Intellectual Property Protection for AI-Generated Output

Robin Döttling, Logan P. Emery, and Shuo Zhao

June 24, 2026

Abstract

We model firms' innovation investment when generative AI can lower investment costs, but only sufficient human input qualifies for IP protection and AI use is unobservable. We characterize the IP system's effect on incentives using two edge cases: when the IP system can "kill" AI use and when low AI costs can "kill" the IP system. If consumers value human-developed products, human-capital use is distorted by adverse selection. The IP policy can mitigate this by acting as a signal of incentives for human-capital use that triggers a feedback loop through consumer beliefs, or by deterring high-cost firms' investment.

Keywords: Generative AI, Innovation, Intellectual Property Protection, Copyright, Adverse Selection

JEL Classification: G31, G38, O31, O34, O38

Robin Döttling is affiliated with Erasmus University Rotterdam, CEPR, and the Finance Theory Group. Logan Emery and Shuo Zhao are affiliated with Erasmus University Rotterdam. For helpful comments and feedback, we thank Fabrizio Core, Shiwei Ye, and Elena Novelli, as well as seminar participants at Erasmus University Rotterdam, the FTG Corporate Governance & Control Group, Tilburg University, the 34th FTG Meeting at the University of Colorado at Boulder, Nova School of Business and Economics, Deutsche Bundesbank, the 2026 FTG Summer Conference, and the Global Entrepreneurship and Innovation Research Conference (GEIRC).

1 Introduction

The advent of Generative Artificial Intelligence (henceforth "AI") has the potential to fundamentally transform corporate innovation. With the ability to generate virtually limitless output at low costs with limited human involvement, many firms are considering transitioning significant portions of their human labor force to AI.[1] However, this raises the question as to who owns AI-generated output. Under current U.S. law, innovation developed without meaningful human input is ineligible for protection by intellectual property (IP) systems such as patents or copyright (Naruto v. Slater, 2018; Thaler v. Vidal, 2022), with most jurisdictions around the world following similar principles. Works produced jointly by humans and AI may qualify, but only when the human contribution is sufficiently substantive and central to the innovation process (Thaler v. Perlmutter, 2023). The exact threshold for required human input remains unclear and is still being debated. Moreover, it may be difficult to verify the exact extent to which AI was used. Nonetheless, these restrictions have important implications for the adoption of AI by innovative firms, as the excludability provided by IP systems has traditionally been regarded as essential for incentivizing innovation (Arrow, 1962; Crouzet et al., 2022).

Further complicating matters, while the legal boundaries of ownership are still being decided in court, consumers have already formed strong views about AI-generated output. In particular, survey and experimental evidence document a preference for human- over AI-developed products.[2] Although these preferences may vary widely across individuals, jurisdictions, and product categories, they significantly influence both firm incentives and the design of IP systems.

In this paper, we analyze how the IP system and consumer preferences shape firms' incentives to invest in innovation supported by AI. Our analysis provides several insights. Starting with a parsimonious baseline model, we characterize how the IP system affects firms' incentives using two interesting edge cases: the IP system "kills'' AI, and AI "kills'' the IP system. While the IP policy can completely disincentivize AI use, low-cost AI can make it unnecessary to grant IP protection to incentivize innovation investment. However, since IP systems currently condition protection on human-capital use, we show that rather than shutting down the system completely, low-cost AI may shift the system's role from innovation subsidy to human-capital subsidy. We then introduce consumer preferences for human-developed products, which interact with the IP system and firms' incentives in non-trivial ways. Most importantly, they create an adverse selection problem that can lead to market collapse. We show that the IP policy can mitigate this problem by deterring investment by high-cost firms. Additionally, the IP system can trigger a positive

[1] Examples include Amazon, Microsoft, Duolingo, Telstra, Lionsgate, and Business Insider.

[2] This has been documented for art (Horton et al., 2023), music (Tubadji et al., 2021), books (Manewitsch and Kalb, 2025), newspaper articles (Kennedy and Yam, 2025), and personal advice (Osborne and Bailey, 2025).

feedback loop through consumer beliefs by acting as a credible signal of incentives for human-capital use, even if consumers never observe whether IP protection is granted.

We develop these insights in a model in which a firm can invest in innovation using a mix of AI and human capital inputs to develop a new intangible asset that may obtain IP protection. Our model can apply to innovation investments that result in different forms of IP. For example, consider a pharmaceutical firm deciding between the use of human pharmacologists or AI when developing a new, patentable vaccine. Alternatively, consider a publishing firm deciding between commissioning human authors or AI to write a new, copyrightable book. In each case, there is an important development stage, which we broadly define as innovation. The resulting intangible is then implemented into a production and commercialization process that results in a product (i.e., the manufacturing and distribution of vaccine doses or the printing and distribution of copies of the book). We discuss potential differences between copyright and patents in the context of our model after deriving our main results.

The investment entails an upfront cost, such that the firm only invests if expected profits exceed this cost. In addition to the binary investment decision, the firm chooses the relative use of human capital versus AI inputs. We assume that AI use can decrease the upfront cost, reflecting the marginal-cost advantage of AI. However, human-capital use may carry positive social value, such as the value of human creativity, the dignity of human work, tacit knowledge spillovers, or skill formation. Additionally, the firm can either be a high- or low-cost type defined by its marginal cost of human capital.

Once the firm has invested and chosen its inputs, an intangible asset is developed. An IP officer then decides whether to grant IP protection in the form of monopoly commercialization rights.[3] Subsequently, the firm enters production and decides on its production scale. If the intangible asset receives protection, the firm operates as a monopolist. Otherwise, competitors can enter the market by copying the resulting product.

Motivated by the policy debates discussed above, we allow the granting of IP protection to be conditional on human-capital use. However, the firm's human-capital use is not observable. Instead, the officer receives a noisy binary signal as to whether the firm used substantial human capital in developing the intangible asset. The officer also cannot observe the firm's type, inhibiting perfect inference of the firm's human-capital use.

We model the granting of IP protection as a probabilistic outcome characterized by two policy parameters: the *intercept* and the *slope*. The intercept parameter represents the baseline probability of granting protection, independent of the signal. It can be interpreted as the requirements for characteristics (other than human-capital use) assessed by the officer when granting protection, such as non-obviousness or distinctiveness from prior works. The slope parameter is defined as the increase in the probability of granting

[3]While we interpret the granting of IP protection as an officer reviewing an application, an equivalent interpretation is that protection is determined through judicial review in a court proceeding.

protection conditional on the signal indicating high human-capital use. It can be interpreted as how strongly the IP policy differentiates between human- and AI-developed intangibles when granting protection. This modeling approach enables us to cleanly isolate the effect of the IP system's baseline leniency (the intercept) and its marginal reward for perceived human-capital use in the innovation process (the slope).

We start by analyzing this parsimonious baseline model to transparently characterize how the IP system affects the firm's incentive to use human capital in the innovation process. To illustrate this effect, we ask whether the IP system can "kill" AI, in the sense that it fully disincentivizes AI use. In our baseline model, the marginal probability of granting protection due to human-capital use determines the IP system's effect on human-capital use. Our baseline model is designed to pin down this marginal probability as the slope policy parameter, so a sufficiently steep slope can "kill" AI. We show that this insight carries over to a continuous-signal setting, where the officer chooses a threshold for the human-capital requirement. This extension shows that greater leniency toward human-capital use can increase *or* decrease human-capital use, depending on whether it raises or lowers the policy's slope. Thus, when assessing the effect of the IP system on human-capital use, it is important to assess the marginal probability of granting protection rather than relying on heuristics such as leniency.

Having explored what disincentivizes the firm from using AI, we next explore what drives this incentive in the other direction. One could imagine that low-cost AI innovation makes it unnecessary to seek monopoly protection in the first place, raising the question of whether AI can "kill" the IP system. To answer this question, we derive the welfare-maximizing IP policy and identify the conditions under which it is optimal to shut down the IP system by setting the intercept and slope to zero. The optimal policy functions as an innovation subsidy by offering potential monopoly profits at the cost of consumer surplus, so the optimal intercept is set to just satisfy the firm's participation constraint. Thus, if AI sufficiently reduces innovation costs such that the firm is willing to invest in innovation without IP protection, then a subsidy is unnecessary and the socially optimal intercept becomes zero. However, our model shows that this does not imply it is optimal to dismantle the IP system altogether. By conditioning IP protection on human-capital use, current IP systems also serve as a Pigouvian-like subsidy for human-capital use in innovation.[4] Therefore, provided that the signal is sufficiently informative,[5] decreasing costs of AI innovation shift the system's role from an innovation subsidy to a human-capital subsidy. This potential shift crucially depends on the social value of human-capital

[4]Given IP protection comes at the cost of monopoly rights, it may be worthwhile to consider alternative policy tools such as direct subsidies to innovation or human-capital use. We discuss such policies after deriving our main results in Section 3.

[5]In an extension, we show that if the signal is imprecise, AI use is hard to detect and the firm may receive protection even if it uses little human capital. It may therefore be optimal to set the slope to zero even when human capital has social value.

use in developing the specific intangible asset, and so warrants careful assessment on a case-by-case basis.

Building on these dynamics, we then introduce a consumer preference for products that commercialize human-developed intangibles (henceforth, human-developed products), which has become central to the debate over AI-generated output. This preference may be particularly prevalent among product categories where consumers connect with a human element of the product. In contrast, consumers may be less likely to express this preference for purely functional innovation, which is more common for intangibles protected by patents. We therefore later explore an extension where consumer preferences may reflect product quality, which can apply to the patent setting if human-developed inventions are of higher quality than AI inventions.

Consumer preferences have a significant impact on the firm's incentives because they lead to a higher willingness to pay, resulting in higher profits — provided that consumers can infer the firm's human-capital use. Throughout, we assume that the IP system parameters are public information, and we separately study the effects of consumers receiving additional information about human-capital use.

Given the imperfect observability of human-capital use, consumer preferences lead to an adverse selection problem: the low-cost firm, which uses more human capital, is pooled with the high-cost firm. For example, imagine consumers trying to distinguish between music produced by a company with access to a network of session musicians, which it draws on when artists require accompaniment, and an otherwise similar company that lacks such a network and therefore uses AI to provide the accompaniment. This adverse selection problem leads to a lower willingness to pay for human-capital use, distorting incentives to use human capital and, at worst, leading to market collapse.

