\left[\bar{A} - \left(\theta_t - a^i_t\right)^2 |S_t, s^{Li}_t\right]$$

Take first order condition with respect to a^i_t :

$$a^i_t = E\left(\theta_t|S_t, s^{Li}_t\right)$$

Given that both S_t and s^{Li}_t are unbiased, normally-distributed signal of θ_t, we can apply the formula for conditional mean with normal distribution to obtain:

$$a^i_t = E\left(\theta_t|S_t, s^{Li}_t\right) = \frac{\gamma_S}{\gamma_S+\gamma_L}S_t + \frac{\gamma_L}{\gamma_S+\gamma_L}s^{Li}_t$$

A similar argument can be applied to derive the optimal action rule of AI.

Proof of Lemma 3.2

From the derivation we know that

$$\begin{aligned} X_t^l &= \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\theta_t + \varepsilon_t^{Li} + \frac{\gamma_S+\gamma_L}{\gamma_L}\varepsilon_t^{Di}\right) \\ &= \theta_t + \frac{1}{N_t}\sum_{i=1}^{N_t}\varepsilon_t^{Li} + \frac{\gamma_S+\gamma_L}{\gamma_L}\frac{1}{N_t}\sum_{i=1}^{N_t}\varepsilon_t^{Di} \end{aligned}$$

Hence

$$\begin{aligned} Var\left(X_t^l|\theta_t\right) &= \frac{1}{N_t^2}N_t \underbrace{\frac{1}{\gamma_L}}_{\text{Variance of } \varepsilon_t^{Li}} + \left(\frac{\gamma_S+\gamma_L}{\gamma_L}\right)^2\frac{1}{N_t^2}N_t \underbrace{\frac{\bar{N}}{\gamma_D}}_{\text{Variance of } \varepsilon_t^{Di}} \\ &= \frac{1}{N_t}\frac{1}{\gamma_L} + \left(\frac{\gamma_S+\gamma_L}{\gamma_L}\right)^2\frac{\bar{N}}{N_t}\frac{1}{\gamma_D} \end{aligned}$$

When both $\bar{N}$ and $N_t \to \infty$, the variance becomes:

$$Var\left(X_t^l|\theta_t\right) \to \left(\frac{\gamma_S+\gamma_L}{\gamma_L}\right)^2\frac{1}{n_t\gamma_D}$$

where $n_t = \frac{N_t}{\bar{N}}$.

Hence the precision of the labor signal X_t^l is $n_t\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D$.

Proof of Lemma 3.3

The AI signal data is given by

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

Given that S_t and $\mu_t = E\left(\theta_t|\Omega_{t-1}\right)$ are public information, the valuable piece in observing D_t^i is

$$\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}S_t^{Ai} + \varepsilon_t^{Di}$$

which is equivalent to

$$S_t^{Ai} + \frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\varepsilon_t^{Di}$$

Hence the AI signal can be summarized as:

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

Hence the variance of the AI signal becomes:

$$\begin{aligned} &\frac{1}{\left(\bar{N}-N_t\right)^2}\left(\bar{N}-N_t\right)\frac{1}{\gamma_A} + \left(\frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\right)^2\frac{1}{\left(\bar{N}-N_t\right)^2}\left(\bar{N}-N_t\right)\frac{\bar{N}}{\gamma_D} \\ = &\frac{1}{\left(\bar{N}-N_t\right)}\frac{1}{\gamma_A} + \left(\frac{\gamma_S+\gamma_A+\gamma_t}{\gamma_A}\right)^2\frac{\bar{N}}{\left(\bar{N}-N_t\right)}\frac{1}{\gamma_D} \end{aligned}$$

When the number of entrepreneurs tends to infinity $\bar{N}$ and $N_t \to \infty$, the variance becomes

$$\left(\frac{\gamma_S + \gamma_A + \gamma_t}{\gamma_A}\right)^2 \frac{1}{(1-n_t)\,\gamma_D}$$

where $n_t = \frac{N_t}{\bar{N}}$. Hence the precision of the AI signal is

$$(1-n_t)\left(\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t}\right)^2 \gamma_D$$

Proof of Lemma 3.4

We start with the recursion:

$$\begin{aligned} Var(\theta_t|\Omega_t) &= Var\left(\theta_t|\Omega_{t-1}, S_t, X_t^l, X_t^A\right) \\ &= \frac{1}{\gamma_t + \gamma_S + n_t\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D + (1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\gamma_D} \end{aligned}$$

where the second equality follows from Lemma 3.2 and 3.3.

$$\frac{1}{\gamma_{t+1}} = Var\left(\theta_{t+1}|\Omega_t\right) = Var\left(\rho\theta_t + \eta_{t+1}|\Omega_t\right) = \rho^2 Var\left(\theta_t|\Omega_t\right) + \frac{1}{\gamma_\eta}$$

Plugging in the expression of $Var(\theta_t|\Omega_t)$, we obtain Lemma 3.4.

Proof of Theorem 3.1

To prove the theorem we need to evaluate the partial derivative of γ_{t+1} with respect to n_t :

$$\frac{\partial\gamma_{t+1}}{\partial n_t} \propto \left[\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 - \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2\right]\gamma_D$$

which is positive if and only if

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

Or

$$\frac{\gamma_L}{\gamma_S+\gamma_L} - \frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t} > 0$$

Or

$$\gamma_L\left(\gamma_S+\gamma_A+\gamma_t\right) > \gamma_A(\gamma_S+\gamma_L)$$

Or

$$\gamma_L\gamma_S + \gamma_L\gamma_t > \gamma_A\gamma_S$$

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

This is the condition 15.

Proof of Theorem 3.4

We take the labor signal quality to be fixed and the AI signal quality $\gamma_A \to \infty$, and therefore $\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t} \to 1$. The law of motion for aggregate data quality becomes:

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

We assume that the labor cost distribution is compressed, so that small changes in the AI productivity would lead to large changes in the share of AI adoption.

