Will Human Creatives Survive Generative AI Shocks?*

Bo Hu† Jiasun Li‡

March 26, 2026

Abstract

The rise of generative AI has sparked urgent questions: With ongoing improvements, will it eventually replace all human creatives? Which tasks will survive, if any? How should workers and policymakers respond? This paper develops a flexible modeling framework to answer these questions while synthesizing the often-mixed empirical evidence. We first use a stylized static model to establish a robust human–AI coexistence that resembles the Grossman and Stiglitz (1980) impossibility of informationally efficient markets: Just as prices cannot fully aggregate private information—otherwise no investor would pay to acquire it, leaving nothing to aggregate—generative AI cannot fully displace human creators—otherwise they would stop producing fresh content, leaving AI to train on stale data that deteriorate performance, creating room for human re-entry. This coexistence is not a knife-edge result: it holds even at the theoretical capability limit of today's generative AI architectures. Building on this static foundation, we then introduce a dynamic renewal microfoundation and enrich the model with preference/skill heterogeneity and task complementarity to identify which human tasks will sustain in the generative-AI era. Overall, we provide a unified framework to analyze multiple aspects of human–generative AI interactions and labor market implications, with clean comparative statics and testable predictions that help reconcile conflicting empirical evidence and offer actionable guidance for creatives and policymakers alike.

Keywords: Generative AI, Grossman-Stiglitz Paradox, Human–AI Interaction, Information Aggregation, Large Language Models, Market Efficiency, Passive vs Active Investment.

*We thank Winston Dou, Zhiguo He, Richard Talley, Laura Veldkamp, Dacheng Xiu, Baozhong Yang, Liyan Yang, Bernard Yeung, and Ziyu Yao, as well as conference participants at Wharton Business & Generative AI Conference and Future Finance Fest (3f), as well as seminar attendants at George Mason University, University of Virginia (Darden), and Federal Reserve Bank of Cleveland for helpful discussions. Li appreciates financial support from U.S. National Science Foundation (Award #2337338). Part of the work was advanced during Li's visit to Carnegie Mellon and Yale, whose hospitality is greatly appreciated.

†bhu5@gmu.edu. George Mason University, 4400 University Drive, MSN 1B8, Fairfax, VA 22030, USA.

‡jli29@gmu.edu. George Mason University, 4400 University Drive, MSN 1B8, Fairfax, VA 22030, USA.

1 Introduction

Much attention has lately been drawn to generative AI technologies, which take natural-language instructions ("*prompts*") from users and generate coherent textual responses, graphical images, or video clips ("*completions*").[1] Although the pre-training processes required for developing large models that support generative AI applications are rather costly, once trained, these models can generate new content in response to prompts far more cost-efficiently than humans. This cost advantage has led many to argue that generative AI poses an existential threat to human content creators, including journalists, graphic designers, film directors, and even academic researchers. Will generative AI models take over creative jobs in the same way that automation and machines have taken over many forms of manual labor? Futuristically, as generative AI models further improve and reach their theoretical performance limit, will they completely replace human content creators? What aspects of human creative tasks may have the hope of surviving generative AI shocks, if any?

In this paper, we develop a simple yet technologically grounded modeling framework to address the various questions above. We first use a stylized static model as a baseline to prove the impossibility of a creatives market that is fully dominated by generative AI, analogous to the impossibility of a fully informationally efficient financial market as established by Grossman and Stiglitz (1980). Specifically, we point out that two well-established empirical patterns of generative AI models render an analogy between the human creatives-generative AI interaction and the relationship between active and passive investors in financial markets: First, the "scaling law" (e.g., Kaplan, McCandlish, Henighan, Brown, Chess, Child, Gray, Radford, Wu and Amodei (2020); Hoffmann, Borgeaud, Mensch, Buchatskaya, Cai,

[1] Examples include conversational AI bot ChatGPT, as well as many other transformer-based Large Language Model (LLM) applications, such as Google's Gemini (formerly Bard), Anthropic's Claude, Meta's Llama, diffusion-based text-to-image solutions like Stable Diffusion or DALL·E 3, and transformer/diffusion-based video generation applications such as OpenAI's Sora (recently shut down to refocus resources). See Section 2 for more details on generative AI technologies.

Rutherford, Casas, Hendricks, Welbl, Clark et al. (2022)), which dictates that AI model performance improves as the size of high-quality training data increases, has the financial market analogy of price informativeness improving as the size of well-researched active investors increases; Second, "model collapse" (or more generally "model decay"), which dictates that AI model performance degrades over time without fresh human content (e.g., Shumailov, Shumaylov, Zhao, Papernot, Anderson and Gal (2024)), has the financial market analogy of unattainable price informativeness without well-researched active investors. Therefore, just like how familiar insights in financial economics predict the coexistence of active and passive investors, we prove that the content market will eventually reach an interior equilibrium in which generative AI and human creatives coexist (even if the rise of generative AI may take over *some* human jobs). Section 2 will provide ample technological backgrounds to motivate the various assumptions underlying a stylized static model, which is analyzed in Section 3.

While our baseline model in Section 3 is highly stylized, its simplicity permits its key result on human–AI coexistence to be robust to various modeling variations. The simplicity of the baseline model also enables it to readily accommodate various extensions to address follow-up questions. To demonstrate such robustness and flexibility, we present in Section 4 three extensions to the baseline model, across which we consistently find that generative AI cannot fully replace human creatives. Besides illustrating the robustness of results and the flexibility of the modeling framework, these extensions also help shed light on many follow-up questions regarding the interaction between human and generative AI (such as what types of human creatives or creative tasks will or will not be replaced by generative AI), or provide dynamic microfoundations for the stylized static model. Results from these extensions also help reconcile the often-mixed empirical evidence so far in the literature, and offer actionable guidance for human creatives and policymakers alike.

Our first extension incorporates content consumers' heterogeneous preferences for content quality and also gives low-preference consumers an option to opt out of the content market.

We first show that for niche content, the introduction of generative AI may not disrupt the content creation market. This finding offers practical guidance to human creatives on how to respond to generative AI shocks. Then, in cases where generative AI does affect the market, we show that AI adoption has a "middle-out" pattern — that is, those with intermediate preferences for content quality are more likely to adopt generative AI, while those with extremely high preferences will stick with human content creation and those with extremely low preferences will continue to opt out. This pattern has both positive and negative implications. On the positive side, generative AI enhances inclusiveness by giving some low-type users access to content (thanks to lower-cost AI-generated content); On the negative side, however, AI adoption may degrade average content quality within the broader knowledge ecosystem. Finally, we find that the cost and value of human content creation may have a non-monotonic effect on equilibrium AI adoption because of the interaction of two countervailing forces: (1) a direct substitution effect between human and AI content creation; and (2) an indirect effect through the positive externality of human content creation on AI quality. We also analyze skill heterogeneity among human content producers and find a "bottom-up" pattern of AI adoption across skill levels, which contrasts with the "middle-out" pattern across consumer preference levels. Further extensions incorporating both consumer preference heterogeneity and producer skill heterogeneity to endogenously determine the price for human content and the market shares of human and AI content creation across both dimensions, we show that both the middle-out and the bottom-up pattern coexist simultaneously in equilibrium, with the former characterizing AI adoption on the demand side (across consumer preference types) and the latter characterizign AI disruption on the supply side (across producer skill levels).

The second extension dives into the multiple complementary sub-tasks often involved in content creation that may respond differently to scaling laws and model decay. For example, some sub-tasks, such as field investigation, require fresh and time-sensitive inputs, so a gen-

erative AI model trained on stale content may degrade quickly in quality; other sub-tasks, such as reasoning and writing, involve capabilities that are more stable over time, so a model trained on older content may remain useful for much longer. We show that in the presence of sub-task complementarity, depending on the prevailing technological parameters, the market may feature equilibrium with different characteristics. In most equilibria, AI quality improvement in one sub-task reinforces AI adoption in the other sub-task. However, when AI is already fully adopted (so that AI training is "saturated") in a time-insensitive sub-task (e.g., writing), further improvement in AI quality for this sub-task may nevertheless enhance the appeal of human efforts in the other more time-sensitive sub-task. For example, when generative AI receives mass adoption for writing, further improvements in AI's writing skills may attract more humans into investigative tasks. In this sense, AI adoption may feature heterogeneous patterns across sub-tasks and non-monotonic paces over time, which explains the often mixed empirical findings so far in the literature, and calls for more empirical findings as generative AI continues to advance.

The third extension provides a dynamic microfoundation for the static baseline model presented in Section 3, capturing the following intertemporal linkage underlying our key insight: how much value generative AI offers today depends on the size and quality of its training corpus, which in turn depends on how intensively humans have produced content in the past; on the other hand, how much value generative AI offers today also influences the extent of the AI adoption today, which in turn influences how much value generative AI will offer in the future. The dynamic extension incorporates this intertemporal linkage and characterizes the market shares of humans and generative AI in content creation as the underlying state of the world evolves stochastically over time, the needs for content arrive dynamically over time, and (in the absence of human inputs) the corpus used for training generative AI grows stale over time. Comparative statics from the resulting equilibrium suggest two broad career strategies for human creatives in the AI era. First, specialize

in "high-volatility" tasks where the underlying state changes quickly and AI grows stale rapidly (for example, investigative work, rapidly evolving policy areas, frontier research, and local or on-the-ground reporting). For such tasks an efficient equilibrium requires frequent human inputs, so demand for expert intervention remains structurally high. Second, human professionals should invest in lowering their own "intervention cost" by building deep domain expertise and by adopting tools and workflows that let them check, correct, and update AI outputs easily; those who can provide high-quality refreshes at low cost remain central to the ecosystem and directly improve both AI performance and overall information quality.

Our analysis provides numerous implications for key stakeholders. First, it delivers a both alarming and reassuring message: generative AI may replace some, but definitely not all, human content creators. We also offer guidance for human creatives on how to strategically shift their focus in the era of generative AI to better safeguard themselves against technological disruption. In particular, niche or time-sensitive content will likely be better sheltered—some may even become more valuable as AI quality in complementary tasks continues to improve. Second, given the possibility of different types of equilibria when content creation involves multiple complementary tasks, and given how these equilibria are influenced jointly by exogenous parameters such as advances in generative AI technologies or changes in consumer preferences and their distribution over time, we caution against over-extrapolating past patterns. This result also offers a starting point for policymakers to think of ways to induce appropriate environments to steer the society toward desirable equilibria. Third, given the potential trade-off between AI's benefits in expanding inclusiveness and its risks of degrading content quality, our model provides a foundation for a more comprehensive policy evaluation framework. Such a framework can help policymakers balance accessibility and quality considerations, especially as human creatives adjust their strategies in response to AI adoption. Finally, by emphasizing the role of consumer preferences in shaping market outcomes, our analysis suggests new avenues for marketing departments to explore consumer

behavior and segmentation in the age of generative AI.

To the best of our knowledge, this paper is the first attempt to use insights from financial and information economics and theoretically analyze the long-term societal impact of generative AI.[2] One example of a direct application is journalism (see Appendix A for a motivating example) where the underlying content space is highly time-sensitive and fast-moving, so generative AI models trained on stale information would quickly lose relevance. Other examples also include human knowledge-sharing platforms such as Wikipedia, Quora, and Stack Overflow (see Appendix B for more background information). Moreover, we expect our core message to also carry over to a broad class of activities involving human creativity and facing potential shocks from the advent of generative AI, while variations of our core model further tease out application-specific implications. To evaluate the scope in which our results apply, note that our theoretical framework operates through two distinct channels. The *scaling channel*—whereby AI quality depends on the volume of human-created training data—applies universally across content domains, from creative writing to software development. The *freshness channel*—whereby AI output degrades as the underlying state evolves—binds most tightly in time-sensitive domains such as news, financial analysis, and scientific research, but is less relevant for tasks with a slowly evolving knowledge base, such as mathematical exposition or literary translation, where freshness barely matters. Together, these channels delineate the spectrum of settings in which the human–AI coexistence mech-

[2]Recently, Agrawal, Gans and Goldfarb (2025) model how transformative AI reshapes labor allocation and displaces routine knowledge workers while human experts specialize in tasks farthest from existing knowledge; their focus is comparative advantage across tasks rather than an equilibrium feedback in which AI performance depends on incentivized human content creation. Demirer, Horton, Immorlica, Lucier, Shahidi and Welcome (2025) model how AI and task chaining reshape job structure and productivity, showing that step adjacency can override comparative advantage, and provides empirical evidence supporting nonlinear automation patterns; Hampole, Papanikolaou, Schmidt and Seegmiller (2025) develop a model that separates direct substitution from indirect labor reallocative effects of AI, and show how task-level exposure to AI affects within-firm labor demand and productivity patterns across occupations. These papers do not consider how the fresh human-created content constrains AI quality and sustains coexistence in equilibrium. In contrast, our paper formalizes a Grossman–Stiglitz-style mechanism in which AI lowers marginal production costs but relies on ongoing human content, delivers sharp coexistence conditions under heterogeneous preferences and task complementarity, and provides a tractable dynamic microfoundation for the human-AI ecosystem.

anism has the strongest bite.

Besides its implications for AI applications, our work also highlights the everlasting power of classic economic intuitions for understanding emerging AI technologies: while generative AI is a new phenomenon, the relevance of foundational insights in financial economics and information economics does not turn old.

Related literature The contributions of this paper are twofold. First, it analytically addresses a question of both immediate community interest and long-term societal significance. Second, it underscores the value of interdisciplinary thinking in the emerging field of generative AI. Straddling distinct areas of generative AI and financial economics, this paper thus contributes to multiple strands of the literature.

First, our results contribute to a rapidly growing literature studying the potential societal and economic impacts of generative AI. For example, on adoption patterns and productivity impacts, using survey data, Bick, Blandin and Deming (2026) find that AI adoption is concentrated among younger and higher-income households but below the very top of the income distribution. Combining evidence from three randomized controlled trials among 4,867 software developers at Microsoft, Accenture, and an anonymous Fortune 100 company, Cui, Demirer, Jaffe, Musolff, Peng and Salz (2025) find that less experienced developers had higher adoption rates and greater productivity gains. They also find that being provided access to an AI-based coding assistant drives a 26.08% increase in completed tasks. The latter result on productivity gains echoes Brynjolfsson, Li and Raymond (2025), who show that AI assistants in customer support raises productivity by 15%, with especially large gains for less experienced workers and modest or even negative quality effects for the most experienced ones. The gains are strongest for "moderately rare" inquiries, where the AI model has sufficient training data but human experience is relatively limited. On labor market impacts, Eisfeldt, Schubert and Zhang (2023) examine how generative AI could

positively affect firm values, with evidence that some exposed workforces are substituted by generative AI.[3] Blank, Schubert and Zhang (2025) use internet browsing data to show that ChatGPT shifts household time toward leisure by making productive activities more efficient, with adoption concentrated among younger and higher-income households. On the other hand, Jiang, Park, Xiao and Zhang (2025) find that greater occupational AI exposure increases work hours and reduces leisure by complementing rather than replacing human labor, especially in competitive markets where productivity gains are not shared with workers. Liu, Papanikolaou, Schmidt and Seegmiller (2025) use LLM-based measures of technology exposure to show that while 20th-century technological progress favored higher-educated, higher-paid, and female-dominated occupations, AI may reverse these trends. All these papers are mainly empirical, and our paper provides a unified analytical framework that helps synthesize and reconcile these empirical patterns and clarifies when and why human content creation and oversight remain essential even as AI becomes highly capable. Our model also shed light on the empirical observation that generative AI reshapes, rather than removes, the role of human workers, with heterogeneous impacts across tasks and skill levels.[4]

Other works focus on policy implications, such as Korinek and Juelfs (2022) on preparing for AI's workforce effects, Eloundou, Manning, Mishkin and Rock (2023) analyzing language models' labor market potential, and Acemoglu and Lensman (2023) discussing regulating transformative AI technologies. From a macro perspective, Johnson and Acemoglu (2023) explore AI's role in economic prosperity over history. Specific to the financial industry, Grennan and Michaely (2020) look at the implication of AI proliferation on financial analyst careers, while Jha, Qian, Weber and Yang (2024) investigate ChatGPT's capability in

[3]Outside of the generative AI context, Zhang (2019) studies general labor-technology substitutions and its implications for asset pricing.

[4]This aspect is miniscent of the task-based framework of Acemoglu and Restrepo (2018, 2019) regarding the impact of automation on the labor market, which emphasizes that automation displaces labor from some tasks while creating new ones (reinstatement).

forecasting corporate decisions. Compared to these studies, our paper features a distinctive perspective from the information economics of financial markets. As generative AI is a breakthrough in data technology, our results also contribute to an emerging literature on the implications of data technology advances, such as Farboodi and Veldkamp (2020), Berg, Burg, Gombović and Puri (2020), Bryzgalova, Huang and Julliard (2023), Horton (2023), Abis and Veldkamp (2024), Farboodi, Singal, Veldkamp and Venkateswaran (2025).

Second, our results also relate to the technical frontiers of generative AI models typically developed in the machine-learning community. Notable contributions to the recent success of generative AI models include work on transformer architecture such as Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin (2017) and Scao, Fan, Akiki, Pavlick, Ilić, Hesslow, Castagné, Luccioni, Yvon, Gallé et al. (2022), model architectures and pre-training objectives such as Wang, Roberts, Hesslow, Le Scao, Chung, Beltagy, Launay and Raffel (2022) and Touvron, Lavril, Izacard, Martinet, Lachaux, Lacroix, Rozière, Goyal, Hambro, Azhar et al. (2023), scaling laws in model training such as Kaplan, McCandlish, Henighan, Brown, Chess, Child, Gray, Radford, Wu and Amodei (2020), and compute-optimal scaling such as Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell et al. (2020) and Hoffmann, Borgeaud, Mensch, Buchatskaya, Cai, Rutherford, Casas, Hendricks, Welbl, Clark et al. (2022) which specifically emphasizes the role of training corpus size within scaling laws. Domain-specific large models such as Wu, Irsoy, Lu, Dabravolski, Dredze, Gehrmann, Kambadur, Rosenberg and Mann (2023) further illustrate how these ideas have been adapted to production-oriented settings. Because our argument also distinguishes between parametric knowledge and access to fresh external information, it is closely related to retrieval-augmented architectures such as Lewis, Perez, Piktus, Petroni, Karpukhin, Goyal, Küttler, Lewis, Yih, Rocktäschel, Riedel and Kiela (2020) and Guu, Lee, Tung, Pasupat and Chang (2020), which combine pretrained language models with explicit document retrieval. In the finance literature, Chen, Didisheim and Somoza

(2024) show that LLMs' conditional probabilities can be economically informative and are more interpretable than their declared confidence, making them a more useful tool for analyzing model priors and biases. Our paper shares a similar spirit, in that we dive into the underlying technologies of generative AI through a unique economics perspective.

While our focus is not on the technical details of these models, our theory builds on several broad properties of production-level generative AI systems. Aside from the scaling laws reviewed above, our mechanism relies on the idea that sustaining the value of generative AI in evolving environments requires continued access to fresh ground-truth human content. In support of this argument, Shumailov, Shumaylov, Zhao, Papernot, Anderson and Gal (2024) demonstrate that AI models collapse when trained on recursively (AI-) generated data.[5] This argument is also supported by the earlier literature on "model drift," which states that if the underlying data distribution evolves from those in the training corpus (data drift) or the functional structure evolves from those learned from ealier training (concept drift), model performance will degrade over time without fresh human input (see e.g., Lu, Liu, Dong, Gu, Gama and Zhang (2018) for a review). Also consistent with our broader premise that generative AI systems rely heavily on human-generated content, Bini, Cong, Huang and Jin (2025) document in experimental settings that such models can inherit human behavioral biases, especially in domains involving preferences.

Third, there is also a growing literature that applies generative AI models to the financial market itself. Among others, Lopez-Lira and Tang (2023) show that applying ChatGPT and other large language models (LLMs) to news headlines can significantly predict subsequent daily stock returns. Chen, Tang, Zhou and Zhu (2024) find that good news extracted by ChatGPT from the front pages of Wall Street Journal can predict the stock market and

[5] Yoon, Hu, Weissburg, Qin and Jeong (2025) provide complementary evidence of model collapse in the context of image generation. Gerstgrasser, Irving and Narayanan (2024) show that model collapse may be contained when synthetic data accumulate alongside rather than replacing real data, which further underscores our emphasis on continued access to fresh ground-truth human content.

is related to macroeconomic conditions. Similarly, Tan, Wu and Zhang (2023) utilize large language models (LLMs) to extract contextualized representations of Chinese news text and predict stock returns in Chinese equity market. Chen, Green, Gulen and Zhou (2024) find that LLMs display human-like behavioral biases, such as overextrapolating trends and producing overconfident return forecasts. Engelberg, Manela, Mullins and Vulicevic (2025) propose a method to mitigate LLM look-ahead bias by using LLMs to remove identifying information from financial text, showing that LLMs can effectively neuter text while maintaining sentiment and predictive content. He, Lv, Manela and Wu (2025) introduce chronologically consistent LLMs that mitigate lookahead bias but maintain strong language understanding and predictive accuracy in financial and social science applications. Lopez-Lira, Tang and Zhu (2025) show that LLMs memorize economic and financial data from their training period, making counterfactual forecasting unreliable, and that masking or prompt instructions cannot fully prevent recall. Andrews and Farboodi (2025) find that major AI model releases in 2023–2024 caused declines in long-term US bond yields, reflecting downward revisions in expected consumption growth or perceived extreme risks. All these papers mainly focus on empirically evaluating the power of generative models in financial markets, while our focus is to theoretically analyze the interaction between human and AI content creators, although our analysis also has implications applicable to the stock market as well.

Finally, our simple model and results infuse new implications into classic theories, especially the *rational expectations equilibrium* models for financial markets, such as Grossman (1976), Grossman and Stiglitz (1980), Hellwig (1980), Diamond and Verrecchia (1981), Admati and Pfleiderer (1990), Kyle (1989), Bond, Edmans and Goldstein (2012), Breon-Drish (2015), Li (2023), Yan, Yang, Zhang and Zhou (2024), Dou, Goldstein, Li and Yang (2025). Our paper demonstrates how foundational insights from financial economics accumulated over the past several decade will continue to survive and thrive in the era of generative AI.

Furthermore, our dynamic extension builds on classical renewal and impulse-control mod-

els in which a decision maker occasionally pays a fixed cost to reset a drifting state. We adopt this modeling framework in a content-refresh setting, which yields a closed-form threshold policy to minimize the long-run average loss, following standard renewal–reward arguments (e.g., Puterman (1994); Ross (1996); Barlow and Proschan (1965)). The impulse-control aspect connects to the economics and operations literature on (s, S) and transaction-cost models with fixed intervention costs and Poisson opportunities (Davis and Norman (1990); Soner and Shreve (1994)). Our dynamic model provides a microfoundation of the static baseline model, preserving analytical transparency and linking standard renewal control to measurable features of AI platforms.

The rest of the paper is structured as follows. Section 2 provides the necessary background for understanding the underlying mechanisms of generative AI models that are relevant to our key argument. Section 3 develops a static baseline model to capture our main argument in a simple reduced-form manner. Section 4 extends the baseline model along three different dimensions, to either provide microfoundations for the simple model of Section 3 or illustrate how the simple model can serve as a workhorse for further enrichment aimed at addressing more specific questions. For ease of reference, Section 5 translates the model's comparative statics into seven falsifiable empirical predictions, discussing how each theoretical object maps to observable data and suggesting concrete data sources for future empirical work. Section 6 then concludes with some further thoughts.

2 Backgrounds

To better appreciate how our theory captures the realities of generative AI, it is essential to understand how generative AI models function. Because the inner workings of these models may not be familiar to non-AI researchers, we provide the necessary background and maintain a high-level overview, offering just enough details to understand the rest of the

paper, while providing references for readers who wish to know more technical aspects.

At its core, a large language model performs next-word prediction based on an estimated probability distribution over possible continuations of text. Formally, suppose $\Phi(\cdot)$ denotes the joint distribution of all possible meaningful content that the human race may ever encounter, then next word prediction aims to find, for any given *prompt* X (which is some text), the most likely *completion* Y (which is also some text) that maximizes

$$\Phi(Y|X).$$

This discussion, however, is highly conceptual. In practice, the probability distribution $\Phi(\cdot)$ is unknown. Indeed, even the population over which $\Phi(\cdot)$ is defined over, that is, all possible meaningful content that the human race may ever encounter, may also be not well specified.

Large language models overcome this difficulty by effectively approximating the distribution function $\Phi(\cdot)$ with another function $\hat{\Phi}(\cdot)$. Suppose for the moment we already have $\hat{\Phi}(\cdot)$ (we will talk about *training*, or the process to learn about $\hat{\Phi}(\cdot)$ later) from a specific large language model, then next word prediction within the model effectively finds, for any given prompt X, the most likely completion Y that maximizes[6]

$$\hat{\Phi}(Y|X).$$

This is also how a typical user engages with a large language model: once the model has been trained, the user provides a prompt, and the model generates a response based on statistical patterns learned from data. This stage is typically called inference. The model itself does not really "understand" the text in the way humans do, but it generates responses based on the massive amount of training data.

[6]In practice, words may be broken into "tokens" to improve performance. Therefore, sometimes large language models may be said to work on *next token prediction*.

In practice, $\hat{\Phi}(\cdot)$ is characterized by specifying, for any prompt X, the conditional probability $\hat{\Phi}(Y|X)$ as a nonlinear function (usually a set of neural networks) with a large number of parameters.[7] The prompt X (which is some text) is typically first converted into some numerical representation (known as *embeddings*) and then fed into the nonlinear function to calculate the most likely output embeddings, which will then be converted back into the most likely completion Y.

For a given large language model (that is, a given form of the nonlinear functions involved), the parameters within the model are obtained during a training process. In this process, an existing body of text, known as the training corpus and typically consisting of content previously written by humans (e.g., Wikipedia articles), is used to optimize the model parameters so that next-word prediction best fits statistical patterns in the corpus. Different large language models may specify different underlying nonlinear functions.[8]

A couple of empirically observed features of AI training are important for the main assumptions behind our theory. First, AI training tends to exhibit so-called *scaling laws*, in that training neural language models with more data, greater parameter counts, and increased computational power leads to significant improvements in performance, often measured as lower test loss (see e.g., Kaplan, McCandlish, Henighan, Brown, Chess, Child, Gray, Radford, Wu and Amodei (2020)). Our theory will specifically rely on the scaling law with respect to more training data, which is often the more binding constraint, as Hoffmann, Borgeaud,

[7] For example, GPT models, especially the latest versions, have billions or even trillions of parameters. The sheer number of parameters makes the model expressive and allows it to retain a vast amount of information.

[8] Training in models like GPT (Generative Pre-trained Transformer) also often undergo a two-step process, *pre-training* and *fine-tuning*. During pre-training, the model is trained on vast amounts of text data (like books, websites, etc.) to predict the next word in a sentence. This helps the model learn grammar, facts about the world, reasoning abilities, and some level of commonsense knowledge. It is a form of unsupervised learning as the model is not given explicit "answers"; instead, it learns patterns from the data. After pre-training, and during fine-tuning, the model is further trained on more specific tasks with labeled data. For example, ChatGPT adapts from base language models through instruction tuning and further refinement such as reinforcement learning from human feedback (RLHF), see e.g., Ouyang, Wu, Jiang, Almeida, Wainwright, Mishkin, Zhang, Agarwal, Slama, Ray, Schulman, Hilton, Kelton, Miller, Simens, Askell, Welinder, Christiano, Leike and Lowe (2022).

Mensch, Buchatskaya, Cai, Rutherford, Casas, Hendricks, Welbl, Clark et al. (2022)) show. Second, it is important to note that increases in training data cannot come solely from recursively generated data in open-ended domains; continued access to fresh ground-truth data remains important. This relates to three distinct AI phenomena: (1) Knowledge cutoff: LLMs cannot directly know about events after the period covered by their training data. (2) Model drift: When the underlying environment changes over time (either due to data drift or concept drift), a model trained on older data may become less reliable even on tasks that are not tied to a single dated fact. Without fresh information supplied through interventions such as re-training, Retrieval-Augmented Generation (RAG), or other external tools, AI outputs may therefore become stale or less aligned with current conditions. (3) Model collapse: when model-generated data increasingly replace real data in recursive training, models may lose tail information, converge toward the center of the distribution, and suffer quality degradation, a phenomenon often described as the curse of recursion (see e.g., Shumailov, Shumaylov, Zhao, Papernot, Anderson and Gal (2024); Alemohammad, Casco-Rodriguez, Luzi, Humayun, Babaei, LeJeune, Siahkoohi and Baraniuk (2023)). All three phenomena motivate our later assumption of low-value AI in the absence of fresh ground-truth content produced by humans.

Based on the above explanation, we can see that while generative AI can generate coherent and contextually relevant text, it may still produce incorrect or nonsensical answers, can be sensitive to the phrasing of the input, does not have personal experiences or emotions, and may fail on events or conditions outside the scope of its training data unless externally updated. These limitations help motivate our later assumption that AI-generated content may involve lower quality or higher verification and alignment costs.

In sum, a key takeaway so far is that while generative AI models can cost-effectively generate novel completions, they do not independently produce fresh, real-world ground-truth content; instead, they rely on patterns learned from existing human-created data. To

visualize this, imagine trying to teach a parrot to converse like a human. If one exposes the parrot to vast amounts of human conversation, it mimics the phrases and structures it hears often. However, if the surrounding environment changes—for example, if new events occur that were not part of its training—the parrot cannot "invent" the news; it remains limited to its "stale" training corpus. Large language models act as sophisticated "parrots" in this regard. While they are highly capable of logical synthesis or "writing," they face a knowledge cutoff and cannot reliably provide time-sensitive updates or field investigation unless the system is refreshed with new ground-truth information, whether through human reporting directly or through external retrieval and updating mechanisms built on such reporting.[9]

Our explanation has skipped most of the technical details within large language models, and lump everything into a non-linear function trained on some sample data.[10] However, it will become clear that our reasoning does not rely on the technical details or actual implementations of any particular generative AI system, that is, on the specific method used to learn $\hat{\Phi}$. Instead, it applies broadly to generative AI systems whose performance depends on learned representations of past data and whose usefulness in evolving environments depends on continued access to fresh ground-truth information. More precisely, one only needs to recognize that the performance of generative AI depends on two broad factors: (1) how well the learned nonlinear function approximates the underlying distribution $\Phi(\cdot)$, and (2) how well the training corpus represents the domain over which the model will later be used. Our later analysis emphasizes the second channel: when the relevant environment evolves quickly,

[9]See also Bender, Gebru, McMillan-Major and Shmitchell (2021) for an early discussion of language models as stochastic parrots.