The IP system can mitigate this distortion through two channels. First, a steep slope parameter can trigger a positive feedback loop via consumer beliefs. Increasing the slope leads consumers to revise upward their expectations of human-capital use. Crucially, this disproportionately increases profits when IP protection is granted, since monopoly rights allow the firm to capture more of the consumers' higher willingness to pay. This strengthens incentives to use human capital, leading to another round of belief updating by consumers, further strengthening incentives to use human capital, and so on. Compared to the baseline model, the resulting feedback loop enhances the effect of the policy's slope on the firm's incentive to use human capital. Since this enhancement is stronger for the low-cost firm, the distortion from adverse selection is mitigated.

Second, the IP system can mitigate adverse selection by deterring investment by the high-cost firm, resulting in a separating equilibrium in which consumers correctly infer the low-cost firm's greater human-capital use. Interestingly, this effect can operate through either the slope or the intercept because both affect firm profits. Therefore, unlike in the baseline model, the baseline leniency of the IP policy can affect human-capital use

by deterring the high-cost firm from investing in innovation. While deterring investment is generally undesirable, we show that it may improve welfare if the distortion resulting from adverse selection is severe enough to cause substantial under-use of human capital by the low-cost firm.

Finally, we consider an extension in which consumers receive their own signal about human-capital use. For example, if product quality is correlated with human-capital use, consumers' assessment of quality constitutes a signal of human input. We show that consumers' signals can mitigate the adverse selection discount because the low-cost firm is more likely to generate positive signals, enabling some degree of differentiation from the high-cost firm. Moreover, the firm's desire to generate positive signals provides an additional incentive to use human capital. This incentive effect exists next to the incentive provided by the IP system, which is in itself strengthened by consumers' signals due to increased benefits from monopoly gains.

This paper contributes to the emerging debate surrounding the interaction of IP protection and AI. IP regulators such as the U.S. Copyright Office and the U.S. Patent and Trademark Office have taken an active role in steering this debate, providing guidance on AI use and an overview of pressing economic questions (U.S. Copyright Office, 2025a,b; U.S. Patent & Trademark Office, 2025). These questions predominantly relate to two areas: the inclusion of protected material in AI training data and the protection of AI-generated output. Presently, the debate has predominantly focused on the former (e.g., Peukert et al., 2025; Samuelson, 2024; Wang et al., 2024).[6] One notable example is Gans (2026), who provides a formal analysis of how copyright protection affects AI training, and shows that an ex-post fair-use mechanism can lead to higher welfare compared to traditional copyright protection. In contrast, our model takes the availability and development of generative AI as given and instead focuses on the protection of AI-generated output. Relatively fewer papers speak to this other part of the debate. Yang and Zhang (2025) build a model of content production and AI development to examine how both fair use for training data and the copyrightability of AI output interact and affect welfare, depending on the availability of pre-training data.[7] Different from their approach, we focus on how IP protection affects firms' innovation investment and the choice of AI use. Furthermore, we introduce an asymmetric information friction about AI use. This enables us to analyze the different roles of the IP policy and its effects on firm investment and human-capital use in the presence of an adverse selection problem that endogenously affects firm profits and incentives.

[6]This debate includes many legal battles over the use of copyrighted materials by AI models, such as Concord Music Group, Inc. v. Anthropic PBC (2024), The New York Times Co v. Microsoft Corp (2025), and Getty Images (US), Inc. v. Stability AI, Ltd. (2025).

[7]Other papers providing discussions of the copyrightability of AI outputs include Lemley (2024), Geiger (2024), Cuntz et al. (2024), Peukert and Windisch (2025), de Rassenfosse et al. (2025), and Liang and Lu (2026). de Rassenfosse et al. (2025) provide a discussion of important issues regarding the patentability of AI outputs.

We also contribute to the broader literature on the economics of AI. Studies most closely related to our focus on the implications of AI for innovation include research on AI's role in economic growth (Agrawal et al., 2018; Aghion et al., 2019) and in driving the direction and quality of innovation (Cong et al., 2025; Wu et al., 2025). Babina et al. (2024) empirically examine firm investment and show that firms investing in AI have increased growth as a result of product innovation. While this literature has primarily considered AI as an outcome of innovation, we provide a complementary perspective by studying how firms use AI as an input for the innovation process and explore theoretically how this affects firms' investment in innovation. Another branch of this literature studies the effects of task automation on labor quantities such as wages, employment, or productivity (e.g., Acemoglu and Restrepo, 2018, 2019, 2020; Eloundou et al., 2024; Brynjolfsson et al., 2025; Colliard and Zhao, 2025; Zhong, 2026). We study the implications of AI substituting for humans in the specific task of innovation, although our focus is on firm investment rather than the labor market.

Finally, we contribute to the literature on firms' investment in innovation. Prior research has collected a wide range of factors affecting investment, such as productivity (Kortum, 2004), competition (Aghion et al., 2005), managerial compensation (Manso, 2011), institutional ownership (Aghion et al., 2013), analyst coverage (He and Tian, 2013), firm size (Phillips and Zhdanov, 2013; Akcigit and Kerr, 2018), information frictions (Akcigit and Liu, 2016), innovative capacity (Acemoglu et al., 2018), and financing (Black and Gilson, 1998; Brown et al., 2009; Ferreira et al., 2014; Kerr et al., 2014). Intellectual property rights are central to this literature and have long been recognized as a main driver of innovation investment (Arrow, 1962; Nordhaus, 1969; Henry, 1974; Johnson, 1985; Scotchmer, 1991; Acemoglu and Akcigit, 2012; Crouzet et al., 2022). However, the incentives created by these rights can distort the types, direction, and disclosure of innovations that receive investment (Budish et al., 2015; Hopenhayn and Squintani, 2016, 2021) and have negative effects on subsequent innovation (Williams, 2013; Galasso and Schankerman, 2015). We contribute to this literature by studying how AI affects the incentives for innovation investment. In particular, we highlight the tension in investment incentives created by the productivity gains of AI and the lack of protection granted to AI outputs.

2 Model Setup

The economy consists of a firm, consumers, and an IP officer. The profit-maximizing firm is deciding whether to invest in developing a single new intangible asset and how much human capital relative to AI to use in developing it. The IP officer establishes an IP system and decides whether to grant IP protection based on receiving a signal about the use of human capital. This decision, along with consumer preferences, determine the

payoffs received by the firm.

The sequence of events is as follows. At $t = 0$, the IP officer commits to an IP system.[8] The IP system is defined as a mapping of a noisy signal of the firm's human-capital use, which the officer receives at $t = 2$, to a binary decision $a = \{G, D\}$, where G stands for "Grant" and D stands for "Deny."

At $t = 1$, nature determines the firm's marginal cost of human capital ψ, of which the firm is privately informed. The firm may be a low-cost type ψ_L with probability $p \in (0, 1)$ or a high-cost type ψ_H with probability $1 - p$, where $\psi_H > \psi_L > 0$. The firm then makes a binary decision whether to make the innovation investment to develop a new intangible asset. If it decides to invest, the firm additionally chooses the share of capital that comes from human capital, $h \in [0, 1]$. The remainder, $1 - h$, comes from AI capital. This structure reflects the share of work between human and AI agents in developing the intangible asset. We interpret $h = 1$ as the case where the firm does not use any AI, reflecting the world before advanced AI became available. Conversely, $h = 0$ represents the case where the firm uses the minimum, cost-minimizing amount of human capital necessary to run and supervise an AI system, reflecting the technological frontier of autonomous AI with minimal human input. Importantly, h is not observable by the officer or the consumers.

The investment comes at a cost $C(h, \psi)$, where $C(h, \psi)$ is increasing and convex in h and satisfies $\frac{\partial^2 C}{\partial h \partial \psi} > 0$ and $C(0, \psi) = \overline{\psi}$. These functional-form assumptions reflect AI enabling low-cost innovation compared to human capital; the parameter ψ captures the marginal cost saving from substituting human capital with AI capital, and $\overline{\psi}$ is the lowest cost the firm can achieve by leveraging AI as much as possible.[9] We use the following functional form to obtain closed-form solutions:

$$C(h, \psi) = \overline{\psi} + \frac{\psi}{2} h^2. \tag{1}$$

If the firm does not invest, no market is created and profits are zero. Therefore, the firm chooses to invest if and only if its expected profits are positive.

At $t = 2$, the officer observes the noisy signal $s(h)$ of the firm's human-capital use h. Following the IP system committed to at $t = 0$, the officer decides to grant IP protection

[8] While in reality officers may have some discretion as to how they apply the IP policy, we assume that the officer has commitment. This could be motivated in a dynamic setting where deviating from the policy ex-post reduces the credibility of the policy for future applicants, potentially disincentivizing future investment in innovation.

[9] Online Appendix C.1 considers an alternative cost function that captures more explicitly the idea that a mix of human and AI inputs is cost-minimizing. This is broadly consistent in spirit with complementarity between human and AI inputs at low levels of human-capital use but substitutability at high levels. Alternatively, AI and human capital may be complements if low-cost AI enables an expansion of output that leads to higher demand for human capital (e.g., see Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018). Our model focuses on how AI can substitute for human capital in creating an intangible asset, and we study the role of the IP policy and information frictions in this context.

based on this signal. While this stage most closely resembles an IP officer assessing an application for IP protection, it can also be interpreted as a judge evaluating the validity of claimed protection under an established IP policy in an infringement dispute.[10] In practice, firms are required to document their use of AI in the innovation process, either on their application or in court, constituting one example of a noisy signal of human-capital use.