We now construct multiple steady state as follows. Imagine a steady state where data quality is low, $\gamma = \gamma_1$, and so that everyone picks labor production:

$$n_t(\gamma_1) = 1$$

Then γ_1 can be solved using the steady state equation:

$$\frac{1}{\gamma_1} = \rho^2 \frac{1}{\gamma_1 + \gamma_S + \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}$$

For another steady state with high data quality $\gamma = \gamma_2$, everyone chooses AI production:

$$n_t(\gamma_2) = 0$$

Hence the equation to solve for γ_2 is:

$$\frac{1}{\gamma_2} = \rho^2 \frac{1}{\gamma_2 + \gamma_S + \gamma_D} + \frac{1}{\gamma_\eta}$$

Thus, to prove the existence of multiple steady states all we need to show is that

$$\gamma_1 < \gamma_2$$

Then for some very compressed labor cost distribution we can have $n_t(\gamma_1) = 1$ and $n_t(\gamma_2) = 0$ given that $n(.)$ is decreasing in γ_t. This confirms the existence of multiple steady states.

To show that it suffices to show that the reasonable root of the following function about x is increasing in coefficient $m > 0$:

$$\frac{1}{x} = \rho^2 \frac{1}{x+m} + \frac{1}{\gamma_\eta} \tag{39}$$

If this can be proved, then given that $\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 < 1$, and thus $\gamma_S + \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D < \gamma_S + \gamma_D$, we can prove the claim that $\gamma_1 < \gamma_2$.

Note that equation 39 can be arranged as the following quadratic equation:

$$x + m = \rho^2 x + \frac{1}{\gamma_\eta} x(x+m)$$

$$\frac{1}{\gamma_\eta} x^2 + \left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) x - m = 0$$

and because $\frac{1}{\gamma_\eta} > 0$ and $m > 0$, this quadratic equation has a negative and a positive root, so the valid root is the positive one, or the larger root:

$$x = \frac{-\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + \sqrt{\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}}}{2\frac{1}{\gamma_\eta}}$$

We need to show that x is increasing in m.

Take the derivative:

$$\begin{aligned} x'(m) &= \frac{-\frac{1}{\gamma_\eta} + \frac{1}{2}\left[\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}\right]^{-\frac{1}{2}}\left(2\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)\frac{1}{\gamma_\eta} + 4\frac{1}{\gamma_\eta}\right)}{2\frac{1}{\gamma_\eta}} \\ &= \frac{-1 + \left[\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}\right]^{-\frac{1}{2}}\left(\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 2\right)}{2} \end{aligned}$$

Thus it suffices to show that

$$-1 + \left[\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}\right]^{-\frac{1}{2}}\left(\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 2\right) > 0$$

Or

$$\left[\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}\right]^{-\frac{1}{2}}\left(\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 2\right) > 1$$

Or

$$\left(\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 2\right) > \left[\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}\right]^{\frac{1}{2}}$$

Or

$$\left(\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 2\right)^2 > \left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}$$

Or

$$\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right) + 4 > \left(\rho^2 - 1 + \frac{m}{\gamma_\eta}\right)^2 + 4\frac{m}{\gamma_\eta}$$

Or

$$\rho^2 - 1 + \frac{m}{\gamma_\eta} + 1 > \frac{m}{\gamma_\eta}$$

Or

$$\rho^2 > 0$$

which is true by assumption. Hence we have proved the claim and the existence of multiple steady states is established.

Proof of Theorem 6.1

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

Hence we need the tax rate to be such that:

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

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

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 $\gamma'(\bar{\varphi})$. Let us determine the sign of it. Write the constraint of the social planner as:

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

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

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

Hence

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

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

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

because

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

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

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

Hence if

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

we can show that

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

, then

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

which means that

$$\tau>0$$

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

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

Then one can show that

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

and hence

$$\tau<0$$

A.1 Derivation of the Cournot Markup

There are N symmetric AI providers, each producing at constant marginal cost c_t. Let Q_j denote the quantity supplied by firm j, and let $Q_{-j}=\sum_{k\neq j}Q_k$ denote the aggregate supply of its competitors. The inverse demand function, denoted $R(\cdot)$, maps total supply to the market-clearing price: $R_t=R(Q_j+Q_{-j})$.

Firm j solves:

$$\max_{Q_j}\ Q_j\big[R(Q_j+Q_{-j})-c_t\big].$$

The first-order condition is:

$$\big[R(Q_j+Q_{-j})-c_t\big]+Q_j\,R'(Q_j+Q_{-j})=0.$$

In the symmetric equilibrium, each firm produces $Q_j=Q$ and total supply is NQ, so $D(R_t,\gamma_t)=NQ$. Substituting:

$$(R_t-c_t)+Q\,R'(NQ)=0.$$

By the inverse function theorem, $R'(NQ)=1/D_R(R_t,\gamma_t)$, where $D_R\equiv\partial D/\partial R_t$. From equation

(22), $D_R = -f(\bar{\varphi}_t)$. Substituting $Q = D(R_t, \gamma_t)/N = [1 - F(\bar{\varphi}_t)]/N$:

$$(R_t - c_t) + \frac{1 - F(\bar{\varphi}_t)}{N} \cdot \frac{1}{-f(\bar{\varphi}_t)} = 0,$$

which rearranges to:

$$R_t - c_t = \frac{1 - F(\bar{\varphi}_t)}{N \cdot f(\bar{\varphi}_t)}.$$

When $N = 1$, this yields the monopoly markup $R_t^m - c_t = \frac{1-F(\bar{\varphi}_t^m)}{f(\bar{\varphi}_t^m)}$. As $N \to \infty$, the markup vanishes and $R_t \to c_t$, recovering the competitive benchmark. □

B Endogenous AI Signal Quality and Robustness

This section offers two extensions to the base model to examine its robustness. We first relax the assumption that entrepreneurs, upon birth, cannot observe any prior history. Instead, we assume that they are born with the ability to observe a certain fraction of past information. We also relax the assumption that the private data generated by AI is unrelated to past history, assuming instead that the quality of the data 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 $\zeta < 1$. 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 super-

script e) is given by:

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

Note that the superscript for the number of noises now spans from 1 to $\zeta\bar{N} < \bar{N}$. 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:

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

where

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

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

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:

$$\gamma_A\left(\gamma_t\right) = a + b\gamma_t \tag{40}$$

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 10 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 $\frac{\gamma_L}{\gamma_S + \gamma_L + \gamma_t^e} s_t^{li}$. Thus, it can be shown that the summary statistics for labor is (taking $\bar{N} \to \infty$):

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

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 40. 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 $\bar{N} \to \infty$):

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

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:

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_s + n_t \left(\frac{\gamma_L}{\gamma_S+\gamma_L+\gamma_t^e} \right)^2 \gamma_D + (1-n_t) \left(\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t} \right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{41}$$

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

$$\frac{1}{\gamma_{t+1}^e} = \rho^2 \frac{1}{\gamma_t^e + \gamma_s + \zeta n_t \left(\frac{\gamma_L}{\gamma_S+\gamma_L+\gamma_t^e} \right)^2 \gamma_D + \zeta(1-n_t) \left(\frac{\gamma_A(\gamma_t)}{\gamma_S+\gamma_A(\gamma_t)+\gamma_t} \right)^2 \gamma_D} + \frac{1}{\gamma_\eta} \tag{42}$$

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 inequal-

ity 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

$$\bar{\varphi}(\gamma_t, \gamma_t^e) = R_t + \frac{1}{\gamma_S + \gamma_A(\gamma_t) + \gamma_t} - \frac{1}{\gamma_S + \gamma_L + \gamma_t^e} \tag{43}$$

Thus model dynamics is fully characterized by equations 41 and 42, 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 43.

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 16 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, we set $\zeta = 0.01$, 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.[34]

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 ζ. The bottom right panel

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

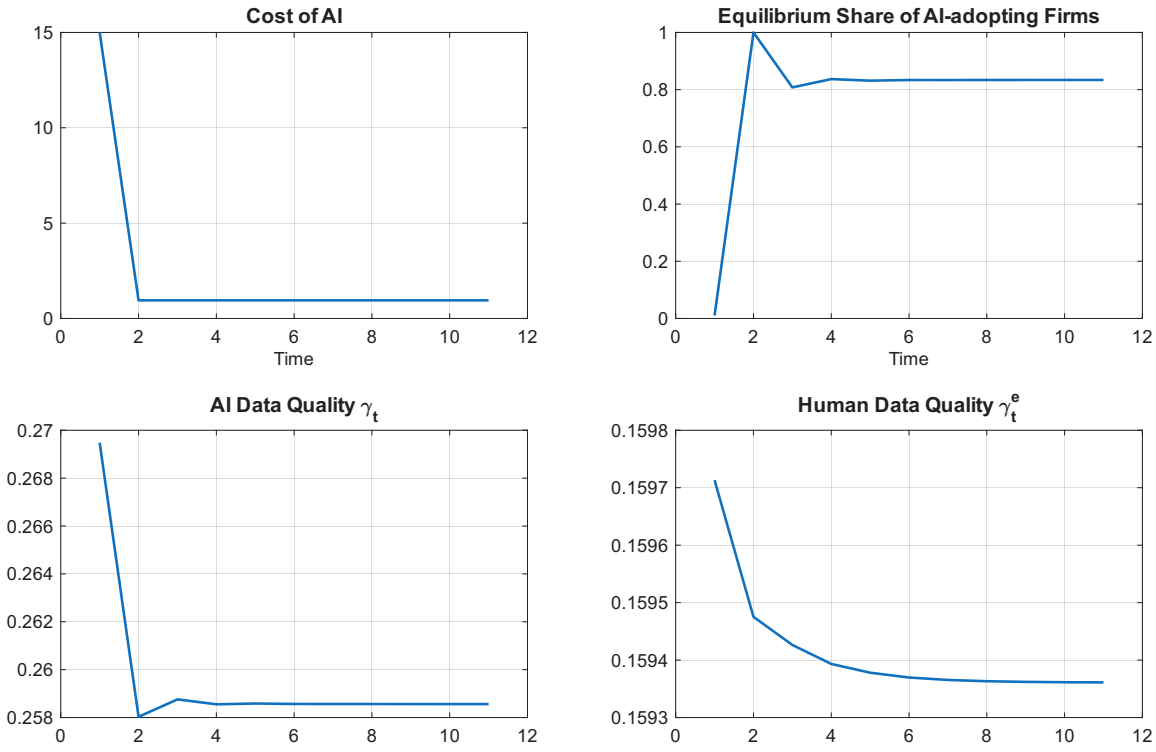

Figure 16: **Model Dynamics with Endogenous AI and Entrepreneur Information Stocks.** The figure extends the ChatGPT experiment to a setting with two endogenous state variables: AI information quality γ_t and entrepreneur information quality γ_t^e. Following the cost reduction, AI adoption rises sharply and then partially reverses, while both information stocks decline relative to their pre-shock levels. The reversal mechanism therefore survives when human-side information accumulation is also endogenous.

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 $\zeta < 1$, and therefore the productivity of labor is less affected by the endogenous evolution of data quality.

C Finite Lookback Window

In the main text, the law of motion for data quality (Lemma 3.4) is derived in the limit where AI has access to the entire history of past data ($T \to \infty$). Here we derive the law of motion when AI is restricted to a finite lookback window of T periods.

With a lookback window of T, the information set at time t consists of the T most recent periods of public signals and data: $\Omega_t^T = \{S_{t-j}, X_{t-j}^l, X_{t-j}^A\}_{j=0,1,\ldots,T-1}$. The key difference from the full-history case is that when transitioning from Ω_{t-1}^T to Ω_t^T, the agent gains the new period-t signals $\{S_t, X_t^l, X_t^A\}$ but loses the oldest signals $\{S_{t-T}, X_{t-T}^l, X_{t-T}^A\}$ that fall outside the window.

We now derive the posterior variance step by step. Start with the posterior of θ_t given the finite-lookback information set:

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

Under the full history ($T = \infty$), this would equal $1/(\gamma_t + \gamma_S + \Delta_t^{\text{new}})$ as in the proof of Lemma 3.4, where $\Delta_t^{\text{new}} = n_t \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\gamma_t}\right)^2 \gamma_D$. With a finite lookback window, however, the prior $\gamma_t = 1/Var(\theta_t|\Omega_{t-1}^T)$ is less precise than under the full history because $\Omega_{t-1}^T \subset \Omega_{t-1}$.