[10]Opening this black box will introduce many technical terms. For example, many models are built on the *transformer* architecture. This design uses stacked blocks (either *encoder*, *decoder*, or both) centered on a key component known as *attention*, which allows the model to weight different parts of the input data differently, enabling it to "pay attention" to the most relevant context in sequential data like sentences or paragraphs. Supporting components include token *embeddings* and *position encoding*, while different model variations may be customized by *hyperparameters* like the overall number of model parameters. Models may also use *iterative learning processes* to improve performances.

a model trained on stale content may become less relevant as $\Phi(\cdot)$ shifts away from $\hat{\Phi}(\cdot)$.

3 A simple baseline model

We first present a simple model to illustrate our core idea about the relationship between generative AI and human creatives. The model is deliberately stylized to capture the essence of why the low-cost yet *derivative* nature of generative AI prevents it from completely replacing human creators. After explaining our core idea in its simplest forms here, Section 4 will extend the baseline model to demonstrate the flexibility of the baseline model and derive further implications.

3.1 Setup

Agents There is a unit-mass continuum of consumers for content. A fraction λ of these consumers subscribe to professional (human) content outlets at a cost $C > 0$, which defrays the costs incurred by human creatives and compensates their efforts in content production. The remaining fraction $1 - \lambda$ of consumers relies on generative AI for content at a lower cost $c < C$. Note that the fraction λ is endogenously determined in equilibrium. In subsequent discussions, we may simply refer to the first group of consumers as human and the second group of consumers as generative AI for brevity. For simplicity, in this section we treat content consumers and producers as vertically integrated—each consumer either produces human content or uses AI-generated content (in other words, they can be called "prosumers" borrowing from Alvin Toffler's terms). Appendix K relaxes this assumption by separating the consumption and production sides of the market.

Preferences Normalizing a consumer's utility when consuming no content (either from human creation or generative AI) to zero, we denote the benefit from consuming high quality

content (as would typically be produced by professional humans) as R. We assume $R > C$, since otherwise the world would be in an uninteresting and unrealistic situation with no content creation or consumption, which is clearly at odds with historical experiences. The payoff to a consumer who subscribes to professional (human) content is therefore

$$\pi_{\text{human}} = R - C > 0. \tag{1}$$

On the other hand, we denote the benefit from consuming AI-generated content as r. Then the payoff to an agent who uses generative AI is

$$\pi_{\text{AI}} = r(\lambda) - c. \tag{2}$$

Here, we write $r(\lambda)$ to highlight that r is a function of λ, because AI performance depends on the volume of human-created content used for training, a phenomenon known as the scaling law (see e.g., Kaplan, McCandlish, Henighan, Brown, Chess, Child, Gray, Radford, Wu and Amodei (2020), Hoffmann, Borgeaud, Mensch, Buchatskaya, Cai, Rutherford, Casas, Hendricks, Welbl, Clark et al. (2022)).[11] The function $r(\cdot)$ can also be interpreted as an aggregator of all features that consumers may value in a given AI model. For instance, consumers may care not only about how accurately AI-generated content reflects common sense (i.e., overall accuracy or quality), but also about how well it captures real-time information and events (i.e., timeliness). Then the function $r(\cdot)$ aggregates all the trade-offs between accuracy/quality and timeliness, and defines an iso-value curve between the two. When there is a trade-off between accuracy/quality and timeliness, the equilibrium outcome depends on how they affect the aggregator $r(\cdot)$: If they cancel out and stay on the same iso-value curve, then, all else equal, the equilibrium will not change. Otherwise, the equilibrium will shift ac-

[11] In practice, this is an intertemporal relationship, in that AI performance *today* depends on the volume of human-created content used for training *in the past*. Section 4.3 will take on this extension and show that results in the static model are consistent with those from a dynamic microfoundation.

cording to which factor dominates. We also note that the way through which λ enters the AI could be either through re-training (which updates model weights) or Retrieval-Augmented Generation (or RAG, which injects current news into the prompt context). Either way preserves the same meaning for λ, which represents the availability of ground-truth data. It is worth emphasizing that the availability of RAG does not undermine this conclusion. RAG allows an AI system to retrieve passages from an existing corpus at inference time, thereby accessing recent information without full re-training. However, the retrievable corpus is itself composed of human-generated content: if humans were to exit the market entirely, the corpus would stop growing and would itself grow stale as the world evolves, just as a re-trained model would. The channel through which λ affects AI quality therefore shifts from model weights to the retrieval database, but the fundamental dependence on ongoing human content production is preserved. If anything, RAG makes the positive relationship between λ and AI quality more transparent, further grounding Assumption 1.

Later we may also parameterize $r(\lambda)$ with some parameter ξ so that $\frac{\partial}{\partial \xi} r(\lambda; \xi) > 0$ for all $\lambda \in [0, 1]$. The parameter ξ captures the advancement level of the AI technology — a higher ξ would give a higher $r(\lambda; \xi)$ for the same level of λ. An increase in ξ then captures the improvement of AI technology. When speaking of a specific AI technology (so ξ is fixed), or when technological advances in AI are not our focus, we will suppress ξ and just write $r(\lambda)$ for notational simplicity.

We hold the following assumptions for the function $r(\lambda)$:

Assumption 1. *The benefit from consuming generative AI-created content, $r(\lambda)$, is an increasing function of λ, the extent to which human professionals produce content.*

Assumption 1 follows from how generative AI models operate. As explained in Section 2, generative AI systems function as sophisticated machines that learn from and depend on human-generated content. Keeping the algorithms of a generative AI model fixed, the per-

formance of this AI model depends on how comprehensive and up-to-date its training sample is. To the extent that increased human content creation (a higher λ) makes a generative AI model's training data more comprehensive and up-to-date, the benefit to consumers, $r(\lambda)$, also increases with the intensity of human content creation. From a scaling law perspective, Assumption 1 also has empirical backings from the artificial intelligence and machine learning literature. For example, Hestness, Narang, Ardalani, Diamos, Jun, Kianinejad, Patwary, Yang and Zhou (2017) provide empirical evidence across numerous tasks that generalization error decreases as a power law in the size of training set. Kaplan, McCandlish, Henighan, Brown, Chess, Child, Gray, Radford, Wu and Amodei (2020) shows that increasing data leads to power-law decreases in loss (i.e., quality improvement) in language models. Also see Viering and Loog (2022) for a review on empirical evidence of learning curves.

Assumption 2. *A generative AI model trained on obsolete data cannot fully match the net value of fresh, professionally produced human content, that is,* $r(0) - c < R - C$.

Assumption 2 also naturally follows how generative AI models operate as previously explained in Section 2. To see this, notice that when $\lambda = 0$, no content is created by humans. AI models would thus have little input for its training purpose and therefore not be able to generate meaningful updates. This is particularly relevant in settings where the underlying content space is fast moving so that without fresh input the value provided by an outdated model will quickly diminish to negligible levels. Indeed, Shumailov, Shumaylov, Zhao, Papernot, Anderson and Gal (2024) also experimentally show that when generative AI models are recursively trained on AI-generated contents (so that the amount of contemporaneous human input is low), its performance quickly collapses.

3.2 Equilibrium characterization

We now proceed to solve for the equilibrium outcomes of the baseline model, characterized by two lemmas that lead to the main proposition.

Lemma 1. *Generative AI completely taking over human creatives cannot be an equilibrium outcome. In other words,* $\lambda > 0$ *in equilibrium.*

Proof. This can be proved by contradiction. Suppose otherwise, then $\lambda = 0$, and each generative AI user gets $r(0) - c$. If one of them were to deviate toward subscribing to human content, then his payoff would become $R - C$ which is higher than $r(0) - c$ by Assumption 2. This is a profitable deviation. Therefore, $\lambda = 0$ cannot hold in equilibrium. □

Lemma 1 addresses the main question posted at the beginning of the paper regarding whether generative AI will completely take over human content creation — the answer is a resounding no, due to generative AI's intrinsic technological constraints.

However, we do expect generative AI to play significant roles in equilibrium, as the next lemma indicates.

Lemma 2. *If the quality of generative AI is sufficiently high (specifically,* $r(\cdot)$ *or* ξ *in* $r(\cdot;\xi)$ *is high enough such that* $r(1) - c > R - C$*), it is not an equilibrium for generative AI to not be adopted. In other words,* $\lambda < 1$ *in equilibrium.*

Proof. We again prove the lemma by contraction. Suppose otherwise, then $\lambda = 1$, and each human content creator gets $R - C$. If he were to deviate toward using generative AI, his payoff would be $r(1) - c > R - C$. Therefore, $\lambda = 1$ cannot hold in equilibrium. □

The equilibrium condition $r(1) - c > R - C$ means that if human creators dominate and the AI model has access to all human-created content (i.e., $\lambda = 1$), the AI model's maximal net performance exceeds that of human creators. Otherwise, generative AI would not be adopted in equilibrium. This assumption and its implication are consistent with

the historical development of generative AI: while the underlying technology has existed for years (dating back to at least Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser and Polosukhin (2017) if not earlier), its use only surged once its output quality (or net benefit) became sufficiently high to be competitive with professional human creators.

Combining Lemma 1 and 2, we have that in a world with adequately powerful AI, the equilibrium must be interior, that is, generative AI and human creatives must coexist. Therefore, in equilibrium we must have $\pi_{\text{human}} = \pi_{\text{AI}}$, or $R - C = r(\lambda) - c$.

Proposition 3. *Given that $r(0) - c < R - C$ and $r(1) - c > R - C$, there exists a unique $\lambda^* \in (0, 1)$ satisfying*

$$R - C = r(\lambda^*) - c, \tag{3}$$

which characterizes the unique equilibrium in our model economy.

Proof. In equation (3), the right-hand side $r(\lambda) - c$ is monotonically increasing in λ, is lower than the left-hand side when $\lambda = 0$, and higher than the left-hand side when $\lambda = 1$. □

Figure 1 illustrates the equilibrium in the baseline model.

3.3 Comparative statics

The definition of λ^* in Equation (3) implies that the equilibrium fraction of consumers who subscribe to human-produced content, λ^*, increases in the value of human-created content R, decreases in the cost of human-produced content C, increases in the marginal cost of AI content generation c, and decreases in the performance of generative AI. Formally, we have:

Proposition 4. *The relative amount of content created by human, λ^*, satisfies $\frac{\partial}{\partial R}\lambda^* > 0$, $\frac{\partial}{\partial C}\lambda^* < 0$, and $\frac{\partial}{\partial c}\lambda^* > 0$. Furthermore, λ^* decreases as the function $r(\cdot)$ increases, that is, for the parameterized $r(\lambda; \xi)$ with $\frac{\partial}{\partial \xi} r(\lambda; \xi) > 0$ for all $\lambda \in [0, 1]$, $\frac{\partial}{\partial \xi}\lambda^* < 0$.*

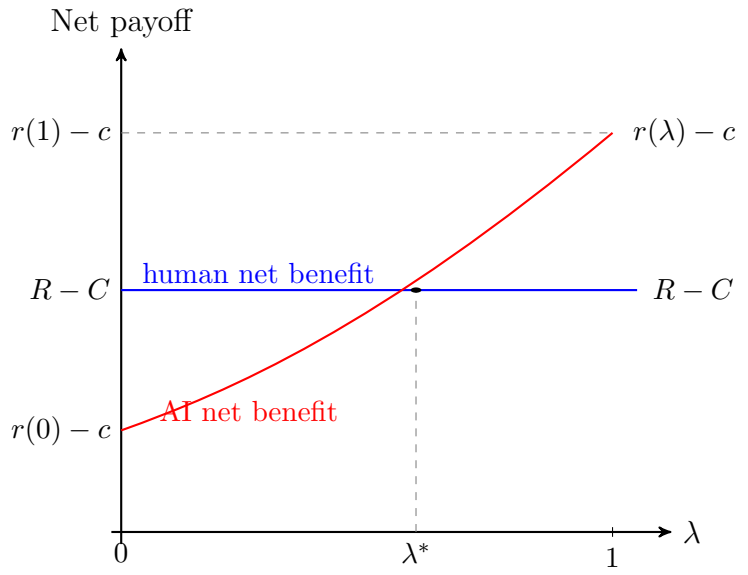


Figure 1: Equilibrium in the baseline model. The horizontal blue line is the human net payoff $R - C$, constant in λ. The upward-sloping red curve is the AI net payoff $r(\lambda) - c$: as more humans produce content ($\lambda\uparrow$), AI training data improve and $r(\lambda)$ rises (scaling law). Assumption 2 ensures $r(0) - c < R - C$ (AI on stale data cannot match human content), and the condition of Lemma 2 ensures $r(1) - c > R - C$ (AI is competitive when all humans produce). The unique equilibrium $\lambda^* \in (0, 1)$ is at the intersection.

Proposition 4 informs us about how improvements in AI capabilities—such as accuracy, creativity, or cost-efficiency—can shift the equilibrium. Overall, as the accuracy or creativity of generative AI improves (as reflected by an increase in $r(\cdot)$), or as its cost-efficiency improves (as reflected by a decrease in c), reliance on human-created content drops. In neither case, however, would generative AI completely replace humans, as we have shown in Lemma 1.

We also note that generative AI's cost c may have rich interpretations. For example, c may include the inference cost of generative AI or the alignment and verification costs associated with using generative AI. The latter may also reflect (some) consumers' preferences for human over AI (due to, say, lack of trust for AI), captured by those consumers' extra (mental) cost in verifying AI content.[12] While it is expected that as generative AI technologies further develop, the inference cost is likely to go down, we may also see the

[12]It is out of the scope of our paper to discuss whether such preferences are well-justified or not.

trend going the other way in terms of alignment or verification costs (which may increase over time). Ultimately, how these costs will evolve depends on the technological development of generative AI as well as changes in consumer tastes, and are thus outside the scope of our economic modeling. That said, the comparative statics analysis from our model can nevertheless accommodate any possible paths forward, and our main conclusion that generative AI cannot fully replace human will remain robust over all possible future paths. This ambiguity cuts both ways: while lower inference costs may accelerate AI adoption, rising verification burdens or regulatory disclosure requirements could partially offset this trend, sustaining demand for human-produced content whose provenance and accuracy are directly verifiable. Our framework therefore remains applicable regardless of which force dominates in a given period or domain.

Different levels of c and $r(\cdot)$ (or ξ in $r(\cdot;\xi)$) could also be interpreted as the applicability of generative AI to different types of tasks involved in content production. For example, as we will further study in Section 4.2, we may think of the production process (e.g., the production of a news piece) to involve two tasks: *investigation*, which involves gathering fresh signals from the world, and *writing*, which involves synthesizing, drafting, and presenting information. Then for each task $i \in \{I, W\}$, where I stands for investigation and W for writing, we can then denote c_i as the cost of task i by generative AI and $r_i(\cdot)$ the benefit created by generative AI for task i. As one may expect generative AI to be more capable of writing than investigating, then we may have a lower c_W than c_I or higher $r_W(\cdot)$ than $r_I(\cdot)$. Then according to the comparative statics in Proposition 4, all else equal, we may expect generative AI to play a bigger role for writing tasks compared to investigation.

3.4 Connection to the Grossman-Stiglitz insight

Our baseline model shares intuition with Grossman and Stiglitz (1980), who study a financial market with two types of investors: active and passive.[13] Active investors may incur a cost C to learn a signal s about the fundamentals of a risky asset and then trade on this information. Their model addresses how an active investor's decision depends on the learned signal, how expected utility depends on both the decision and the realized market outcome, and how the equilibrium price depends on these decisions. For our purposes, we only need to note two key insights: (1) learning the fundamentals (relative to not learning) increases an active investor's expected utility, and we denote this gain by R; and (2) the equilibrium price incorporates the signal s, with the mapping from the signal to the price denoted by $P(s)$.

Passive investors do not engage in fundamental research; instead, they base their investment decisions solely on the price P. Interpreting P may incur a small cost $c < C$, although in their original model c is implicitly assumed to be zero. Grossman and Stiglitz (1980) then argue that if the market is informationally efficient, knowing $P(s)$ will be as informative as knowing s itself. In this case, passive investors enjoy the same utility gains from investment, R. Therefore, they strictly prefer remaining passive to becoming active, because the former yields $R-c$ while the latter yields $R-C < R-c$. Consequently, all rational investors would choose to be passive. However, if this were the case, no investor would acquire the signal s in the first place, then how could the price $P(s)$ fully reflect s? This creates a paradox.

Grossman and Stiglitz (1980) then introduce uninformative "noise" into the model, so that P can no longer fully reflect the signal s. To distill the essence of their argument and fit into our notation, this "noise" effectively reduces the gains to passive investors to some value

[13] Note that the description below will skip many modeling details of Grossman and Stiglitz (1980) and only focus on their key insights that are analogous to human-generative AI interactions. Notation will also deviate from their original paper and conform to ours. Since the purpose of this section is just to draw connections and help with understanding, readers who would like to learn the full model of Grossman and Stiglitz (1980) should consult the original reference, and readers familiar with Grossman and Stiglitz (1980) and already see the connection can skip this section without loss of continuity.

r below R. They further show that the more active investors are present in the market, the more informative the price P becomes (and thus the higher is r). In this "noisy" version, they show that the model features an equilibrium in which active and passive investors coexist, with the marginal investor being indifferent between being active and passive.

Back to our baseline model, human content creation resembles active investing: it incurs a high cost C and generates a high benefit R. The content that gets created is analogous to the signal s in Grossman and Stiglitz (1980), while the implicit joint distribution function $\hat{\Phi}(\cdot)$ of a generative AI model is akin to the price function $P(\cdot)$. Based on well-established scaling laws in AI training, the benefit brought by a generative AI model increases with the size of the training sample (or the amount of human-created content). In cases where the underlying $\Phi(\cdot)$ evolves rapidly, additional training data from the latest $\Phi(\cdot)$ are also required to prevent $\hat{\Phi}(\cdot)$ from becoming stale. Thus, analogous to Grossman and Stiglitz (1980), we show that human content creators and generative AI coexist in equilibrium, so that the latter cannot fully take over the former's jobs.

4 Model extensions

Section 3 delivers our core message that human creatives and generative AI will coexist. Despite its simplicity, the baseline model provides rather robust results and can serve as a backbone workhorse for various extensions. As illustrations, this section presents three extensions to the baseline model that respectively address the following questions: (1) How do consumer preferences for content quality affect the equilibrium level of generative AI penetration? (2) What if generative AI complements rather than substitutes some tasks performed by human creatives? (3) What is the dynamic microfoundation for the static baseline model, and how does the human–AI ecosystem evolve over time in a stochastic world? We demonstrate that these model extensions, while aimed at answering different

questions, nevertheless all reinforce the core observations from the baseline model. Although these extensions are certainly not exhaustive, they illustrate the usefulness of our baseline as a workhorse for further analyses of the relationship between humans and generative AI.

4.1 Heterogeneous consumer preferences

Our baseline model in Section 3 assumes that all consumers have homogeneous preferences, and they consume content created by either human or generative AI. This simplicity prevents us from analyzing the role of consumer preferences or production skills, especially their heterogeneity and distribution within the population, in shaping equilibrium outcomes. Furthermore, as stated earlier, the treatment there has also assumed content consumers and producers are vertically integrated, which simplifies the analysis but also prevents us from making sharp intepretations for the labor market implications of introducing generative AI. This subsection (and subsequently Appendix K) will relax these restrictions one by one and derive richer implications.

We first consider the possibility of heterogeneous consumer preferences for content quality. As in the baseline model, there is a unit mass of consumers. To capture their heterogeneous preferences, we index all potential consumers by a type parameter $\theta \in [0, +\infty)$, which is distributed according to an atomless cumulative distribution function (cdf) F over $[0, +\infty)$. For a type-θ consumer, his payoff from consuming human-generated content is

$$\pi(\theta) = \theta \cdot R - C,$$

while his payoff from consuming AI-generated content is

$$\pi_{AI}(\theta) = \theta \cdot r(\lambda) - c,$$

where λ again denotes the endogenous amount of human generated content.[14]

Unlike in the baseline model discussed in Section 3, here low-type consumers may also choose to opt out (that is, not consume any content, whether generated by humans or by generative AI) and receive a payoff that is normalized to 0. Intuitively, the parameter θ captures an intrinsic taste for quality. Alternatively, we may interpret the inverse of θ as capturing a consumer's sensitivity to cost.

To illustrate the impact of introducing generative AI, we proceed to first analyze the pre-AI equilibrium, and then compare it to the equilibrium after the rise of generative AI.

Equilibrium without generative AI In the pre-generative-AI world, a potential consumer with type θ will participate if only if his surplus $\theta R - C$ is higher than 0 (when he opts out). Therefore, if we define the cutoff:

$$\theta_{H0} \equiv \frac{C}{R}, \tag{4}$$

then consumers with $\theta \in [0, \theta_{H0})$ will strictly prefer to opt out, while consumers with $\theta \in [\theta_{H0}, +\infty)$ will choose to purchase content.[15]

Equilibrium with the introduction of generative AI With the introduction of generative AI, a potential consumer with type θ gets

$$\theta R - C, \ \theta r(\lambda) - c, \text{ and } 0$$

[14] Assuming payoffs take the multiplicative form $\theta \cdot q$ is without loss of generality. In general, a consumer of type θ may derive gross benefit $v(\theta, q)$ from content of value q, where v is strictly increasing in θ for any fixed $q > 0$. Since preference depends only on the value obtained rather than whether it is delivered by human content ($q = R$) or AI content ($q = r(\lambda)$) — the same function v applies to both. Defining $\hat{\theta} \equiv v(\theta, q)/q$ yields $v(\theta, q) = \hat{\theta} \cdot q$ by construction, with F reinterpreted as the distribution of $\hat{\theta}$.

[15] Technically speaking, consumers with $\theta = \theta_{H0}$ are indifferent between opting out or purchasing content. Since they are of measure zero, it does not really matter for later analysis how we specify their behavior. We will follow the convention that they behave the same as those immediately right to them on the type axis.

from consuming human-generated content, consuming AI-generated content, and opting out, respectively. In equilibrium, each consumer chooses the option that gives him the highest payoff. We further define two cutoffs:

$$\theta_{A0}(\lambda) \equiv \frac{c}{r(\lambda)}, \tag{5}$$

$$\theta_{HA}(\lambda) \equiv \frac{C-c}{R-r(\lambda)}. \tag{6}$$

Proposition 5 then summarizes the equilibrium outcome after introducing generative AI:

Proposition 5. *With the introduction of heterogeneous consumer tastes θ and consumer choices (including the right to opt out), if*

$$\frac{C}{R} \leq \frac{c}{r\left(1-F\left(\frac{C}{R}\right)\right)}, \tag{7}$$

then human remain intact from AI: the equilibrium has consumers with $\theta \in [0, \theta_{H0})$ opt out, while consumers with $\theta \in [\theta_{H0}, +\infty)$ chooses human, where $\theta_{H0} = C/R$ as given in (4).

If (7) is violated, that is,

$$\frac{C}{R} > \frac{c}{r\left(1-F\left(\frac{C}{R}\right)\right)},$$

then there exists a (unique) λ^* satisfying

$$\lambda^* = 1 - F\left(\frac{C-c}{R-r(\lambda^*)}\right), \tag{8}$$

and the equilibrium has consumers with $\theta \in [0, \theta_{A0}(\lambda^*))$ opt out, consumers with $\theta \in [\theta_{A0}(\lambda^*), \theta_{HA}(\lambda^*))$ choose AI, and consumers with $\theta \in [\theta_{HA}(\lambda^*), +\infty)$ chooses human, where θ_{A0} and θ_{HA} are defined by (5) and (6), respectively.

Proof. See Appendix C. □

The cutoff condition (7) has a rather intuitive interpretation. The left-hand side of (7)

is the cost per unit of benefit under human production, while the right-hand side is AI's cost per benefit, computed at the pre-AI human participation mass $1 - F(C/R)$. If the right-hand side is no less than the left-hand side, then AI's performance is too low to cause any change to the content market such that the pre-AI equilibrium persists; Otherwise (if AI is cost-effective), some consumers will adopt AI-generated content in equilibrium.

Proposition 5 offers several implications on how generative AI affects the content market.

First, all else equal, condition (7) is more likely to hold when $F(\cdot)$ is large. This yields practical guidance for human professionals seeking to resist the impact of generative AI: focus on content spheres with a large $F(\cdot)$. What kinds of content have a large $F(\cdot)$? To the extent that a large $F(\cdot)$ reflects a thick left tail—meaning that most of the population cares little about the content and only a small fraction values it highly—such content can be interpreted as niche content. In sum, human creatives may find better shelter by devoting themselves to niche content production in the era of generative AI. In journalism, for example, niche content may include local news, highly specialized topics, or media tailored to minority communities. If human creatives devote more effort to niche content markets, not only would they be better protected from the shock of generative AI, the society overall may also benefit from increased content diversity.

Second, AI adoption has a "middle-out" pattern, that is, AI captures the intermediate segment of the market. It attracts consumers who are on the margin of opting out (expanding access) as well as those who previously consumed human content but find AI a "good enough" low-cost substitute. Meanwhile, consumers with extremely low preferences continue to opt out, and those with extremely high preferences stick with human professionals. This pattern is consistent with what we observe from other generative AI adoption cases. For example, struggling students who used to give up on certain problems (but not too far away from working out) would now adopt generative AI to help them finish, while those who did manage to finish with substantial effort may now choose the easy path and have generative AI do

their work — students who struggle so much and completely have no clue of the problem or those who are talented and can finish the problems with a breeze may be the last to use AI (for the specific problems being discussed here, not necessarily overall AI adoption). Similar middle-out adoption patterns also arise in journalism, design, legal services, and software development.[16]

Third, the "middle-out" pattern has both positive and negative implications. On the positive side, since $\theta_{A0} < \theta_{H0}$, the rise of AI could enhance inclusiveness by giving some low-type consumers access to content via lower-cost AI-generated material. However, on the negative side, since $\theta_{HA} > \theta_{H0}$, AI may take over the jobs of some human content creators, potentially generating risks to knowledge ecosystems, for example, by degrading content quality. To see this, notice that the total value created by human-generated content (which could serve as AI training samples) used to be $R\left(1 - F(\theta_{H0})\right)$ before AI's rise, which after the adoption of generative AI strictly decreases to $R\left(1 - F(\theta_{HA})\right)$.[17]

Finally, redrawing attention to the case in which the introduction of generative AI has no impact on the equilibrium outcome, we note that the possibility of all human creatives resisting the impact of AI and retaining their jobs does not arise in our baseline model in Section 3. This is because the baseline model does not feature heterogeneous tastes and does not allow consumers to opt out of content consumption; instead, we directly rule out the case in which no content consumption happens even without AI.

Comparative statics By Equation (8) in Proposition 5, we immediately draw the following comparative statics:

[16] For example, Cui, Demirer, Jaffe, Musolff, Peng and Salz (2025) find higher adoption of AI among newer and more junior employees at Microsoft, Accenture, and an anonymous Fortune 100 company. The newer and more junior employees in these firms correspond to the "middle" types in our model, long-tenure and more senior employees in these firms correspond to the "high" types in our model, and those "opted-out" low types in our model would correspond to those not employed in those firms or even non software developers.

[17] Even if we account for all content created by both humans and AI, their total value $R(1-F(\theta_{HA}(\lambda))) + r(\lambda)\left(F(\theta_{HA}(\lambda)) - F(\theta_{A0}(\lambda))\right)$ can still be less than $R(1-F(\theta_{H0}))$ depending on the shapes of $r(\cdot)$ and $F(\cdot)$.

Corollary 6. *The relative amount of human-created content λ^* increases as generative AI (inference, alignment, or verification) cost c increases, human creation cost C decreases, human content value R increases, and generative AI's quality $r(\cdot)$ (or ξ) decreases.*

Corollary 6 is intuitive: as the value (cost) of human creation increases (decreases), or as the value (cost) of AI creation decreases (increases), more customers choose human content.

Note that the amount of AI consumption is $F(\theta_{HA}(\lambda)) - F(\theta_{A0}(\lambda))$, or more explicitly

$$F\left(\frac{C-c}{R-r(\lambda)}\right) - F\left(\frac{c}{r(\lambda)}\right). \tag{9}$$

Therefore we can also draw comparative statics as follows:

Corollary 7. *The relative amount of AI-generated content increases as generative AI (inference or alignment/verification) cost c decreases or the quality of generative AI, as reflected by $r(\cdot)$ (or ξ), increases. The relationship with human creation cost C and human content value R may be non-monotonic.*

Proof. The proof involves a series of straightforward but tedious applications of the implicit function theorem, so we delegate it to Appendix D. □

The non-monotonic dependence of AI adoption on R and C results from two countervailing forces at play. On the one hand, a decrease in C (or an increase in R) directly improves the attractiveness of human-generated content and draws away consumers who would otherwise adopt AI (at the high end). On the other hand, the increase in human-generated content also improves the quality of AI, which lures consumers who would otherwise opt out (at the low end) to adopt AI. Which force dominates depends on both the quality of the AI model, as reflected in $r(\cdot)$, and the distribution of consumer types, as reflected in $F(\cdot)$.

To illustrate the effects of increasing the AI cost c or reducing the human cost C on the human-AI adoption pattern, we take two specifications for numerical examples in Figure 2:

We assume that the consumer type θ is uniformly distributed on $[0, \bar{\theta}]$ so that $F(\theta) = \theta/\bar{\theta}$, and the AI quality r is a linear function of the amount of human-created content, $r(\lambda) = a + b\lambda$.

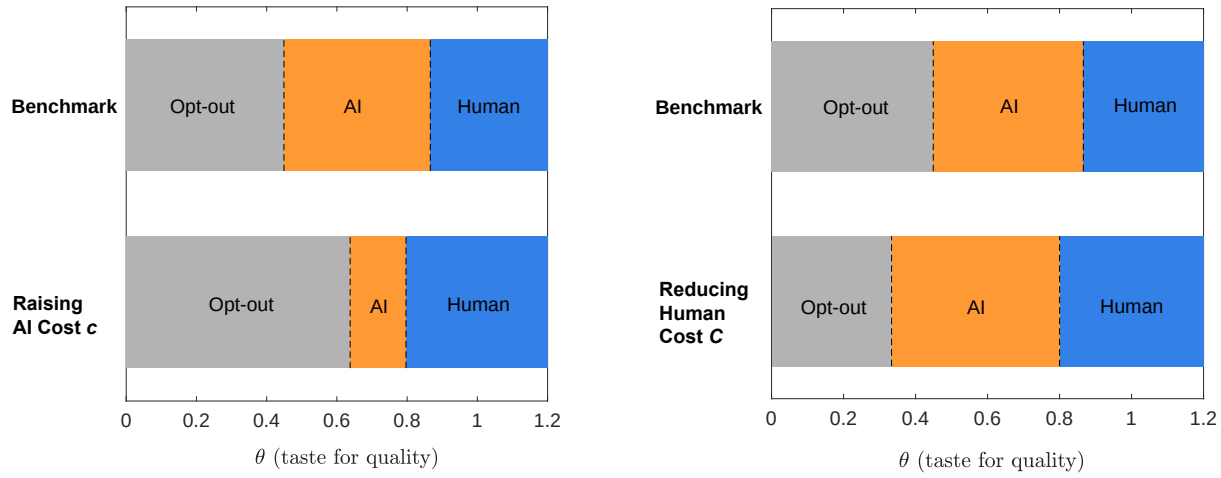


Figure 2: Equilibrium adoption patterns under heterogeneous consumer preferences (θ uniform on $[0, \bar{\theta}]$, $r(\lambda) = a + b\lambda$). In both panels, the upper portion displays the benchmark pattern given that $R = 0.80$, $C = 0.60$, $c = 0.10$, and $r(\lambda) = 0.8\lambda$, with θ uniformly distributed over $[0, 1.2]$. The left panel shows how raising the AI cost ($c = 0.16$) contracts AI adoption and expands the opt-out segment. The right panel shows how reducing the human cost ($C = 0.50$) shrinks the opt-out segment and expands human creation.