In our baseline model, we assume that the signal follows a Bernoulli distribution, where the officer receives an "above" signal ($s = A$) with probability h and a "below" signal ($s = B$) with probability $1 - h$.[11] Given this signal structure, we characterize the IP system as the two policy variables $(q, \Delta q)$, which determine the probability of granting IP protection conditional on the signal as

$$\Pr[a = G|s = A] = q + \Delta q, \qquad \Pr[a = G|s = B] = q,$$

with $q \in [0, 1]$ and $\Delta q \in [0, 1-q]$.[12] In this characterization of the IP system, q resembles the baseline leniency of the IP system. A high value of q can be interpreted as high leniency in terms of characteristics orthogonal to AI use. For example, the IP system could set a low bar for non-obviousness or distinctiveness from prior works, irrespective of AI use. The variable Δq captures how strongly the IP system rewards positive signals about human-capital use, reflecting the extent to which the IP system establishes differential policies for human- and AI-generated intellectual property. This characterization enables cleanly isolating the effect of the baseline leniency of the IP system from the marginal effect of positive signals about human capital on obtaining IP protection Δq.

At $t = 3$, the firm enters the production stage, where it decides on the scale of production and profits from commercialization are realized. If the firm is granted IP protection, it operates as a monopolist and earns profits π_G, while consumers earn a surplus u_G. If not, several competitors enter the market by copying the product, the firm earns $\pi_D < \pi_G$, and consumer surplus is $u_D > u_G$. We derive these quantities under a standard Cournot competition model in Online Appendix B.1. To streamline the main text, here we simply impose the following parameter assumptions consistent with the profits and consumer surplus arising endogenously in the Cournot model:

Assumption 1. *Profits and consumer surplus satisfy*

(i) $\pi_G > \pi_D \geq 0$,

[10] For example, in many European countries, copyright protection is assumed when the work is created and only assessed if challenged in court.

[11] In Section 5.1, we explore the alternative assumption that the signal is normally distributed and the IP system is characterized by a threshold such that IP protection is granted for signal values above this threshold.

[12] One interpretation of the uncertainty in the granting decision is that there is a distribution of officers who vary in how strict they are in applying the system's standards (Cockburn et al., 2002; Lemley and Sampat, 2012; Sampat and Williams, 2019; Farre-Mensa et al., 2020).

(ii) $\pi_D + u_D > \pi_G + u_G \geq C(h, \psi_H),\ \forall h \in [0, 1]$.

Online Appendix B.1 captures our main interpretation of the innovation investment in our model, namely the development of a new intangible asset. In Online Appendix B.2, we show that Assumption 1 is also consistent with an alternative version of the micro-foundation in which the investment resembles step-by-step innovation à la Aghion et al. (2001) that enables the firm to gain a competitive edge in an existing product market. Additionally, in Section 4, we allow π_a and u_a to also be a function of h, capturing consumers' greater willingness to pay for human-developed products, separate from its social value. This also enables us to study consumers' inference problem about AI use.

Welfare W is the sum of the firm's payoff and consumer surplus.[13] Additionally, we allow for the possibility that there is a social value to human inputs $v(h) = \gamma h$, which enables us to characterize the welfare-maximizing IP policy as a function of the marginal social value of the firm's human-capital use $\gamma \geq 0$. This social value can be interpreted as capturing positive externalities associated with human involvement in the firm's innovation process, such as the value and dignity of human work and creativity, skill formation, or tacit knowledge spillovers.

If the firm invests, welfare is

$$W_a(h, \psi) = \pi_a + u_a + v(h) - C(h, \psi). \tag{2}$$

If the firm does not invest, then welfare is 0. Under Assumption 1, $W_D(h, \psi) > W_G(h, \psi) \geq 0$. That is, conditional on the firm investing, it would be best not to grant protection so that competition drives down prices, increasing consumer surplus. However, monopoly profits may be necessary for the firm to have an incentive to invest in the first place.

First-Best Allocation. From Assumption 1 and the functional form assumption on $C(h, \psi)$, it immediately follows that, in the baseline model, the first-best allocation is characterized by $a^{fb} = D$ and h^{fb} such that the marginal social benefit of human-capital use equals the marginal cost, $v'(h) = C'(h)$, implying $h^{fb} = \gamma/\psi$.

3 Baseline Model

We now solve the baseline model to understand how the design of the IP system affects firms' incentive to use AI.

[13] Welfare also includes competing firms' profits. For simplicity, these are included in u_D.

3.1 Granting Decision

At $t = 2$, the officer receives a signal $s(h) \in \{A, B\}$ about the use of human capital. The signal and IP system $(q, \Delta q)$ map to the decision of whether to grant IP protection $a \in \{G, D\}$ as follows. The probability of granting protection conditional on observing s is $q + \Delta q$ if $s = A$(bove) and q if $s = B$(elow). This implies that the ex-ante probability of being granted protection, as a function of h, is

$$\Pr[a = G] = q + \Delta q \cdot h. \tag{3}$$

That is, the IP system can be represented as a cumulative distribution function $\Pr[a = G] = q + \Delta q \cdot h$, where q is the "intercept" representing the baseline probability of being granted protection, and Δq is the "slope." There are two notable edge cases: if $q = 1$ and $\Delta q = 0$, then protection is always granted, and if $q = \Delta q = 0$, then it is never granted.

3.2 Firm Decisions

Human-Capital Use. At $t = 1$, given the IP system, the firm chooses h to maximize expected profit

$$\max_h \ \Pi(h, \psi) = \Pr[a = G]\pi_G + (1 - \Pr[a = G])\pi_D - C(h, \psi), \quad s.t.\ h \in [0, 1].$$

For the interior solution $h^* \in [0, 1]$, the first order condition of this problem is

$$\frac{\partial \Pr[a = G]}{\partial h}\Delta\pi = \frac{\partial C(h, \psi)}{\partial h}, \tag{4}$$

with $\Delta\pi \equiv \pi_G - \pi_D$. This condition states that the firm increases human-capital usage until the marginal cost equals the marginal benefit, which is the expected profit premium from obtaining IP protection.

Using the functional form assumptions, the optimal human-capital use is given by

$$h^* = \min\left\{1, \frac{\Delta q \Delta\pi}{\psi}\right\}. \tag{5}$$

Investment Decision. The firm decides to invest in innovation if and only if $\Pi(h^*, \psi) \geq 0$. This highlights that the prospect of earning monopoly profits may be necessary for the firm to invest.[14]

Lemma 1. *A firm with cost type ψ invests in innovation if and only if its expected profit*

[14] In our model, the innovation is guaranteed to be successfully created if the firm chooses to invest. Introducing risky innovation only affects the level of expected profits, so we omit innovation risk to simplify our analysis.

$\Pi(h^*, \psi)$ *is non-negative. Given an IP policy* $(q, \Delta q)$*, the participation constraint is*

$$\pi_D + q\Delta\pi + \frac{(\Delta q)^2(\Delta\pi)^2}{2\psi} \geq \overline{\psi}. \tag{6}$$

Proof. See Appendix A.1. □

3.3 Results

3.3.1 The Effect of the IP System on AI Use

Eq. (5) shows that h^* is affected by the slope of the policy, Δq, but not the intercept, q. The two quantities that determine the IP policy, q and Δq, therefore separately govern the firm's two choices: Δq determines human-capital use, allowing q to satisfy the participation constraint. Importantly, this separability no longer holds when we introduce a consumer preference for human-developed products in Section 4 due to the effect of q on the consumer's expectation of the firm's human-capital use.

From Eq. (5), we can also evaluate how the leniency of the IP policy affects the firm's incentive to use human capital versus AI in the innovation process. The overall leniency of the policy is represented by $\Pr[a = G] = q + \Delta q \cdot h$. This implies that an increase in Δq is associated with a more lenient IP system, in that protection is granted more often. Moreover, a higher Δq creates greater incentives to use human capital, implying a positive association between IP-system leniency and human-capital use. By contrast, if the officer always or never grants IP protection, then $\Delta q = 0$, and the firm chooses the corner solution $h^* = 0$ despite polar opposite leniency in these two cases. This implies that the effect of leniency is ambiguous.

In Section 5.1, we use a continuous-signal setting to demonstrate that this result generalizes to leniency specifically with respect to human-capital use. While the binary signal is designed to pin down this quantity as the constant Δq, the continuous-signal setting allows us to introduce a threshold for minimum required human-capital use and a non-monotonic policy slope. We show that increases to this threshold can either increase *or* decrease the use of human capital, depending on how it affects the slope of the policy. Thus, when assessing the effect of the IP system on human-capital use, it is important to assess the marginal probability of granting protection (i.e., $\frac{\partial \Pr[a=G]}{\partial h}$) rather than relying on heuristics such as leniency.

3.3.2 Killing AI

Based on the previous discussion, the IP system can fully disincentivize AI use if it strongly rewards positive signals about human-capital use.

Proposition 1. *If*

$$\Delta q \Delta \pi \geq \max\left\{\psi, \sqrt{2\psi(\overline{\psi} - \pi_D - q\Delta\pi)}\right\},$$

then a firm of cost type ψ invests in innovation and chooses $h^* = 1$, thus "killing" AI.

Proof. See Appendix A.2. □

In addition to the slope Δq, the profit premium $\Delta\pi$ also plays a significant role in driving Proposition 1, and the profit premium is itself significantly driven by the entry of competitors should the firm not receive IP protection. However, the extent of competitive entry is highly dependent on the characteristics of the product that results from commercializing the intangible asset. For example, copyrightable works may be easy to copy and therefore face many market entrants when not granted copyright, whereas patentable inventions may not be disclosed when the patent is rejected and so face relatively few market entrants. Even within IP types, some products, such as digital music, may be easily copied by competitors, while others, such as brand-specific advertising campaigns, may have a high degree of natural excludability. Therefore, the potential for the IP system to kill AI can vary significantly across products.

3.3.3 Killing the IP System

We next analyze whether the use of AI in innovation has the potential to make the IP system obsolete. In doing so, we first derive the optimal IP policy conditional on both firm types investing in innovation, which is characterized in the following proposition.[15]

Proposition 2. *Conditional on inducing both firm types to invest in innovation, the welfare-maximizing IP policy is*

$$q^* = \frac{\overline{\psi} - \pi_D}{\Delta\pi} - \frac{(\Delta q^*)^2 \Delta\pi}{2\psi_H},$$

$$\Delta q^* = \frac{\gamma\left(\frac{p}{\psi_L} + \frac{1-p}{\psi_H}\right)}{-\left((\Delta\pi + 2\Delta u)\left(\frac{p}{\psi_L} + \frac{1-p}{\psi_H}\right) - \frac{\Delta\pi + \Delta u}{\psi_H}\right)},$$

with $\Delta u \equiv u_G - u_D$, and $\Delta q^* \geq 0$ under Assumption 1.