To characterize γ_t under a finite window, note that Ω_{t-1}^T contains signals from periods $t-1, t-2, \ldots, t-T$, while Ω_{t-2}^T contains signals from $t-2, t-3, \ldots, t-T-1$. Thus going from Ω_{t-2}^T to Ω_{t-1}^T, the agent gains $\{S_{t-1}, X_{t-1}^l, X_{t-1}^A\}$ (with combined precision $\gamma_S + \Delta_{t-1}^{\text{new}}$) but loses $\{S_{t-T-1}, X_{t-T-1}^l, X_{t-T-1}^A\}$. Since signals from period $t-T-1$ provide information about θ_{t-T-1}, and $\theta_{t-1} = \rho^T \theta_{t-T-1} + \sum_{j=0}^{T-1} \rho^j \eta_{t-1-j}$, the lost signals carry only ρ^{2T} times their original precision for predicting θ_{t-1}. Therefore:

$$Var(\theta_{t-1}|\Omega_{t-1}^T) \quad = \quad \frac{1}{\gamma_{t-1} + \gamma_S + \Delta_{t-1}^{\text{new}} - \rho^{2T}\left[\gamma_S + \Delta_{t-T-1}^{\text{old}}\right]}.$$

Applying the AR(1) transition once more, $Var(\theta_{t+1}|\Omega_t^T) = \rho^2 Var(\theta_t|\Omega_t^T) + 1/\gamma_\eta$, we obtain the

law of motion for data quality under a finite lookback window:

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_S + \Delta_t^{\text{new}} - \rho^{2T}\left[\gamma_S + \Delta_{t-T}^{\text{old}}\right]} + \frac{1}{\gamma_\eta}, \tag{44}$$

where

$$\begin{aligned}
\Delta_t^{\text{new}} &= n_t \left(\frac{\gamma_L}{\gamma_S + \gamma_L}\right)^2 \gamma_D + (1 - n_t)\left(\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_t}\right)^2 \gamma_D, \\
\Delta_{t-T}^{\text{old}} &= n_{t-T} \left(\frac{\gamma_L}{\gamma_S + \gamma_L}\right)^2 \gamma_D + (1 - n_{t-T})\left(\frac{\gamma_A}{\gamma_S + \gamma_A + \gamma_{t-T}}\right)^2 \gamma_D
\end{aligned}$$

denote the data precision contributions at dates t and $t - T$, respectively.

As $T \to \infty$, the correction term $\rho^{2T}[\gamma_S + \Delta_{t-T}^{\text{old}}] \to 0$ because $\rho < 1$, and equation (44) reduces to the law of motion in Lemma 3.4. Thus, the infinite-lookback limit is well-defined and the tractable recursive form we employ in the main text provides a good approximation whenever T is sufficiently large.

D Data Access Policy

In addition to the AI adoption policy analyzed in the main text, the social planner can also consider restricting AI's access to historical data. Specifically, the planner can control the subset of historical data that AI is allowed to access:

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

where Ω_t^A represents the subset of historical data available to AI. The AI's decision problem under restricted data access is:

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

The following lemma highlights the potential tradeoff faced by the social planner when considering restricting AI's access to historical data:

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

$$a_t^i = E(\theta_t|\Omega_{t-1}^A, S_t, s_t^{Ai}) = \frac{\gamma_S}{\gamma_S + \gamma_A + \kappa\gamma_t} S_t + \frac{\kappa\gamma_t}{\gamma_S + \gamma_A + \kappa\gamma_t}\mu_t + \frac{\gamma_A}{\gamma_S + \gamma_A + \kappa\gamma_t} s_t^{Ai}, \tag{45}$$

and the value of AI production is given by:

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

where κ is the ratio of conditional precision given Ω_{t-1} and Ω^A_{t-1}:

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

where the last inequality is due to $\Omega^A_{t-1} \subseteq \Omega_{t-1}$.

The direct cost of limiting AI's access to data is that it lowers AI productivity, as seen in equation 45, where introducing $\kappa < 1$ reduces the value of V^A. What is important is not the specific data points AI can access but the overall quality of the dataset. Therefore, the planner's decision comes down to choosing the optimal value of κ.

From an individual perspective, there is no advantage to restricting AI's access to data because V^A increases with κ. However, from a societal standpoint, there is a potential benefit. Restricting AI's access to public data (lowering κ) forces the AI to rely more on its privately-generated information. This can be seen in equation 45, where the weight on AI's private signal s^{Ai}_t, $\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}$, increases as κ decreases. Limiting AI's access to public data "forces" it to focus more on its own private information, potentially enhancing the quality of the overall dataset, especially in an economy where AI plays a significant role:

Lemma D.2 *Given the current data quality γ_t and some share of labor adoption n_t, from the law of motion with κ:*

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

reducing κ leads to higher future data quality (increasing γ_{t+1}).

From an individual AI user's perspective, it is always optimal to use 100% of the data available:

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

However, the social planner recognizes that the choice of κ also impacts aggregate data quality:

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

where $\gamma'(\kappa) < 0$ 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 D.2).

Thus optimal κ solves:

$$\max_\kappa \kappa\gamma(\kappa),$$

which suggests that setting κ to less than 100% can potentially improve welfare. The first-order condition is straightforward:

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

The first term, γ, represents the cost of restricting AI's access to data—the higher the quality of the available dataset, the greater the cost of limiting its use. The second term reflects the societal benefit of restricting AI's data access, as this can improve overall data quality and increase the productivity of future AI systems.

Theorem D.1 *With data corruption, long-run data quality decreases. Given this reduction in quality, the cost of limiting AI's access to data becomes smaller, making such a policy more likely to be socially desirable.*

In the presence of data corruption, the quality of data deteriorates over time. As a result, the cost of restricting AI's access to data decreases. On the other hand, the benefit of limiting data access is less sensitive to changes in data quality. Therefore, as the value of γ declines, the net benefit of constraining AI's access to data increases, making it more likely that such restrictions can improve social welfare.

Endogenous AI signal quality. A natural concern is that restricting AI's data access ($\kappa < 1$) could also impair AI's ability to generate novel private signals—that is, γ_A itself may depend on the richness of the data environment available to AI. For instance, modern generative AI systems trained on larger and more diverse datasets may develop stronger capabilities for producing genuinely new insights (as emphasized in recent advances in reasoning models). If so, restricting data access could simultaneously (i) improve data quality through the composition channel identified in Theorem D.1, but also (ii) reduce γ_A, degrading AI's innovative capacity.