As shown in the left panel of Figure 2, raising AI cost c not only compresses the band of AI adoption but also expands the human opt-out region. In contrast, as shown in the right panel, reducing human cost C has two effects: a direct substitution effect and an indirect quality effect via λ. While making human creation less expensive shrinks the AI region at the high end, the resulting increase in high-quality human content feeds back to improve the AI quality, thereby expanding AI adoption at the low end.

Policy implications. The comparative statics offer timely guidance for how different regulatory and platform policies can reshape the human-AI ecosystem. Specifically, a key debate has been concerning whether AI model training should rely primarily on strict copyright

enforcement or on licensed access with remuneration for rights holders.[18]

Cast in the context of our comparative statics, we notice that policies such as tightening intellectual property (IP) restrictions to constrain training data or raise penalties for unlicensed training can effectively increase AI's cost c or reduce AI quality (lower $r(\cdot)$). These changes increase both $\theta_{A0} = c/r(\lambda)$ and $\theta_{HA} = (C-c)/(R-r(\lambda))$, thereby shrinking the AI-adoption region and raising the human share. However, a tighter IP policy also increases opt-out: as AI becomes less attractive and human content remains priced at C, more low-type consumers prefer not to consume any content. In addition, suppressing $r(\cdot)$ eliminates the positive feedback through which a healthy human sector (higher λ) could otherwise raise long-run AI quality. Thus, although strict IP protection protects human employment, it may do so at the cost of consumer access and slower improvement in compliant AI system, as illustrated in the left panel of Figure 2.

On the other hand, another policy is to permit AI training through licensing arrangements, with fees remitted to human creators. Licensing revenue can be used to reduce human creators' effective costs (lower C) or raise their compensation and quality (higher R), both of which expand the region in which humans remain intact. The resulting fixed point λ^* can be larger than under either laissez-faire or strict-IP policies, enabling $r(\cdot)$ to scale with licensed access to fresh human content, preserving the "virtuous cycle" in which a robust human sector improves AI quality, which in turn supports sustainable coexistence. Therefore, consistent with the right panel of Figure 2, licensed access with remuneration supports a more inclusive market, stronger human incentives, and a healthier long-run human–AI ecosystem.[19]

[18]High-profile litigation, such as The New York Times' 2023 lawsuit against OpenAI and Microsoft, frames large-scale ingestion of journalistic content as a potential copyright violation, while major publishers increasingly pursue licensing agreements with AI developers to receive compensation. This raises a policy-relevant question: which approach, copyright enforcement or licensing with remuneration, more effectively balances the interests of AI developers and human creators.

[19]That said, when licensing is infeasible to enforce or unlicensed training is pervasive, strict IP rule may be the only short-run tool for preventing an AI-driven collapse in human supply. Likewise, in highly sensitive

Taken together, our model suggests that licensed access with human remuneration is the welfare-improving policy when credible enforcement and revenue-sharing mechanisms are available. Strict IP protection preserves human share but at the cost of higher opt-out and weaker AI quality. Our model thus provides an equilibrium-based rationale for licensing and remuneration as key strategies for a healthy, sustainable human–AI content ecosystem.

Alternative dimensions of heterogeneity. Consumer preferences is just one dimension where heterogeneity can be introduced to the baseline model. Appendix K will further consider other dimensions of heterogeneity. Appendix K.1 examines the parallel case where heterogeneity lies on the production side (in terms of producer skills). Appendix K.2 then will then unify both cases by introducing an endogenous market price that clears supply and demand simultaneously.

4.2 Generative AI: substitutes or complements?

Another dimension in which we can extend the model is to recognize that content generation is not necessarily a monolithic process. For example, the production process of a news piece can involve multiple complementary tasks, such as *investigation* — which involves gathering fresh signals from the real world and *writing* — which involves synthesizing, drafting, and presenting information. Generative AI may have different impacts on different types of tasks. For example, one may expect generative AI to be more capable of writing than investigating, in which case the rise of generative AI may *substitute* the writing part of content creation, but may on the other hand also *complement* the investigation part.[20] Such heterogeneous impacts may also shed further light on what kind of jobs be threatened by generative AI, and provide career guidance for human professionals today on what they should focus on

niches (large $F(C/R)$), a temporary IP restriction may buy time for human remuneration mechanisms to be developed and implemented.

[20]For example, Jiang, Park, Xiao and Zhang (2025) indeed find empirical evidence of ChatGPT complementing human labor using data from the American Time Use Survey (ATUS).

moving forward.

In this section, we consider a simple extension to our baseline model in Section 3 to illustrate the intricate forces mentioned above. We do not claim that the model presented here exhausts all relevant economic forces, but hopefully it illustrates many first-order considerations.

4.2.1 Setup

Suppose each piece of content creation requires both investigation and writing. Each task may be performed either by a human agent or by AI. The value R of a piece of content depends on both the quality of investigation and that of writing, following the Cobb-Douglas form defined above, where $\alpha \in (0, 1)$. The Cobb-Douglas functional form reflects the complementarity between investigation and writing. From an economic perspective, investigation requires gathering novel signals from the real world, while writing is more about style and packaging, which may not be sensitive to real world evolutions. Therefore, while we use the labels *investigation* versus *writing* here solely for concreteness and expositional ease, the message of this section is not restricted to only these two tasks but also applies broadly to any complementary tasks that feature different paces of how the underlying knowledge base changes over time.

Denote $C_I > 0$ and $C_W > 0$ as the respective costs incurred by doing investigation and writing without AI (human production), while having AI do investigation and writing costs $0 < c_I < C_I$ and $0 < c_W < C_W$, respectively. We assume that the cost of investigation and the cost of writing is additive for any particular content.

Denote Q_I and Q_W as the respective quality of human investigation and human writing, while AI-investigation and AI-writing has respective quality $q_I(\lambda_I)$ and $q_W(\lambda_W)$, where $\lambda_I \in [0, 1]$ and $\lambda_W \in [0, 1]$ denote the aggregate fractions of investigation and writing tasks

performed by humans, respectively. We assume that $\frac{\partial}{\partial \lambda_I} q_I(\lambda_I) > 0$ and $\frac{\partial}{\partial \lambda_W} q_W(\lambda_W) \geq 0$.[21] We also assume $q_I(0) = 0$ to reflect an extreme level of model decay, in that the content market is supposedly fast changing so that without sufficient fresh human input AI investigation outcomes will degrade very fast.[22]

In sum, there are four possible choices for content production by using either human or AI for each of the two tasks, investigation and writing. The corresponding payoffs are

1. *Human investigation + human writing:*

$$\pi_{HH} \equiv Q_I^{\alpha} Q_W^{1-\alpha} - C_I - C_W > 0. \tag{10}$$

We assume (10) is strictly positive; otherwise no content would have been produced prior to the emergence of generative AI, which does not line up with the reality.

2. *Human investigation + AI writing:*

$$\pi_{HA} \equiv Q_I^{\alpha} \big(q_W(\lambda_W)\big)^{1-\alpha} - C_I - c_W. \tag{11}$$

3. *AI investigation + human writing:*

$$\pi_{AH} \equiv \big(q_I(\lambda_I)\big)^{\alpha} Q_W^{1-\alpha} - c_I - C_W. \tag{12}$$

[21] For simplicity, our specification here shuts down cross spillovers from writing to investigation or vice versa, that is, we do not consider $q_I(\lambda_I, \lambda_W)$ and $q_W(\lambda_I, \lambda_W)$ with nonzero cross-derivatives. This assumption effectively rules out "transfer learning" and treats the generative AI models for investigation and writing as task-specific rather than foundation models (see e.g., Bommasani, Hudson, Adeli, Altman, Arora, von Arx, Bernstein, Bohg, Bosselut, Brunskill et al. (2021)). Appendix L.1 will take on such further extensions.

[22] This condition is assumed for simplicity only. Later analysis will go through even if we relax it to only require $q_I(0) < Q_I$, $q_I(0)^{\alpha} Q_W^{1-\alpha} - c_I - C_W < 0$, and $q_I(0)^{\alpha} q_W(1)^{1-\alpha} - c_I - c_W < 0$, that is, in case no human does investigation the AI model for investigation will decay so much that it offers lower quality than human investigation and negative overall value regardless of whether one writes by human or AI.

4. *AI investigation + AI writing:*

$$\pi_{AA} \equiv \big(q_I(\lambda_I)\big)^{\alpha}\big(q_W(\lambda_W)\big)^{1-\alpha} - c_I - c_W. \tag{13}$$

4.2.2 Equilibrium characterization

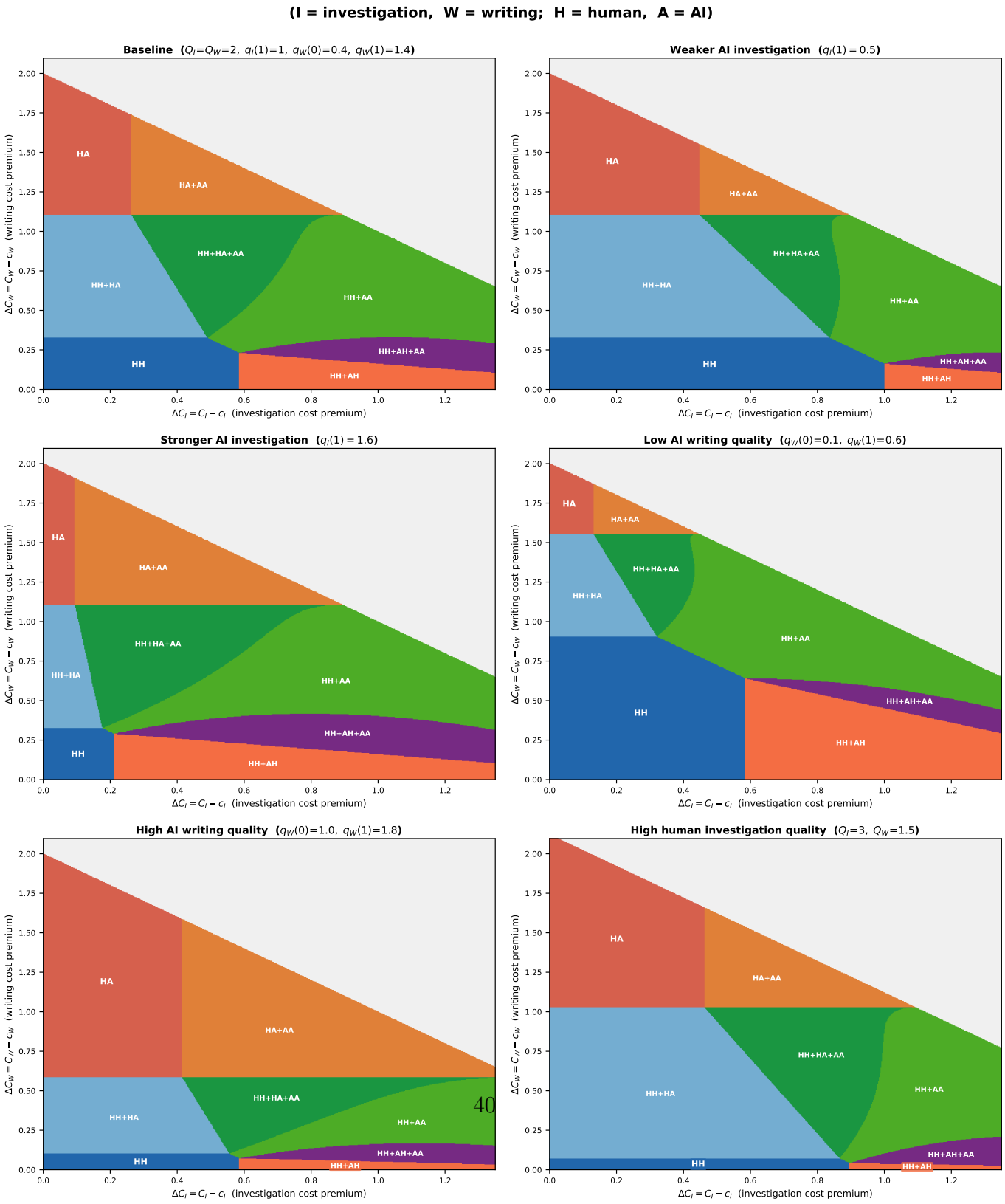
The following proposition summarizes all possible equilibria across parameter regimes. Since the full, general characterization is lengthy, we give a compact statement here with some quick remarks and relegate details to Appendix E. Our focus will be on the next subsection, where we illustrate the general result through two empirically relevant special examples.

Proposition 8. *Depending on the relative magnitudes of the parameters Q_I, Q_W, $q_W(0)$, $q_I(1)$, $q_W(1)$, C_I, C_W, c_I, and c_W, one of the following eight possible equilibrium outcomes emerges in the content production process: (1) only human investigators and human writers; (2) only human investigators and AI writers; (3) only human investigators and both human and AI writers; (4) only human writers and both human and AI investigators; (5) human investigators who are also human writers and AI investigators who are also AI writers; (6) only AI writers and both human and AI investigators; (7) all combinations except human investigators who are AI writers (8) all combinations except AI investigators who are human writers.*

Proof. See Appendix E. □

Figure 3 illustrates the equilibrium phase diagram for different levels of parameters. As Proposition 8 suggests (and Figure 3 illustrates), not all but only 8 of the total $2^4 - 1 = 15$ possible outcomes can sustain in equilibrium. Notably, consistent with our main result in the baseline model of Section 3, there is no equilibrium in which AI completely takes over both tasks, and indeed no equilibrium with AI completely taking over investigation. This

Figure 3: Equilibrium Phase Diagram in $(\Delta C_I, \Delta C_W)$ Parameter Space. Each region corresponds to a distinct equilibrium configuration; existence conditions for the four pure-assignment equilibria are mutually incompatible, so within each region the equilibrium is generically unique. $\Delta C_I \equiv C_I - c_I$ and $\Delta C_W \equiv C_W - c_W$ denote the human cost advantages for investigation and writing, respectively.



observation suggests that among all tasks within the content production process, the type of task with a fast-moving underlying knowledge base is difficult for AI to replace.

The eight equilibrium configurations are pairwise mutually exclusive—each prevails in a distinct parameter region (as shown in the proof of Proposition 8 in Appendix E and illustrated by Figure 3).[23] This structure reflects two forces: the fast-decay constraint ($q_I(0) = 0$), which rules out pure AI-investigation outcomes and restricts the feasible configurations; and the endogenous feedback between human participation and AI quality ($q_I(\lambda_I)$, $q_W(\lambda_W)$), which generates interior equilibria alongside corner ones.

We also note that in some equilibria, better AI writing can increase human investigation. In this sense, advances in AI for one task may actually enhance human's jobs for another task. We will elaborate on this finding in the special examples below.

4.2.3 Two special cases

We illustrate our general result through two special examples. The first one looks into the case in which AI is simply incapable of investigation, that is, $q_I(\lambda_I) = 0$ for all $\lambda_I \in [0, 1]$. The second example considers the case in which generative AI is so capable of writing that it takes over all human writing tasks — we will show that in this case a further improvement in AI writing quality may actually trigger more adoption of human investigation.

Incapable AI for investigation. As a benchmark, the special case in which AI is completely incapable of investigation is relevant if the investigation involves intensive physical activities (e.g., site visits) for which generative AI does not help at all. In this case, $q_I(\lambda_I) = 0$ for all $\lambda_I \in [0, 1]$, and everyone has to investigate in person, leading to $\lambda_I = 1$. Then immediately from the proof of Proposition 8, there are only three possible equilibria:

[23] Once we allow "transfer learning" between investigation and writing, multiple equilibria may occur, due to familiar coordination forces. See Appendix L.1 for details.

1. If $Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) \geq C_W - c_W$, there is an equilibrium with $\lambda_I = \lambda_W = 1$ that features only human investigators and human writers;

2. If $Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) \leq C_W - c_W$, there is an equilibrium with $\lambda_I = 1$ and $\lambda_W = 0$ that features only human investigators and AI writers;

3. If $Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) < C_W - c_W < Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(0))$, there is an equilibrium with $\lambda_I = 1$ and λ_W solving $q_W^{1-\alpha}(\lambda_W) = Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^\alpha}$. This equilibrium features only human investigators and both human and AI writers.

Capable AI for writing. Given that generative AI has demonstrated impressive writing skills, we further examine a special case in which generative AI is so capable of writing that it completely takes over human in all writing tasks involved in content production. In this case, we show that while a further improvement in AI's writing quality may solidify AI's advantage over humans for writing tasks, it may nevertheless also incentivize people to do more human investigation.

In this case, as no human writes anymore, $\lambda_W = 0$ and thus $q_W(\lambda_W) = q_W(0)$.[24] Immediately from the proof of Proposition 8, we have only two possible equilibria:

1. If generative AI is adequately capable of writing in that $q_W(0)$ is adequately high, or specifically if $q_W^{1-\alpha}(0) \geq Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^\alpha}$, $q_W^{1-\alpha}(0) > \frac{q_I^\alpha(1)Q_W^{1-\alpha} + (C_I - c_I) - (C_W - c_W)}{Q_I^\alpha}$, and $q_W^{1-\alpha}(0) > \frac{C_I - c_I}{Q_I^\alpha - q_I^\alpha(1)}$, there is an equilibrium with $\lambda_I = 1$ and $\lambda_W = 0$. This equilibrium features only human investigators and AI writers;

2. If generative AI is moderately capable of writing in that $q_W(0)$ is adequately high but

[24] It is also reasonable to focus on $q_W(0)$ only here because in the case of an already well-trained large language model, it is accurate to assume $q_W(\lambda_W) \approx q_W(0)$ for all $\lambda_W \in [0, 1]$. This assumption builds on the fact that what makes good writing tend to be stable over time, so that it is possible for an already well-trained large language model to perform well even without new training samples (a phenomenon known as "saturation" in the generative AI community) — this is in stark contrast with other tasks such as investigation, which involve a fast-moving world and quickly deteriorating quality once left out of fresh human inputs.

not too high, or specifically if $q_W^{1-\alpha}(0) \geq Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^\alpha}$ and $q_W^{1-\alpha}(0) \leq \frac{C_I - c_I}{Q_I^\alpha - q_I^\alpha(1)}$, there is an equilibrium with $\lambda_W = 0$ and $\lambda_I \in (0, 1)$ solving

$$q_I^\alpha(\lambda_I) = Q_I^\alpha - \frac{C_I - c_I}{q_W^{1-\alpha}(0)}. \tag{14}$$

This equilibrium features only AI writers and both human and AI investigators.

A comparative statics exercise with respect to Equation (14) gives the following finding:

Proposition 9. *If generative AI is adequately capable of writing, then the amount of human investigation (i.e., λ_I) increases as AI's writing quality further improves (i.e., $q_W(0)$ increases). Specifically, when generative AI is moderately capable of writing so that Equation (14) holds, λ_I increases as $q_W(0)$ increases; As generative AI's writing quality further improves above a threshold, all investigation tasks will remain with human.*

Proposition (9) renders an interesting finding: When generative AI becomes so proficient at writing that everyone adopts it for this task, a further improvement to AI's writing quality $q_W(0)$ will actually incentivize greater efforts toward human investigation, causing λ_I to increase. Intuitively, this result comes from the complementarity in the production function between writing and investigation. Specifically, the equilibrium indifference condition (14) requires that the quality gap between human and AI investigation, scaled by AI writing quality, equals the cost advantage of AI investigation: $(Q_I^\alpha - q_I^\alpha(\lambda_I)) \cdot q_W^{1-\alpha}(0) = C_I - c_I$. When $q_W(0)$ rises, writing quality acts as a larger multiplicative amplifier on investigation quality differences, raising the overall utility gains from choosing high-quality human investigation over AI investigation. This makes human investigation more attractive, drawing more humans in until the indifference condition is restored. When AI quality is not yet high enough and there are still human writers (when $\lambda_W > 0$), the effect of an increase in AI writing quality (i.e., an uplift of the function $q_W(\cdot)$ itself) may be offset by a decrease in the

equilibrium level of λ_W, but this offsetting effect does not occur when λ_W is already zero. This observation reconfirms the core thesis of our paper: generative AI will not fully replace humans. Moreover, it offers a practical lesson to human professionals who are concerned about job security: As generative AI increasingly dominates writing tasks, human professionals can find shelter in investigative work, because an ever-improving generative AI (in writing) may actually increase the demand for human-led investigation.[25]

4.2.4 Comparative statics

While our setting features multiple equilibria in general (see Proposition 8), they all exhibit rather consistent comparative statics, as summarized below:

Corollary 10. *In each equilibrium, all interior values of λ_I and λ_W strictly increase in Q_I and Q_W (except in the "Capable AI for writing" case discussed in Section 4.2.3, where λ_I strictly increases in $q_W(0)$ instead of Q_W). All interior values of λ_I (λ_W) also strictly decrease in C_I (C_W) and strictly increase in c_I (c_W). In some equilibria, λ_I (λ_W) strictly decrease in C_W (C_I) and strictly increase in c_W (c_I), while in other equilibria λ_I (λ_W) are unaffected by C_W (C_I) or c_W (c_I).*

Proof. See details contained in the proof of Proposition 8 in Appendix E. □

The intuition for Corollary 10 is two-fold: First, as a straightforward direct effect, for either task (investigation or writing), a lower quality from humans, higher cost to humans, or lower cost to AI will lead to higher AI adoption. Second, there is also an indirect effect coming from the complementarity between the two tasks: A lower quality from humans, higher cost to humans, or lower cost to AI in one task i (say writing) may also lead to higher AI adoption in another task j (say investigation), even though our model does not directly assume cross

[25] This finding is consistent with the empirical evidence from Goldfarb, Taska and Teodoridis (2023), who use online job posting data to demonstrate that AI job demands complement demands for other types of jobs. They use this evidence to argue that AI is a general-purpose technology (GPT).

spillovers in the AI quality functions. This indirect effect comes from the complementarity embedded in the production function. To see this, take as an example any equilibrium in which λ_I and λ_W take interior values. We first note that in such an equilibrium AI must have strictly worse quality than human for both investigation and writing — otherwise, suppose AI has a quality no worse than human in task $i \in \{I, W\}$, then since AI already has a strictly lower cost, it will be strictly dominant to adopt AI for task i, leading to $\lambda_i = 0$ and a contradiction. Hence, *ceteris paribus*, a decrease in human quality in task i reduces the overall utility gains from choosing high-quality human over AI in task j: since task i and task j quality are complements, lower task-i quality acts as a smaller multiplier on the quality gap in task j, diminishing the reward to human task j and thereby motivating more AI adoption in task j. Similarly, when human quality lowers, human cost increases, or AI cost lowers for task i, fewer human efforts will be put into i, which leads to lower AI quality for i due to training sample shrinkage. This reduced AI quality in task i in turn lowers the overall utility gains from choosing high-quality human over AI in task j, again because complementarity means that lower task-i quality diminishes the value of the quality contest in task j, and thus motivates more AI adoption in task j.

4.3 Dynamic Human-AI Interactions

Although the simple static model in Section 3 delivers our general insights, it does not reflect the intrinsically dynamic interactions between AI and human adoption.[26] As a microfoundation, this section presents a dynamic model that describes the co-evolution of AI and human adoption, and show that its implications are well-captured by our static baseline model.

The dynamic extension considers a stochastically evolving world. At any point in time,

[26] In financial markets, the interaction between passive and active investors are also intrinsically dynamic. The REE literature bypasses this problem because complicated dynamics could be captured by contingent actions through the submission of limit orders. Human-AI interactions, however, do not enjoy such a luxury, so it is imperative to derive a dynamic microfoundation for the reduced-form static model, as we do here.

one may either incur costly human efforts to acquire up-to-date information about the state of the world, or rely on generative AI at a lower cost (e.g., zero cost in the extreme) for the same purpose. A key tension is that, although generative AI has a lower marginal cost of use, it tends to be more outdated as it is trained on stale content, that is, signals produced by other agents' past information-acquisition efforts.

Our model reflects this tension. Specifically, in a repeated-renewal model, each costly human information acquisition resets the AI's staleness age to zero (denoted as a "refresh"). An agent who does not incur the human cost, however, has to rely on a cached AI estimate generated at the most recent past refresh, and the cached AI estimate incurs an error that increases in the volatility of the underlying state and grows over time since the last refresh.

4.3.1 Setup

Suppose the underlying information state ("truth"), denoted S_t, follows a Brownian motion

$$dS_t = \sigma_t dw_t.$$

The volatility σ_t can be time-varying and stochastic. For simplicity, we focus on the case of constant volatility $\sigma_t = \sigma$. Consider a generative AI model which is initially calibrated to the true state S_0. Without human input or intervention ("refresh"), the AI system will become increasingly outdated over time as the underlying environment S_t evolves.

There are $N \geq 1$ human agents, each of whom encounters decision opportunities that arrive according to an independent Poisson process with rate $\eta > 0$. Upon a decision opportunity at time t, a human agent chooses one of the following two actions:

1. (human effort): incurs a cost $C > 0$ to learn about the true state S_t and act upon it, and the action also adds to the training corpus that directly improves the AI model;

2. (AI usage): incurs a smaller cost $c < C$ on the generative AI output which nevertheless

tends to grow stale over time.

At each decision opportunity t, the arrival agent chooses an action a_t and receives a realized utility of $-(a_t - S_t)^2$. The agent who has incurred costly human effort would choose $a_t = S_t$, while those who adopt generative AI would choose $a_t = \arg\max_a \mathbb{E}[-(a - S_t)^2 | \mathcal{I}_t] = \mathbb{E}[S_t | \mathcal{I}_t]$, where $\mathcal{I}_t$ denotes the training set available to generative AI. Due to staleness, if the agent does *not* refresh, the action used is the cached AI estimate $a_t = S_t^-$, where t^- denotes the last refresh time.[27] Denote the *staleness age* of AI model by $x = t - t^-$, which grows with time until being reset to zero whenever any human agent chooses to refresh. This captures the empirically grounded feature of *model decay* in AI systems. Then, under constant volatility, the running loss at time t is linear in the age $x = t - t^-$:

$$\ell(x) \;=\; \mathbb{E}\big[(S_t - S_{t^-})^2\big] \;=\; \mathbb{E}\left[\int_{t^-}^{t} \sigma_u^2 \, du\right] = \sigma^2 x.$$

This is borne by all agents until the next agent arrives and chooses to refresh the AI system. The linearity of $\ell(x)$ in age x is a direct consequence of the independent-increments property of Brownian motion: the variance of $S_t - S_{t^-}$ accumulates additively over time. Under more general specifications—such as mean-reverting dynamics or jump-diffusions—the loss function would be concave or piecewise-linear in age, but the qualitative threshold structure of the equilibrium is preserved.

4.3.2 Equilibrium characterization

We explore a symmetric Nash equilibrium in which each agent follows a threshold strategy that targets an age Δ^* and refreshes at the *first* decision arrival after the staleness age x

[27] The assumption that the AI learns the true state of the world without noise at each human refresh time is innocuous for our purposes. Because one of our key arguments concerns human–AI coexistence, allowing the AI to learn without noise can only bias the analysis against finding our results.

exceeds Δ^*.[28] Specifically, under a symmetric threshold Δ, the staleness age grows from 0 to Δ deterministically; after the age hits Δ, the first of the N independent Poisson processes hits after an exponential waiting time $\tau \sim \text{Exp}(N\eta)$, at which point a refresh occurs and the age is reset to 0. Figure 4 illustrates the dynamics of the staleness age $x(t)$ under a threshold policy, with a characteristic "sawtooth" pattern: linear ramps up to Δ followed by random overshoots due to the stochastic waiting time, and discrete resets at refresh times.

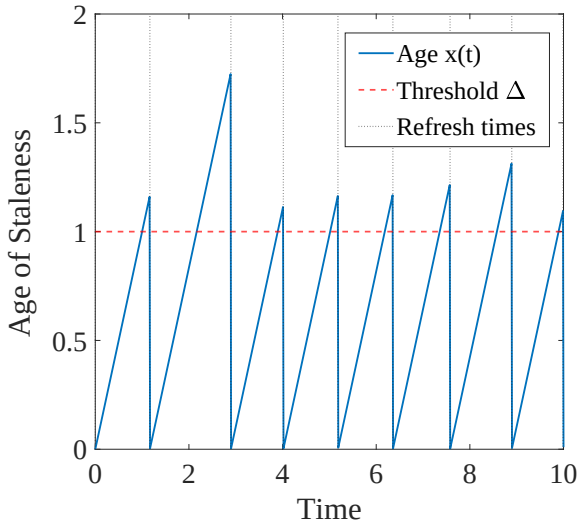


Figure 4: Illustration of the staleness age under a chosen threshold $\Delta = 1$.

Consider agent i, and take the strategies of the other $N-1$ agents as given. Let $\{\tau_n^i\}$ denote agent i's own decision times, which arrive at a Poisson rate η. If agent i uses AI at time τ_n^i, she incurs the cost $c + \sigma^2 x_{\tau_n^i}$. If instead she learns about the world by herself, then she pays C and resets the age of AI model to zero. Therefore, given a time discount rate $\rho > 0$, her objective function is

$$V(x) = \inf_{\{a_n^i\}_{n\geq 1}} \mathbb{E}_x \left[\sum_{n\geq 1} e^{-\rho \tilde{\tau}_n^i} \Big((c + \sigma^2 \tilde{x}_{\tau_n^i}) \, \mathbf{1}\{a_n^i = \text{AI}\} + C \, \mathbf{1}\{a_n^i = \text{refresh}\} \Big) \right]. \tag{15}$$

[28] We focus on symmetric equilibria as a natural benchmark: agents are ex ante identical (same cost C, same arrival rate η), so the symmetric profile is the equilibrium in which no agent is singled out for a distinguished role. Asymmetric equilibria, in which a strict subset of agents shoulder all refreshing, may also exist but require an exogenous coordination device to assign roles; we leave this analysis for future work.

This allows us to derive the Hamilton-Jacobi-Bellman (HJB) equation for the dynamic equilibrium problem. The HJB equation determines the equilibrium threshold of age at which each agent becomes indifferent between free-riding AI and refreshing it.