Proof. See Appendix A.3. □

[15] In Appendix A.3, we provide the complete characterization of the optimal IP policy, including the condition under which it is optimal to induce investment by both firm types as opposed to only the low-cost type. In Proposition 2, we focus on the case where both firm types invest. This allows us to show the impact of AI on the optimal IP policy without complicating the analysis with different investment scenarios. We return to the implications of inducing investment by only the low-cost type in Section 4, where consumer preferences make the analysis of firm investment particularly relevant.

In our model, and consistent with prior literature (e.g., Gilbert and Shapiro, 1990; Hopenhayn et al., 2006; Acemoglu and Akcigit, 2012), one important role of the optimal IP policy is to incentivize investment in innovation by providing a subsidy via monopoly commercialization rights. The officer determines this subsidy by setting q^*. A natural intuition may therefore be that, if AI sufficiently drives down the cost of innovation, then it is no longer necessary to grant monopoly rights. In such a case, the IP system's role as an innovation subsidy becomes obsolete. Consistent with this intuition, Proposition 2 shows that the optimal interior baseline probability of granting IP protection, q^*, decreases as the fixed cost of AI innovation, $\overline{\psi}$, decreases.[16] Consequently, the corner solution $q^* = 0$ applies if

$$\overline{\psi} \leq \pi_D + \frac{(\Delta q^* \Delta\pi)^2}{2\psi_H}.$$

However, the same technology that reduces the fixed cost of implementing AI may also reduce the cost of copying the product if the underlying intangible asset did not receive protection. As more firms enter the competitive market, profits are driven downwards, increasing the relative monopoly gains $\Delta\pi$, and raising the need for the IP system to fulfill its role as an innovation subsidy.[17] Reflecting this tension, the above inequality shows that the corner solution $q^* = 0$ applies with smaller values of $\overline{\psi}$ as π_D decreases. Thus, the net effect of technological improvements on the system's role as an innovation subsidy depends on the relative rates at which innovation costs and entry costs are driven down.

Even if innovation costs decrease at a faster rate than entry costs, though, it still may not be optimal to completely dismantle the IP system. This is because the other important role of the optimal IP policy is to incentivize the socially optimal level of human-capital use in innovation. By conditioning the granting of IP protection on positive signals of human-capital use, the IP policy effectively provides a human-capital subsidy, which the officer determines by setting Δq^*. Therefore, we refer to this second role of the IP system as a "human-capital subsidy," analogous to its role as an innovation subsidy. Since human-capital use beyond the cost-minimizing level (represented by $h = 0$) does not provide any benefits to the firm in the baseline model, the officer's choice to incentivize human-capital use is driven by its social value parameter γ. Consistent with this intuition, Proposition 2 shows that Δq^* increases in the social benefit γ.

[16] For example, as AI lowers the cost of producing innovation of a given quality, the IP system may respond by raising the non-obviousness threshold, thereby reducing q^*. These standards could also be applied differentially across AI- and human-developed intangibles, implying $\Delta q^* > 0$. One such proposal is developed in Liang and Lu (2026), who argue that infringement for AI-generated output should depend on whether the output could have been generated without the inclusion of the copyrighted material in the model's training data.

[17] In the microfoundation using Cournot competition in Online Appendix B.1, we model this effect explicitly by allowing endogenous entry such that the number of firms N^* entering the market is a function of entry costs, which are in turn a function of the same technological advancements that may reduce $\overline{\psi}$.

Combining the effects of AI on both roles of the IP system leads to the following corollary to Proposition 2 characterizing the conditions under which the IP system becomes obsolete.

Corollary 1. *If $\pi_D \geq \overline{\psi}$ and $\gamma > 0$, then $q^* = 0$ but $\Delta q^* > 0$. Thus, the IP system shifts its role to a Pigouvian-like subsidy for human-capital use in innovation. However, if $\pi_D \geq \overline{\psi}$ and $\gamma = 0$, then $q^* = \Delta q^* = 0$ and AI "kills" the IP system.*

Proof. See Appendix A.4. □

Corollary 1 shows that low-cost AI is not a sufficient condition for making the entire IP system obsolete. While low-cost AI implies it is optimal to minimize the expected loss in consumer welfare due to monopoly rights by setting $q^* = 0$ (eliminating the system's role as an innovation subsidy), the optimal IP policy is nevertheless characterized by $\Delta q^* > 0$ as long as there is social value to human involvement in the innovation process, $\gamma > 0$. This implies that, rather than "killing" the IP system, sufficiently low-cost AI shifts emphasis within the IP system from its role as an innovation subsidy to its role as a human-capital subsidy. Due to the social value of human capital in innovation, human-capital use constitutes a positive externality, motivating a Pigouvian-like corrective subsidy for human-capital use in the innovation process.

Recall that γ (or $v(h)$) captures only the externalities of human involvement; the effect of human involvement on consumer preferences, consumer surplus, and firm profits are considered separately in Section 4. As we show there, consumer preferences give rise to another reason for setting a positive Δq^*, even when $\gamma = 0$ (see Corollary 2).

The case $\gamma > 0$ applies if, for example, society attaches intrinsic value to preserving human creativity and originality, or when human engagement fosters skill formation or sustains innovation ecosystems that depend on tacit knowledge passed between people. In other words, even when AI can deliver outputs cheaply, the IP system is still socially valuable if human involvement generates positive externalities. By contrast, $\gamma = 0$ describes settings with no such value, and so low-cost AI "kills" the IP system in that it becomes optimal to dismantle it entirely. This holds when outputs are purely functional, such as routine data summaries or standardized reports, or when AI-generated translations, generic images, and background music are viewed as perfect substitutes for human-created counterparts, leaving no reason to subsidize human participation. More broadly, whenever creative labor generates no meaningful positive externalities, IP protection becomes obsolete in the baseline model without consumer preferences.

The interpretation and value of γ therefore differs across different types of IP protection. As we discuss in more detail in Section 5.4, this implies our model's implications may differ for the patent system compared to the copyright system.[18]

[18] In Section 5.4, we also discuss the patent system's role in incentivizing the disclosure of techni-

Signal Noise. One caveat to the result in Corollary 1 is that the observability of AI use may vary across different types of intangibles. To capture this notion, in Section 5.2 we present an extension to study the effect of signal noise. This extends our result in Corollary 1 by showing that it may be optimal to "kill" the IP system by setting $q^* = \Delta q^* = 0$ if the signal is sufficiently noisy, despite a positive social value of human capital $\gamma > 0$. Intuitively, if the signal is imprecise, the firm may receive protection even if it uses little human capital. This creates the social cost of monopoly rights without effectively stimulating human-capital use. Just as a lack of social value ($\gamma = 0$) eliminates the need to incentivize human capital, a lack of signal informativeness can eliminate the IP system's ability to provide such incentive efficiently.

Dynamic AI Training. One consideration absent from our analysis is the dynamic interactions among the IP policy, firms' human-capital use, and future AI costs. For example, the use of human capital in innovation today can have significant implications for the cost of using AI in innovation in the future via its effects on data used to train AI models. Similarly, today's granting decisions can affect the availability of data for AI training. While a full investigation of these interactions is beyond the scope of our paper, we briefly sketch their implications for incentives in Online Appendix C.4. Specifically, we embed our model in a two-period economy corresponding to two separate innovation rounds. The IP officer sets a single IP policy for both periods at the beginning of the first period. Importantly, the second-period cost $\overline{\psi}_2$ is a function of both human-capital use and the IP decision in the first round. Therefore, the optimal IP policy slope differs between the dynamic and static settings, with the officer setting a larger slope in the dynamic setting if the net effect of the previously mentioned interactions results in the slope decreasing future AI costs. However, the sign of this net effect is still an open question and further empirical evidence is needed to assess the relative strength of these channels. Nonetheless, Online Appendix C.4 shows that the incentives derived from our baseline analysis provide the foundation for decisions even in a dynamic setting.

Direct Subsidies. When considering the role of the IP system in granting both innovation and human-capital subsidies, one might consider an alternative system in which direct subsidies are granted rather than monopoly commercialization rights. While this would circumvent the loss in consumer surplus due to monopoly rights, it would require the regulator, rather than the market, to determine the appropriate level of subsidies. Still, if monopoly profits (π_G) and competitive profits (π_D) could be perfectly forecast, then direct subsidies could exactly satisfy the firm's participation constraint and incentivize human-capital use while always establishing a competitive market. However,

cal information that facilitates follow-on innovation, which may prevent the patent system from being "killed."

forecasting the market for a new product is often difficult and has therefore been something IP systems have explicitly avoided (Bleistein v. Donaldson Lithographing Co., 1903; U.S. Copyright Office, 2025b). Given the considerable uncertainty in forecasting profits, granting monopoly commercialization rights gives the firm a residual claim on the market it creates, aligning incentives for producing valuable innovation and using the appropriate level of human capital.[19] Thus, although we do not include profit-forecast uncertainty in our model, we focus on an IP system that provides subsidies through the granting of monopoly commercialization rights.

4 Consumer Preference for Human Capital

In the baseline model, firm profits and consumer surplus are independent of the firm's human-capital use. However, consumers may prefer products that commercialize human-developed intangibles (i.e., human-developed products). This preference is most salient for IP where consumers connect with a human creator, rather than purely functional innovations. For example, experimental and survey evidence show that participants like various products less once they learn it was generated by AI, independent of product quality, indicating an inherent preference for human-created output. This has been documented for art (Horton et al., 2023), music (Tubadji et al., 2021), books (Manewitsch and Kalb, 2025), newspaper articles (Kennedy and Yam, 2025), and personal advice (Osborne and Bailey, 2025).[20]

In this section, we introduce consumer preference for human-developed products to our model. This enables us to study consumers' inference problem for AI use and its interaction with the IP system in the presence of adverse selection. Initially, we conceptualize this preference as an intrinsic taste independent of product quality. It is therefore more likely to apply to works that are protected by copyright, such as the examples listed above. Nonetheless, in Section 5.3, we modify the consumer's information set such that this preference can be interpreted as a preference for product quality. This interpretation may therefore be more relevant for patent-protected inventions, provided that human-developed inventions are of higher quality than AI inventions. We discuss the different implications for copyright- and patent-protected innovations in more detail in Section 5.4.