To analyze this channel formally, we adopt the functional form introduced in Appendix B and allow AI's signal quality to depend on the data environment AI has access to:

$$\gamma_A(\kappa\gamma_t) = a + b \cdot \kappa\gamma_t, \tag{46}$$

where $a \geq 0$ represents AI's baseline innovative capacity (independent of historical data) and $b \geq 0$ captures the degree to which richer data environments enhance AI's ability to generate novel signals. When $b = 0$, we recover the baseline model where γ_A is exogenous and the analysis

of Theorem D.1 applies directly. When $b > 0$, restricting data access ($\kappa < 1$) reduces AI's signal quality, creating an additional cost of data restriction.

With endogenous γ_A, the law of motion for data quality (Lemma D.2) becomes

$$\frac{1}{\gamma_{t+1}} = \rho^2 \frac{1}{\gamma_t + \gamma_S + n_t \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n_t)\left(\frac{a+b\kappa\gamma_t}{\gamma_S+a+(1+b)\kappa\gamma_t}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta}, \tag{47}$$

where we have substituted $\gamma_A(\kappa\gamma_t) = a + b\kappa\gamma_t$ into the AI information term. The key quantity in this law of motion is the informational content of each AI-generated observation:

$$\left(\frac{a + b\kappa\gamma_t}{\gamma_S + a + (1+b)\kappa\gamma_t}\right)^2 \gamma_D. \tag{48}$$

This expression captures the amount of novel information each AI data point contributes to the aggregate dataset. To understand how data restriction ($\kappa < 1$) affects data quality through the law of motion (47), consider how κ changes this informational content. Differentiating the ratio inside the square with respect to κ:

$$\frac{\partial}{\partial\kappa}\left(\frac{a + b\kappa\gamma}{\gamma_S + a + (1+b)\kappa\gamma}\right) = \frac{\gamma(b\gamma_S - a)}{(\gamma_S + a + (1+b)\kappa\gamma)^2}. \tag{49}$$

The sign of this expression is determined by whether $b\gamma_S - a$ is positive or negative, giving rise to two cases:

Case 1: $a > b\gamma_S$ (baseline innovative capacity dominates). The derivative in equation (49) is negative, so reducing κ increases the informational content in equation (48). Intuitively, AI's signal quality $\gamma_A = a + b\kappa\gamma$ has a large exogenous component a that does not depend on data access. Restricting data forces AI to rely more on this inherent capability, and the numerator $a + b\kappa\gamma$ falls by less (proportionally) than the denominator $\gamma_S + a + (1+b)\kappa\gamma$. Thus each AI observation becomes more informative, raising γ_{t+1} in equation (47). The qualitative conclusion from Theorem D.1 carries over: restricting data access improves data quality, and the benefit of restriction grows as data corruption worsens.

Case 2: $a < b\gamma_S$ (signal quality highly sensitive to data environment). The derivative in equation (49) is positive, so reducing κ decreases the informational content in equation (48). Here AI's signal quality depends heavily on the data it has access to (b is large relative to a). Restricting data sharply reduces γ_A, causing the numerator to fall proportionally more than the denominator. Each AI observation becomes *less* informative, potentially lowering γ_{t+1}. This works against the composition-channel benefit identified in Theorem D.1, and data restriction may reduce welfare.

The role of γ_S in the condition $a \gtrless b\gamma_S$ deserves brief comment. In the weight AI places on its private signal, $\frac{a+b\kappa\gamma}{\gamma_S+a+(1+b)\kappa\gamma}$, the public signal precision γ_S appears in the denominator but not in the numerator, and it does not respond to κ. When κ is reduced, both the numerator and parts of the denominator shrink—but γ_S stays fixed. When γ_S is large, this fixed component anchors the denominator, preventing it from shrinking proportionally as much as the numerator. As a result, the weight falls and restriction makes AI *less* reliant on its private signal. Conversely, when γ_S is small, the denominator responds strongly to κ, allowing the weight to rise so that restriction works as intended.

The condition $a \gtrless b\gamma_S$ thus determines whether the main results on data access policy are reinforced or overturned by endogenous signal quality. Turning to the planner's optimal choice of κ, note that the total precision available to AI under endogenous γ_A is $\gamma_S + a + (1+b)\kappa\gamma$. Since both the direct-access term ($\kappa\gamma$) and the signal-quality term ($\gamma_A = a + b\kappa\gamma$) scale with κ, the planner's problem is equivalent to maximizing $(1+b)\kappa\gamma(\kappa)$ (up to the constant $\gamma_S + a$). Differentiating and simplifying, the first-order condition reduces to the same form as the exogenous case:

$$\underbrace{\gamma}_{\text{Cost of Limiting Data Access}} + \underbrace{\kappa\frac{\partial\gamma}{\partial\kappa}}_{\text{Benefit of Limiting Data Access}} = 0, \tag{50}$$

because the $(1+b)$ factor multiplies both terms and cancels. The structure of the trade-off is unchanged: the first term is the cost of restricting data access, and the second term is the benefit through improved data quality. However, the sign and magnitude of $\frac{\partial\gamma}{\partial\kappa}$ now depend on which case applies. In Case 1, $\frac{\partial\gamma}{\partial\kappa} < 0$ as in the exogenous model, so restriction is beneficial. In Case 2, $\frac{\partial\gamma}{\partial\kappa}$ may be positive, making full access ($\kappa = 1$) optimal.

Proof of Lemma D.1. Denote $\gamma' = \frac{1}{Var\left(\theta_t|\Omega^A_{t-1}\right)}$. Hence we have

$$\kappa = \frac{\gamma'}{\gamma_t}$$
$$\gamma' = \kappa\gamma_t$$

Then the optimal action by AI is:

$$\begin{aligned} a^i_t &= E\left(a^i_t|\Omega^A_{t-1}, S_t, s^{Ai}_t\right) \\ &= \frac{\gamma_S}{\gamma_S+\gamma_A+\gamma'}S_t + \frac{\gamma'}{\gamma_S+\gamma_A+\gamma_t}\mu_t + \frac{\gamma_A}{\gamma_S+\gamma_A+\gamma'}S^{Ai}_t \\ &= \frac{\gamma_S}{\gamma_S+\gamma_A+\kappa\gamma_t}S_t + \frac{\kappa\gamma_t}{\gamma_S+\gamma_A+\gamma_t}\mu_t + \frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}S^{Ai}_t \end{aligned}$$

Similarly we can easily derive:

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

□

Proof of Lemma D.2. Define data quality in the same way as before:

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

Following a similar process as in the Lemma 3.3, one can show that X_t^A is an unbiased signal of θ_t with precision

$$(1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}\right)^2\gamma_D$$

where $\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}$ is the loading of AI's action on its privately generated data (see Lemma D.1). Thus

$$\begin{aligned}
\frac{1}{\gamma_{t+1}} &= Var\left(\theta_{t+1}|\Omega_t\right) \\
&= \rho^2 Var\left(\theta_t|\Omega_t\right) + \frac{1}{\gamma_\eta} \\
&= \rho^2 Var\left(\theta_t|\Omega_{t-1}, S_t, X_t^l, X_t^A\right) + \frac{1}{\gamma_\eta} \\
&= \rho^2 \frac{1}{\gamma_t + \gamma_S + n_t\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D + (1-n_t)\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma_t}\right)^2\gamma_D} + \frac{1}{\gamma_\eta}
\end{aligned}$$

Given that κ only shows up in the precision of the AI signal X_t^A, it is straightforward to see that γ_{t+1} increases when κ decreases, holding fixed other equilibrium quantities. □

Proof of Theorem D.1. The tradeoff of κ is that reducing κ tends to reduce the value of AI adoption, but at the same time allows AI to act more aggressively upon its own private information, thus integrating more information into aggregate knowledge, raising the value of γ. In other words:

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

The law of motion for γ can be written compactly as $\gamma(\bar{\varphi}, \kappa)$. The optimal κ solves $\max_\kappa \kappa\gamma(\bar{\varphi}, \kappa)$, yielding:

$$\underbrace{\gamma}_{\text{Cost of reducing } \kappa} + \underbrace{\kappa\frac{\partial\gamma}{\partial\kappa}}_{\text{Benefit of reducing } \kappa} = 0.$$

We now derive $\frac{\partial \gamma}{\partial \kappa}$ explicitly. Define the law of motion for γ as:

$$H(\gamma, \bar{\varphi}, \kappa) = \rho^2 \frac{1}{\gamma + \gamma_S + n * \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n) * \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D} + \frac{1}{\gamma_\eta} - \frac{1}{\gamma} = 0$$

$$\frac{\partial \gamma}{\partial \kappa} = -\frac{\frac{\partial H}{\partial \kappa}}{\frac{\partial H}{\partial \gamma}} = -\frac{\rho^2 \frac{1}{\left[\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right]^2} (1-n) * \gamma_A^2 \frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^4} 2\left(\gamma_S+\gamma_A+\kappa\gamma\right) * \gamma * \gamma_D}{\rho^2 \frac{-1}{\left(\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right)^2} + \frac{1}{\gamma^2}}$$

Note that

$$\frac{\partial H}{\partial \kappa} = \rho^2 \frac{1}{\left[\gamma + \gamma_s + n * \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n) * \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right]^2} (1-n) * \gamma_A^2 \frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^4} 2\left(\gamma_S+\gamma_A+\kappa\gamma\right)$$

$$\frac{\partial H}{\partial \gamma} = \rho^2 \frac{-\left(1-2\left(1-n\right)\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\frac{\gamma_A}{(\gamma_S+\gamma_A+\kappa\gamma)^2}\kappa\gamma_D\right)}{\left(\gamma + \gamma_S + n * \left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D + (1-n) * \left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right)^2} + \frac{1}{\gamma^2} > 0$$

Hence

$$\frac{\partial \gamma}{\partial \kappa} < 0$$

Thus the optimal condition is given by

$$\frac{\partial \gamma}{\partial \kappa} = -\frac{\gamma}{\kappa}$$

Let's simplify the derivative $\frac{\partial \gamma}{\partial \kappa}$ and see how it changes with γ :

$$\frac{\partial \gamma}{\partial \kappa} = -\frac{\rho^2 \frac{2(1-n)*\gamma_A^2 \frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^3}\gamma*\gamma_D}{\left[\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right]^2}}{\rho^2 \frac{-\left(1-2(1-n)\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\frac{\gamma_A}{(\gamma_S+\gamma_A+\kappa\gamma)^2}\kappa\gamma_D\right)}{\left(\gamma+\gamma_s+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right)^2} + \frac{1}{\gamma^2}}$$

$$= -\frac{\gamma}{\kappa}\frac{\kappa}{\gamma}\frac{\rho^2 \frac{2(1-n)*\gamma_A^2 \frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^3}\gamma*\gamma_D}{\left[\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right]^2}}{\rho^2 \frac{-\left(1-2(1-n)\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\frac{\gamma_A}{(\gamma_S+\gamma_A+\kappa\gamma)^2}\kappa\gamma_D\right)}{\left(\gamma+\gamma_s+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right)^2} + \frac{1}{\gamma^2}}$$

$$= -\frac{\gamma}{\kappa}\frac{\rho^2 \frac{2(1-n)*\gamma_A^2 \frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^3}\gamma*\gamma_D*\kappa}{\left[\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right]^2}}{\rho^2 \frac{-\left(\gamma-2(1-n)\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\frac{\gamma_A}{(\gamma_S+\gamma_A+\kappa\gamma)^2}\kappa\gamma\gamma_D\right)}{\left(\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2 \gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2 \gamma_D\right)^2} + \frac{1}{\gamma}}$$

The denominator becomes:

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

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

where N is the numerator: $N = \rho^2 \frac{2(1-n)*\gamma_A^2\frac{1}{(\gamma_S+\gamma_A+\kappa\gamma)^3}\gamma*\gamma_D*\kappa}{\left[\gamma+\gamma_S+n*\left(\frac{\gamma_L}{\gamma_S+\gamma_L}\right)^2\gamma_D+(1-n)*\left(\frac{\gamma_A}{\gamma_S+\gamma_A+\kappa\gamma}\right)^2\gamma_D\right]^2}$

Note from the law of motion of γ :

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

Thus

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

Substitute:

$$-\rho^2\gamma\left(\frac{\frac{1}{\gamma}-\frac{1}{\gamma_\eta}}{\rho^2}\right)^2+\frac{1}{\gamma}$$

Hence the derivative becomes:

$$\frac{\partial\gamma}{\partial\kappa} = -\frac{\gamma}{\kappa}\frac{N}{\frac{1}{\gamma}-\rho^2\gamma\left(\frac{\frac{1}{\gamma}-\frac{1}{\gamma_\eta}}{\rho^2}\right)^2+N}$$