Proposition 11 (Symmetric Equilibrium Threshold)**.** *For $N \geq 1$, $\eta > 0$, $C > c > 0$, $\rho \geq 0$, and $\sigma > 0$, the optimal symmetric threshold Δ^* must solve the following nonlinear equation:*

$$C - c - \sigma^2 \Delta = \left(V(0) - \frac{\eta c}{\rho} \right) \left(e^{\rho \Delta} - 1 \right) - \frac{\eta \sigma^2}{\rho^2} \left[e^{\rho \Delta} - (1 + \rho \Delta) \right], \tag{16}$$

where

$$V(0) = \frac{(\rho + N\eta)(c + \sigma^2 \Delta - C) + \eta C}{\rho}. \tag{17}$$

Existence of this symmetric equilibrium requires[29] $(\rho V(0) - \eta c) e^{\rho \Delta} - \frac{\eta \sigma^2}{\rho} \left(e^{\rho \Delta} - 1 \right) + \sigma^2 \geq 0.$

Proof. See Appendix F. □

From equation (16), it is easy to verify that $\Delta^* = 0$ and $V(0) = \eta c / \rho$ when $C = c$. The full expression of $V(x)$ is provided in Appendix F. In particular, the value function becomes age-independent above the threshold ($x > \Delta^*$), because what matters is only the waiting time to the next arrival with the aggregate Poisson rate $N\eta$, regardless of how far above the threshold the current age is.

Two equilibrium quantities are of particular interest, as we discuss below:

[29] This condition ensures that a finite threshold $\Delta^* > 0$ exists. An agent who has just refreshed (age $x = 0$) faces the lowest possible staleness cost; the condition requires that, at this point, waiting is preferable to refreshing immediately again. It thereby rules out configurations in which agents prefer to refresh continuously—since refreshing is itself a human activity, this degenerate case ($\Delta^* = 0$) corresponds to agents producing human content at every opportunity, the dynamic analog of the all-human equilibrium ($\lambda = 1$) in the static model. Just as Lemma 2 ($r(1) - c > R - C$) rules out that case by ensuring AI content is valuable enough to be adopted, the existence condition here ensures the benefit of tolerating some AI staleness outweighs the cost of continuous human renewal.

Proposition 12 (Human Share and AI Quality)**.** *Under the optimal threshold Δ^*, the average share of decisions handled by humans rather than AI is given by*

$$\Lambda_{\text{human}}(\Delta^*) \;=\; \frac{1/(N\eta)}{\Delta^* + 1/(N\eta)} = \frac{1}{1 + N\eta\Delta^*} \in (0, 1]. \tag{18}$$

The quality of the AI system can be measured by the inverse of the long-run average loss with zero discount (i.e., the long-run average precision) under the optimal threshold strategy:

$$Q_{\text{AI}} = \left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \sigma^2 x_t \, dt \right)^{-1} = \frac{2N\eta\,\Lambda_{\text{human}}}{\sigma^2(1 + \Lambda_{\text{human}}^2)}. \tag{19}$$

which increases in the human share $\Lambda_{\text{human}} \in (0, 1]$ and reaches the maximum at $\Lambda_{\text{human}} = 1$.

Proof. See Appendix G. □

Proposition 12 shows that the dynamic model provides a microfoundation for the baseline static model discussed in Section 3. Recall that the static model has two key primitives, the amount of human content λ and the AI quality $r(\lambda)$, both of which can now be interpreted in more structural terms here. In the static model, $\lambda \in (0, 1]$ is the (endogenous) fraction of users who rely on human content and, through training, sustains AI quality $r(\lambda)$. In the dynamic model, a group of N human agents repeatedly intervenes to refresh the AI system, generating a stationary cycle with threshold Δ^*. This implies the long-run fraction of human decisions, $\Lambda_{\text{human}} = \frac{1}{1+N\eta\Delta^*} \in (0, 1]$, which microfounds λ, the mass of human content in the static model. The corresponding AI quality can be quantified by the long-run precision of the AI system, Q_{AI}, which indeed increases in Λ_{human} and thus agrees with the empirically observed scaling law that AI quality increases with the size of human training corpus. We can thus map λ monotonically to Λ_{human} and map $r(\lambda)$ to $Q_{\text{AI}}(\Lambda_{\text{human}})$ between these models.

Given the same discount rate $\rho > 0$, we can solve the social planner's problem that

minimizes the long-run average cost of all agents in this model.

Proposition 13 (Social planner's refresh policy)**.** *The social planner's problem admits a unique optimal threshold Δ_P, determined by*

$$C - c = \frac{\sigma^2}{\rho^2}\left[(N\eta + \rho)\rho\,\Delta_P - N\eta\big(1 - e^{-\rho\,\Delta_P}\big)\right]. \tag{20}$$

Moreover, Δ_P can be written in closed form as

$$\Delta_P = \frac{1}{\rho}\left[\frac{N\eta}{N\eta+\rho} + \frac{(C-c)\rho^2}{(N\eta+\rho)\sigma^2} + W\left(-\frac{N\eta}{N\eta+\rho}\exp\left\{-\frac{N\eta}{N\eta+\rho} - \frac{(C-c)\rho^2}{(N\eta+\rho)\sigma^2}\right\}\right)\right], \tag{21}$$

where $W(\cdot)$ denotes the principal branch of the Lambert W function. The planner's threshold satisfies $\Delta_P \geq \Delta^*$ with equality holds at $N = 1$. The comparative statics of Δ_P is below:

$$\frac{\partial \Delta_P}{\partial C} > 0, \qquad \frac{\partial \Delta_P}{\partial c} < 0, \qquad \frac{\partial \Delta_P}{\partial N} < 0, \qquad \frac{\partial \Delta_P}{\partial \eta} < 0, \qquad \frac{\partial \Delta_P}{\partial \sigma} < 0, \qquad \frac{\partial \Delta_P}{\partial \rho} > 0.$$

Proof. See Appendix H. □

At $N = 1$, the planner's threshold equation (20) coincides with the individual's equilibrium threshold equation (16) such that $\Delta^* = \Delta_P$. However, when $N > 1$ we have $\Delta^* > \Delta_P$. The social planner internalizes the full benefit of refreshing earlier for all future users of the system. In contrast, decentralized agent internalizes only her private gain from refresh and, once the age is high, also free-rides on the possibility that another agent will refresh soon.

4.3.3 Comparative statics

Analytical results in Propositions 11 and 12 imply the following properties:

Corollary 14. *The individual's optimal threshold Δ^* characterized by the equilibrium equation (16) increases in C, N, and ρ, and decreases in c, η, and σ; the share of human*

decisions Λ_{human} *decreases with* C*,* N*,* η*, and* ρ*, and increases with* c *and* σ*; the AI quality* Q_{AI} *decreases with* C *and* ρ*, increases with* c *and* η*, and is generally ambiguous in* σ *and* N*.*

Proof. See Appendix I. □

The comparative static analysis in Corollary 14 yield multiple policy-relevant implications. First, reducing the human refresh cost C is the most powerful strategy: a lower C reduces the optimal threshold Δ^*, raises the share of human activities Λ_{human} and enhances the AI quality Q_{AI}; in other words, policies that make high-skill human intervention cheaper (e.g., via training or remuneration) can simultaneously support human work and improve AI performance.[30] Second, all else equal, increasing the decision arrival rate η moves the system toward more frequent use of AI on the margin (lower Λ_{human}) and improves the AI performance Q_{AI}; this suggests that high-frequency environments naturally support AI-heavy workloads when good refresh opportunities are available. Finally, higher volatility σ in the underlying space lowers the threshold Δ^*, increases Λ_{human} and may worsen AI quality.[31] Thus, when environments are changing more rapidly, efficient operation calls for a higher level of human involvement in terms of the flow of human-generated content.

Taken together, these results suggest that policies aiming to reduce the effective cost of human creatives (by providing better tools and remuneration for example) are not merely defensive measures for preserving human jobs, but they also enhance the performance and sustainability of the AI system itself. Moreover, human professionals' comparative advantage can concentrate in states of high volatility which naturally raise the human share of decisions. Therefore, human career focus may tilt toward high volatility domains (such as breaking events, regulatory changes, scientific outbreaks, and product innovations) where the marginal

[30] A similar pattern is observed in our static extension with heterogeneous preferences in Section 4.1, where C governs the entry cutoff $\theta_{H0} = C/R$ for human content. In that setting, reducing human cost C can simultaneously strengthen the human side, raise the AI quality, and widen the adoption of AI.

[31] A higher σ directly magnifies the error associated with a given age process and tends to lower AI quality, but it also induces earlier refresh by lowering Δ^*, which tends to improve AI quality.

value of human intervention is high.

4.3.4 Special case with vanishing discount rate $\rho \to 0$

We further illustrate the above analysis with the case of zero-discount ($\rho = 0$), which features closed-form expressions.

Corollary 15. *In the zero-discount limit ($\rho \to 0$), the threshold equation (16) reduces to*

$$\eta\sigma^2 \left(N - \frac{1}{2}\right) \Delta^2 + \left[\sigma^2 - \eta(N-1)(C-c)\right]\Delta - (C-c) = 0. \tag{22}$$

and the optimal threshold admits a closed-form solution:

$$\lim_{\rho \to 0} \Delta^* = \frac{-\left[\sigma^2 - \eta(N-1)(C-c)\right] + \sqrt{\left[\sigma^2 - (C-c)\eta(N-1)\right]^2 + 2\eta\sigma^2(2N-1)(C-c)}}{\eta\sigma^2(2N-1)}. \tag{23}$$

For comparison, the social planner's optimal threshold under zero discount rate solves

$$\frac{\eta\sigma^2}{2}\Delta_P^2 + \frac{\sigma^2}{N}\Delta_P - \frac{C-c}{N} = 0. \tag{24}$$

and the unique solution is

$$\lim_{\rho \to 0} \Delta_P = -\frac{1}{N\eta} + \sqrt{\frac{1}{(N\eta)^2} + \frac{2(C-c)}{N\eta\sigma^2}}, \tag{25}$$

which depends on the aggregate decision rate $N\eta$ and the relative cost $C - c$.

Proof. See Appendix J. □

Given the closed-form solutions (23) and (25), it is easy to verify that when $N = 1$ the socially optimal threshold coincides with the individually optimal threshold, i.e., $\Delta_P = \Delta^*$. When $N > 1$, we have $\Delta_P < \Delta^*$, because everyone waits longer, hoping that others will

refresh. This observation echoes the free-riding phenomena in our baseline model (AI-users free-riding humans) and those in Grossman and Stiglitz (1980) (passive investors free-riding active investors). In welfare terms, this gap implies that the equilibrium features under-provision of human content: agents collectively tolerate a higher average staleness age than is socially optimal, resulting in lower aggregate AI quality. The welfare loss is increasing in N, as more agents amplify the incentive to free-ride on others' refreshing effort.

5 Empirical Implications and Testable Predictions

The model delivers sharp comparative statics and equilibrium predictions that, in principle, can be taken to data. This section makes these predictions explicit and discusses how a researcher could map the model's theoretical objects to observable quantities. We organize the predictions by their source in the model, progressing from the baseline to the extensions. Table 1 provides a compact summary; the subsections below elaborate on each prediction.

Prediction 1: Human–AI coexistence (never full displacement). *Statement.* Even as generative AI improves ($\xi \uparrow$), the equilibrium human content share λ^* remains strictly positive. Full AI displacement ($\lambda^* = 0$) is never an equilibrium (Lemma 1).

Operationalization. The variable λ^* maps to the share of content produced by humans in a given market segment. Researchers can proxy this with: (i) the fraction of creative-service contracts awarded to human freelancers on platforms such as Upwork or Fiverr, where AI improvement can be proxied by time elapsed since major model releases (GPT-4: March 2023; GPT-4o: May 2024); (ii) the share of human-authored versus AI-generated articles on news or content platforms, detectable via AI-content classifiers (e.g., GPTZero, Originality.ai) applied to publisher archives; (iii) employment levels in content-intensive occupations from the BLS Occupational Employment and Wage Statistics (OES), including Writers and

Table 1: Summary of testable empirical predictions

#	Prediction	Model source	Key variable(s)	Suggested data
1	Human–AI coexistence	Lemma 1	$\lambda^* > 0$ always	BLS OES (SOC 27-30xx); Upwork/Fiverr transaction data; AI-content classifiers on publisher archives
2	AI quality ↑ reduces human share	Prop. 4	$\partial\lambda^*/\partial\xi < 0$	MMLU/HumanEval benchmarks (quarterly index); OpenAI API pricing history; Lightcast job-posting volume
3	Niche/high-value content survives longer	Prop. 4; §4.1	λ^* increasing in R; small $F(\theta)$	Substack creator revenue; SSRN vs. Medium article counts; Google Trends long-tail query traffic
4	Middle-out adoption on demand side	§4.1	Hump-shaped adoption in θ	Bick et al. (2026) household data; Pew AI Adoption Surveys 2023–2025; WTP choice experiments
5	Bottom-up displacement on supply side	§4.1	Low-skill workers exit first	Upwork/Fiverr pre-GPT ratings & hourly rates; O*NET routine-task intensity; Lightcast credentials
6	Human creators persist in volatile domains	§4.3	Intervention ↑ with σ	GDELT event counts; Wikipedia revision frequency by category; BLS newspaper publisher subsectors
7	AI writing saturation raises demand for investigation	§4.2	Non-monotone cross-task adoption	Lightcast (reporter vs. editor role postings); Reuters Institute Digital News Report; INN nonprofit funding flows

Authors (SOC 27-3040), Graphic Designers (27-1024), and Photographers (27-4021), tracked over time.

Falsification. The prediction is falsified if human employment in any broad creative category converges to zero after AI adoption saturates. Consistent with coexistence, Eisfeldt,

Schubert and Zhang (2023) document firm-level AI exposure but find continued human employment in content roles, and Bick, Blandin and Deming (2026) find AI adoption below 50% even among heavy users.

Prediction 2: AI quality negatively predicts human content share. *Statement.* As AI capability ξ rises, λ^* falls (Proposition 4).

Operationalization. Proxy ξ with: (i) aggregate model benchmark scores over time (MMLU, HumanEval, or HELM leaderboards), aggregated to a quarterly index; (ii) the log-inverse of API pricing per 1,000 tokens (OpenAI API pricing history), as falling inference costs proxy for both lower c and higher ξ; (iii) a post-release indicator for major model introductions, supporting a regression-discontinuity or difference-in-differences design around the ChatGPT launch (November 2022) or GPT-4 release (March 2023). The outcome λ^* can be measured as monthly job-posting volume for writing and design roles (Lightcast/Burning Glass data) relative to total postings, or as the platform-level human content shares described in Prediction 1.

Falsification. The prediction is falsified if human content share rises monotonically with AI capability improvements. Consistent with the prediction, Blank, Schubert and Zhang (2025) document that ChatGPT adoption shifts time from productive to leisure activities, suggesting substitution of AI for human effort.

Prediction 3: Higher-value and niche content retains more human production. *Statement.* λ^* increases in R (Proposition 4). With heterogeneous consumer preferences (Section 4.1), niche content—consumed by a small mass of high-θ types—is also more insulated from AI disruption than mainstream commodity content.

Operationalization. Proxy R with: (i) content pricing or subscription revenue per piece (Substack creator revenue data; Bloomberg/Reuters paywall conversion rates by category); (ii) content complexity (Flesch-Kincaid reading grade level, domain-specific terminology den-

sity). Identify niche content as: low-circulation specialist outlets versus mainstream publishers; long-tail Google search-query traffic (bottom decile by volume); or specialist platforms (academic preprints on SSRN/arXiv versus mass-market articles on Medium). A difference-in-differences design comparing the post-AI-shock change in human content share between high-R/niche versus low-R/mainstream categories would test this prediction.

Falsification. The prediction is falsified if niche or premium categories show larger declines in human production than commodity categories following AI adoption. Consistent evidence comes from Brynjolfsson, Li and Raymond (2025), who find that AI gains are largest for "moderately rare" customer inquiries where AI has sufficient training data—i.e., mainstream rather than niche content—mirroring our model's prediction that AI is most disruptive where $r(\lambda)$ is highest for a given λ.

Prediction 4: "Middle-out" adoption on the demand side. *Statement.* With heterogeneous consumer preferences (Section 4.1), AI adoption is concentrated among consumers with intermediate taste for quality θ: high-θ consumers remain loyal to human content, low-θ consumers who previously opted out do not enter the market, and middle-θ consumers switch to AI.

Operationalization. Proxy consumer type θ with: (i) household income or education level (higher income/education proxying for higher willingness-to-pay for quality); (ii) professional versus casual consumption mode (subscription status, reading depth, or content engagement metrics). Data sources include: (a) Bick, Blandin and Deming (2026), who provide household-level AI adoption data linked to income and education; (b) Pew Research Center AI Adoption Surveys (2023–2025), which contain cross-sectional adoption rates by demographic group; (c) survey or field experiments presenting subjects with a choice between human-written and AI-written content at varying prices can directly elicit θ and map it to adoption decisions.

Falsification. The prediction is falsified if adoption is monotonically increasing or decreasing in income/preference type, rather than hump-shaped in the middle. Consistent with the middle-out pattern, Bick, Blandin and Deming (2026) document that AI adoption is highest among younger and higher-income households *but below the very top* of the income distribution.

Prediction 5: AI replaces low-skill producers first ("bottom-up" pattern on the supply side). *Statement.* With producer skill heterogeneity (Section 4.1), AI adoption first displaces the lowest-skill human producers, while high-skill producers persist or even earn higher rents as AI quality in complementary tasks improves.

Operationalization. Proxy worker skill with: (i) prior earnings, reputation scores, or hourly rates on freelance platforms (Upwork/Fiverr) measured before the ChatGPT launch; (ii) educational credentials matched to job postings (Lightcast); (iii) occupation-level routine-task intensity from O*NET task descriptors, as an inverse proxy for skill. A difference-in-differences design comparing earnings changes for low- versus high-skill creative workers around AI adoption events tests this prediction. For platform-level analysis, Upwork's anonymized transaction histories or Fiverr's public seller data can be scraped and linked to pre-shock ratings.

Falsification. The prediction is falsified if high-skill creative workers experience larger earnings declines than low-skill workers following AI adoption. Consistent with the bottom-up pattern, Brynjolfsson, Li and Raymond (2025) and Cui, Demirer, Jaffe, Musolff, Peng and Salz (2025) both find that less-experienced workers gain most from AI assistance, consistent with AI compressing skill premia at the lower end of the distribution.

Prediction 6: Human creators survive longer in high-volatility, time-sensitive domains. *Statement.* From the dynamic extension (Section 4.3), human intervention frequency increases with state volatility σ: in fast-moving domains where the world state

changes rapidly, AI grows stale quickly and human refreshes are required more often.

Operationalization. Proxy domain volatility σ with: (i) news update frequency in a topic category, measured via the GDELT Global Database of Events, Language, and Tone, or via the count of revisions to Wikipedia articles per category per month; (ii) social media volume velocity for a topic (Twitter/X firehose or Google Trends daily change); (iii) financial-market analogy: equity return volatility by sector, with higher-volatility sectors proxying for more time-sensitive information environments. Compare human content production persistence across high-σ domains (financial news, political reporting, sports) versus low-σ domains (literary essays, historical analysis, mathematical exposition). BLS subsector employment data for newspaper publishers (NAICS 51111) can separate fast-cycle from long-cycle journalism.

Falsification. The prediction is falsified if human content production declines faster in volatile or time-sensitive domains than in stable domains following AI adoption.

Prediction 7: AI improvement in writing complements demand for human investigation. *Statement.* From the task-complementarity extension (Section 4.2), when AI is already highly adopted for writing-type sub-tasks (the "saturation" regime), further improvement in AI writing quality *increases* equilibrium demand for human investigative effort—a non-monotonic pattern of AI adoption across tasks.

Operationalization. Proxy writing-task AI adoption with: (i) adoption rates of AI writing assistants among journalists or researchers (Reuters Institute Digital News Report; academic survey data); (ii) fraction of articles disclosing AI writing assistance (publisher-level post-disclosure mandates). Proxy investigative task demand with: (iii) job postings for investigative reporters, fact-checkers, and field correspondents versus copy editors and content writers (Lightcast data, parsed by O*NET task descriptors); (iv) funding flows to investigative journalism nonprofits (INN member financial data). A cross-outlet panel comparing organizations that adopted AI writing tools early versus late, testing whether investigative-

role headcount diverges over time, would provide a clean test.

Falsification. The prediction is falsified if AI improvement in writing leads to a monotonic decline in demand for all human creative sub-tasks, with no evidence of complementarity or rising demand for investigative or field-reporting roles.

6 Conclusion

This paper at its core makes a simple point: Generative AI will not completely take over human creatives. Our idea resembles the tension between active and passive investors in financial markets, and the result demonstrates that some classic insight in information economics remains relevant when we embrace new technologies such as generative AI. We not only illustrate our insight using a simple baseline model, but also analyze multiple extensions to show how heterogeneous consumer preferences, complementarity/substitution between tasks, and dynamic human interaction with the generative AI and the physical world may affect the equilibrium outcome in the human–AI ecosystem, providing rich implications for various stakeholders, including human creatives, policy makers, firm managers, and investors.

A key friction behind why generative AI cannot completely replace humans is that AI cannot automatically perform the task of *investigation* (or information acquisition) regarding the physical world. Therefore, existing generative models need to rely on the first-hand inputs created by humans to generate second-hand outputs. The conclusion of our paper may change, however, if this fundamental friction is no longer present. One foreseeable way how this friction may disappear is when generative AIs can directly learn from the physical world (e.g., via sensors or robots), effectively performing the *investigation* task themselves. The fact that many AI models are now increasingly multimodal or even omni-model shows some promise to qualify our conclusion — however, that vision still looks a bit far away at the moment and would be a discussion of certain more advanced AI paradigms that we have

not yet seen in production (e.g., terms like "physical AI" have been floating in some circles in the AI community, although they often tend to be associated with AI-equipped robots and we are not aware of any concrete products with mass adoption on this front yet). All that said, the core message of this current paper at least infuses more economic substantiation for the repeated argument by Yann LeCun (Turing Award winner and one of the founding fathers of contemporaneous AI development) that current large language models (LLMs) like ChatGPT are not sufficient for achieving true Artificial General Intelligence (AGI).[32]

References

Abis, Simona, and Laura Veldkamp. 2024. "The changing economics of knowledge production." *The Review of Financial Studies*, 37(1): 89–118. 10

Acemoglu, Daron, and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American Economic Review*, 108(6): 1488–1542. 9

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and new tasks: How technology displaces and reinstates labor." *Journal of Economic Perspectives*, 33(2): 3–30. 9

Acemoglu, Daron, and Todd Lensman. 2023. "Regulating Transformative Technologies." National Bureau of Economic Research. 9

Admati, Anat R, and Paul Pfleiderer. 1990. "Direct and indirect sale of information." *Econometrica : journal of the Econometric Society*, 901–928. Publisher: JSTOR. 12

Agrawal, Ajay K, Joshua S Gans, and Avi Goldfarb. 2025. "Genius on demand: The value of transformative artificial intelligence." National Bureau of Economic Research. 7

[32]See for example an interview with Yann LeCun.

Alemohammad, Sina, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G Baraniuk. 2023. "Self-Consuming Generative Models Go MAD." *arXiv preprint arXiv:2307.01850*. 16

Andrews, Isaiah, and Maryam Farboodi. 2025. "Do Markets Believe in Transformative AI?" National Bureau of Economic Research. 12

Barlow, Richard E., and Frank Proschan. 1965. *Mathematical Theory of Reliability*. New York:John Wiley & Sons. 13

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" 610–623. 17

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. "On the rise of fintechs: Credit scoring using digital footprints." *The Review of Financial Studies*, 33(7): 2845–2897. 10

Bick, Alexander, Adam Blandin, and David J Deming. 2026. "The rapid adoption of generative AI." *Management Science*. 8, 56, 57, 58

Bini, Pietro, Lin William Cong, Xing Huang, and Lawrence J Jin. 2025. "Behavioral Economics of AI: LLM Biases and Corrections." *Available at SSRN 5213130*. 11

Blank, Michael, Gregor Schubert, and Miao Ben Zhang. 2025. "The Household Impact of Generative AI: Evidence from Internet Browsing Behavior." *working paper*. 9, 56

Bommasani, Rishi, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut,

Emma Brunskill, et al. 2021. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258*. 38, 117

Bond, Philip, Alex Edmans, and Itay Goldstein. 2012. "The real effects of financial markets." *Annual Review of Financial Economics*, 4(1): 339–360. Publisher: Annual Reviews. 12

Breon-Drish, Bradyn. 2015. "On existence and uniqueness of equilibrium in a class of noisy rational expectations models." *Review of Economic Studies*, forthcoming. Publisher: Oxford University Press. 12

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. "Language models are few-shot learners." *Advances in neural information processing systems*, 33: 1877–1901. 10

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, 140(2): 889–942. 8, 57, 58

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard. 2023. "Bayesian solutions for the factor zoo: We just ran two quadrillion models." *The Journal of Finance*, 78(1): 487–557. 10

Chen, Hui, Antoine Didisheim, and Luciano Somoza. 2024. "Out of the (black) box: AI as conditional probability." *Available at SSRN 5012852*. 10

Chen, Jian, Guohao Tang, Guofu Zhou, and Wu Zhu. 2024. "ChatGPT, Stock Market Predictability and Links to the Macroeconomy." *working paper*. 11

Chen, Shuaiyu, T Clifton Green, Huseyin Gulen, and Dexin Zhou. 2024. "What Does ChatGPT Make of Historical Stock Returns? Extrapolation and Miscalibration in LLM Stock Return Forecasts." *arXiv preprint arXiv:2409.11540*. 12

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2025. "The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers." *Available at SSRN 4945566*. 8, 32, 58, 108

Davis, Mark H. A., and Andrew R. Norman. 1990. "Portfolio selection with transaction costs." *Mathematics of Operations Research*, 15(4): 676–713. 13

Demirer, Mert, J Horton, Nicole Immorlica, Brendan Lucier, Peyman Shahidi, and Preliminary Draft-Comments Welcome. 2025. "The economic impacts of generative AI on the structure of work." Working Paper. 7

Diamond, Douglas W, and Robert E Verrecchia. 1981. "Information aggregation in a noisy rational expectations economy." *Journal of Financial Economics*, 9(3): 221–235. Publisher: Elsevier. 12

Dou, Winston Wei, Itay Goldstein, Zigang Li, and Liyan Yang. 2025. "Human Edge, Machine Limits: AI-Human Competition in Financial Markets." *Machine Limits: AI-Human Competition in Financial Markets (November 18, 2025)*. 12

Eisfeldt, Andrea L, Gregor Schubert, and Miao Ben Zhang. 2023. "Generative ai and firm values." National Bureau of Economic Research. 8, 55

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "Gpts are gpts: An early look at the labor market impact potential of large language models." *arXiv preprint arXiv:2303.10130*. 9

Engelberg, Joseph, Asaf Manela, William Mullins, and Luka Vulicevic. 2025. "Entity Neutering." *Available at SSRN 5182756*. 12

Farboodi, Maryam, and Laura Veldkamp. 2020. "Long-run growth of financial data technology." *American Economic Review*, 110(8): 2485–2523. 10

Farboodi, Maryam, Dhruv Singal, Laura Veldkamp, and Venky Venkateswaran. 2025. "Valuing financial data." *The Review of Financial Studies*, 38(3): 938–980. 10

Gerstgrasser, Matthias, Geoffrey Irving, and Arvind Narayanan. 2024. "Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data." *arXiv preprint arXiv:2404.01413*. 11

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis. 2023. "Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings." *Research Policy*, 52(1): 104653. 44

Grennan, Jillian, and Roni Michaely. 2020. "Artificial intelligence and high-skilled work: Evidence from analysts." *Swiss Finance Institute Research Paper*, , (20-84). 9

Grossman, Sanford J. 1976. "On the efficiency of competitive stock markets where trades have diverse information." *Journal of Finance*, 31(2): 573–585. Publisher: Wiley Online Library. 12

Grossman, Sanford J, and Joseph E Stiglitz. 1980. "On the impossibility of informationally efficient markets." *American Economic Review*, 70(3): 393–408. Publisher: JSTOR. 1, 2, 12, 26, 27, 54

Guu, Kelvin, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. "REALM: Retrieval-Augmented Language Model Pre-Training." 3929–3938. 10

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. 2025. "Artificial intelligence and the labor market." National Bureau of Economic Research. 7

Hellwig, Martin F. 1980. "On the aggregation of information in competitive markets." *Journal of Economic Theory*, 22(3): 477–498. Publisher: Academic Press. 12

He, Songrun, Linying Lv, Asaf Manela, and Jimmy Wu. 2025. "Chronologically Consistent Large Language Models." *arXiv preprint arXiv:2502.21206*. 12

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou. 2017. "Deep learning scaling is predictable, empirically." *arXiv preprint arXiv:1712.00409*. Empirical study showing error follows power-law in data size on many tasks. 21

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. 2022. "Training compute-optimal large language models." *arXiv preprint arXiv:2203.15556*. Chinchilla scaling. 2, 10, 15, 19

Horton, John J. 2023. "Large language models as simulated economic agents: What can we learn from homo silicus?" National Bureau of Economic Research. 10

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang. 2024. "ChatGPT and corporate policies." National Bureau of Economic Research. 9

Jiang, Wei, Junyoung Park, Rachel Jiqiu Xiao, and Shen Zhang. 2025. "AI and the Extended Workday: Productivity, Contracting Efficiency, and Distribution of Rents." National Bureau of Economic Research. 9, 36

Johnson, Simon, and Daron Acemoglu. 2023. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Hachette UK. 9

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. "Scaling laws for neural language models." *arXiv preprint arXiv:2001.08361*. 2, 10, 15, 19, 21

Korinek, Anton, and Megan Juelfs. 2022. "Preparing for the (non-existent?) future of work." National Bureau of Economic Research. 9

Kyle, Albert S. 1989. "Informed speculation with imperfect competition." *Review of Economic Studies*, 56(3): 317–355. Publisher: Oxford University Press. 12

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems*, 33: 9459–9474. 10

Li, Jiasun. 2023. "Information aggregation via contracting." *Journal of Finance*, 78(2): 935–965. Publisher: Wiley Online Library. 12

Liu, Huben, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. 2025. "Technology and Labor Markets: Past, Present, and Future; Evidence from Two Centuries of Innovation." National Bureau of Economic Research. 9

Lopez-Lira, Alejandro, and Yuehua Tang. 2023. "Can chatgpt forecast stock price movements? return predictability and large language models." *arXiv preprint arXiv:2304.07619*. 11

Lopez-Lira, Alejandro, Yuehua Tang, and Mingyin Zhu. 2025. "The Memorization Problem: Can We Trust LLMs' Economic Forecasts?" *arXiv preprint arXiv:2504.14765*. 12

Lu, Jie, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. 2018. "Learning under concept drift: A review." *IEEE transactions on knowledge and data engineering*, 31(12): 2346–2363. 11

Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. "Training Language Models to Follow Instructions with Human Feedback." *Advances in Neural Information Processing Systems*, 35: 27730–27744. 15

Puterman, Martin L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York:John Wiley & Sons. 13

Ross, Sheldon M. 1996. *Stochastic Processes.* . 2 ed., New York:John Wiley & Sons. 13

Scao, Teven Le, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. "Bloom: A 176b-parameter open-access multilingual language model." *arXiv preprint arXiv:2211.05100*. 10

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024. "AI models collapse when trained on recursively generated data." *Nature*, 631(8022): 755–759. 3, 11, 16, 21

Soner, H. Mete, and Steven E. Shreve. 1994. "Optimal investment and consumption with transaction costs." *The Annals of Applied Probability*, 4(3): 609–692. 13

Tan, Lin, Huihang Wu, and Xiaoyan Zhang. 2023. "Large Language Models and Return Prediction in China." *Available at SSRN 4712248*. 12

Touvron, Hugo, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. "Llama: Open and efficient foundation language models." *arXiv preprint arXiv:2302.13971*. 10

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is all you need." *Advances in neural information processing systems*, 30. 10, 23

Viering, Tom, and Marco Loog. 2022. "The shape of learning curves: a review." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 7799–7819. Survey of empirical and theoretical learning curve behaviour. 21

Wang, Thomas, Adam Roberts, Daniel Hesslow, Teven Le Scao, Hyung Won Chung, Iz Beltagy, Julien Launay, and Colin Raffel. 2022. "What language model architecture and pretraining objective works best for zero-shot generalization?" 22964–22984, PMLR. 10

Wu, Shijie, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. "Bloomberggpt: A large language model for finance." *arXiv preprint arXiv:2303.17564*. 10

Yan, Hongjun, Liyan Yang, Xueyong Zhang, and Deqing Zhou. 2024. "To Dissimulate or Not to Dissimulate? Insider Trading When Anticipating Future Information." *Insider Trading When Anticipating Future Information (February 24, 2024)*. 12

Yoon, Youngseok, Dainong Hu, Iain Weissburg, Yao Qin, and Haewon Jeong. 2025. "Model Collapse in the Self-Consuming Chain of Diffusion Finetuning: A Novel Perspective from Quantitative Trait Modeling." *ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models*. 11

Zhang, Miao Ben. 2019. "Labor-technology substitution: Implications for asset pricing." *The Journal of Finance*, 74(4): 1793–1839. 9

Appendix

A Anecdotes from the institutional press

The concern of generative AI replacing journalists is not merely an academic debate, but rather an imminent reality. The following excerpt from a *SeekingAlpha* article "Wall Street Breakfast: AI Generation" on August 14th, 2023 gives a vivid example:

> *With generative AI poised to reshape the publishing industry, companies are scrambling to take positions on the latest technology. What does it mean for their operations? How will it impact their workforce? And when can they begin to expect to benefit from, or be harmed by, the new developments? These questions have rattled publishers in nearly every sector of the media industry, from online content and the printed word to movies and even music.*
>
> *Snapshot: The knee-jerk reaction of many firms was to band together to stave off any threats to their bottom line. Hollywood actors and writers have gone on strike in part over fear that they might lose or share revenue with machines, with the walkout now going on for more than 100 days. Publishers like The New York Times (NYT), Wall Street Journal's News Corp. (NWS) and Dotdash Meredith owner IAC (IAC) also started talks to establish a united front against Google (GOOG) (GOOGL) and Microsoft (MSFT). The coalition explored suing Big Tech over "content scraping" and training their AI models on data produced by the organizations, but the NYT has reportedly dropped out of the group, suggesting that they might reach separate deals with tech giants that wield tremendous power over distribution and visibility. Check out recent subscriber numbers for the NYT and News Corp. in charts.*
>
> *While corporations are debating their cash flow streams, the non-profit sector is*

having an easier time making decisions. In July, the Associated Press reached a two-year deal with OpenAI, the parent company of ChatGPT, by declaring the first official news-sharing agreement with an artificial intelligence firm. That direction could benefit others as they scale their offerings and revenue, such as News Corp Australia (NWS), which is using a team of four staff to produce 3,000 articles per week using generative AI. The items cover daily topics like local weather, fuel prices, and traffic conditions, and have driven a surge in subscriptions and the subsidiary's best year in a decade.