4.1 New Model Elements

The preference for human-developed products is captured by taking consumer surplus, u_a, as an increasing function of human capital h, i.e., $u_a(h)$. Consequently, the firm's

[19] Similar arguments are supported, both empirically and theoretically, in the literature on executive compensation (e.g., Holmström, 1979; Garen, 1994; Coles et al., 2006).

[20] Also broadly consistent with this notion, experimental evidence documents an aversion to algorithms (Dietvorst et al., 2015; Castelo et al., 2019; Dargnies et al., 2024).

profits π_a also increase with h, i.e., $\pi'_a(h) > 0$, as a higher h increases consumer willingness to pay. We assume $\pi'_G(h) > \pi'_D(h)$, which means that the firm has a larger marginal benefit from favorable consumer beliefs when IP protection is granted. Economically, this assumption reflects that monopoly rights enable firms to extract a greater share of consumers' willingness to pay for human-developed products. Online Appendix B.1 provides a microfoundation for this assumption under Cournot competition with a representative consumer who has a preference for human capital h.[21]

Consumer Beliefs and Information Set. Since the firm's human-capital use is unobservable, the consumer forms conjectures $\hat{h}_L$ and $\hat{h}_H$ regarding the firm's human-capital use and forms a posterior belief $\mathbb{E}[h] = p\hat{h}_L + (1-p)\hat{h}_H$. For now, we assume that the consumer does not receive an independent signal, such that the information set only consists of the IP system variables $(q, \Delta q)$. In Section 5.3, we relax this assumption by allowing the consumer to observe an independent signal of h, which can, for example, be interpreted as the product's observable quality being correlated with human-capital use (alternatively, the consumer may observe the granting decision $a \in \{G, D\}$). This extension produces the same main insights from this section along with additional results.

Human-Capital Use. A firm of type ψ chooses its human-capital use $h \in [0, 1]$ to maximize expected profit $\Pi(h, \psi)$, taking the consumer's posterior belief $\mathbb{E}[h]$ and the officer's policy $(q, \Delta q)$ as given. The firm's expected profit is

$$\Pi(h, \psi) = \Pr[a = G]\pi_G(\mathbb{E}[h]) + \Pr[a = D]\pi_D(\mathbb{E}[h]) - C(h, \psi). \tag{7}$$

A rational expectations equilibrium requires the conjectures to be aligned with actions, i.e., $\hat{h} = h^*$. Solving the firm's maximization problem, the equilibrium level of human-capital use h^* is

$$h^* = \min\left\{1, \frac{\Delta q \Delta\pi(\mathbb{E}[h^*])}{\psi}\right\}, \tag{8}$$

with $\Delta\pi(\mathbb{E}[h^*]) \equiv \pi_G(\mathbb{E}[h^*]) - \pi_D(\mathbb{E}[h^*])$ and where $\mathbb{E}[h^*]$ is formed using h^*_L and h^*_H. Compared to the human-capital use in the baseline model in Eq. (5), consumer preferences create a demand-side benefit for human-capital use h^*. This benefit operates entirely through the consumer beliefs $\mathbb{E}[h^*]$. Given the assumption $\pi'_G(h) > \pi'_D(h) > 0$, the profit premium from IP protection widens as consumer beliefs improve, $\Delta\pi'(\mathbb{E}[h^*]) > 0$.

[21] We also show that allowing for endogenous entry fixes π_D such that $\pi'_D = 0$, amplifying the difference between $\pi'_G(h)$ and $\pi'_D(h)$.

4.2 Results

4.2.1 Adverse Selection Problem

In the presence of consumer preferences for human-developed products, asymmetric information about the firm's cost type $\psi \in \{\psi_L, \psi_H\}$ generates an adverse selection problem in the product market. The consumer cannot distinguish between the low-cost firm, which optimally uses more human capital, and the high-cost firm, which uses less. Consequently, the rational consumer forms a pooled equilibrium belief $\mathbb{E}[h^*]$ about the expected level of human-capital use that determines the willingness to pay. This pooled belief benefits the high-cost firm, whose actual use is lower than the belief ($h_H^* < \mathbb{E}[h^*]$), while penalizing the low-cost firm, whose actual use exceeds it ($h_L^* > \mathbb{E}[h^*]$).

This adverse selection problem affects the firm's profit π_a, which in turn has two effects. First, it distorts incentives for how much human capital each firm uses:

Lemma 2. *Define* $h^{**} = \min\left\{1, \frac{\Delta q \Delta\pi(h^{**})}{\psi}\right\}$ *as the firm's human-capital use if consumers could perfectly observe the firm's type. In a pooling equilibrium in which both firms invest, the low-cost firm's human-capital use is below this benchmark and the high-cost firm's human-capital use is above this benchmark:* $h_L^* < h_L^{**}$ *and* $h_H^* > h_H^{**}$.

Proof. See Appendix A.5. □

Second, the adverse selection problem can lead to inefficient innovation investment decisions. Given a consumer belief $\mathbb{E}[h^*]$ and policy $(q, \Delta q)$, a firm of type ψ invests if $\Pi^*(h^*, \psi) \geq 0$. This condition implies that investment only occurs if $\mathbb{E}[h^*]$ is sufficiently large. The adverse selection problem can therefore potentially lead to the low-cost firm inefficiently forgoing innovation investments when it is pooled with the high-cost firm.

We next analyze how the IP system can mitigate these effects. Working backwards through the model timeline, we first assume that both firm types invest and show how the policy slope Δq can set off a positive feedback loop in consumer beliefs that reduces the distortion in human-capital use. We then study how the IP system $(q, \Delta q)$ can resolve the adverse-selection problem by deterring the high-cost firm from investing.

4.2.2 Feedback Loop

In order to characterize how the IP policy affects the distortion of human-capital use, we first derive the effect of the policy slope Δq on human-capital use conditional on both types investing. In the presence of consumer preferences, the slope has two such effects. For an interior solution $h^* \in (0, 1)$, we can decompose the marginal effect of Δq as follows:

$$\frac{dh^*}{d\Delta q} = \underbrace{\frac{1}{\psi}\Delta\pi(\mathbb{E}[h^*])}_{\text{Direct incentive}} + \underbrace{\frac{\Delta q}{\psi}\frac{d\Delta\pi(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}}_{\text{Indirect feedback loop}}. \tag{9}$$

The first term shows that Δq provides a direct IP system-based incentive by raising the expected profit premium associated with obtaining IP protection ($\Delta\pi(\mathbb{E}[h^*])$). This incentive exists regardless of consumer preferences for human-capital use, mirroring the effect of the baseline model, where $dh^*/d\Delta q = \Delta\pi/\psi$ (see Eq. (5)).

The second term is unique to this setting due to consumer preferences for human-developed products. It captures a feedback loop initiated by the policy's role as a credible signal to the consumer. Because the IP policy is observable, the slope Δq signals the anticipated equilibrium level of human-capital use. A higher Δq signals stronger firm incentives, leading the consumer to form higher beliefs about the firm's human-capital use (i.e., $\frac{d\,\mathbb{E}[h^*]}{d\Delta q} > 0$). Since the profit premium from IP protection widens as consumer beliefs improve, this increase in $\mathbb{E}[h^*]$ raises the firm's expected profit premium associated with IP protection (i.e., $\Delta\pi'(\mathbb{E}[h^*]) > 0$). This larger profit premium then feeds back into the firm's incentives and amplifies the marginal benefit of investing in h, which in turn increases consumer beliefs. That is, consumer preferences generate a feedback loop: an increase in Δq increases consumers' beliefs about h, leading to a higher profit premium $\Delta\pi$, which in turn leads to a higher human-capital use h, further increasing consumers' beliefs about h, and so on.

It is important to note that this effect of Δq operates even if the consumer does not observe the officer's final decision $a \in \{G, D\}$. Instead, it suffices that the consumer knows that the IP policy rewards positive signals about human-capital use; this leads to an upward revision in the consumer's belief about h, setting off the feedback loop. Economically, this feedback loop implies that an IP policy that is more sensitive to human-capital inputs (i.e., a higher Δq) induces the consumer to believe that firms are using more human capital, while a policy that is less responsive to human-capital inputs (i.e., lower Δq) signals lower human-capital use. For example, China's copyright policy grants protection to AI-generated works when a human has carefully engineered the prompt provided to the model. In contrast, U.S. copyright policy views such works as having insufficient human input, and so does not grant protection. As a result, when assessing a particular work, consumers may be more likely to infer that AI was used in creating the work if it was created in China as opposed to the U.S.

The following proposition summarizes how the feedback loop amplifies the effect of the policy slope on human-capital use:

Proposition 3. *Consumer preferences for human-developed products give rise to a feedback loop that amplifies the effect of Δq on the firm's human-capital use as long as $\pi'_G(h) > \pi'_D(h)$. That is, $\frac{dh^*}{d\Delta q}$ is greater compared to a counterfactual in which $\pi'_G(h) = \pi'_D(h)$. This amplification effect is stronger for the low-cost firm's human-capital use.*

Proof. See Appendix A.6. □

Proposition 3 has important implications for the distortion of human-capital use due

to adverse selection. Recall that Lemma 2 shows that adverse selection causes the low-cost firm to under-utilize human capital while the high-cost firm over-utilizes it. Because both effects in Eq. (9) are scaled by ψ, the effect of Δq on h^* is stronger for the low-cost firm. This implies that as Δq increases, it increases h^* for the low-cost firm more than for the high-cost firm, which results in a net reduction in distortion. A larger Δq therefore mitigates the distortion in human-capital use. Moreover, since the feedback loop amplifies the effect of Δq on h^*, it further contributes to this mitigation.