Now, substitute in the first order condition:

$$-\frac{\gamma}{\kappa}\frac{N}{\frac{1}{\gamma}-\rho^2\gamma\left(\frac{\frac{1}{\gamma}-\frac{1}{\gamma_\eta}}{\rho^2}\right)^2+N} = -\frac{\gamma}{\kappa}$$

We have the following equation to pin down γ :

$$\frac{1}{\gamma}-\rho^2\gamma\left(\frac{\frac{1}{\gamma}-\frac{1}{\gamma_\eta}}{\rho^2}\right)^2 = 0$$

Hence this equation pins down the optimal γ when κ is endogenous. The optimum can be solved in a quadratic equation

$$(1-\rho^2)\left(\frac{1}{\gamma}\right)^2-2\frac{1}{\gamma_\eta}\frac{1}{\gamma}+\left(\frac{1}{\gamma_\eta}\right)^2 = 0$$

Hence

$$\begin{aligned}\left(\frac{1}{\gamma}\right)^* &= \frac{\frac{2}{\gamma_\eta} \pm \sqrt{\frac{4}{\gamma_\eta^2} - 4\left(1-\rho^2\right)\left(\frac{1}{\gamma_\eta}\right)^2}}{2\left(1-\rho^2\right)} \\ &= \frac{\frac{2}{\gamma_\eta} \pm \frac{2}{\gamma_\eta}\rho}{2\left(1-\rho^2\right)} = \frac{1}{\gamma_\eta}\frac{1}{1\pm\rho}\end{aligned}$$

The root should be

$$\left(\frac{1}{\gamma}\right)^* = \frac{1}{\gamma_\eta}\frac{1}{1-\rho}$$

because

$$\frac{1}{\gamma} > \frac{1}{\gamma_\eta}$$

Hence to maintain optimal level of γ^*, with data corruption, the value of κ needs to go down. Hence

$$\gamma^* = \gamma_\eta\left(1-\rho\right)$$

and we can check the conditions under which this is satisfied or not.

We start with a steady state where the cost of AI adoption is very high. Hence the share of AI adoption is minimal ($1 - n \approx 0$). In this steady state, the data quality is relatively high. Denote it by γ_h. Assume that $\gamma_h > \gamma^*$. To reduce the value of γ, we need to increase κ, but the maximum value of κ is 1. Hence at this steady state $\kappa = 1$.

We now examine a steady state with data corruption. From the numerical simulation we know that with data corruption, the AI adoption rate is positive: $\bar{\varphi} < \infty$ and $n < 100\%$. If the data corruption channel is sufficiently strong, without data regulation the value of steady state γ would drop by a lot. If $\gamma < \gamma^*$ as defined by previous equation, we need to set the value of κ to increase the value of γ. From the law of motion of γ and Lemma D.2, we know that:

$$\frac{\partial\gamma}{\partial\kappa} < 0$$

Hence to raise the value of γ we need κ to decrease, which implies that the social planner potentially needs to constrain AI's access to data to achieve the desirable level of γ. □

E Supplementary Quantitative Results

E.1 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 17, 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's action reveals less private information and contributes 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.

E.2 The Role of Public Information

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, we set the precision of the public signal γ_S to 1. With this change, other parameters

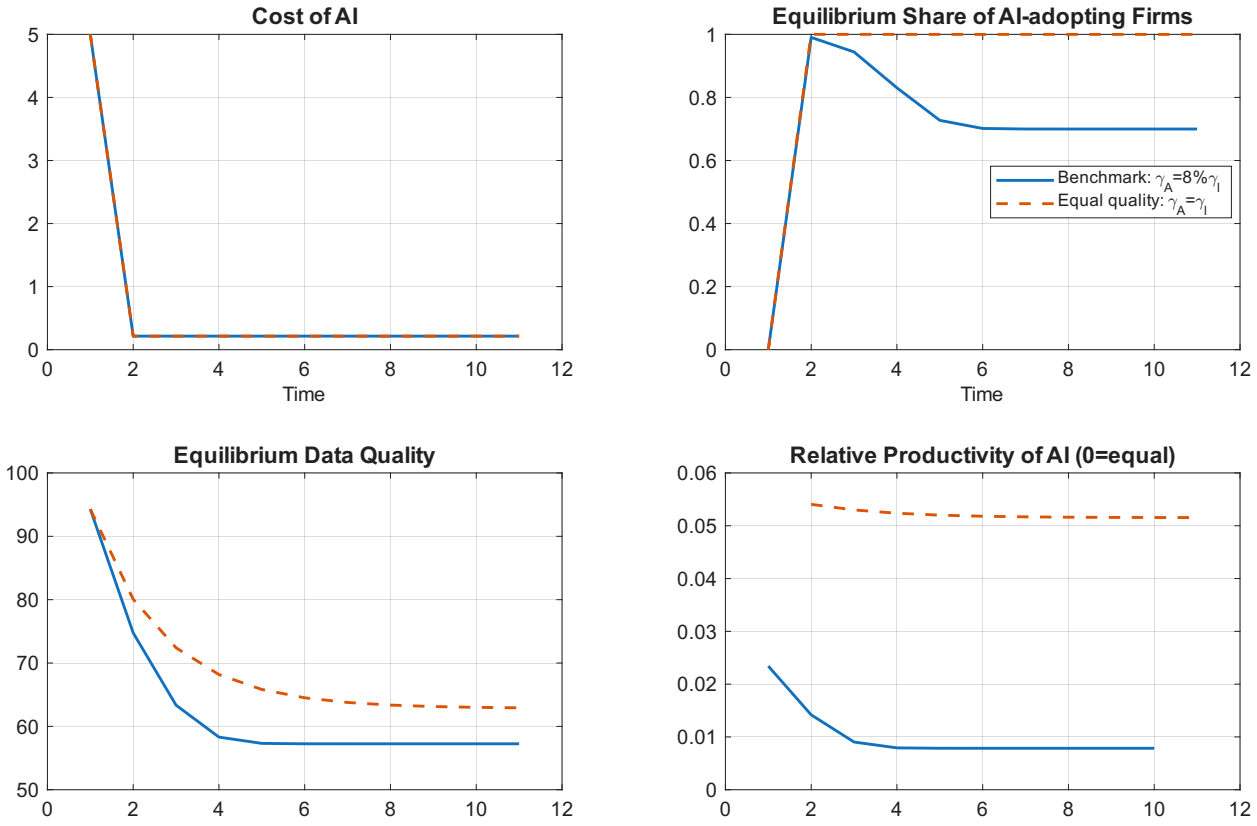

Figure 17: **The ChatGPT Experiment with Equal Information Quality.** Blue solid: baseline ($\gamma_A = 8$, data corruption). Red dashed: counterfactual ($\gamma_A = \gamma_L = 100$, equal quality). With equal quality, AI adoption remains near 100% with no reversal, and data quality declines only modestly. The 30% reversal in labor displacement observed in the baseline is entirely attributable to the data corruption channel.