Outlook: Consumers are also debating what artificial intelligence means for their media consumption. Should they be informed whether a robot wrote an article, or is the creative process continuing to evolve in a similar fashion to what algo-generated tunes and CGI did to the music and movie industries? Things get even more complicated if generative AI is only used to conduct research or help write an article, or prompt topics that enhance value and quality. Trust also plays a role here, but those lines might be lost or could one day be strengthened, depending on the future of brands and their audiences.

B The Generative AI Shock to Content Creators

The emergence of generative AI has ignited a profound economic conflict between the technology sector and global content creation industries. Proponents argue that generative AI represents a democratization of creativity and promises a future in which AI-augmented tools lower barriers to entry for complex content production. Opponents, by contrast, view this transition as a predatory disruption, characterized by practices such as the unauthorized use of human-authored data. Indeed, since the public release of ChatGPT in late November 2022, generative AI has been widely described as a disruptive force capable of fundamen-

tally restructuring the professional class of content creators. This "Generative AI Shock" is manifesting across four primary dimensions: the displacement of routine human labor, the erosion of intellectual property, the collapse of referral-based business models, and the risk of degradation of the information ecosystem. While the discussion below focuses on journalism, the underlying points are broadly applicable across many other creative industries.

The Displacement of Routine Human Labor. Generative AI has enabled the granular displacement of routine tasks such as transcription, copy-editing, and basic information synthesis. The automation of these functions contributes to a shrinking job market for junior professionals, as entry-level positions are increasingly hollowed out; see the *Wall Street Journal* commentary "*AI Means the End of Entry-Level Jobs*" on December 22, 2025. This shift forces a transformation of roles in which human creatives act more as managers than producers, with primary value residing in higher-level judgment, ethical oversight, and lived experience. While AI can significantly enhance human productivity across many tasks, it simultaneously increases the cognitive burden of verification, as human professionals remain legally and ethically responsible for AI-augmented outputs.

The Erosion of Intellectual Property. The most prominent conflict is exemplified by *The New York Times v. OpenAI* litigation, in which the plaintiff alleges that Large Language Models (LLMs) free-ride on The New York Times's massive investments in its journalism; see the news article "*The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*" on December 27th, 2023. Mainstream outlets have highlighted that LLMs do not merely learn from data but often mimic and regurgitate copyrighted material verbatim, effectively creating derivative works that compete directly with the original sources. This dynamic has produced a fragmented industry response, split between strategic partnership/licensing deals and aggressive litigation, both aimed at preserving the economic value of human-authored content.

The Collapse of Referral-based Business Model. Another immediate threat is the transition of search engines into "answer" engines. This shift creates a *zero-click* environment in which AI-generated summaries satisfy user queries directly within the search interface, bypassing the referral traffic that sustains digital business models; see, for example, *Wall Street Journal* articles "*News Publishers See Google's AI Search Tool as a Traffic-Destroying Nightmare*" (December 14, 2023) and "*News Sites Are Getting Crushed by Google's New AI Tools*" (June 10, 2025). Analysts warn that news publishers may face a substantial decline in traffic as *Google's AI Overviews* becomes the primary interface for information consumption. This dynamic extracts informational value from the publishers who produced the content and threatens the viability of traditional referral-based business models.

The Risk of Degrading Information Ecosystem Generative AI may also endanger the information ecosystem, as illustrated by the rise of so-called "pink slime" sites—fully automated outlets that churn out low-quality content to capture programmatic advertising revenue; see the *Reuters Institute* article "*AI-generated slop is quietly conquering the internet. Is it a threat to journalism or a problem that will fix itself?*" (November 26, 2024). The AI-driven commoditization of general news has pushed traditional creators to pivot toward higher-value, human-produced content, such as original investigations and on-the-ground reporting, where humans retain a comparative advantage. In response to the generative AI shock, mainstream reporting identifies a strategic shift toward personality-led journalism and trusted human branding. As the internet is increasingly flooded with synthetic content, human-created materials, which are witnessed, verified, and contextualized by humans, may retain enduring value.

The Reuters Institute recently published an article, "*How will AI reshape the news in 2026? Forecasts by 17 experts from around the world*", which identified five recurring themes in expert forecasts:

1. Audiences will increasingly access news through AI
2. There will be increased demand for verification work
3. Automation and agents will reshape newsrooms
4. Newsrooms will upskill and build AI infrastructure
5. AI will further empower data journalists

Audience responses also indicated that "*AI will take on a role in commissioning and setting the news agenda, changes in business models to cope with a collapse in direct traffic to news websites, and a shift in trust and authority from brands to individuals.*" Together, these points effectively capture the broader context and implications of the generative AI shock in news media.

As the economic threats to content creation became manifest, legislative interventions offer potential safeguards for human creators (such as journalists and artists), while curating technological innovation and progress. Two notable examples of such legislative responses are outlined below:[33]

The EU AI Act: The *European Union's Artificial Intelligence Act*, which entered into force on August 1, 2024 is the first comprehensive law that regulates artificial intelligence. It aims to ensure that AI systems are safe, trustworthy, respectful of fundamental rights, and human-centric, while also fostering innovation across the EU. In particular, it requires *General-Purpose AI Models* meet specific transparency obligations and governance rules (e.g., requiring developers to disclose detailed summaries of the copyrighted data used for training), especially when these models carry systemic risks. This framework also provides avenues for human creators to seek remuneration through collective bargaining or litigation.

[33]Despite these legislative efforts, the regulatory lag remains a significant concern, given that AI models evolve faster than the judicial and legislative frameworks designed to contain them.

The U.S. NO FAKES Act: The *NO FAKES Act* (short for *Nurture Originals, Foster Art, and Keep Entertainment Safe Act*) is proposed federal legislation in the United States aimed at addressing harms from AI-generated "deepfakes" and other unauthorized digital replicas. As part of broader congressional efforts to modernize legal frameworks for generative AI, the Act would establish a federal legal right granting individuals (and their designated rights holders) exclusive control over the use of their voice, name, and visual likeness in digital replicas, including those produced by AI without consent. The legislation is intended to prevent deceptive or exploitative AI deepfakes that misappropriate someone's identity.

In summary, the concern that generative AI may replace human content creators is not merely an academic debate but an imminent reality. Generative AI is likely to disrupt the traditional *factory* model of content creation; still, it calls for a *human-in-the-loop* approach, in which AI handles scalable, repetitive tasks while humans provide the judgment, ethical oversight, and lived experience that LLMs inherently lack. This shift emphasizes a transition from being solely content producers to acting as information curators and investigators. Accordingly, human creators who survive the generative AI shocks will be those who pivot successfully from competing with anonymous AI on routine tasks to outperforming it on depth, trust, and originality.

C Proof of Proposition 5

We first prove by contradiction that $R > r(\lambda)$ in equilibrium: Suppose otherwise so that $R \leq r(\lambda)$, then $\theta R - C < \theta r(\lambda) - c$ for all θ. Therefore, consuming content created by human is strictly dominated by consuming content created by generative AI for all consumers, so there should be no human content production (i.e., $\lambda = 0$). But this leads to $r(\lambda) = 0 < R$, leading to a contradiction.

Now with $R > r(\lambda)$, we further consider two possibilities:

1. $\theta_{H0} \leq \theta_{A0}(\lambda)$: In this case, we further separate $\theta_{H0} = \theta_{A0}(\lambda)$ and $\theta_{H0} > \theta_{A0}(\lambda)$. When $\theta_{H0} = \theta_{A0}(\lambda)$, we have $\theta_{HA} = \theta_{H0} = \theta_{A0}(\lambda) > 0$, so consumers with θ above or equal to $\theta_{HA} = \theta_{H0} = \theta_{A0}(\lambda) > 0$ choose human, and otherwise opt out.

   When $\theta_{H0} < \theta_{A0}(\lambda)$, we have $0 < \theta_{HA} < \theta_{H0} < \theta_{A0}(\lambda)$, so consumers with $\theta \in [0, \theta_{H0})$ choose to opt out, and consumers with $\theta \in [\theta_{H0}, +\infty)$ chooses human.

   Therefore, overall we have that in the case of $\theta_{H0} \leq \theta_{A0}(\lambda)$,

   - consumers with $\theta \in [0, \theta_{H0})$ choose to opt out,
   - consumers with $\theta \in [\theta_{H0}, +\infty)$ chooses human.

2. $\theta_{H0} > \theta_{A0}(\lambda)$: In this case $0 < \theta_{A0}(\lambda) < \theta_{H0} < \theta_{HA}$, hence

   - consumers with $\theta \in [0, \theta_{A0}(\lambda))$ choose to opt out,
   - consumers with $\theta \in [\theta_{A0}(\lambda), \theta_{HA}(\lambda))$ choose AI, and
   - consumers with $\theta \in [\theta_{HA}(\lambda), +\infty)$ chooses human.

We proceed to solve the equilibrium by separately studying case 1 and case 2.

In case 1, $\theta_{H0} \leq \theta_{A0}(\lambda)$ requires $\frac{C}{R} \leq \frac{c}{r(\lambda)}$, and $\lambda = 1 - F(\theta_{H0})$ gives

$$\lambda = 1 - F\left(\frac{C}{R}\right). \tag{26}$$

Hence, $\lambda^* = 1 - F\left(\frac{C}{R}\right)$ satisfies $\frac{C}{R} \leq \frac{c}{r(\lambda)}$ if and only if $\frac{c}{C} R \geq r\left(1 - F\left(\frac{C}{R}\right)\right)$.

In case 2, $\theta_{H0} > \theta_{A0}(\lambda)$ requires $\frac{C}{R} > \frac{c}{r(\lambda)}$, and $\lambda = 1 - F(\theta_{HA}(\lambda))$ gives

$$\lambda = 1 - F\left(\frac{C - c}{R - r(\lambda)}\right). \tag{27}$$

Since the left-hand side increases in λ while the right-hand side decreases in λ, and at $\lambda = 1$ the left-hand side is larger than the right-hand side as $1 > 1 - F\left(\frac{C-c}{R-r(1)}\right)$, there exists a

unique λ^* that satisfies (8) if and only if the left-hand side of (8) is smaller than its right-hand side at $r^{-1}\left(\frac{cR}{C}\right)$, that is, $r^{-1}\left(\frac{cR}{C}\right) < 1 - F\left(\frac{C}{R}\right)$.

D Proof of Corollary 7

We study the comparative statics one by one.

Comparative statics for c. Applying the implicit function theorem to (8), we have that

$$d\lambda = -f\left(\frac{C-c}{R-r(\lambda)}\right)\left(\frac{C-c}{(R-r(\lambda))^2}r'(\lambda)d\lambda - \frac{1}{R-r(\lambda)}dc\right)$$

$$\Rightarrow \left(\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}\right)d\lambda = dc.$$

Hence,

$$\begin{aligned}
\frac{d}{dc}(9) &= \frac{\partial}{\partial c}(9) + \frac{\partial}{\partial \lambda}(9) \times \frac{d\lambda}{dc} && (28)\\
&= -f\left(\frac{C-c}{R-r(\lambda)}\right)\frac{1}{R-r(\lambda)} - f\left(\frac{c}{r(\lambda)}\right)\frac{1}{r(\lambda)} && (29)\\
&+ \frac{f\left(\frac{C-c}{R-r(\lambda)}\right)\frac{C-c}{(R-r(\lambda))^2}r'(\lambda) + f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r^2(\lambda)}r'(\lambda)}{\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} && (30)\\
&= -f\left(\frac{c}{r(\lambda)}\right)\frac{1}{r(\lambda)} + \frac{-1 + f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r^2(\lambda)}r'(\lambda)}{\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} && (31)\\
&= \frac{-1 + f\left(\frac{c}{r(\lambda)}\right)\frac{1}{r(\lambda)}\left(\frac{c}{r(\lambda)} - \frac{C-c}{R-r(\lambda)}\right)r'(\lambda) - f\left(\frac{c}{r(\lambda)}\right)\frac{1}{r(\lambda)}\frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}}{\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} < 0. && (32)
\end{aligned}$$

Comparative statics for R. Similarly, applying the implicit function theorem to (8) gives

$$d\lambda = -f\left(\frac{C-c}{R-r(\lambda)}\right)\left(\frac{C-c}{(R-r(\lambda))^2}r'(\lambda)d\lambda - \frac{C-c}{(R-r(\lambda))^2}dR\right)$$

$$\Rightarrow \left(\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)} \right) d\lambda = \frac{C-c}{R-r(\lambda)} dR.$$

Hence,

$$\begin{aligned}
\frac{d}{dR}(9) &= \frac{\partial}{\partial R}(9) + \frac{\partial}{\partial \lambda}(9) \times \frac{d\lambda}{dR} && (33)\\
&= -f\left(\frac{C-c}{R-r(\lambda)}\right) \frac{C-c}{(R-r(\lambda))^2} && (34)\\
&+ \frac{f\left(\frac{C-c}{R-r(\lambda)}\right) \frac{C-c}{(R-r(\lambda))^2} r'(\lambda) + f\left(\frac{c}{r(\lambda)}\right) \frac{c}{r^2(\lambda)} r'(\lambda)}{\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} \frac{C-c}{R-r(\lambda)} && (35)\\
&= \frac{f\left(\frac{c}{r(\lambda)}\right) \frac{c}{r^2(\lambda)} r'(\lambda) - 1}{\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} \frac{C-c}{R-r(\lambda)}, && (36)
\end{aligned}$$

which may be either positive or negative depending on the sign of $f\left(\frac{c}{r(\lambda)}\right) \frac{c}{r^2(\lambda)} r'(\lambda) - 1$.

Comparative statics for C. Similarly, applying the implicit function theorem to (8) gives

$$d\lambda = -f\left(\frac{C-c}{R-r(\lambda)}\right) \left(\frac{C-c}{(R-r(\lambda))^2} r'(\lambda) d\lambda + \frac{1}{R-r(\lambda)} dC \right)$$

$$\Rightarrow \left(\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)} \right) d\lambda = -dC.$$

Hence,

$$\begin{aligned}
\frac{d}{dC}(9) &= \frac{\partial}{\partial C}(9) + \frac{\partial}{\partial \lambda}(9) \times \frac{d\lambda}{dC} && (37)\\
&= f\left(\frac{C-c}{R-r(\lambda)}\right) \frac{1}{R-r(\lambda)} - \frac{f\left(\frac{C-c}{R-r(\lambda)}\right) \frac{C-c}{(R-r(\lambda))^2} r'(\lambda) + f\left(\frac{c}{r(\lambda)}\right) \frac{c}{r^2(\lambda)} r'(\lambda)}{\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} && (38)\\
&= \frac{1 - f\left(\frac{c}{r(\lambda)}\right) \frac{c}{r^2(\lambda)} r'(\lambda)}{\frac{C-c}{R-r(\lambda)} r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}} && (39)
\end{aligned}$$

which may be either positive or negative depending on the sign of $1 - f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r^2(\lambda)}r'(\lambda)$.

Comparative statics for $r(\cdot)$. Similarly, applying the implicit function theorem to (8) gives

$$d\lambda = -f\left(\frac{C-c}{R-r(\lambda)}\right)\left(\frac{C-c}{(R-r(\lambda))^2}r'(\lambda)d\lambda + \frac{C-c}{(R-r(\lambda))^2}dr(\lambda)\right)$$

$$\Rightarrow \left(\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}\right)d\lambda = -\frac{C-c}{R-r(\lambda)}dr(\lambda).$$

Hence,

$$\begin{aligned}
\frac{d}{dr(\lambda)}(9) &= \frac{\partial}{\partial r(\lambda)}(9) + \frac{\partial}{\partial \lambda}(9) \times \frac{d\lambda}{dr(\lambda)} && (40)\\
&= f\left(\frac{C-c}{R-r(\lambda)}\right)\frac{C-c}{(R-r(\lambda))^2} + f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r(\lambda)^2} && (41)\\
&- \frac{f\left(\frac{C-c}{R-r(\lambda)}\right)\frac{C-c}{(R-r(\lambda))^2}r'(\lambda) + f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r^2(\lambda)}r'(\lambda)}{\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}}\frac{C-c}{R-r(\lambda)} && (42)\\
&= \frac{1 + f\left(\frac{c}{r(\lambda)}\right)\frac{c}{r(\lambda)^2}\frac{R-r(\lambda)}{C-c}\frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}}{\frac{C-c}{R-r(\lambda)}r'(\lambda) + \frac{R-r(\lambda)}{f\left(\frac{C-c}{R-r(\lambda)}\right)}}\frac{C-c}{R-r(\lambda)} > 0. && (43)
\end{aligned}$$

We can also replace all $r(\cdot)$ by $r(\cdot;\xi)$ and study $\frac{d}{d\xi}\lambda$. The result would be identical.

E Proof of Proposition 8

For ease of exposition, with slight abuse of notation, we further define λ_{ij} where $i, j \in \{H, A\}$ as the population who choose i for investigation and j for writing. Then we have that $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$ and $\lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$.

Also for notation ease, we use, for example, HH to indicate that in equilibrium everyone uses human for investigation and human for writing (therefore in this case $\lambda_{HA} = \lambda_{AH} = \lambda_{AA} = 0$). We also use, for example, $HH + HA$ to indicate that in equilibrium everyone

uses human for investigation and some of them uses human for writing while others use AI for writing (therefore in this case $\lambda_{AH} = \lambda_{AA} = 0$). We note that any equilibrium must fall into one of the following 15 cases:

1. HH: In this case $\lambda_I = \lambda_W = 1$, and (10)>(11), (10)>(12), (10)>(13).

2. HA: In this case $\lambda_I = 1$, $\lambda_W = 0$, and (11)>(10), (11)>(12), (11)>(13).

3. AH: In this case $\lambda_I = 0$, $\lambda_W = 1$, and (12)>(10), (12)>(11), (12)>(13).

4. AA: In this case $\lambda_I = 0$, $\lambda_W = 0$, and (13)>(10), (13)>(11), (13)>(12).

5. $HH + HA$: In this case $\lambda_I = 1$, $\lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}}$, and (12)<(10)=(11)>(13).

6. $HH + AH$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}}$, $\lambda_W = 1$, and (11)<(10)=(12)>(13).

7. $HH + AA$: In this case $\lambda_I = \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AA}}$ and (12)<(10)=(13)>(11).

8. $HA + AH$: In this case $\lambda_I = \frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AH}}$, $\lambda_W = \frac{\lambda_{AH}}{\lambda_{HA}+\lambda_{AH}}$, and (10)<(11)=(12)>(13).

9. $HA + AA$: In this case $\lambda_I = \frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AA}}$, $\lambda_W = 0$, and (10)<(11)=(13)>(12).

10. $AH + AA$: In this case $\lambda_I = 0$, $\lambda_W = \frac{\lambda_{AH}}{\lambda_{AH}+\lambda_{AA}}$, and (10)<(12)=(13)>(11).

11. $HA + AH + AA$: In this case $\lambda_I = \frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$, $\lambda_W = \frac{\lambda_{AH}}{\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$, and (10)<(11)=(12)=(13).

12. $HH + AH + AA$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}} \leq \lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}}$, and (10)=(12)=(13)>(11).

13. $HH + HA + AA$: In this case $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}} \geq \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}}$, and (10)=(11)=(13)>(12).

14. $HH + HA + AH$: In this case $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}}$, $\lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}}$, and (10)=(11)=(12)>(13).

15. $HH + HA + AH + AA$: In this case $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$, $\lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AH}+\lambda_{AA}}$, and (10)=(11)=(12)=(13).

To proceed, we first note that in any equilibrium we must have $q_I(\lambda_I) < Q_I$, that is, generative AI cannot match human in investigation skills. If this were not true (i.e., $q_I(\lambda_I) \geq Q_I$), then generative AI can do no worse than human for investigation yet enjoy a strict cost advantage. Then everyone will switch to AI investigation leading to $\lambda_I = 0$. But since $q_I(0) < Q_I$, we get a contradiction with $q_I(\lambda_I) \geq Q_I$. With this result, we can further prove that (the proofs should be immediate):

- if (10)=(11), then (12)<(13).
- if (12)≥(13), then (10)>(11).

Therefore, case 8 (HA+AH), 11 ($HA + AH + AA$), 14 ($HH + HA + AH$), and 15 ($HH + HA + AH + AA$) are not possible. Furthermore, when $\lambda_I = 0$, we have (12)<(13)< 0 <(10). Hence, case 3 (AH), 4 (AA), and 10 (AH+AA) are also not possible. We are therefore left with 8 possible equilibria:

1. HH: In this case $\lambda_I = \lambda_W = 1$, and (10)>(11), (10)>(12), (10)>(13).
2. HA: In this case $\lambda_I = 1$, $\lambda_W = 0$, and (11)>(10), (11)>(12), (11)>(13).
3. $HH + HA$: In this case $\lambda_I = 1$, $\lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}}$, and (10)=(11)>(13), (10)>(12).
4. $HH + AH$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}}$, $\lambda_W = 1$, and (10)=(12)>(13), (10)>(11).
5. $HH + AA$: In this case $\lambda_I = \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AA}}$ and (10)=(13)>(11), (10)>(12).
6. $HA + AA$: In this case $\lambda_I = \frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AA}}$, $\lambda_W = 0$, and (10)<(11)=(13)>(12).
7. $HH + AH + AA$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}} \leq \lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}}$, and (10)=(12)=(13)>(11).

8. $HH + HA + AA$: In this case $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}} \geq \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}}$, and (10)=(11)=(13)>(12).

Expanding the conditions, simplifying, and removing dominated ones give

1. HH: In this case $\lambda_I = \lambda_W = 1$, and

$$\begin{cases} Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) > 0, \\ Q_W^{1-\alpha}(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) > 0, \\ Q_I^\alpha Q_W^{1-\alpha} - C_I - C_W > q_I^\alpha(1) q_W^{1-\alpha}(1) - c_I - c_W. \end{cases}$$

2. HA: In this case $\lambda_I = 1$, $\lambda_W = 0$, and

$$\begin{cases} Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0, \\ Q_I^\alpha q_W^{1-\alpha}(0) - C_I - c_W > q_I^\alpha(1) Q_W^{1-\alpha} - c_I - C_W, \\ q_W^{1-\alpha}(0)(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) > 0. \end{cases}$$

3. $HH + HA$: In this case $\lambda_I = 1$, $\lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}}$, and

$$\begin{cases} Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(\lambda_W)) - (C_W - c_W) = 0, \\ q_W^{1-\alpha}(\lambda_W)(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) > 0, \\ Q_W^{1-\alpha}(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) > 0. \end{cases}$$

4. $HH + AH$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}}$, $\lambda_W = 1$, and

$$\begin{cases} Q_W^{1-\alpha}(Q_I^\alpha - q_I^\alpha(\lambda_I)) - (C_I - c_I) = 0, \\ q_I^\alpha(\lambda_I)(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) > 0, \\ Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) > 0. \end{cases}$$

5. $HH + AA$: In this case $\lambda_I = \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AA}}$ and

$$\begin{cases} Q_I^{\alpha} Q_W^{1-\alpha} - C_I - C_W = q_I^{\alpha}(\lambda_I) q_W^{1-\alpha}(\lambda_W) - c_I - c_W, \\ q_W^{1-\alpha}(\lambda_W)(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I)) - (C_I - c_I) < 0, \\ Q_W^{1-\alpha}(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I)) - (C_I - c_I) > 0. \end{cases}$$

6. $HA + AA$: In this case $\lambda_I = \frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AA}}$, $\lambda_W = 0$, and

$$\begin{cases} Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0, \\ q_W^{1-\alpha}(0)(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I)) - (C_I - c_I) = 0, \\ q_I^{\alpha}(\lambda_I)(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0. \end{cases}$$

7. $HH + AH + AA$: In this case $\lambda_I = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}} \leq \lambda_W = \frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}}$, and

$$\begin{cases} Q_W^{1-\alpha}(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I)) - (C_I - c_I) = 0, \\ q_I^{\alpha}(\lambda_I)(Q_W^{1-\alpha} - q_W^{1-\alpha}(\lambda_W)) - (C_W - c_W) = 0. \end{cases}$$

8. $HH + HA + AA$: In this case $\lambda_I = \frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}} \geq \lambda_W = \frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}}$, and

$$\begin{cases} Q_W^{1-\alpha} > q_W^{1-\alpha}(\lambda_W), \\ Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(\lambda_W)) - (C_W - c_W) = 0, \\ q_W^{1-\alpha}(\lambda_W)(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I)) - (C_I - c_I) = 0. \end{cases}$$

Therefore, with further simplification and removing dominated conditions, we get the following full characterization of all equilibria.

1. If
$$\begin{cases} Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) > 0, \\ Q_W^{1-\alpha}(Q_I^{\alpha} - q_I^{\alpha}(1)) - (C_I - c_I) > 0, \\ Q_I^{\alpha}Q_W^{1-\alpha} - C_I - C_W > q_I^{\alpha}(1)q_W^{1-\alpha}(1) - c_I - c_W, \end{cases}$$
there is an equilibrium with $\lambda_I = \lambda_W = 1$. This equilibrium features HH.

2. If
$$\begin{cases} Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0, \\ Q_I^{\alpha}q_W^{1-\alpha}(0) - C_I - c_W > q_I^{\alpha}(1)Q_W^{1-\alpha} - c_I - C_W, \\ q_W^{1-\alpha}(0)(Q_I^{\alpha} - q_I^{\alpha}(1)) - (C_I - c_I) > 0, \end{cases}$$
there is an equilibrium with $\lambda_I = 1$ and $\lambda_W = 0$. This equilibrium features HA.

3. If
$$\begin{cases} Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) \geq 0, \\ Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) \leq 0, \\ \left(Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}\right)(Q_I^{\alpha} - q_I^{\alpha}(1)) - (C_I - c_I) > 0, \end{cases}$$
there is an equilibrium with $\lambda_I = 1$ and λ_W solving
$$q_W^{1-\alpha}(\lambda_W) = Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}.$$
This equilibrium features $HH + HA$ and $\frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}} = \lambda_W$.

4. If
$$\begin{cases} Q_W^{1-\alpha}(Q_I^{\alpha} - q_I^{\alpha}(1)) - (C_I - c_I) \leq 0, \\ \left(Q_I^{\alpha} - \frac{C_I - c_I}{Q_W^{1-\alpha}}\right)(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) > 0, \end{cases}$$
there is an equilibrium with λ_I solving
$$q_I^{\alpha}(\lambda_I) = Q_I^{\alpha} - \frac{C_I - c_I}{Q_W^{1-\alpha}},$$

and $\lambda_W = 1$. This equilibrium features $HH + AH$ and $\frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}} = \lambda_I$.