Implications for the Optimal IP Policy. In Online Appendix B.3, we re-derive the optimal IP policy and find that the presence of consumer preferences drives up the use of Δq, independent of the social value of human capital γ. The reason is that the firm's optimal human-capital use does not internalize the positive effect of higher beliefs on consumer surplus and profits (the latter because the firm takes the consumer's belief as given). Consequently, the result in Corollary 1 ("AI kills the IP system") is altered by consumer preferences. Even if the social value of human capital γ is zero, the IP system may remain valuable in providing a credible signal to the consumer.

Corollary 2. *In the presence of consumer preferences for human-developed products, it may be optimal to maintain a positive slope $\Delta q^* > 0$ even if the social value of human capital γ is zero.*

Proof. See Appendix A.7. □

4.2.3 Investment Decision Under Adverse Selection

The IP policy can also shape and potentially mitigate the adverse selection problem by affecting which firm decides to invest, thereby determining whether a pooling or separating equilibrium emerges. Recall that the base grant probability q acts as an innovation subsidy, raising the expected payoff Π for both types without affecting their marginal incentive to invest in h^*. The officer can therefore use q to counteract the profit reduction caused by low beliefs $\mathbb{E}[h^*]$. Because the expected profit of a low-cost firm $\Pi^*(\psi_L)$ is weakly higher than that of the high-cost firm $\Pi^*(\psi_H)$, the low-cost firm invests at a lower threshold q than the high-cost firm. The following proposition characterizes how q shapes the equilibrium structure.[22]

Proposition 4. *For a fixed Δq, there exist thresholds q_L, q_H, and q_B that satisfy the strict ordering $q_L < q_H < q_B$. These thresholds are defined in Appendix A.8 and characterize the unique Pareto-dominant equilibrium as follows:*

[22] To streamline the analysis, we focus on the case where the base grant probability q required to induce investment under the most pessimistic belief ($\mathbb{E}[h] = 0$) is strictly higher than that required to sustain the equilibrium where both types invest. This limits the number of cases we need to consider, see Appendix A.8.

(i) If $q < q_L$, no investment occurs.

(ii) If $q_L \leq q < q_H$, only the low-cost firm invests.

(iii) If $q_H \leq q < q_B$, no investment occurs.

(iv) If $q \geq q_B$, both types of firm invest.

Proof. See Appendix A.8. □

The interval $q \in [q_H, q_B)$ presents a region of market breakdown caused by the adverse selection problem. In this region, the base grant probability q is sufficiently high to attract the high-cost firm to invest if the consumer holds the favorable belief h_L^*, which then disrupts the equilibrium where only the low-cost firm invests. However, once consumer beliefs adjust downward to $\mathbb{E}[h^*]$ in response to this potential investment by the high-cost firm, the base grant probability q becomes insufficient to make investment profitable for the high-cost type. Thus, the potential investment by the high-cost type drives down consumer beliefs sufficiently to cause the market to collapse.

An important policy insight is that the officer can resolve this failure by strategically decreasing q to make the IP system stricter. By setting q below q_H, the officer effectively screens out the high-cost firm. This exclusion allows the consumer to maintain the favorable belief h_L^*, restoring the investment incentives for the low-cost firm. The following proposition formalizes this insight by comparing expected welfare in the equilibrium with $q = q_L$ versus the equilibrium with $q = q_B$, which under Assumption 1 represent the welfare-maximizing policies for inducing participation from the low-cost type alone and both types, respectively.

Proposition 5. *Let the difference in the marginal cost be defined as $\Delta\psi \equiv \psi_H - \psi_L$. For a fixed Δq, there exists a unique cutoff $\widetilde{\Delta\psi}$ such that for all $\Delta\psi > \widetilde{\Delta\psi}$, the equilibrium with $q = q_L$, in which only the low-cost type invests, yields strictly higher expected welfare than the equilibrium with $q = q_B$, in which both types invest.*

Proof. See Appendix A.9. □

While Assumption 1 implies that it is in itself undesirable to prevent the high-cost firm from investing, the value of its innovation investment needs to be weighed against the cost of the adverse selection discount it imposes on the low-cost firm. When the cost difference $\Delta\psi$ is small, the high-cost firm is relatively efficient in using human capital. In this case, the welfare loss from its lower human-capital use is outweighed by the value of its investment in innovation, making the equilibrium in which both types invest desirable. However, as $\Delta\psi$ increases, the high-cost firm becomes increasingly inefficient. The officer must offer an increasingly large subsidy q_B to sustain its investment, which generates welfare loss due to monopoly rights. In addition, the inclusion of the high-cost firm

degrades consumer beliefs, lowering the incentives for the low-cost firm to use human capital. Once the cost asymmetry exceeds the cutoff $\widetilde{\Delta\psi}$, these costs dominate, and it becomes socially optimal to screen out the high-cost firm entirely to prevent the adverse selection discount for the low-cost type.

This result highlights a novel function of the base grant probability q that is absent in our baseline model. In the baseline analysis without consumer preferences, the base grant probability q acts solely as an innovation subsidy but has no impact on how much human capital h^* the firm uses once it has invested. By contrast, in the presence of consumer preferences, tightening the base grant probability (lowering q below q_B) screens out the high-cost firm from investing and resolves the adverse selection problem. This leads the consumer to update their belief from the average $\mathbb{E}[h^*]$ to the higher separating belief h_L^*, resulting in a higher willingness to pay and profit premium $\Delta\pi$. Consequently, the low-cost firm responds to this belief update by increasing its human-capital use h_L^*. Thus, consumer preferences create a link between the base grant probability q and the firm's human-capital use.

The Role of Δq. Proposition 4 shows how the intercept q shapes the equilibrium investment decision for a given slope Δq. Online Appendix B.4.1 complements this analysis by deriving how Δq affects the thresholds in Proposition 4. Two insights emerge. First, an increase in Δq benefits firm profits and relaxes the participation constraints, lowering the thresholds q_L, q_H, and q_B. Second, this effect is differential across types. Because Δq rewards human capital at the margin and the low-cost firm uses more human capital, a higher Δq disproportionately slackens the low-cost firm's participation constraint. This widens the window of q over which only the low-cost firm invests, i.e., widening the $q \in [q_L, q_H)$ region. This demonstrates that Δq has a screening role complementary to its role to incentivize human-capital use highlighted in Eq. (8).

Moreover, in Online Appendix B.4.2, we show that this screening role of Δq is particularly important if we relax the assumption that the AI cost floor $\overline{\psi}$ is uniform across the two types. Specifically, if the firm with high human-capital cost (ψ_H) has a lower AI cost floor $\overline{\psi}$, then it could be the more profitable firm. If it is, then the participation constraint for the firm with ψ_H becomes the easier of the two to satisfy, and lowering q can result in screening out the firm with low human-capital cost (ψ_L) instead. By contrast, Δq retains its ability to screen out the firm with ψ_H because it disproportionately benefits the profits of the firm with ψ_L. In addition, a high Δq can reinstate the regime where the firm with ψ_L is more profitable, recovering the ability for q to screen out the firm with ψ_H. In this case, a high Δq may also become a screening prerequisite if the IP officer prefers the equilibrium in which only the firm with ψ_L invests.

5 Model Extensions and Discussion

5.1 Continuous Signal

In our baseline model, the officer receives a binary signal about whether the use of human capital exceeds the threshold for IP protection, which is implicitly set by q and Δq. While this setting provides tractability and cleanly separates the effects of the "level" and "slope" of the IP policy, it is more difficult to directly map into a threshold of human-capital use.

To capture this mapping, we consider a version of our model where the officer sets a human-capital-use threshold for IP protection, $\bar{h} \in [0, 1]$, and observes a continuous signal $s = h + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$, about the firm's use of human capital. The officer then forms a rational expectation about the firm's human-capital use. In Online Appendix C.3, we show that this setup reduces the granting decision to a threshold rule on s, where the officer grants IP protection if the signal exceeds a unique threshold $\bar{s}$, which is an increasing function of $\bar{h}$. Therefore, as in the baseline model, the IP policy is defined as a cumulative distribution function (CDF) that is a function of human-capital use (cf. Eq. (3)). In the continuous setting, this means the policy follows the CDF of a normal distribution and the slope of the policy is given by the probability density function (PDF) of a normal distribution.

We then explore the firm's optimal human-capital use in this setting. The general formula for the firm's first order condition for human-capital use given in Eq. (4) in the baseline model continues to hold. That is, the effect of the IP policy on human-capital use is entirely driven by the slope of the policy, i.e., the marginal probability of being granted IP protection (PDF). This implies that the effect of the threshold on optimal human-capital use depends on how changes in $\bar{h}$ affect the PDF. In Online Appendix C.3, we show that this effect is ambiguous, with $\frac{dh^*}{d\bar{h}} \geq 0$ if and only if $\bar{s} \leq h^*$. Thus, leniency with respect to human-capital use can either increase *or* decrease human-capital use, and so is not a sufficient heuristic for how the IP policy affects firms' incentives to use human capital.

5.2 Signal Noise

In the baseline model, the noisiness of the signal is fixed and higher human-capital use h translates one-for-one to a higher probability of the officer receiving a positive signal. However, in reality, signal noise depends on how difficult it is to detect AI use and may therefore vary across types of intangibles and over time. To study how noise affects human-capital use and the two roles of the IP system, we introduce a signal informativeness parameter $\rho \in [0, 1]$.

Let the probability of observing a positive signal $s = A$ given human capital h be

defined as a weighted average of the true human-capital use and random noise:

$$\Pr[s = A|h] = \rho h + (1 - \rho)\frac{1}{2}.$$

This specification nests our baseline model and a pure noise case. Specifically, setting $\rho = 1$ recovers the baseline where the probability of a positive signal equals h. Setting $\rho = 0$ represents an uninformative pure-noise case where the probability of a positive signal is $\frac{1}{2}$ and is uncorrelated with h. The probability of obtaining IP protection, which is given by Eq. (3) in the baseline model, is now given by:

$$\Pr[a = G] = q + \Delta q \left(\rho h + \frac{1 - \rho}{2}\right).$$

The Effect of Signal Noise on Human-Capital Use. The firm chooses h to maximize expected profits, resulting in the optimal human-capital use

$$h^* = \frac{\rho \Delta q \Delta \pi}{\psi}. \tag{10}$$

As the signal becomes noisier ($\rho \to 0$), the firm's optimal human-capital use h^* approaches zero for any fixed slope Δq. This effect captures the dilution of incentives, as h translates less effectively into a positive signal $s = A$. This highlights that the effect of noise in the granting decision on the optimal human-capital use only depends on how noise changes the slope of the IP system, $\frac{\partial \Pr[a=G]}{\partial h}$, which is given by $\rho \Delta q$ in this model.