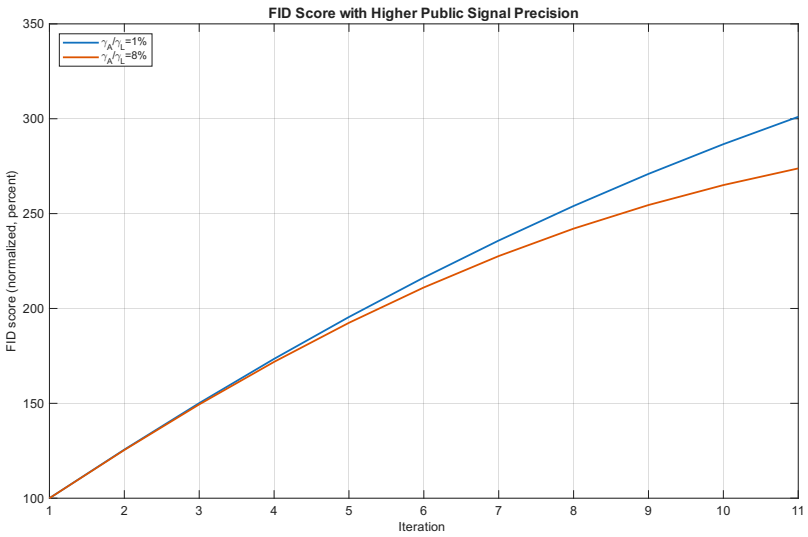

Figure 18: **FID Score with Higher Public Signal Precision ($\gamma_S = 1$).** Red: relative AI quality at baseline 8%—the FID increase falls short of 300% because higher public signal precision narrows the gap between human and AI data contributions. Blue: recalibrated AI quality at 1%, which recovers the 300% FID increase target.

need to be recalibrated, particularly the relative quality of information generated by AI versus human. Figure 18 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 18). All other parameters remain the same as the benchmark calibration.

Figure 19 presents the numerical results when we conduct the ChatGPT experiment under this scenario of 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, with the equilibrium share of AI-adopting firms stabilizing around 75%. This suggests that even under

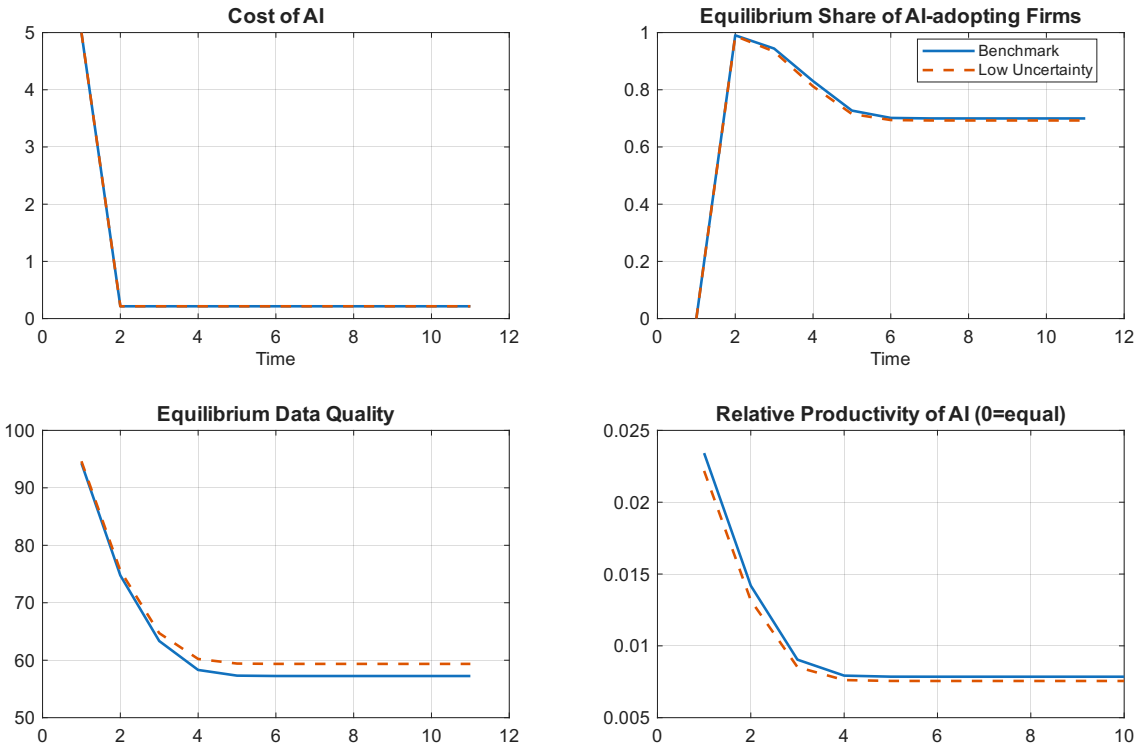

Figure 19: **Model Dynamics with Low Uncertainty ($\gamma_S = 1$).** Blue solid: benchmark calibration. Red dashed: low-uncertainty calibration with $\gamma_S = 1$ and recalibrated γ_A. Despite different parameter values, the labor displacement reversal is quantitatively similar (≈30%), confirming that the data corruption channel's magnitude is robust once calibrated to match FID score evidence.

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.

Table 1: **Selected Works on Data Corruption and AI Model Collapse.** The table summarizes representative papers from May 2023 to July 2024 that document recursive-training failures, synthetic-data degradation, and related quality losses in generative AI systems.

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 Models 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 is not meant to be exhaustive.