5. If
$$\begin{cases} Q_I^\alpha Q_W^{1-\alpha} - C_I - C_W \leq q_I^\alpha(1) q_W^{1-\alpha}(1) - c_I - c_W, \\ q_W^{1-\alpha}(\lambda^*)(Q_I^\alpha - q_I^\alpha(\lambda^*)) - (C_I - c_I) < 0, \\ Q_W^{1-\alpha}(Q_I^\alpha - q_I^\alpha(\lambda^*)) - (C_I - c_I) > 0, \end{cases}$$
there is an equilibrium with $\lambda_I = \lambda_W = \lambda^*$ where λ^* solves
$$Q_I^\alpha Q_W^{1-\alpha} - C_I - C_W = q_I^\alpha(\lambda^*) q_W^{1-\alpha}(\lambda^*) - c_I - c_W.$$
This equilibrium features $HH + AA$ and $\frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AA}} = \lambda^*$.

6. If
$$\begin{cases} Q_I^\alpha(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0, \\ q_W^{1-\alpha}(0) Q_I^\alpha - (C_I - c_I) \geq 0, \\ q_W^{1-\alpha}(0)(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) \leq 0, \\ \left(Q_I^\alpha - \frac{C_I - c_I}{q_W^{1-\alpha}(0)}\right)(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) < 0, \end{cases}$$
there is an equilibrium with λ_I solving
$$q_I^\alpha(\lambda_I) = Q_I^\alpha - \frac{C_I - c_I}{q_W^{1-\alpha}(0)},$$
and $\lambda_W = 0$. This equilibrium features $HA + AA$ and $\frac{\lambda_{HA}}{\lambda_{HA}+\lambda_{AA}} = \lambda_I$.

7. If
$$\begin{cases} Q_W^{1-\alpha}(Q_I^\alpha - q_I^\alpha(1)) - (C_I - c_I) \leq 0, \\ \left(Q_I^\alpha - \frac{C_I - c_I}{Q_W^{1-\alpha}}\right)(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) \geq 0 \\ \left(Q_I^\alpha - \frac{C_I - c_I}{Q_W^{1-\alpha}}\right)(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) \leq 0, \end{cases}$$

there is an equilibrium with λ_I and λ_W respectively solving

$$\begin{cases} q_I^{\alpha}(\lambda_I) = Q_I^{\alpha} - \frac{C_I - c_I}{Q_W^{1-\alpha}}, \\ q_W^{1-\alpha}(\lambda_W) = Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha} - \frac{C_I - c_I}{Q_W^{1-\alpha}}}, \end{cases} \quad \text{if } \lambda_I \leq \lambda_W.$$

This equilibrium features $HH + AH + AA$, $\frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}} = \lambda_I$, and $\frac{\lambda_{HH}+\lambda_{AH}}{\lambda_{HH}+\lambda_{AH}+\lambda_{AA}} = \lambda_W$.

8. If

$$\begin{cases} Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(0)) - (C_W - c_W) \geq 0, \\ Q_I^{\alpha}(Q_W^{1-\alpha} - q_W^{1-\alpha}(1)) - (C_W - c_W) \leq 0, \\ \left(Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}\right) Q_I^{\alpha} - (C_I - c_I) \geq 0, \\ \left(Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}\right) (Q_I^{\alpha} - q_I^{\alpha}(1)) - (C_I - c_I) \leq 0, \end{cases}$$

there is an equilibrium with λ_I and λ_W respectively solving

$$\begin{cases} q_W^{1-\alpha}(\lambda_W) = Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}, \\ q_I^{\alpha}(\lambda_I) = Q_I^{\alpha} - \frac{C_I - c_I}{Q_W^{1-\alpha} - \frac{C_W - c_W}{Q_I^{\alpha}}}, \end{cases} \quad \text{if } \lambda_I \geq \lambda_W.$$

This equilibrium features $HH + HA + AA$, $\frac{\lambda_{HH}+\lambda_{HA}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}} = \lambda_I$ and $\frac{\lambda_{HH}}{\lambda_{HH}+\lambda_{HA}+\lambda_{AA}} = \lambda_W$.

It is easily verified that the existence conditions for equilibria 1-4 are mutually incompatible with each other, so none of equilibria 1-4 can simultaneously exist. We can also check that the existence conditions for equilibria 1-4 are incompatible with those of equilibrium 6, 7, or 8, respectively.

From the equations characterizing each equilibrium we can also immediately get the comparative statics for λ_I and λ_W with respect to Q_I, Q_W, C_I, C_W, c_I, and c_W. The math is rather straightforward so is skipped, and the results are summarized in Corollary 10.

F Proof of Proposition 11

Fix a candidate common threshold Δ, and suppose every other agent refreshes at her own decision opportunity if and only if the current age satisfies $x \geq \Delta$. We derive the Bellman equation over a short interval $[0, dt]$ based on the objective function of each agent:

$$V(x) = \inf_{\{a_n^i\}_{n\geq 1}} \mathbb{E}_x \left[\sum_{n\geq 1} e^{-\rho \tilde{\tau}_n^i} \Big((c + \ell(\tilde{x}_{\tau_n^i}))\, \mathbf{1}\{a_n^i = \text{AI}\} + C\, \mathbf{1}\{a_n^i = \text{refresh}\} \Big) \right]. \tag{44}$$

Suppose the current age at time 0 is x. Over the interval $[0, dt]$, up to terms of order $o(dt)$, one of three events can happen: (1) no decision opportunity arrives; (2) agent i receives a decision opportunity; (3) one of the other $N-1$ agents receives a decision opportunity. Since arrivals are independent Poisson processes, we have $\Pr(\text{agent } i \text{ arrives}) = \eta\, dt + o(dt)$, and $\Pr(\text{one of the other agents arrives}) = (N-1)\eta\, dt + o(dt)$.

Continuation if no decision arrival occurs. If no one arrives during $[0, dt]$, the age drifts to $x + dt$. Therefore the continuation value is

$$V(x + dt) = V(x) + V'(x)\, dt + o(dt).$$

Agent i's arrival. If agent i receives a decision opportunity during $[0, dt]$, she compares:

- *wait*: immediate cost $c + \sigma^2 x$, continuation value $V(x)$, total $c + \sigma^2 x + V(x)$;
- *refresh*: immediate cost C, continuation value $V(0)$, total $C + V(0)$.

Hence conditional on an own arrival, the optimal continuation contribution is

$$\min\{c + \sigma^2 x + V(x),\ C + V(0)\}.$$

Other agents' arrivals. If any one of the other $N-1$ agents receives a decision opportunity during $[0, dt]$, then there are two scenarios depending on the staleness age:

if $x < \Delta$, that agent waits, so the state remains at x;

if $x \geq \Delta$, that agent refreshes, so the age of AI system resets to zero.

Hence the continuation contribution from the other agents is

$$(N-1)\eta\, dt \cdot \Big[\mathbf{1}_{\{x<\Delta\}} V(x) + \mathbf{1}_{\{x\geq\Delta\}} V(0)\Big].$$

Because future costs are discounted at rate ρ, the Bellman equation over $[0, dt]$ is

$$V(x) = e^{-\rho dt}\Bigg\{\big(1 - N\eta\, dt\big)V(x+dt) + \eta\, dt \cdot \min\{c + \sigma^2 x + V(x),\ C + V(0)\} \\ + (N-1)\eta\, dt \cdot \Big[\mathbf{1}_{\{x<\Delta\}} V(x) + \mathbf{1}_{\{x\geq\Delta\}} V(0)\Big]\Bigg\} + o(dt). \tag{45}$$

Using $e^{-\rho dt} = 1 - \rho dt + o(dt)$, expanding $V(x+dt)$, canceling $V(x)$, dividing by dt, and letting $dt \to 0$, we obtain

$$\rho V(x) = V'(x) + \eta \min\{c + \sigma^2 x,\ C + V(0) - V(x)\} - (N-1)\eta\, \mathbf{1}_{\{x\geq\Delta\}}\big(V(x) - V(0)\big). \tag{46}$$

At one of agent i's own decision opportunities, refreshing is optimal if and only if

$$C + V(0) \leq c + \sigma^2 x + V(x).$$

Thus a threshold strategy with cutoff Δ is a best response if and only if

$$V(x) - V(0) \leq C - c - \sigma^2 x \quad \text{for } x < \Delta, \qquad V(x) - V(0) \geq C - c - \sigma^2 x \quad \text{for } x \geq \Delta, \tag{47}$$

with the indifference condition:

$$V(\Delta) - V(0) = C - c - \sigma^2 \Delta. \tag{48}$$

Below the threshold. For $x < \Delta$, waiting is optimal and the other agents also wait. Hence (46) reduces to

$$V'(x) = \rho V(x) - \eta(c + \sigma^2 x). \tag{49}$$

The solution below the threshold Δ is

$$V(x) = Ae^{\rho x} + \frac{\eta}{\rho}(c + \sigma^2 x), \tag{50}$$

where the coefficient A is to be determined.

Above the threshold. For $x \geq \Delta$, refreshing is optimal at own arrivals and the other agents also refresh at theirs. Hence (46) becomes

$$V'(x) = (\rho + N\eta)V(x) - \eta C - N\eta\, V(0). \tag{51}$$

Imposing the transversality condition, the unique bounded solution is constant:

$$V(x) = \frac{\eta C + N\eta\, V(0)}{\rho + N\eta}, \qquad x \geq \Delta. \tag{52}$$

At the optimal threshold Δ^*, indifference condition requires $V(\Delta^*) - V(0) = C - c - \sigma^2 \Delta^*$. Combining it with (52) at Δ^*, we have $V(\Delta^*) = \frac{\eta C + N\eta\, V(0)}{\rho + N\eta} = V(0) + C - c - \sigma^2 \Delta^*$. Thus,

$$V(0) = \frac{(\rho + N\eta)(c + \sigma^2 \Delta^* - C) + \eta C}{\rho}. \tag{53}$$

Since equation (50) at $x = 0$ implies $V(0) = A + \frac{\eta}{\rho} + \frac{\eta\sigma^2}{\rho^2}$, we get

$$A(\Delta^*) = V(0) - \frac{\eta c}{\rho} - \frac{\eta\sigma^2}{\rho^2} = \frac{(\rho + (N-1)\eta)(c + \sigma^2\Delta^* - C) + \eta\sigma^2\Delta^*}{\rho} - \frac{\eta\sigma^2}{\rho^2}. \tag{54}$$

Substituting (53) into (50) and imposing (48) yields the optimal threshold equation

$$C - c - \sigma^2\Delta^* = \left(V(0) - \frac{\eta c}{\rho}\right)\left(e^{\rho\Delta^*} - 1\right) - \frac{\eta\sigma^2}{\rho^2}\left(e^{\rho\Delta^*} - (1 + \rho\Delta^*)\right). \tag{55}$$

Equation (55) implicitly determines the symmetric equilibrium threshold Δ^*.

Based on the above results, the equilibrium value function is

$$V(x) = \begin{cases} V(0)e^{\rho x} - \frac{\eta c}{\rho}\left(e^{\rho x} - 1\right) - \frac{\eta\sigma^2}{\rho^2}\left(e^{\rho x} - (1 + \rho x)\right), & x < \Delta^*, \\ \frac{\eta C + N\eta V(0)}{\rho + N\eta}, & x \geq \Delta^*, \end{cases} \tag{56}$$

where

$$V(0) = \frac{(\rho + N\eta)(c + \sigma^2\Delta^* - C) + \eta C}{\rho}. \tag{57}$$

The candidate threshold Δ is a true equilibrium threshold if and only if

$$V(x) - V(0) - \left(C - c - \sigma^2 x\right) \leq 0 \qquad \text{for all } x \in [0, \Delta]. \tag{58}$$

At $x = 0$, this inequality strictly holds since $C > c$; At $x = \Delta$, equality holds in (58) which becomes exactly the indifference condition (48). Therefore, as long as the derivative of the left-hand side of (58) is strictly positive, then the condition (58) holds for all $x \in [0, \Delta]$. By the expression of $V(x)$ in (56), the existence condition of this symmetric equilibrium is

$$\left(\rho V(0) - \eta c\right)e^{\rho x} - \frac{\eta\sigma^2}{\rho}\left(e^{\rho x} - 1\right) + \sigma^2 \geq 0. \tag{59}$$

G Proof of Proposition 12

Under our specification,

$$\ell(x) = \sigma^2 x,$$

and the measure of AI quality

$$Q_{\mathrm{AI}} = \left(\sigma^2 \cdot \lim_{T\to\infty} \frac{1}{T} \int_0^T x_t \, dt \right)^{-1}. \tag{60}$$

we use results in the zero-discount limit to compute Q_{AI}. Under a stationary threshold policy Δ^*, the age process is regenerative. A cycle begins immediately after refresh, the age rises linearly from 0 to Δ^*, and then the system waits for the first arrival after threshold crossing, which occurs after an independent exponential time, $\tilde{\tau} \sim \mathrm{Exp}(N\eta)$. Thus the cycle length is

$$T_{cyc} = \Delta^* + \tilde{\tau}, \qquad \mathrm{E}[T_{cyc}] = \Delta^* + \frac{1}{N\eta}.$$

During the cycle, the age path is simply $x_t = t$ for $t \in [0, \Delta^* + \tilde{\tau}]$. Therefore the total error accumulated over one cycle is

$$\int_0^{\Delta^*+\tilde{\tau}} \ell(x_t)\, dt = \int_0^{\Delta^*+\tilde{\tau}} \sigma^2 t \, dt = \sigma^2 \frac{(\Delta^* + \tilde{\tau})^2}{2}.$$

Taking expectations and using $\mathbb{E}[\tilde{\tau}] = \frac{1}{N\eta}$ and $\mathbb{E}[\tilde{\tau}^2] = \frac{2}{(N\eta)^2}$, we obtain

$$\mathbb{E}\left[\int_0^{\Delta^*+\tilde{\tau}} \ell(x_t)\, dt \right] = \sigma^2 \, \mathbb{E}\left[\frac{(\Delta^* + \tilde{\tau})^2}{2} \right] = \sigma^2 \left(\frac{(\Delta^*)^2}{2} + \frac{\Delta^*}{N\eta} + \frac{1}{(N\eta)^2} \right). \tag{61}$$

By the renewal-reward theorem, the long-run time-averaged error is therefore

$$\frac{\mathbb{E}\left[\int_0^{\Delta^*+\tilde{\tau}} \ell(x_t)\, dt \right]}{\mathbb{E}[\Delta^* + \tilde{\tau}]} = \frac{\sigma^2 \left(\frac{(\Delta^*)^2}{2} + \frac{\Delta^*}{N\eta} + \frac{1}{(N\eta)^2} \right)}{\Delta^* + \frac{1}{N\eta}}. \tag{62}$$

Hence the corresponding AI quality measure is

$$Q_{\mathrm{AI}}(\Delta^*) = \frac{\Delta^* + \frac{1}{N\eta}}{\sigma^2\left(\frac{(\Delta^*)^2}{2} + \frac{\Delta^*}{N\eta} + \frac{1}{(N\eta)^2}\right)}. \tag{63}$$

The share of human refresh in the threshold equilibrium is

$$\Lambda_{\mathrm{human}} = \frac{\frac{1}{N\eta}}{\Delta^* + \frac{1}{N\eta}}, \tag{64}$$

which allows us to simplify the expected loss over a cycle:

$$\begin{aligned}\frac{(\Delta^*)^2}{2} + \frac{\Delta^*}{N\eta} + \frac{1}{(N\eta)^2} &= \frac{1}{2}\left(\frac{1}{N\eta\Lambda_{\mathrm{human}}} - \frac{1}{N\eta}\right)^2 + \frac{1}{N\eta}\left(\frac{1}{N\eta\Lambda_{\mathrm{human}}} - \frac{1}{N\eta}\right) + \frac{1}{(N\eta)^2} \\ &= \frac{1}{2(N\eta\Lambda_{\mathrm{human}})^2} + \frac{1}{2(N\eta)^2}.\end{aligned} \tag{65}$$

Therefore, we can rewrite the AI quality measure as

$$Q_{\mathrm{AI}} = \frac{(N\eta\Lambda_{\mathrm{human}})^{-1}}{\frac{\sigma^2}{2(N\eta)^2}\left(\frac{1}{\Lambda_{\mathrm{human}}^2} + 1\right)} = \frac{2N\eta\ \Lambda_{\mathrm{human}}}{\sigma^2(1 + \Lambda_{\mathrm{human}}^2)}, \tag{66}$$

which is an increasing function of $\Lambda_{\mathrm{human}} \in (0, 1]$ and reaches maximum at $\Lambda_{\mathrm{human}} = 1$. Thus, this expression provides a microscopic view about the AI quality: a higher human fraction is associated with a higher aggregate refresh rate and therefore with higher AI quality.

H Proof of Proposition 13

Fix a threshold $\Delta \geq 0$, before the age reaches Δ, each of the N agents receives decision opportunities at rate η, and each such decision uses AI and incurs per-agent expected loss $c + \sigma^2 x$ when the current age is x. Once the age reaches Δ, the next arrival triggers refresh immediately. Therefore, if the system starts from a fresh state, the planner's *per-agent*

discounted cost satisfies the renewal equation:

$$J_P(\Delta) = \int_0^{\Delta} e^{-\rho t}\eta(c+\sigma^2 t)\,dt + \mathbb{E}\big[e^{-\rho(\Delta+\tilde{\tau})}\big]\left(\frac{C}{N} + J_P(\Delta)\right), \tag{67}$$

where $\tilde{\tau} \sim \text{Exp}(N\eta)$ is the waiting time from threshold crossing to the next arrival.

Since

$$\mathbb{E}\big[e^{-\rho(\Delta+(\tilde{\tau})}\big] = e^{-\rho\Delta}\frac{N\eta}{N\eta+\rho},$$

equation (67) becomes

$$J_P(\Delta) = \int_0^{\Delta} e^{-\rho t}\eta(c+\sigma^2 t)\,dt + e^{-\rho\Delta}\frac{N\eta}{N\eta+\rho}\left(\frac{C}{N} + J_P(\Delta)\right). \tag{68}$$

Evaluating the integral gives

$$J_P(\Delta) = \frac{\frac{\eta c}{\rho}\big(1-e^{-\rho\Delta}\big) + \frac{\eta\sigma^2}{\rho^2}\big(1-e^{-\rho\Delta}(1+\rho\Delta)\big) + e^{-\rho\Delta}\frac{\eta C}{N\eta+\rho}}{1-e^{-\rho\Delta}\frac{N\eta}{N\eta+\rho}}. \tag{69}$$

Differentiate (68) with respect to Δ. It is convenient to differentiate the renewal equation directly. At an interior optimum,

$$0 = \eta e^{-\rho\Delta}(c+\sigma^2\Delta) - \rho e^{-\rho\Delta}\frac{N\eta}{N\eta+\rho}\left(\frac{C}{N} + J_P(\Delta)\right).$$

Using (68) to eliminate $J_P(\Delta)$, and simplifying, yields (20). The right-hand side of (20) is strictly increasing in Δ, since

$$\frac{\partial}{\partial\Delta}\left\{\frac{\sigma^2}{\rho^2}\left[(N\eta+\rho)\rho\Delta - N\eta\big(1-e^{-\rho\Delta}\big)\right]\right\} = \frac{\sigma^2}{\rho}\left[N\eta+\rho-N\eta e^{-\rho\Delta}\right] > 0.$$

Hence the optimizer is unique. To obtain the closed form, rearrange (20) as

$$e^{-\rho\Delta} = 1 + \frac{(C-c)\rho^2}{N\eta\,\sigma^2} - \frac{N\eta+\rho}{N\eta}\rho\Delta.$$

This is a standard Lambert-W equation, and solving it gives the closed-form expression (21).

Comparative Statics of Δ_P Define

$$F(\Delta; C, c, N, \eta, \sigma, \rho) \equiv \frac{\sigma^2}{\rho^2}\left[(N\eta+\rho)\rho\Delta - N\eta\big(1-e^{-\rho\Delta}\big)\right] - (C-c).$$

Then $F(\Delta_P;\cdot) = 0$, and

$$F_\Delta(\Delta_P;\cdot) = \frac{\sigma^2}{\rho}\left[N\eta+\rho-N\eta e^{-\rho\Delta_P}\right] > 0. \tag{70}$$

Hence the implicit-function theorem applies.

For C and c,

$$F_C = -1, \qquad F_c = 1,$$

so

$$\frac{\partial\Delta_P}{\partial C} = -\frac{F_C}{F_\Delta} > 0, \qquad \frac{\partial\Delta_P}{\partial c} = -\frac{F_c}{F_\Delta} < 0.$$

For N,

$$F_N = \frac{\eta\sigma^2}{\rho^2}\left[\rho\Delta_P - \left(1-e^{-\rho\Delta_P}\right)\right] > 0,$$

because $1-e^{-x} < x$ for all $x > 0$. Therefore

$$\frac{\partial\Delta_P}{\partial N} < 0.$$

Similarly,

$$F_\eta = \frac{N\sigma^2}{\rho^2}\left[\rho\Delta_P - \left(1 - e^{-\rho\Delta_P}\right)\right] > 0,$$

hence

$$\frac{\partial \Delta_P}{\partial \eta} < 0.$$

For σ,

$$F_\sigma = \frac{2\sigma}{\rho^2}\left[(N\eta + \rho)\rho\Delta_P - N\eta\left(1 - e^{-\rho\Delta_P}\right)\right].$$

Using (20), this becomes

$$F_\sigma = \frac{2(C-c)}{\sigma} > 0,$$

so

$$\frac{\partial \Delta_P}{\partial \sigma} < 0.$$

Finally, for ρ, differentiating F and substituting (20) yields

$$F_\rho = -\frac{N\eta\sigma^2}{\rho^3}\Big[\rho\Delta_P e^{\rho\Delta_P} + \rho\Delta_P - 2e^{\rho\Delta_P} + 2\Big]e^{-\rho\Delta_P}.$$

Let $H(x) \equiv xe^x + x - 2e^x + 2$. Obviously, $H(0) = H'(0) = 0$, while $H''(x) = xe^x > 0$ for $x > 0$. Hence $H(x) > 0$ for every $x > 0$. Therefore $F_\rho < 0$, and so

$$\frac{\partial \Delta_P}{\partial \rho} = -\frac{F_\rho}{F_\Delta} > 0.$$

Comparing Δ_P with Δ^*: Next, we prove that $\Delta^* \geq \Delta_P$ with equality holds at $N = 1$.

When $N = 1$, the equilibrium threshold equation (16) becomes

$$C - c - \sigma^2\Delta = \frac{\rho(c-C) + (\rho+\eta)\sigma^2\Delta}{\rho}\left(e^{\rho\Delta} - 1\right) - \frac{\eta\sigma^2}{\rho^2}\left(e^{\rho\Delta} - 1 - \rho\Delta\right). \tag{71}$$

which can be simplified to the following

$$C - c = \frac{\sigma^2}{\rho^2}\left[(\eta+\rho)\rho\Delta - \eta\left(1-e^{-\rho\Delta}\right)\right],$$

exactly the planner's threshold equation (20) at $N = 1$. Therefore, the individual's equilibrium threshold and the planner's optimal threshold coincide.

When $N > 1$, the social planner's threshold Δ_P solves (20). Define

$$D_P(\Delta) \equiv \frac{\sigma^2}{\rho^2}\left[(N\eta+\rho)\rho\Delta - N\eta\left(1-e^{-\rho\Delta}\right)\right]. \tag{72}$$

Then Δ_P is characterized by

$$D_P(\Delta_P) = C - c.$$

The decentralized equilibrium threshold Δ^* solves (16), which can be written as

$$\sigma^2\Delta + \frac{\left[(\rho+N\eta)\sigma^2\Delta - (\rho+(N-1)\eta)D\right]\left(e^{\rho\Delta}-1\right)}{\rho} - \frac{\eta\sigma^2}{\rho^2}\left(e^{\rho\Delta}-1-\rho\Delta\right) = C - c. \tag{73}$$

Define the left hand side of (73) as $D_{EQ}(\Delta)$ which can be simplified as

$$D_{EQ}(\Delta) \equiv \sigma^2\frac{\rho\Delta + \eta\Delta\left[N-(N-1)e^{-\rho\Delta}\right] - \frac{\eta}{\rho}\left(1-e^{-\rho\Delta}\right)}{\rho+(N-1)\eta\left(1-e^{-\rho\Delta}\right)}. \tag{74}$$

Hence Δ^* is characterized by

$$D_{EQ}(\Delta^*) = C - c.$$

We proceed in two steps:

Step 1: $D_P(\Delta)$ is strictly increasing in Δ.

Differentiating (72) gives

$$D'_P(\Delta) = \frac{\sigma^2}{\rho}\Big[(N\eta + \rho) - N\eta e^{-\rho\Delta}\Big]. \tag{75}$$

Since $e^{-\rho\Delta} \le 1$, we have

$$(N\eta + \rho) - N\eta e^{-\rho\Delta} \ge \rho > 0.$$

Therefore

$$D'_P(\Delta) > 0 \qquad \text{for all } \Delta \ge 0. \tag{76}$$

Step 2: for every $\Delta > 0$, we have $D_P(\Delta) > D_{EQ}(\Delta)$.

Subtracting (74) from (72) and simplifying yields

$$D_P(\Delta) - D_{EQ}(\Delta) = \frac{\eta\sigma^2(N-1)e^{-\rho\Delta}\big(\rho\Delta e^{\rho\Delta} - e^{\rho\Delta} + 1\big)\big(N\eta e^{\rho\Delta} - N\eta + \rho e^{\rho\Delta}\big)}{\rho^2\Big((N-1)\eta(e^{\rho\Delta} - 1) + \rho e^{\rho\Delta}\Big)}. \tag{77}$$

We now check the sign of each factor. Because $N > 1$, $\eta > 0$, $\sigma^2 > 0$, $\rho > 0$, and $e^{-\rho\Delta} > 0$, all prefactors are positive. Also,

$$N\eta e^{\rho\Delta} - N\eta + \rho e^{\rho\Delta} = N\eta\big(e^{\rho\Delta} - 1\big) + \rho e^{\rho\Delta} > 0,$$

and

$$(N-1)\eta(e^{\rho\Delta} - 1) + \rho e^{\rho\Delta} > 0.$$

Let $H(z) \equiv ze^z - e^z + 1 = e^z(z-1) + 1$ with $z = \rho\Delta > 0$. Then $H(0) = 0$ and $H'(z) = ze^z > 0$ for $z > 0$. Hence $H(z) > 0$ for every $z > 0$, that is,

$$\rho\Delta e^{\rho\Delta} - e^{\rho\Delta} + 1 > 0 \qquad \text{for every } \Delta > 0.$$

Thus every factor on the right-hand side of (77) is strictly positive, and therefore

$$D_P(\Delta) > D_{EQ}(\Delta) \qquad \text{for every } \Delta > 0. \tag{78}$$

Since $D_{EQ}(\Delta^*) = C - c$. By (78),

$$D_P(\Delta^*) > D_{EQ}(\Delta^*) = C - c.$$

Since $D_P(\Delta)$ is strictly increasing in Δ by (76), and the planner's threshold Δ_P is defined by $D_P(\Delta_P) = C - c$, it follows that

$$\Delta_P < \Delta^*.$$

I Proof of Corollary 14

Note that the optimal threshold equation depends only on the cost gap $C - c$. Therefore, the comparative static of Δ^* with c is exactly opposite to that with C, and we can treat C as the effective cost and absorb c to save notation in the following proof.

For a given discount rate $\rho > 0$, the equilibrium threshold Δ^* is characterized by

$$F(\Delta^*; C, N, \eta, \sigma, \rho) = 0, \tag{79}$$

where

$$F \equiv \frac{\Big[(\rho + N\eta)\sigma^2\Delta - (\rho + (N-1)\eta)C\Big]\left(e^{\rho\Delta} - 1\right)}{\rho} - \frac{\eta\sigma^2}{\rho^2}\Big(e^{\rho\Delta} - 1 - \rho\Delta\Big) + \sigma^2\Delta - C. \tag{80}$$

Suppose the threshold Δ^* is indeed the best response. Then it follows

$$\left.\frac{\partial F(\Delta; C, N, \eta, \sigma, \rho)}{\partial \Delta}\right|_{\Delta=\Delta^*} > 0, \tag{81}$$

and the implicit function theorem applies. At the threshold, indifference implies

$$V(\Delta^*) - V(0) = C - \sigma^2 \Delta^*.$$

Since the stale state is weakly worse than the refreshed state, one has

$$C - \sigma^2 \Delta^* \geq 0. \tag{82}$$

This quantity will repeatedly appear in the comparative statics below.

Comparative statics of Δ^*: First, differentiating (80) with respect to C gives

$$F_C(\Delta^*; \cdot) = -1 - \frac{\rho + (N-1)\eta}{\rho}\left(e^{\rho \Delta^*} - 1\right) < 0. \tag{83}$$

Hence, given (81), we have

$$\frac{\partial \Delta^*}{\partial C} = -\frac{F_C}{F_\Delta} > 0. \tag{84}$$

Equivalently,

$$\frac{\partial \Delta^*}{\partial C} > 0, \qquad \text{which implies} \qquad \frac{\partial \Delta^*}{\partial c} < 0. \tag{85}$$

Second, differentiating (80) with respect to N yields

$$F_N(\Delta^*; \cdot) = -\frac{\eta}{\rho}\left(e^{\rho \Delta^*} - 1\right)\left(C - \sigma^2 \Delta^*\right) \leq 0. \tag{86}$$

Therefore, by the implicit function theorem,

$$\frac{\partial \Delta^*}{\partial N} = -\frac{F_N}{F_\Delta} \geq 0, \tag{87}$$

with strict inequality whenever $C - \sigma^2 \Delta^* > 0$ holds.

Third, differentiating (80) with respect to η and using the equilibrium threshold equation $F(\Delta^*; \cdot) = 0$ gives the identity

$$F_\eta(\Delta^*; \cdot) = e^{\rho \Delta^*} \left(C - \sigma^2 \Delta^*\right) \geq 0. \tag{88}$$

Hence,

$$\frac{\partial \Delta^*}{\partial \eta} \leq 0, \tag{89}$$

again strict inequality holds if and only if $C - \sigma^2 \Delta^* > 0$ holds.

Fourth, differentiating (80) with respect to σ gives

$$F_\sigma(\Delta^*; \cdot) = \frac{2\sigma}{\rho^2} \Big[N\eta\rho\Delta^* \big(e^{\rho\Delta^*} - 1\big) + \eta\rho\Delta^* + \rho^2 \Delta^* e^{\rho\Delta^*} - \eta\big(e^{\rho\Delta^*} - 1\big) \Big]. \tag{90}$$

Since $e^x - 1 < xe^x$ for every $x > 0$, the bracketed term is strictly positive, so that

$$F_\sigma(\Delta^*; \cdot) > 0 \qquad \Longrightarrow \qquad \frac{\partial \Delta^*}{\partial \sigma} < 0. \tag{91}$$

Finally, the effect of ρ is intuitive: a larger discount rate (impatience) makes the future gains from refresh less valuable relative to the current cost C, so the agent waits longer before refreshing. Thus

$$\frac{\partial \Delta^*}{\partial \rho} > 0. \tag{92}$$

The above observations can be summarized as follows.