5.2.1 Killing the IP System With Noise

Corollary 1 establishes that when the social benefit of human capital, γ (or $v(h)$), is zero, the slope Δq of the IP system is "killed." The following corollary shows that noise in the signal can lead to the same result, potentially killing the IP system even when $\gamma > 0$.

Corollary 3. *Let $\Lambda \equiv -(\Delta\pi + \Delta u) > 0$ be the welfare loss due to monopoly rights. There exists a threshold $\bar{\rho} = \frac{\Lambda}{2\gamma\Delta\pi\mathbb{E}[1/\psi]+\Lambda}$ such that, if signal informativeness falls below this level, i.e., $\rho \leq \bar{\rho}$, then it is welfare-maximizing to set $\Delta q^* = 0$.*

Proof. See Appendix A.10. □

This result reinforces the insight from Corollary 1 regarding the slope Δq. Just as a lack of social value ($\gamma = 0$) eliminates the need to incentivize human capital, a lack of information ($\rho \leq \bar{\rho}$) eliminates the IP system's ability to provide such an incentive efficiently. In both cases, the optimal IP policy is to set the slope to zero ($\Delta q^* = 0$).

To understand this result, it is instructive to examine the first-order condition of

expected welfare with respect to Δq:

$$\frac{d\mathbb{E}[W]}{d\Delta q} = \gamma\rho\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right] - \rho^2\Delta\pi\Big(- (\Delta\pi + 2\Delta u)\Big)\mathbb{E}\left[\frac{1}{\psi}\right]\Delta q - \frac{\Lambda(1-\rho)}{2} = 0. \tag{11}$$

This condition characterizes the trade-off between the marginal social benefit and the marginal social cost of increasing the slope Δq. The first term represents the linear marginal benefit of human-capital use. The second term captures the marginal cost of human-capital use (consisting of the direct cost $C(h, \psi)$ and the indirect cost of monopoly rights caused by increases in h). These first two terms mirror the baseline model, provided that the firm's human-capital choice h^* is scaled by the signal informativeness ρ. The third term is unique to this extension. It represents the linear marginal cost caused by the additional noise of the signal s, which is the social cost of granting monopoly rights based on uninformative noise rather than h.

If $\rho = 1$, the third term in Eq. (11) is equal to zero. As shown in Corollary 1, in this case it is socially optimal to induce some human-capital involvement as long as $\gamma > 0$. However, as ρ declines, the signal becomes noisier and the cost of granting monopoly rights based on uninformative noise, captured by the third term, increases. At the same time, the marginal benefit, captured by the first term, declines. For sufficiently low ρ, the cost exceeds the benefit, thus defining the threshold $\bar{\rho}$ in Corollary 3.

To summarize, this extension builds on our result in Corollary 1 as follows: If the cost of AI is sufficiently low (i.e., $\pi_D \geq \overline{\psi}$) such that it is unnecessary for the IP system to act as an innovation subsidy, and additionally either $\gamma = 0$ or $\rho \leq \bar{\rho}$, then AI "kills" the IP system in that the optimal policy is to entirely dismantle the IP system ($q^* = \Delta q^* = 0$).

5.3 Consumers' Information

In Section 4, we analyze the impact of consumer preferences for human-developed products under the assumption that the consumer cannot observe the firm's human-capital use directly. In this extension, we relax this assumption by introducing an informative signal s_c, which is directly observed by the consumer and allows the consumer to update their beliefs regarding the firm's input. We provide the formal analysis in Online Appendix C.2. It is natural that the consumer receives some information about h, for example, through observable characteristics. One interpretation of this signal is the realized quality of the product observed by the consumer. This interpretation applies if higher product quality is associated with more human-capital use, allowing the consumer to infer the extent of human input from the observed quality. Another interpretation is that the consumer observes the realized IP decision $a \in \{G, D\}$. In this case, the granting decision itself is informative about the firm's human-capital use because higher h makes a grant $a = G$ more likely.

In the model in Section 4, the firm cannot signal its human-capital use to the consumer. In contrast, in this setting, the firm's choice of h determines the distribution of the signal s_c and, consequently, the consumer's willingness to pay. This has two important implications. First, the signal mitigates the adverse selection problem identified in Section 4.2.1 because it enables differentiation. As the low-cost firm optimally uses more human capital, it generates favorable signals for the consumer more often than the high-cost firm. As shown in Online Appendix C.2, this mechanism ensures that the low-cost firm achieves strictly higher expected consumer beliefs and profits than the high-cost firm, reducing the adverse selection discount.

Second, the presence of the consumer's signal s_c introduces a new incentive for the firm to use human capital, which we term the "Information Incentive." Specifically, the firm's human-capital use is determined by the following condition:

$$\psi h^* = \underbrace{\Delta q \cdot \mathbb{E}_{s_c}[\pi_G(\mathbb{E}[h|s_c]) - \pi_D(\mathbb{E}[h|s_c])]}_{\text{IP Incentive}} + \underbrace{\int_{s_c} \frac{\partial f(s_c|h)}{\partial h} \Big((q + \Delta q h)\pi_G(\mathbb{E}[h|s_c]) + \big(1 - (q + \Delta q h)\big)\pi_D(\mathbb{E}[h|s_c]) \Big) ds_c}_{\text{Information Incentive}} . \tag{12}$$

The firm's optimal human-capital use equates the marginal cost ψh to the sum of two components. The first component is the "IP Incentive," which is the expected profit premium from obtaining IP protection. This term mirrors the one in the consumer-preference model in Eq. (8). The second component is the "Information Incentive" and is unique to this extension. This term captures the marginal increase in expected profits resulting from the shift in consumer beliefs driven by a more favorable signal distribution. Importantly, this component incentivizes the firm to use human capital independently of the profit premium provided by monopoly rights. This occurs because increasing human-capital use makes favorable signals more likely (i.e., $\frac{\partial f(s_c|h)}{\partial h} > 0$), resulting in higher consumer beliefs $\mathbb{E}[h|s_c]$ and willingness to pay, which in turn increases the firm's profits $\pi_a(\mathbb{E}[h|s_c])$.[23]

Moreover, as derived in Online Appendix C.2, the "Information Incentive" amplifies the firm's responsiveness to the slope Δq. Beyond the effect of Δq in the model without consumer signals, here Δq interacts with the "Information Incentive" by increasing the firm's marginal benefit of generating favorable signals. This interaction occurs because favorable signals lead to higher consumer beliefs, which increase the profit premium $\Delta\pi(\cdot)$ given the assumption $\pi'_G(\cdot) > \pi'_D(\cdot)$. Since the slope Δq determines the marginal proba-

[23] In the special case where the consumer's signal is the officer's granting decision itself, i.e. $s_c = a$, the probability of granting protection ($a = G$) is the probability that the consumer observes the favorable signal ($s = G$), so the "IP Incentive" and the "Information Incentive" are no longer separable. Online Appendix C.2 provides this binary-signal example and shows how the firm's marginal benefit can nevertheless be decomposed into these two components.

bility of capturing this profit premium, the effect of Δq is amplified when this premium is larger. Because the profit premium is higher when signals are more favorable, a higher Δq increases the firm's marginal benefit of generating favorable signals. Consequently, the firm's human-capital use becomes more responsive to Δq. This increased responsiveness implies that the IP officer can achieve the socially optimal level of human-capital use with a flatter policy slope. Therefore, the presence of informative consumer signals lowers the socially optimal slope Δq relative to the case without consumer information.

5.4 Copyright Versus Patents

Our baseline model is designed to broadly accommodate both the copyright and patent settings for IP protection. However, model parameters may differ substantially across these settings, leading to substantially different implications. In particular, the social value of human-capital use and consumer preferences for human-developed products are likely to diverge between the copyright and patent contexts. In this section, we interpret the model parameters within each setting to contrast the resulting implications for firm investment and innovation.

Copyrightable works such as music or paintings are more likely to generate a positive social value of human-capital use. For example, if creativity provides an inherent joy that enriches people's lives, the copyright office may want to design the system to encourage such opportunities. Furthermore, survey evidence suggests that consumers prefer these types of works to be created by humans, implying a stronger adverse-selection problem. Both of these effects drive the copyright office toward a more strongly differentiated policy (higher Δq, see Propositions 2 and 6), with the system taking on increased roles in subsidizing, signaling, and screening for human-capital use (see Corollary 1, Corollary 2, and Proposition 5). If the social value of human-capital use and the adverse-selection problem are especially large, the copyright system may altogether "kill" firms' incentives to use AI for such works (see Proposition 1). The relatively low degree of natural excludability for these works further amplifies firms' sensitivity to protection, strengthening the copyright system's ability to do so (see Section 3.3.2).

By comparison, patentable inventions are less likely to either generate social value from human-capital use or carry an intrinsic consumer preference for human involvement. For example, in the development of a new vaccine, efficacy and time-to-market are likely more important than incentivizing the participation of humans in the development process. However, it remains unclear how AI- and human-developed inventions compare in terms of quality. With AI models continuing to improve, it is easy to foresee a future in which AI is able to produce superior inventions at faster speed and lower cost. Alternatively, the reliance of AI models on prior knowledge may limit their creative capacity, and large quantities of AI output could inhibit sufficient quality assurance. If the latter

concerns dominate, then consumers may have a preference for human-capital use that reflects innovation quality (see Section 5.3), and the IP system's role in mitigating potential adverse selection becomes critical due to the consequences of innovation quality for future economic growth. Nonetheless, at present, the preference for human-capital use appears weaker for patentable inventions compared to copyrightable works. This implies a less differentiated policy (lower Δq) for the patent system and greater protection for AI-generated output (see Propositions 2 and 6). Recent changes in the USPTO's guidance for AI-assisted inventions align with this prediction. The agency has shifted from a separate standard for AI-assisted inventions to a uniform standard that treats AI as a tool in the inventive process, provided that the human inventor(s) can demonstrate conception (U.S. Patent & Trademark Office, 2025).