Comparative statics of Δ^*: Under (81), the equilibrium threshold satisfies

$$\frac{\partial \Delta^*}{\partial C} > 0, \qquad \frac{\partial \Delta^*}{\partial c} < 0, \qquad \frac{\partial \Delta^*}{\partial N} \geq 0, \qquad \frac{\partial \Delta^*}{\partial \eta} \leq 0, \qquad \frac{\partial \Delta^*}{\partial \sigma} < 0, \qquad \frac{\partial \Delta^*}{\partial \rho} > 0.$$

The inequalities with respect to N and η are strict whenever $C - \sigma^2 \Delta^* > 0$ holds.

Thus, a higher refresh cost C, a larger human group N, or a higher discount rate ρ raises the equilibrium threshold and therefore delays human refresh. By contrast, a higher AI-use cost c, a higher arrival rate η, or a more volatile information environment σ lowers the threshold and induces earlier human intervention.

Comparative statics of the human decision share Given the equilibrium threshold, define the aggregate human refresh rate by

$$\Psi_{\text{human}} \equiv \frac{1}{\Delta^* + \frac{1}{N\eta}}, \tag{93}$$

then the human fraction of activities is

$$\Lambda_{\text{human}} \equiv \frac{\Psi_{\text{human}}}{N\eta} = \frac{1}{1 + N\eta\Delta^*}. \tag{94}$$

From (93), any parameter that raises Δ^* lowers Ψ_{human}, and any parameter that lowers Δ^* raises Ψ_{human}. In addition, a rise in N or η shortens the expected waiting time $1/(N\eta)$, which tends to increase Ψ_{human}. Therefore,

$$\frac{\partial \Psi_{\text{human}}}{\partial C} < 0, \qquad \frac{\partial \Psi_{\text{human}}}{\partial c} > 0, \qquad \frac{\partial \Psi_{\text{human}}}{\partial \eta} > 0, \qquad \frac{\partial \Psi_{\text{human}}}{\partial \sigma} > 0, \qquad \frac{\partial \Psi_{\text{human}}}{\partial \rho} < 0. \tag{95}$$

The sign of $\partial \Psi_{\text{human}}/\partial N$ is generally ambiguous, because a larger human group not only increases free-riding (raising Δ_{EQ}) but also reduces the post-threshold waiting time $1/(N\eta)$.

The comparative statics of Λ_{human} with respect to C, c, σ, and ρ follow directly from (94). In addition, because N enters multiplicatively in the denominator, one has

$$\frac{\partial \Lambda_{\text{human}}}{\partial N} < 0. \tag{96}$$

Next, we will prove $\Lambda_{\text{human}} > 0$ decreases in η. The equilibrium threshold equation is

$$\eta\sigma^2 \left(N - \frac{1}{2}\right) (\Delta^*)^2 + \left[\sigma^2 - C\eta(N-1)\right]\Delta^* - C = 0. \tag{97}$$

Define

$$y \equiv \eta\Delta^*. \tag{98}$$

Substituting $d^* = y/\eta$ into (97) and multiplying through by η yields

$$\sigma^2 \left(N - \frac{1}{2}\right) y^2 + \sigma^2 y - C\eta\left[(N-1)y + 1\right] = 0. \tag{99}$$

Hence

$$\eta = \frac{\sigma^2 \left[\left(N - \frac{1}{2}\right) y^2 + y\right]}{C\left[(N-1)y + 1\right]}. \tag{100}$$

Differentiate (100) with respect to y. A direct calculation gives

$$\frac{d\eta}{dy} = \frac{\sigma^2}{C} \frac{2(N-1)^2y^2 + (N-1)y^2 + 4(N-1)y + 2y + 2}{2\left[(N-1)y + 1\right]^2} > 0. \tag{101}$$

Therefore η is strictly increasing in y, and hence

$$\frac{dy}{d\eta} > 0. \tag{102}$$

Since $\Lambda_{\text{human}} = 1/(1+N\eta\Delta^*) = 1/(1+Ny)$, we have $\partial\Lambda_{\text{human}}/\partial\eta < 0$. In summary,

$$\frac{\partial\Lambda_{\text{human}}}{\partial C} < 0, \quad \frac{\partial\Lambda_{\text{human}}}{\partial c} > 0, \quad \frac{\partial\Lambda_{\text{human}}}{\partial \eta} < 0, \quad \frac{\partial\Lambda_{\text{human}}}{\partial \sigma} > 0, \quad \frac{\partial\Lambda_{\text{human}}}{\partial N} < 0, \quad \frac{\partial\Psi_{\text{human}}}{\partial \rho} < 0. \tag{103}$$

Comparative statics of AI quality We use the inverse long-run time-averaged error to measure the AI quality measure:

$$Q_{\text{AI}} \equiv \frac{\Delta^* + \frac{1}{N\eta}}{\sigma^2\left(\frac{(\Delta^*)^2}{2} + \frac{\Delta^*}{N\eta} + \frac{1}{(N\eta)^2}\right)} = \frac{2\Psi_{\text{human}}}{\sigma^2\left[1+\left(\frac{\Psi_{\text{human}}}{N\eta}\right)^2\right]}. \tag{104}$$

Since Q_{AI} is decreasing in Δ^*, the comparative statics of Δ^* imply

$$\frac{\partial Q_{\text{AI}}}{\partial C} < 0, \qquad \frac{\partial Q_{\text{AI}}}{\partial c} > 0, \qquad \frac{\partial Q_{\text{AI}}}{\partial \rho} < 0.$$

Next, we prove the AI quality Q_{AI} also increases with η. Define the aggregate arrival rate

$$\lambda \equiv N\eta.$$

Given a threshold Δ, the long-run time-averaged AI loss can be written as

$$\mathcal{E}(\Delta, \lambda, \sigma) \equiv \sigma^2 \frac{\frac{\Delta^2}{2} + \frac{\Delta}{\lambda} + \frac{1}{\lambda^2}}{\Delta + \frac{1}{\lambda}}. \tag{105}$$

Accordingly,

$$Q_{\text{AI}} = \frac{1}{\mathcal{E}(\Delta^*, N\eta, \sigma)}. \tag{106}$$

Thus, for comparative statics, it is sufficient to analyze $\mathcal{E}$ because Q_{AI} moves in the opposite direction. Differentiating (105) yields

$$\frac{\partial \mathcal{E}}{\partial \Delta} = \frac{\lambda \sigma^2 \, \Delta(\lambda \Delta + 2)}{2(\lambda \Delta + 1)^2} > 0, \tag{107}$$

and

$$\frac{\partial \mathcal{E}}{\partial \lambda} = -\frac{\sigma^2(\lambda^2 \Delta^2 + 4\lambda \Delta + 2)}{2\lambda^2(\lambda \Delta + 1)^2} < 0. \tag{108}$$

Hence, holding everything else fixed, a larger threshold Δ worsens AI quality, while a higher aggregate arrival rate λ improves AI quality, because it shortens the expected post-threshold waiting time until refresh. Since $\lambda = N\eta$, the total derivative of $\mathcal{E}$ with respect to η is

$$\frac{d\mathcal{E}}{d\eta} = \frac{\partial \mathcal{E}}{\partial \Delta} \frac{\partial \Delta^*}{\partial \eta} + \frac{\partial \mathcal{E}}{\partial \lambda} N. \tag{109}$$

By (107), (108), and $\partial \Delta^* / \partial \eta < 0$, both terms on the right-hand side of (109) are strictly negative. Therefore

$$\frac{d\mathcal{E}}{d\eta} < 0, \qquad \Longrightarrow \qquad \frac{dQ_{AI}}{d\eta} > 0. \tag{110}$$

The effect of N is generally ambiguous, because a larger human group not only reduces waiting time but also encourages free-riding. Finally, the effect of σ is also mixed: a larger σ lowers the threshold and raises human refresh intensity, but it also directly makes informational errors larger. When the direct effect dominates, Q_{AI} should decrease with σ.

J Proof of Corollary 15

In the zero-discount limit ($\rho \to 0$), using the Taylor expansions

$$e^{\rho \Delta} - 1 = \rho \Delta + \frac{\rho^2 \Delta^2}{2} + o(\rho^2), \qquad e^{\rho \Delta} - (1 + \rho \Delta) = \frac{\rho^2 \Delta^2}{2} + o(\rho^2),$$

we can simplify equation (55) into

$$C - c - \sigma^2\Delta = (\rho V(0) - \eta c)\left(\Delta + \frac{\rho\Delta^2}{2}\right) - \eta\sigma^2\frac{\Delta^2}{2} + o(1).$$

As $\rho \to 0$, we have

$$\rho V(0) \to N\eta\sigma^2\Delta - (N-1)\eta(C-c) + \eta c.$$

Therefore, the limit of (55) bcomes

$$C - c - \sigma^2\Delta = \Big(N\eta\sigma^2\Delta - (N-1)\eta(C-c)\Big)\Delta - \eta\sigma^2\frac{\Delta^2}{2},$$

which rearranges to the optimal threshold equation:

$$\eta\sigma^2\left(N - \frac{1}{2}\right)\Delta^2 + \big(\sigma^2 - (N-1)\eta(C-c)\big)\Delta - (C-c) = 0. \tag{111}$$

This is exactly equation (22) and admits the unique solution (23).

For comparison, we also consider and solve the social planner's problem that aims to minimize the long-run average cost:

$$\min_{\Delta\geq 0}\mathcal{L}_P = \min_{\Delta\geq 0}\frac{\mathbb{E}\left[N\int_0^\Delta (c + \sigma^2\tilde{x}_{\tau_n})\mathbf{1}\{a_n = \text{AI}\}\,dt\right] + C}{\Delta + \frac{1}{N\eta}} = \min_{\Delta\geq 0}\frac{cN\eta\Delta + N\eta\sigma^2\frac{\Delta^2}{2} + C}{\Delta + \frac{1}{N\eta}}, \tag{112}$$

where $L_P(\Delta)$ measures the long-run average of the total cost over a cycle which has an expected length $\Delta + \frac{1}{N\eta}$ under a candidate threshold strategy Δ. Equation (24) follows immediately from computing the derivative $\frac{d\mathcal{L}_p(\Delta)}{d\Delta} = 0$ by the criterion (112).

Alternatively, one can check that the quadratic equation (24) is exactly the limit of (20) as $\rho \to 0$, and the planner's general solution (21) also reduces to the planner's zero-discount solution (25) when $\rho \to 0$.

K Heterogeneity in production and consumption

This appendix develops two extensions that respectively complements and further relaxes the homogeneity consumer taste analysis of Section 4.1. Instead of heterogeneous consumer tastes, Appendix K.1 introduces heterogeneous producer skills and shows that AI adoption follows a bottom-up pattern, with the lowest-skill producers switching to AI first. Appendix K.2 then unifies both demand- and supply-side heterogeneity in a single framework with an endogenous market price. By separating content consumers and producers, the analysis also allows us to more explicit state the labor market implications of introducing generative AI.

K.1 Heterogeneity in human production costs

We now consider heterogeneity on the production side rather than the demand side. The setup mirrors Section 4.1, except that agents are *producers* indexed by skill $\phi \in [0, +\infty)$, distributed according to an atomless CDF G over $[0, +\infty)$. A type-ϕ producer's payoff from producing human content is $R - C/\phi$,[34] so that higher ϕ means lower cost and hence higher profit from human production; his payoff from adopting AI is $r(\lambda) - c$; and his payoff from opting out is 0.

We define the pre-AI cutoff $\phi_{H0} \equiv C/R$, and the post-AI cutoff

$$\phi_{HA}(\lambda) \equiv \frac{C}{R - r(\lambda) + c}. \tag{113}$$

[34] Assuming the cost of a type-ϕ producer is C/ϕ is without loss of generality. In general, one may specify an arbitrary strictly decreasing cost function $\mathcal{C} : [0, +\infty) \to (0, +\infty)$, with a type-$\phi$ producer bearing cost $\mathcal{C}(\phi)$. Define $\hat{\phi} \equiv C/\mathcal{C}(\phi)$ for any reference constant $C > 0$; since $\mathcal{C}$ is strictly decreasing and positive, this is a strictly increasing relabeling of producers, and the cost of a type-$\hat{\phi}$ producer is exactly $C/\hat{\phi}$. All equilibrium conditions depend on $\mathcal{C}$ and G only through the composite $G(\mathcal{C}^{-1}(\cdot))$, which equals $\hat{G}(C/(\cdot))$ under the relabeling, where $\hat{G}$ is the distribution of $\hat{\phi}$ induced by G. No economic content is altered. For boundary behavior: if $\mathcal{C}$ has bounded range $[\underline{\mathcal{C}}, \bar{\mathcal{C}}] \subset (0, +\infty)$, the support of $\hat{G}$ is $[C/\bar{\mathcal{C}}, C/\underline{\mathcal{C}}]$, extended to $[0, +\infty)$ by the standard CDF convention; if $\bar{\mathcal{C}} = +\infty$ or $\underline{\mathcal{C}} = 0$, the support naturally reaches 0 or $+\infty$ respectively. This parallels the normalization $\hat{\theta} = v(\theta, q)/q$ for preferences for in Section 4.1.

Proposition 16. *With heterogeneous producer skills ϕ, if*

$$r\left(1 - G\left(\frac{C}{R}\right)\right) < c, \tag{114}$$

then the pre-AI equilibrium persists: producers with $\phi \in [0, \phi_{H0})$ opt out, while those with $\phi \in [\phi_{H0}, +\infty)$ produce human content.

If (114) is violated, there exists a unique λ^* satisfying

$$\lambda^* = 1 - G\left(\frac{C}{R - r(\lambda^*) + c}\right), \tag{115}$$

and the equilibrium has producers with $\phi \in [0, \phi_{HA}(\lambda^*))$ adopt AI and those with $\phi \in [\phi_{HA}(\lambda^*), +\infty)$ produce human content.

The key contrast with Section 4.1 is twofold. First, in the case of skill heterogeneity instead of taste heterogeneity, AI adoption is *bottom-up* rather than middle-out: it is the lowest-skill producers (for whom C/ϕ is highest, i.e., ϕ is smallest) who switch to AI, since AI is most attractive to those for whom human production is most costly. This is consistent with Cui, Demirer, Jaffe, Musolff, Peng and Salz (2025), who find higher AI adoption among more junior employees—here, junior employees map directly to the low-skill (bottom) producers. Second, there is no opt-out segment in the disruption case: since everyone can adopt AI at cost c with benefit $r(\lambda)$, every producer for whom human content is unprofitable prefers AI to opting out as long as $r(\lambda) > c$.

Similar to Section 4.1, we can derive the following comparative statics:

Corollary 17. *The relative amount of human-created content λ^* increases as generative AI cost c increases, human content value R increases, or generative AI quality $r(\cdot)$ decreases. Since all producers either use AI or produce human content, AI adoption moves in the opposite direction.*

K.2 Separating content consumption and production

We further develop an extension that separates consumers and producers to shed more direct light on the labor market impact of genarative AI. The extension also unifies the consumer taste heterogeneity of Section 4.1 with producer skill heterogeneity. The defining departure from Section 4.1 and Appendix K.1 is that we now *explicitly separate* the roles of content consumers and content producers: consumers are indexed by a taste parameter θ capturing heterogeneity in preference for content quality (as in Section 4.1), while producers are indexed by a skill parameter ϕ capturing heterogeneity in human production costs (as in Appendix K.1). This combined framework allows us to study how demand-side and supply-side heterogeneity jointly shape the human–AI coexistence equilibrium.

Setup. There is a unit mass of *consumers*, indexed by type $\theta \in [0, +\infty)$, distributed according to an atomless CDF F on $[0, +\infty)$. There is also a unit mass of *producers*, indexed by type $\phi \in [0, +\infty)$, distributed according to an atomless CDF G on $[0, +\infty)$. Higher ϕ denotes higher skill: the cost of producing one unit of human content for a type-ϕ producer is C/ϕ, which is strictly decreasing in ϕ (higher skill implies lower cost).

Human-generated content has intrinsic quality $R > 0$. AI-generated content is available to consumers at cost $c > 0$ and has quality $r(\lambda)$, where λ denotes the equilibrium mass of human-generated content. We maintain $r'(\cdot) \geq 0$ reflecting the scaling law.

The content market is competitive with an endogenous (labor market clearing) price $P > 0$ for human content. Each producer chooses to either supply one unit of human content (earning price P and incurring cost C/ϕ, yielding profit $P - C/\phi$) or opt out (earning 0). Each consumer chooses among: (i) human content at price P, yielding payoff $\theta R - P$; (ii) AI-generated content at cost c, yielding payoff $\theta r(\lambda) - c$; or (iii) opt out, yielding 0. In equilibrium, the price P adjusts so that the mass of consumers purchasing human content equals the mass of producers supplying it.

To illustrate the impact of introducing generative AI, we first analyze the pre-AI equilibrium and then compare it to the equilibrium after the rise of generative AI.

Equilibrium without generative AI. Without AI, consumer θ purchases human content if and only if $\theta R - P \geq 0$, i.e., $\theta \geq P/R \equiv \theta_{H0}(P)$. Producer ϕ supplies human content if and only if $P - C/\phi \geq 0$, i.e., $\phi \geq C/P \equiv \phi_{H0}(P)$. Market clearing requires

$$1 - F\left(\frac{P_0}{R}\right) \;=\; 1 - G\left(\frac{C}{P_0}\right), \tag{116}$$

which determines the pre-AI equilibrium price P_0 and human content mass $\lambda_0 \equiv 1 - G(C/P_0)$.[35] Consumers with $\theta \geq P_0/R$ purchase human content; all others opt out. Producers with $\phi \geq C/P_0$ supply human content; all others opt out.

Equilibrium with the introduction of generative AI. With generative AI available, define two consumer cutoffs:

$$\theta_{A0}(\lambda) \;\equiv\; \frac{c}{r(\lambda)}, \tag{117}$$

$$\theta_{HA}(\lambda, P) \;\equiv\; \frac{P - c}{R - r(\lambda)}, \tag{118}$$

where θ_{A0} is the consumer type indifferent between AI and opting out, and θ_{HA} is the type indifferent between human and AI content (assuming $R > r(\lambda)$, which is verified below). In equilibrium: consumers with $\theta \in [0, \theta_{A0})$ opt out, those with $\theta \in [\theta_{A0}, \theta_{HA})$ consume AI-generated content, and those with $\theta \in [\theta_{HA}, +\infty)$ purchase human content.

On the supply side, the mass of human content produced at price P is $\lambda(P) = 1 - G(C/P)$.

[35] The existence and uniqueness of P_0 satisfying (116) is guaranteed by the monotonicity and continuity of the functions involved. The left-hand side of (116) is strictly decreasing in P_0 with range $(0, 1)$, while the right-hand side is strictly increasing in P_0 with range $(0, 1)$. Therefore, there exists a unique P_0 such that the two sides are equal.

Inverting, the equilibrium price consistent with supply λ is

$$\Psi(\lambda) \equiv \frac{C}{G^{-1}(1-\lambda)}. \tag{119}$$

Since $G^{-1}(1-\lambda)$ is decreasing in λ, we have $\Psi'(\lambda) > 0$: attracting a larger mass of lower-skill (higher-cost) producers requires a higher market price.

Market clearing between the demand for human content and its supply yields the fixed-point condition

$$\lambda = 1 - F\left(\frac{\Psi(\lambda) - c}{R - r(\lambda)}\right). \tag{120}$$

Proposition 18 characterizes the equilibrium.

Proposition 18. *By the same argument as in Appendix C, $R > r(\lambda)$ holds in any equilibrium. With the introduction of generative AI:*

(i) ***Human-intact case.*** *If*

$$\frac{P_0}{R} \leq \frac{c}{r(\lambda_0)}, \tag{121}$$

then the pre-AI equilibrium persists: producers with $\phi \in [C/P_0, +\infty)$ supply human content, consumers with $\theta \in [0, P_0/R)$ opt out, and consumers with $\theta \in [P_0/R, +\infty)$ purchase human content. Generative AI is not adopted in equilibrium.

(ii) ***AI-disruption case.*** *If condition (121) is violated, i.e.,*

$$\frac{P_0}{R} > \frac{c}{r(\lambda_0)},$$

then there exists a λ^* satisfying the fixed-point condition (120), with equilibrium price $P^* = \Psi(\lambda^*) < P_0$ and $\lambda^* < \lambda_0$. In this equilibrium:

- *Producers with $\phi \in [C/P^*, +\infty)$ supply human content; $\lambda^* = 1 - G(C/P^*)$.*

- *Consumers with* $\theta \in [0,\ \theta_{A0}(\lambda^*))$ *opt out.*
- *Consumers with* $\theta \in [\theta_{A0}(\lambda^*),\ \theta_{HA}(\lambda^*, P^*))$ *consume AI-generated content.*
- *Consumers with* $\theta \in [\theta_{HA}(\lambda^*, P^*), +\infty)$ *purchase human content.*

Proof. We first argue that $R > r(\lambda)$ in equilibrium. Suppose for contradiction that $R \leq r(\lambda)$. Then $\theta R - P < \theta r(\lambda) - c$ for all θ (since any consumer would prefer AI to human content at any positive price P, given $c < P$ is needed for producers to participate). Therefore no consumer purchases human content, implying $\lambda = 0$. But since $r(0) < R$, we get a contradiction.

Human-intact case. Suppose condition (121) holds, i.e., $c/r(\lambda_0) \geq P_0/R$. We verify that the pre-AI equilibrium (P_0, λ_0) survives with the introduction of AI.

For consumers currently purchasing human content ($\theta \geq P_0/R$), they prefer human over AI if $\theta(R - r(\lambda_0)) \geq P_0 - c$, i.e., $\theta \geq \theta_{HA}(\lambda_0, P_0) = (P_0 - c)/(R - r(\lambda_0))$. The condition $P_0/R \geq \theta_{HA}(\lambda_0, P_0)$ holds if and only if $P_0(R - r(\lambda_0)) \geq R(P_0 - c)$, i.e., $Rc \geq P_0 r(\lambda_0)$, i.e., $c/r(\lambda_0) \geq P_0/R$. This is exactly condition (121). Hence all current human consumers (with $\theta \geq P_0/R$) still prefer human content.

For consumers currently opting out ($\theta < P_0/R$), they prefer AI over opting out if $\theta r(\lambda_0) \geq c$, i.e., $\theta \geq c/r(\lambda_0) \geq P_0/R$, contradicting $\theta < P_0/R$. Hence no opt-out consumer adopts AI. The pre-AI equilibrium thus persists.

AI-disruption case. Suppose condition (121) is violated: $P_0/R > c/r(\lambda_0)$, equivalently $P_0 r(\lambda_0) > Rc$.

Define $\Gamma(\lambda) \equiv 1 - F\big((\Psi(\lambda) - c)/(R - r(\lambda))\big)$ on $[0, \bar{\lambda}]$ where $\bar{\lambda} = 1 - G(\phi_{\min})$ is the maximum feasible supply (with $\phi_{\min} = \inf \operatorname{supp}(G)$). We show $\Gamma(\lambda_0) \leq \lambda_0$:

At $\lambda = \lambda_0$ and $\Psi(\lambda_0) = P_0$, the condition $P_0/R > c/r(\lambda_0)$ is equivalent to $P_0 r(\lambda_0) > Rc$, i.e., $\theta_{HA}(\lambda_0, P_0) = (P_0 - c)/(R - r(\lambda_0)) \geq P_0/R$. Therefore $F(\theta_{HA}(\lambda_0, P_0)) \geq F(P_0/R) =$

$1 - \lambda_0$, so $\Gamma(\lambda_0) = 1 - F(\theta_{HA}(\lambda_0, P_0)) \leq \lambda_0$.

Moreover, under mild regularity (e.g., $\lim_{\lambda \to 0} \Psi(\lambda) < c + R \cdot r(0^+)/c$ or a positive mass of consumers with low θ), $\Gamma(0) > 0$. Since Γ is continuous and $\Gamma(\lambda_0) \leq \lambda_0$ while $\Gamma(0) > 0$, the intermediate value theorem applied to $\lambda - \Gamma(\lambda)$ guarantees existence of a fixed point $\lambda^* \in (0, \lambda_0]$.

At λ^*, set $P^* = \Psi(\lambda^*) \leq \Psi(\lambda_0) = P_0$ (since Ψ is increasing and $\lambda^* \leq \lambda_0$). The supply satisfies $\lambda^* = 1 - G(C/P^*)$ by construction of Ψ. The market clearing condition (120) holds by the fixed-point property. The consumer segmentation (opt out, AI, human) follows directly from the cutoff definitions, with $\theta_{A0}(\lambda^*) < \theta_{HA}(\lambda^*, P^*)$ whenever $c(R - r(\lambda^*)) < r(\lambda^*)(P^* - c)$, i.e., $cR < r(\lambda^*)P^*$, which holds when $P^*/R > c/r(\lambda^*)$. □ □

The human-intact condition (121) is the direct analog of condition (7) in Section 4.1: the ratio P_0/R plays the role of C/R, but here P_0 is endogenous and reflects producer skill heterogeneity through the pre-AI market clearing condition (116). Three new features distinguish the combined model from its special cases. First, the equilibrium price P^* adjusts endogenously: a more skilled producer pool (stochastically higher G, so $G^{-1}(1-\lambda)$ larger for all λ) lowers $\Psi(\lambda) = C/G^{-1}(1 - \lambda)$ and thus P^*, making human content more attractive to consumers. Second, the consumer segmentation retains the "middle-out" pattern from Section 4.1, but the boundary θ_{HA} now depends jointly on AI quality $r(\lambda^*)$ and the endogenous price P^*. Third, unlike Appendix K.1 (where all agents choose between human and AI with no opt-out in the disruption case), the combined model reintroduces the opt-out segment via consumer taste heterogeneity. Fourth, the endogenous price P^* serves as a transmission mechanism linking the two sides of the market. When generative AI draws consumers away from human content, demand falls, P^* declines below the pre-AI level P_0, and the cost threshold $\phi_{HA} = C/P^*$ rises, displacing the least-skilled producers and amplifying the supply-side contraction. This cross-market amplification channel is absent from Section 4.1 (which has no competitive price) and Appendix K.1 (which has no consumer heterogeneity);

it is the defining new feature of explicitly separating consumer and producer roles.

Equilibrium partition in the (θ, ϕ) type space. The three cutoffs from Proposition 18—$\theta_{A0}(\lambda^*)$, $\theta_{HA}(\lambda^*, P^*)$, and $\phi_{HA}(P^*) = C/P^*$—partition the joint (θ, ϕ) type space into six rectangular regions. The horizontal consumer cutoffs (θ_{A0}, θ_{HA}) carry over from Section 4.1; the vertical producer cutoff (ϕ_{HA}) carries over from Appendix K.1. What is new here is their combination: because each producer's supply decision depends only on ϕ and P^* (independent of θ) and each consumer's demand decision depends only on θ, λ^*, and P^* (independent of ϕ), the partition is always a *rectangular grid* regardless of the joint distribution of (θ, ϕ). The market-clearing condition $\lambda^* = 1 - G(C/P^*) = 1 - F(\theta_{HA})$ links the two dimensions, pinning down P^* jointly from supply and demand and connecting the horizontal and vertical cutoffs.

Figure 5 illustrates the partition using the same numerical parameters as Figure 2 extended to heterogeneous producers: F uniform on $[0, \bar{\theta}{=}1.2]$, G uniform on $[0, \bar{\phi}{=}2]$, $r(\lambda) = 0.8\lambda$, $R = 0.80$, $C = 0.60$.

Comparative statics. We derive comparative statics as follows.

Corollary 19. *In the AI-disruption case, the equilibrium human content mass λ^* increases as: generative AI cost c increases, human content quality R increases, generative AI quality $r(\cdot)$ decreases, or the producer skill distribution G improves in the sense of first-order stochastic dominance. The relationship between λ^* and AI adoption may be non-monotonic in R and G due to countervailing direct and indirect effects, by arguments analogous to Corollary 7.*

Proof. Define $\Gamma(\lambda) \equiv 1 - F\big((\Psi(\lambda) - c)/(R - r(\lambda))\big)$. Since Ψ and r are both increasing in λ, the ratio $(\Psi(\lambda) - c)/(R - r(\lambda))$ is increasing, so Γ is decreasing. Implicit differentiation at the fixed point $\lambda^* = \Gamma(\lambda^*)$ gives: increasing c lowers $\Psi(\lambda) - c$, raising $\Gamma(\lambda)$ and thus λ^*; increasing

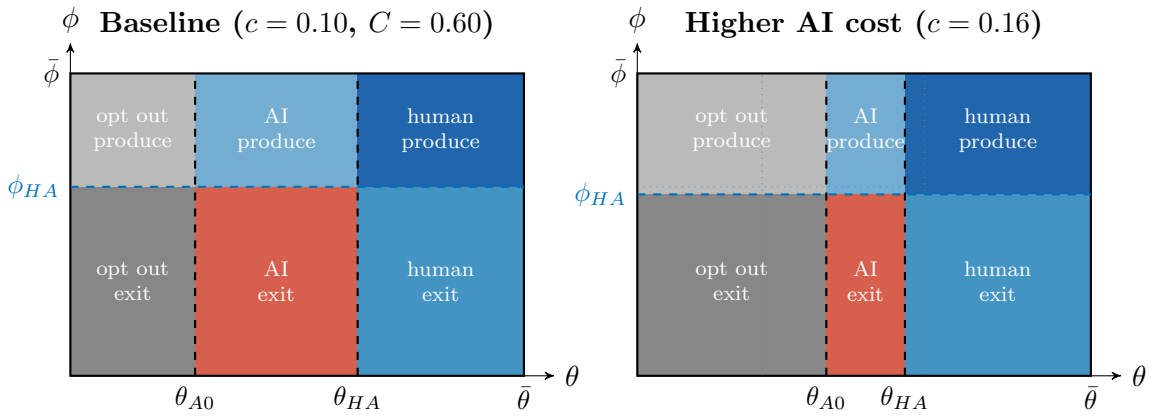


Figure 5: Equilibrium partition in the joint (θ, ϕ) type space (θ uniform on $[0, \bar{\theta}{=}1.2]$, ϕ uniform on $[0, \bar{\phi}{=}2]$, $r(\lambda) = 0.8\lambda$, $R = 0.80$, $C = 0.60$; same parameters as Figure 2 extended to heterogeneous producers). The x-axis is consumer type θ; the y-axis is producer type ϕ. Vertical dashed lines mark consumer cutoffs θ_{A0} (opt-out/AI boundary) and θ_{HA} (AI/human boundary), carried over from Section 4.1; the horizontal blue dashed line marks producer cutoff $\phi_{HA} = C/P^*$ (produce/exit boundary), carried over from Appendix K.1. New to this subsection: their combination into a rectangular grid, with the market-clearing condition $\lambda^* = 1 - G(C/P^*) = 1 - F(\theta_{HA})$ linking the two cutoffs. The right panel shows the effect of raising AI cost to $c = 0.16$; dotted lines reproduce baseline cutoffs. Raising c widens the opt-out region (θ_{A0} moves right), narrows the AI adoption band, and lowers ϕ_{HA} (the higher P^* enables lower-skill producers to re-enter).

R raises $R - r(\lambda)$, similarly raising λ^*; improving $r(\cdot)$ shrinks $R - r(\lambda)$, lowering λ^*; a first-order stochastic improvement in G raises $G^{-1}(1-\lambda)$ for all λ, reducing $\Psi(\lambda) = C/G^{-1}(1-\lambda)$ and thus raising λ^*. Non-monotonicity in R and G follows from the same two countervailing forces as in Corollary 7, with the additional channel that improvements in G shift both human supply and its market price. □

The amount of AI consumption in equilibrium, $F\big((\Psi(\lambda^*) - c)/(R - r(\lambda^*))\big) - F\big(c/r(\lambda^*)\big)$, generalizes expression (9) by replacing the exogenous production cost C/ϕ with the endogenous market price $P^* = \Psi(\lambda^*) = C/G^{-1}(1 - \lambda^*)$.

Corollary 20. *The equilibrium amount of AI consumption increases as AI cost c decreases or AI quality $r(\cdot)$ increases—the same directions as in Section 4.1 (Corollary 7) and Appendix K.1. Its dependence on R, C, and the producer distribution G may be non-monotonic, owing to two countervailing forces already present in Section 4.1: (1) a* direct substitution

effect *(higher R or lower C makes human content more attractive, contracting AI adoption from the top via lower θ_{HA}) and (2) an* indirect quality effect *(the resulting higher λ^* raises $r(\lambda^*)$, drawing in low-θ consumers who previously opted out via lower θ_{A0}). The combined model adds a* third channel *absent from both Section 4.1 and Appendix K.1: an improvement in the producer skill distribution G lowers the equilibrium price $P^* = \Psi(\lambda^*)$, compressing the human–AI price gap and contracting AI adoption at the top margin (lower θ_{HA}), while simultaneously raising λ^* and AI quality, expanding adoption at the bottom margin (lower θ_{A0}). These two sub-effects of G work in opposite directions, making the net impact on the AI adoption band $F(\theta_{HA}) - F(\theta_{A0})$ ambiguous.*

Proof. Monotonicity in c and $r(\cdot)$ follows from Corollary 19: a lower c directly reduces $\theta_{A0} = c/r(\lambda^*)$ and raises $\theta_{HA} = (\Psi(\lambda^*) - c)/(R - r(\lambda^*))$, widening the adoption band; a higher $r(\cdot)$ raises λ^*, which lowers θ_{A0} and raises θ_{HA} similarly. For R, C, and G, the non-monotonicity follows by implicit differentiation of the AI consumption expression using the fixed-point condition (120). The two countervailing forces for R and C are the same as in Corollary 7. For G: a first-order stochastic improvement raises λ^* (Corollary 19) but simultaneously lowers $\Psi(\lambda^*) = C/G^{-1}(1-\lambda^*)$ at any given λ, so the effect on $\theta_{HA} = (\Psi(\lambda^*) - c)/(R - r(\lambda^*))$ and the indirect effect on $\theta_{A0} = c/r(\lambda^*)$ via λ^* run in opposite directions, leaving the net change in $F(\theta_{HA}) - F(\theta_{A0})$ ambiguous. □

L Further extensions to Section 4.2

We present two further extensions to Section 4.2. The first (Section L.1) incorporates transfer learning and foundation models, showing that the paper's core results survive and are in fact strengthened. The second (Section L.2) considers general aggregation functions beyond the Cobb–Douglas form, showing that the core insights about human–AI substitution and complementarity generalize.

L.1 Transfer Learning and Foundation Models

Modern generative AI systems are typically built as *foundation models*—large models pre-trained on vast, undifferentiated corpora spanning many tasks simultaneously (Bommasani, Hudson, Adeli, Altman, Arora, von Arx, Bernstein, Bohg, Bosselut, Brunskill et al., 2021). A key implication is *transfer learning*: representations learned from human content in one task (e.g., writing) improve AI performance in another task (e.g., investigation), and vice versa. In this section we extend the model of Section 4.2 to incorporate this feature, and show that the paper's core results survive and are in fact strengthened.

Setup. We generalize the AI quality functions to $q_I = q_I(\lambda_I, \lambda_W)$ and $q_W = q_W(\lambda_I, \lambda_W)$, both increasing in both arguments:

$$\frac{\partial q_I}{\partial \lambda_I} > 0, \quad \frac{\partial q_I}{\partial \lambda_W} > 0, \quad \frac{\partial q_W}{\partial \lambda_I} > 0, \quad \frac{\partial q_W}{\partial \lambda_W} > 0.$$

Human effort in investigation now spills over into AI writing quality, and human writing effort spills over into AI investigation quality, capturing the joint-corpus training of foundation models. The baseline model of Section 4.2 is the special case in which the cross-derivatives are zero ($\partial q_I/\partial \lambda_W = \partial q_W/\partial \lambda_I = 0$), so that $q_I(\lambda_I, \lambda_W) = q_I(\lambda_I)$ and $q_W(\lambda_I, \lambda_W) = q_W(\lambda_W)$. All other elements of the model remain as in Section 4.2.

We maintain the following analogs of the assumptions in Section 4.2.

Assumption 3. *$q_I(0, \lambda_W) = 0$ for all $\lambda_W \in [0, 1]$.*[36]

[36] This condition is assumed for simplicity only. The analysis goes through under the weaker conditions: $q_I(0, \lambda_W) < Q_I$, $q_I^{\alpha}(0, \lambda_W)\, Q_W^{1-\alpha} - c_I - C_W < 0$, and $q_I^{\alpha}(0, \lambda_W)\, q_W^{1-\alpha}(0, \lambda_W) - c_I - c_W < 0$ for all $\lambda_W \in [0, 1]$. These are the natural generalization to the transfer learning setting of the conditions in footnote 22. The first condition reflects fast model decay: writing-to-investigation cross-spillovers may improve the AI model's language and reasoning capabilities, but cannot substitute for fresh real-world signal acquisition, so AI investigation quality remains strictly below human quality even at maximum writing effort. The second and third conditions ensure AI investigation is unprofitable when no human investigates, for all levels of human writing effort.

Assumption 4. *The function* $H(\lambda_I) \equiv (Q_I^\alpha - q_I^\alpha(\lambda_I, 0)) \cdot q_W^{1-\alpha}(\lambda_I, 0)$ *is strictly decreasing in* λ_I*, i.e.,*

$$\frac{\partial q_I^\alpha / \partial \lambda_I}{Q_I^\alpha - q_I^\alpha(\lambda_I, 0)} > \frac{\partial q_W^{1-\alpha} / \partial \lambda_I}{q_W^{1-\alpha}(\lambda_I, 0)}, \qquad \forall\, \lambda_I \in [0, 1].$$

That is, the proportional rate at which human investigation erodes the quality gap $(Q_I^\alpha - q_I^\alpha)$ *exceeds the proportional rate at which it improves AI writing quality* $q_W^{1-\alpha}$ *through cross-spillover.*[37]

Unchanged results. Several results from Section 4.2 remain intact when we introduce transfer learning.

Proposition 21. *Under Assumption 3:*

1. *The same eight equilibrium types characterized in Proposition 8 remain the only possible equilibrium configurations.*
2. *AI investigation never fully displaces human investigation in any equilibrium* ($\lambda_I > 0$ *always).*

Proof. We first establish that $q_I(\lambda_I^*, \lambda_W^*) < Q_I$ at any equilibrium state $(\lambda_I^*, \lambda_W^*)$. Suppose to the contrary that $q_I(\lambda_I^*, \lambda_W^*) \geq Q_I$. Since AI investigation then matches or exceeds human investigation quality while costing strictly less ($c_I < C_I$), every agent strictly prefers AI investigation, driving $\lambda_I^* = 0$. But Assumption 3 gives $q_I(0, \lambda_W^*) = 0 < Q_I$, contradicting $q_I(\lambda_I^*, \lambda_W^*) \geq Q_I$. Hence $q_I(\lambda_I^*, \lambda_W^*) < Q_I$ in any equilibrium.

With this, the key lemma of the proof of Proposition 8 carries over unchanged: at any equilibrium state,

$$\pi_{HH} = \pi_{HA} \;\Rightarrow\; Q_I^\alpha \big(Q_W^{1-\alpha} - q_W^{1-\alpha}\big) = C_W - c_W$$

[37] Assumption 4 holds trivially in the baseline model, where $\partial q_W / \partial \lambda_I = 0$ makes the right-hand side zero. It is the natural generalization of that property to the transfer learning setting, requiring that own-task effects dominate cross-task spillovers in the investigation equilibrium condition.

$$\Rightarrow\ q_I^{\alpha}\left(Q_W^{1-\alpha} - q_W^{1-\alpha}\right) < C_W - c_W \ \Leftrightarrow\ \pi_{AH} < \pi_{AA},$$

where the second implication uses $q_I(\lambda_I^*, \lambda_W^*) < Q_I$ established above. This rules out HA+AH and related configurations (cases 8, 11, 14, 15 in the original 15-case enumeration).

Furthermore, when $\lambda_I = 0$, Assumption 3 gives $q_I(0, \lambda_W) = 0$, so $\pi_{AH} < \pi_{AA} < 0 < \pi_{HH}$ for any λ_W, ruling out pure AH, AA, and AH+AA equilibria by the same argument as in Proposition 8. The remaining eight equilibrium types are therefore the only possible configurations, and each features $\lambda_I > 0$. □

Transfer learning strengthens human investigation incentives We now show that cross-spillovers reinforce rather than weaken the incentive for human investigation. We focus on the "Capable AI for writing" equilibrium ($\lambda_W = 0$, $\lambda_I \in (0, 1)$) from Section 4.2.3, which provides the sharpest illustration. With $\lambda_W = 0$, the equilibrium condition (14) generalizes to

$$\underbrace{\left(Q_I^{\alpha} - q_I^{\alpha}(\lambda_I,\, 0)\right)}_{\text{investigation quality gap}} \cdot \underbrace{q_W^{1-\alpha}(\lambda_I,\, 0)}_{\text{AI writing quality}} \ =\ C_I - c_I. \tag{122}$$

The left-hand side now embeds two forces as λ_I rises: the investigation quality gap shrinks (as in the baseline), while AI writing quality $q_W(\lambda_I, 0)$ rises through the investigation-to-writing cross-spillover. Assumption 4 ensures the former dominates, so $H(\lambda_I)$ remains strictly decreasing and equation (122) has a unique solution.

Proposition 22. *Suppose Assumptions 3 and 4 hold. Then the human investigation share in the "Capable AI for writing" equilibrium under transfer learning, λ_I^{TL}, weakly exceeds that in the baseline model without cross-spillovers, λ_I^{base}:*

$$\lambda_I^{TL} \ \geq\ \lambda_I^{base},$$

with strict inequality whenever $q_W(\lambda_I^{base}, 0) > q_W(0)$. *Moreover, an improvement in AI writing quality raises* λ_I^{TL} *through two reinforcing channels:*

1. *the direct Cobb-Douglas multiplier channel of Proposition 9; and*
2. *an additional feedback channel whereby higher* λ_I *further improves* $q_W(\lambda_I, 0)$ *through cross-spillover, amplifying the initial effect.*

Proof. Since the baseline is the special case with $\partial q_W/\partial \lambda_I = 0$, we have $q_W(\lambda_I, 0) = q_W(0)$ for all λ_I in the baseline. The investigation-to-writing cross-spillover means $q_W(\lambda_I, 0) \geq q_W(0)$ for all $\lambda_I \geq 0$ in the transfer learning model, with strict inequality for $\lambda_I > 0$.

Evaluate H at the baseline equilibrium λ_I^{base}, which satisfies $(Q_I^\alpha - q_I^\alpha(\lambda_I^{base})) \cdot q_W^{1-\alpha}(0) = C_I - c_I$:

$$\begin{aligned} H(\lambda_I^{base}) &= \left(Q_I^\alpha - q_I^\alpha(\lambda_I^{base})\right) \cdot q_W^{1-\alpha}(\lambda_I^{base}, 0) \\ &\geq \left(Q_I^\alpha - q_I^\alpha(\lambda_I^{base})\right) \cdot q_W^{1-\alpha}(0) \\ &= C_I - c_I, \end{aligned}$$

where the inequality uses $q_W(\lambda_I^{base}, 0) \geq q_W(0)$, with strict inequality when $\lambda_I^{base} > 0$.[38] Since H is strictly decreasing (Assumption 4), the unique solution to $H(\lambda_I) = C_I - c_I$ satisfies $\lambda_I^{TL} \geq \lambda_I^{base}$, with strict inequality when $q_W(\lambda_I^{base}, 0) > q_W(0)$.

For the amplification result: an upward shift in AI writing quality increases $H(\lambda_I)$ at every λ_I (raising the writing quality multiplier on the investigation quality gap), so λ_I^{TL} increases — this is channel (i). The resulting rise in λ_I^{TL} further increases $q_W(\lambda_I^{TL}, 0)$ through $\partial q_W/\partial \lambda_I > 0$, shifting H up again and reinforcing the initial response — this is channel (ii). Channel (ii) is absent in the baseline (where $\partial q_W/\partial \lambda_I = 0$), so the total effect on λ_I^{TL} is

[38] In writing the first line we use that $q_I(\lambda_I^{base}, 0) = q_I(\lambda_I^{base})$, since the baseline model is the special case of the transfer learning model with zero cross-derivatives, so the direct investigation effect is identical across the two models.

weakly larger than in the baseline. □

Proposition 22 delivers a clear message: foundation model cross-spillovers strengthen rather than undermine the case for human investigation. In the baseline, human investigation supports content quality solely through the Cobb-Douglas production function. Under transfer learning, human investigation additionally improves AI writing quality, which in turn raises the overall utility gains from choosing high-quality human over AI investigation (since writing quality serves as a multiplier on the value of the investigation quality gap). This self-reinforcing loop means that a world with stronger transfer learning features *more* human investigative effort in equilibrium, not less.

The writing-to-investigation cross-spillover ($\partial q_I/\partial \lambda_W > 0$) plays a complementary role in equilibria where $\lambda_W > 0$: it raises AI investigation quality at the equilibrium writing share, increasing the stakes of the investigation quality contest and further incentivizing human investigators to maintain their quality advantage over AI. While this can shift which of the eight equilibrium types prevails across parameter regimes, Proposition 21 ensures human investigation is never fully displaced.

Strong transfer learning and coordination multiplicity The foregoing analysis maintains Assumption 3, which ensures that writing cross-spillovers cannot fully offset the decay of AI investigation quality when human investigators exit. We now examine what happens when this assumption is violated and cross-spillovers are strong enough that

$$q_I(0,1) \geq Q_I, \tag{123}$$

$$q_W(1,0) \geq Q_W. \tag{124}$$

Condition (123) says that human writing effort alone suffices to raise AI investigation quality to match or exceed the human benchmark; condition (124) says that human investigation

effort alone suffices to raise AI writing quality to match or exceed the human benchmark.

Under these conditions, the key lemma in the proof of Proposition 21 breaks down: $q_I(\lambda_I^*, \lambda_W^*) \geq Q_I$ becomes possible (specifically at $\lambda_I^* = 0$, $\lambda_W^* = 1$), so the argument ruling out AH configurations no longer applies. Two self-sustaining equilibria can then coexist:

- A *human-investigation equilibrium* $(\lambda_I, \lambda_W) = (1, 0)$, in which all agents choose HA (human investigation, AI writing). The high investigation effort raises AI writing quality via cross-spillover (124), validating the choice of AI writing.

- A *human-writing equilibrium* $(\lambda_I, \lambda_W) = (0, 1)$, in which all agents choose AH (AI investigation, human writing). The high writing effort raises AI investigation quality via cross-spillover (123), validating the choice of AI investigation.

Each equilibrium is self-fulfilling: the cross-spillover it generates sustains it. Sufficient conditions for both to exist simultaneously are that, at $(\lambda_I, \lambda_W) = (1, 0)$: $\pi_{HA} \geq \max(\pi_{HH}, \pi_{AH}, \pi_{AA})$; and at $(\lambda_I, \lambda_W) = (0, 1)$: $\pi_{AH} \geq \max(\pi_{HH}, \pi_{HA}, \pi_{AA})$. Under conditions (123)–(124) and standard regularity on q_I and q_W, both sets of inequalities can be simultaneously satisfied, yielding genuine equilibrium multiplicity.

This multiplicity is qualitatively different from the parameter-driven diversity of the eight types in Proposition 8: there, a given parameter configuration uniquely selects an equilibrium type; here, a single parameter configuration admits two coexisting equilibria, with the outcome depending on initial conditions and coordinated expectations—a standard feature of games with strategic complementarities.

As discussed in Section 6, conditions (123)–(124) require AI systems with autonomous real-world sensing capabilities—such as physical agents, satellite imagery, or IoT networks—that can substitute for human investigators even without fresh human-generated content. This lies outside the scope of current language-based generative AI, for which Assumption 3 applies and Propositions 21–22 are the relevant results.

L.2 General Supermodular Aggregator

This appendix shows that the analysis of Section 4.2 extends beyond the Cobb-Douglas specification. We replace the aggregator $Q_I^{\alpha} Q_W^{1-\alpha}$ with a general function $R(Q_I, Q_W)$ satisfying $R(0, Q_W) = R(Q_I, 0) = 0$, and

$$\frac{\partial R}{\partial x}(x, y) > 0, \qquad \frac{\partial R}{\partial y}(x, y) > 0, \qquad \frac{\partial^2 R}{\partial x \, \partial y}(x, y) > 0, \tag{125}$$

for all $x, y > 0$. The first two conditions say that content value is strictly increasing in both investigation quality and writing quality. The third condition—strict supermodularity—says that higher investigation quality raises the marginal value of writing quality, and vice versa, capturing the complementarity between the two tasks that is central to Section 4.2.

Setup. The setup is identical to Section 4.2, except that the value of a piece of content with investigation quality q_1 and writing quality q_2 is $R(q_1, q_2)$ rather than $q_1^{\alpha} q_2^{1-\alpha}$. All other notation—costs C_I, C_W, c_I, c_W; human qualities Q_I, Q_W; AI quality functions $q_I(\lambda_I)$, $q_W(\lambda_W)$ with $q_I'(\lambda_I) > 0$, $q_W'(\lambda_W) \geq 0$; and the aggregate participation rates $\lambda_I, \lambda_W \in [0, 1]$—is unchanged. We retain the assumption $q_I(0) = 0$, so that in the absence of human investigation the AI investigation quality degrades completely, ensuring $R(q_I(0), Q_W) - c_I - C_W < 0$ and $R(q_I(0), q_W(1)) - c_I - c_W < 0$.[39]

The four payoffs are:

1. *Human investigation + human writing:*

$$\pi_{HH} \equiv R(Q_I, Q_W) - C_I - C_W > 0. \tag{126}$$

[39] As in Section 4.2, the condition $q_I(0) = 0$ is for simplicity. The analysis goes through under the weaker requirement that $R(q_I(0), Q_W) - c_I - C_W < 0$ and $R(q_I(0), q_W(1)) - c_i - c_w < 0$, i.e., when no human investigates, the degraded AI investigation quality renders every content-production option unprofitable or strictly dominated.

We assume (126) is strictly positive; otherwise no content would have been produced prior to the emergence of generative AI.

2. *Human investigation + AI writing:*

$$\pi_{HA} \equiv R\big(Q_I,\, q_W(\lambda_W)\big) - C_I - c_W. \tag{127}$$

3. *AI investigation + human writing:*

$$\pi_{AH} \equiv R\big(q_I(\lambda_I),\, Q_W\big) - c_I - C_W. \tag{128}$$

4. *AI investigation + AI writing:*

$$\pi_{AA} \equiv R\big(q_I(\lambda_I),\, q_W(\lambda_W)\big) - c_I - c_W. \tag{129}$$

Equilibrium characterization An equilibrium requires that each agent chooses the option with the highest payoff, given the aggregate fractions (λ_I, λ_W) that are consistent with these individual choices. Because R is strictly increasing in its first argument and AI has a strict cost advantage ($c_I < C_I$), the same argument as in Appendix E establishes that no equilibrium exists in which AI fully takes over investigation: if $q_I(\lambda_I) \geq Q_I$, then AI investigation would strictly dominate human investigation for every agent, forcing $\lambda_I = 0$, but at $\lambda_I = 0$ we have $q_I(0) = 0 < Q_I$—a contradiction.

Interior equilibria are characterized by indifference conditions. Writing q_W^* and q_I^* for the equilibrium AI quality levels:

- *Indifference over investigation* (given writing quality q_W^*):

$$R(Q_I,\, q_W^*) - R\big(q_I(\lambda_I),\, q_W^*\big) = C_I - c_I.$$

- *Indifference over writing* (given investigation quality q_I^*):

$$R(q_I^*,\, Q_W) - R\big(q_I^*,\, q_W(\lambda_W)\big) = C_W - c_W.$$

Proposition 23. *Under the conditions in (125), the same eight classes of equilibrium outcomes as in Proposition 8 are possible. They are characterized by replacing each Cobb-Douglas expression $x^\alpha y^{1-\alpha}$ with $R(x, y)$ in the conditions stated in Appendix E. Specifically, the eight equilibria feature:*

1. *Only HH ($\lambda_I = \lambda_W = 1$): $R(Q_I, Q_W) - R(Q_I, q_W(1)) \geq C_W - c_W$, $R(Q_I, Q_W) - R(q_I(1), Q_W) \geq C_I - c_I$, and $R(Q_I, Q_W) - C_I - C_W \geq R(q_I(1), q_W(1)) - c_I - c_W$.*

2. *Only HA ($\lambda_I = 1$, $\lambda_W = 0$): $R(Q_I, Q_W) - R(Q_I, q_W(0)) \leq C_W - c_W$, $R(Q_I, q_W(0)) - R(q_I(1), q_W(0)) \geq C_I - c_I$, and $R(Q_I, q_W(0)) - C_I - c_W > R(q_I(1), Q_W) - c_I - C_W$.*

3. *HH + HA ($\lambda_I = 1$, $\lambda_W \in (0, 1)$): $q_W(\lambda_W)$ solves*

$$R(Q_I, Q_W) - R\big(Q_I,\, q_W(\lambda_W)\big) = C_W - c_W.$$

4. *HH + AH ($\lambda_W = 1$, $\lambda_I \in (0, 1)$): $q_I(\lambda_I)$ solves*

$$R(Q_I, Q_W) - R\big(q_I(\lambda_I),\, Q_W\big) = C_I - c_I.$$

5. *HH + AA ($\lambda_I = \lambda_W = \lambda^*$): λ^* solves*

$$R(Q_I, Q_W) - C_I - C_W = R\big(q_I(\lambda^*),\, q_W(\lambda^*)\big) - c_I - c_W.$$

6. *HA + AA ($\lambda_W = 0$, $\lambda_I \in (0,1)$): $q_I(\lambda_I)$ solves*

$$R\big(Q_I,\, q_W(0)\big) - R\big(q_I(\lambda_I),\, q_W(0)\big) = C_I - c_I.$$

7. *HH + AH + AA ($\lambda_I \leq \lambda_W$): $q_I(\lambda_I)$ solves $R(Q_I, Q_W) - R(q_I(\lambda_I), Q_W) = C_I - c_I$, and $q_W(\lambda_W)$ solves $R(q_I(\lambda_I), Q_W) - R(q_I(\lambda_I), q_W(\lambda_W)) = C_W - c_W$.*

8. *HH + HA + AA ($\lambda_I \geq \lambda_W$): $q_W(\lambda_W)$ solves $R(Q_I, Q_W) - R(Q_I, q_W(\lambda_W)) = C_W - c_W$, and $q_I(\lambda_I)$ solves $R(Q_I, q_W(\lambda_W)) - R(q_I(\lambda_I), q_W(\lambda_W)) = C_I - c_I$.*

No equilibrium features $\lambda_I = 0$.

Proof. The proof follows the same case analysis as Appendix E, replacing every instance of $x^\alpha y^{1-\alpha}$ with $R(x, y)$. The strict monotonicity and supermodularity of R in (125) are the only properties needed to replicate each step; no special structure of the Cobb-Douglas form is used in Appendix E. □

Two special cases.

Incapable AI for investigation. If AI is completely incapable of investigation ($q_I(\lambda_I) = 0$ for all λ_I), every agent must investigate in person, so $\lambda_I = 1$. The three possible equilibria are:

1. If $R(Q_I, Q_W) - R(Q_I, q_W(1)) \geq C_W - c_W$, there is an equilibrium with $\lambda_W = 1$ (only HH);

2. If $R(Q_I, Q_W) - R(Q_I, q_W(0)) \leq C_W - c_W$, there is an equilibrium with $\lambda_W = 0$ (only HA);

3. If $R(Q_I, Q_W) - R(Q_I, q_W(1)) < C_W - c_W < R(Q_I, Q_W) - R(Q_I, q_W(0))$, there is an equilibrium with $\lambda_W \in (0, 1)$ solving

$$R(Q_I, Q_W) - R\big(Q_I,\, q_W(\lambda_W)\big) = C_W - c_W. \tag{130}$$

This equilibrium features only human investigators and both human and AI writers.

Existence and uniqueness of λ_W in case 3 follow from the strict monotonicity of $R(Q_I, \cdot)$ (guaranteed by $\partial R/\partial y > 0$), which ensures a unique solution for $q_W(\lambda_W)$ and hence for λ_W.

Capable AI for writing. Consider the case in which generative AI is so capable of writing that it completely takes over all human writing ($\lambda_W = 0$, so $q_W(\lambda_W) = q_W(0)$). The equilibrium conditions for investigation reduce to:

1. If $R(Q_I, q_W(0)) - R(q_I(1), q_W(0)) \geq C_I - c_I$ (and other conditions ensuring $\lambda_W = 0$ is sustainable), there is an equilibrium with $\lambda_I = 1$ (only human investigators, only AI writers);

2. If $R(Q_I, q_W(0)) - R(q_I(1), q_W(0)) \leq C_I - c_I$ and the market is in a mixed-investigation regime, there is an equilibrium with $\lambda_W = 0$ and $\lambda_I \in (0, 1)$ solving

$$R\big(Q_I,\, q_W(0)\big) - R\big(q_I(\lambda_I),\, q_W(0)\big) = C_I - c_I. \tag{131}$$

The following proposition generalizes Proposition 9 to any supermodular R.

Proposition 24. *Under the conditions in* (125)*, when generative AI is moderately capable of writing so that* (131) *holds, the equilibrium level of human investigation λ_I strictly increases as AI writing quality $q_W(0)$ improves.*

Proof. Differentiate (131) implicitly with respect to $q_W(0)$:

$$\underbrace{\frac{\partial R}{\partial y}\big(Q_I, q_W(0)\big) - \frac{\partial R}{\partial y}\big(q_I(\lambda_I), q_W(0)\big)}_{>\,0} = \frac{\partial R}{\partial x}\big(q_I(\lambda_I), q_W(0)\big) \cdot q_I'(\lambda_I) \cdot \frac{d\lambda_I}{dq_W(0)}.$$

The left-hand side is strictly positive: since $Q_I > q_I(\lambda_I)$ in every interior equilibrium and $\partial^2 R/\partial x\,\partial y > 0$, the function $\partial R/\partial y(\cdot, q_W(0))$ is strictly increasing, so its value at Q_I exceeds its value at $q_I(\lambda_I)$. The right-hand side has $\partial R/\partial x > 0$ and $q_I' > 0$, so $d\lambda_I/dq_W(0) > 0$. □

The economic intuition is unchanged from Proposition 9: strict supermodularity of R means that an improvement in AI writing quality raises the marginal return to investigation quality. Because human investigation quality Q_I exceeds AI investigation quality $q_I(\lambda_I)$, humans benefit disproportionately from this increase, drawing more agents into investigative work. This result holds for any supermodular R satisfying (125) and does not depend on the Cobb-Douglas parametric form.

Comparative statics

Corollary 25. *Under the supermodular conditions in* (125)*, Corollary 10 continues to hold: in each equilibrium, all interior values of λ_I and λ_W strictly increase in Q_I and Q_W (with the exception noted in Corollary 10 for the "capable AI for writing" case, where λ_I increases in $q_W(0)$ instead of Q_W); all interior values of λ_I (λ_W) strictly decrease in C_I (C_W) and strictly increase in c_I (c_W). Cross-task effects—whereby the human quality or cost in one task affects equilibrium human adoption in the other—are governed by the strict supermodularity condition $\partial^2 R/\partial x\,\partial y > 0$.*

Proof. We verify the two types of effects.

Own effects. For any interior equilibrium condition of the form $R(Q_I, q_W^*) - R(q_I(\lambda_I), q_W^*) =$

$C_I - c_I$ with q_W^* fixed, implicit differentiation with respect to Q_I yields

$$\frac{\partial R}{\partial x}(Q_I, q_W^*) \;=\; \frac{\partial R}{\partial x}\big(q_I(\lambda_I), q_W^*\big) \cdot q_I'(\lambda_I) \cdot \frac{d\lambda_I}{dQ_I}.$$

Since $\partial R/\partial x > 0$ and $q_I' > 0$, we obtain $d\lambda_I/dQ_I > 0$. The effects of C_I and c_I follow directly from differentiating with respect to $C_I - c_I$. Symmetric arguments apply for λ_W, Q_W, C_W, and c_W.

Cross-task effects. For the equilibrium condition $R(Q_I, Q_W) - R(q_I(\lambda_I), Q_W) = C_I - c_I$ (with $\lambda_W = 1$), differentiating with respect to Q_W yields

$$\frac{\partial R}{\partial y}(Q_I, Q_W) - \frac{\partial R}{\partial y}\big(q_I(\lambda_I), Q_W\big) \;=\; \frac{\partial R}{\partial x}\big(q_I(\lambda_I), Q_W\big) \cdot q_I'(\lambda_I) \cdot \frac{d\lambda_I}{dQ_W}.$$

By strict supermodularity ($\partial^2 R/\partial x\,\partial y > 0$) and $Q_I > q_I(\lambda_I)$, the left-hand side is strictly positive, so $d\lambda_I/dQ_W > 0$. Identical reasoning applies to other equilibrium types and to λ_W.

Equilibria in which λ_I (or λ_W) is unaffected by cross-task parameters arise when one task is entirely handled by humans or entirely by AI, so the relevant indifference condition does not depend on the other task's quality parameter—exactly as in Corollary 10. □