As a result, patent-dependent firms face weaker incentives to use human capital. However, lower AI costs $\overline{\psi}$ still reduce the baseline policy leniency q, implying that firms face a stricter standard for patenting as AI makes innovation easier (see Section 3.3.3). Nonetheless, with limited social value of human-capital use, negligible welfare losses due to adverse-selection distortions, and a higher degree of natural excludability due to the difficulty of copying inventions without technical information, firms may become less dependent on patent protection as AI models improve. Consequently, AI is more likely to "kill" the patent system (see Corollary 1).[24]

6 Conclusion

We develop a model to analyze how the lack of IP protection for AI-generated output affects firms' incentives to invest in innovation and use AI in the innovation process. While AI offers substantial cost advantages, IP systems currently deny protection to intangible assets developed without meaningful human input. This creates a fundamental tension between the cost advantages of AI-based innovation and the monopoly profits granted by IP protection to human-developed intangibles.

Our analysis yields several insights. First, we characterize how the IP system affects firms' incentives through two interesting edge cases. On the one hand, the IP policy can completely disincentivize AI use, "killing" AI in innovation. On the other hand, when AI costs are sufficiently low, granting IP protection is no longer necessary to incentivize innovation investment. But rather than "killing" the IP system, we show that this may

[24]One consideration absent from our model is the patent system's role in incentivizing the disclosure of technical information through published patents. This disclosure generates positive knowledge spillovers that may incentivize the patent officer to set a positive q even if the conditions of Corollary 1 are satisfied. However, it remains unclear how improvements in AI models will affect the value of disclosure. For example, AI models may become better at reverse-engineering inventions or developing subsequent inventions without patent disclosures. If these possibilities manifest, then AI may "kill" the disclosure role of patents as well.

shift the system from an innovation subsidy to a human capital subsidy. Second, when consumers value human-developed products but cannot directly observe human-capital use, an adverse selection problem arises that distorts firms' incentives and can lead to market collapse. The IP system can mitigate this distortion by acting as a credible signal of incentives for human-capital use or by deterring investment by high-cost firms.

Taken together, these results illustrate how corporate innovation and the role of IP systems may evolve in the age of generative AI. Whether IP systems "kill'' the use of AI in innovation, AI "kills'' the need for IP systems, or something in between depends on important parameters such as AI costs, the value of the product market being created, the social value of human capital, the observability of AI use, and consumer preferences for human-developed products. Our analysis provides a characterization of how these parameters interact within the innovation process, facilitating the assessment of specific cases by future research.

A Proofs

A.1 Proof of Lemma 1

The expected profit of a firm with cost type ψ is

$$\begin{aligned}\Pi(h^*, \psi) &= \Pr[a = G]\pi_G + (1 - \Pr[a = G])\pi_D - C(h^*, \psi) \\ &= (q + h^*\Delta q)\pi_G + (1 - q - h^*\Delta q)\pi_D - \overline{\psi} - \frac{\psi}{2}(h^*)^2.\end{aligned}$$

Substituting the interior solution $h^* = \frac{\Delta q \Delta \pi}{\psi}$ from Eq. (5):

$$\begin{aligned}\Pi(h^*, \psi) &= \left(q + \frac{\Delta q \Delta \pi}{\psi}\Delta q\right)\pi_G + \left(1 - q - \frac{\Delta q \Delta \pi}{\psi}\Delta q\right)\pi_D - \overline{\psi} - \frac{\psi}{2}\left(\frac{\Delta q \Delta \pi}{\psi}\right)^2 \\ &= \pi_D + q(\pi_G - \pi_D) + \frac{(\Delta q)^2 \Delta \pi}{\psi}(\pi_G - \pi_D) - \overline{\psi} - \frac{(\Delta q)^2(\Delta \pi)^2}{2\psi} \\ &= \pi_D + q\Delta\pi + \frac{(\Delta q)^2(\Delta \pi)^2}{\psi} - \overline{\psi} - \frac{(\Delta q)^2(\Delta \pi)^2}{2\psi} \\ &= \frac{(\Delta q)^2(\Delta \pi)^2}{2\psi} + q\Delta\pi + \pi_D - \overline{\psi}.\end{aligned}$$

Thus, investment occurs if and only if:

$$\pi_D + q\Delta\pi + \frac{(\Delta q)^2(\Delta \pi)^2}{2\psi} \geq \overline{\psi}.$$

A.2 Proof of Proposition 1

From Eq. (5), if $\Delta q \Delta \pi \geq \psi$, then the solution to the first-order condition is greater than or equal to 1. Since h is constrained to be in $[0, 1]$, the firm chooses the corner solution $h^* = 1$.

To ensure investment, Condition (6) in Lemma 1 needs to hold. Rearranging this condition to solve for the marginal benefit of human capital, $\Delta q \Delta \pi$, yields the minimum required benefit for investment (assuming the term inside the square root is non-negative):

$$\Delta q \Delta \pi \geq \sqrt{2\psi(\overline{\psi} - \pi_D - q\Delta\pi)}.$$

A.3 Proof of Proposition 2

Proof. In this proof, we first derive the optimal IP policy and expected welfare for two cases: (i) the pooling equilibrium, where the officer induces investment by both firm types, and (ii) the separating equilibrium, where the officer induces investment only by the low-cost type. We then derive the condition under which the pooling equilibrium dominates the separating equilibrium, which is the case we focus on in Section 3.

Pooling Equilibrium. In the pooling equilibrium, the IP officer maximizes expected social welfare subject to the constraint that both firm types invest in innovation. Since the high-cost firm (ψ_H) is less profitable than the low-cost firm (ψ_L), ensuring the high-cost firm invests guarantees that the low-cost firm also invests. The officer's maximization problem is given by:

$$\begin{aligned}\max_{q,\Delta q} &\mathbb{E}[W] = pW(\psi_L, q, \Delta q) + (1-p)W(\psi_H, q, \Delta q)\\ \text{s.t.} \quad &\Pi(\psi_H) \geq 0.\end{aligned}$$

Where the participation constraint for the firm with cost ψ_H is:

$$\Pi(\psi_H) = \pi_D + q\Delta\pi + \frac{(\Delta q \Delta\pi)^2}{2\psi_H} - \overline{\psi} \geq 0.$$

Substituting the social benefit function $v(h) = \gamma h$ and the firm's optimal human-capital use $h^* = \frac{\Delta q \Delta\pi}{\psi}$, the expected welfare simplifies to:

$$\mathbb{E}[W] = \pi_D + u_D - \overline{\psi} + q(\Delta\pi + \Delta u) + \Delta\pi\gamma\mathbb{E}\left[\frac{1}{\psi}\right]\Delta q + \Delta\pi\left(\frac{\Delta\pi}{2} + \Delta u\right)\mathbb{E}\left[\frac{1}{\psi}\right](\Delta q)^2$$

where $\mathbb{E}\left[\frac{1}{\psi}\right] = \frac{p}{\psi_L} + \frac{1-p}{\psi_H}$.

We first solve the firm's participation constraint for q. Given that the expected welfare decreases with the base granting probability q, the optimal q is obtained by solving the binding participation constraint for the firm with cost ψ_H,

$$q^* = \frac{\overline{\psi} - \pi_D}{\Delta\pi} - \frac{(\Delta q)^2 \Delta\pi}{2\psi_H},$$

as stated in Proposition 2. Substituting q^* back into the expected welfare function reduces the dimensionality of the officer's maximization problem to Δq,

$$\begin{aligned}\max_{\Delta q} \mathbb{E}[W] =& \pi_D + u_D - \overline{\psi} + \left(\frac{\overline{\psi} - \pi_D}{\Delta\pi} - \frac{(\Delta q)^2\Delta\pi}{2\psi_H}\right)(\Delta\pi + \Delta u)\\ &+ \Delta\pi\gamma\mathbb{E}\left[\frac{1}{\psi}\right]\Delta q + \Delta\pi\left(\frac{\Delta\pi}{2} + \Delta u\right)\mathbb{E}\left[\frac{1}{\psi}\right](\Delta q)^2.\end{aligned}$$

The first-order condition with respect to Δq is

$$\frac{d\mathbb{E}[W]}{d\Delta q} = \Delta\pi\gamma\mathbb{E}\left[\frac{1}{\psi}\right] + 2\Delta q\left(\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right]\left(\frac{\Delta\pi}{2} + \Delta u\right) - \frac{\Delta\pi}{2\psi_H}(\Delta\pi + \Delta u)\right) = 0.$$

Re-arranging, the optimal slope Δq^* is:

$$\Delta q^* = \frac{\gamma \mathbb{E}\left[\frac{1}{\psi}\right]}{-2\left(\mathbb{E}\left[\frac{1}{\psi}\right]\left(\frac{\Delta\pi}{2}+\Delta u\right)-\frac{\Delta\pi+\Delta u}{2\psi_H}\right)} = \frac{\gamma \mathbb{E}\left[\frac{1}{\psi}\right]}{-\left(\mathbb{E}\left[\frac{1}{\psi}\right](\Delta\pi+2\Delta u)-\frac{\Delta\pi+\Delta u}{\psi_H}\right)},$$

as stated in Proposition 2. Therefore, the expected welfare under pooling equilibrium is:

$$\mathbb{E}[W]^*_{pool} = \left(u_D + \frac{\Delta u}{\Delta\pi}(\overline{\psi}-\pi_D)\right) + \frac{\Delta\pi\gamma^2\left(\mathbb{E}\left[\frac{1}{\psi}\right]\right)^2}{2\left(-\mathbb{E}\left[\frac{1}{\psi}\right](\Delta\pi+2\Delta u)+\frac{\Delta\pi+\Delta u}{\psi_H}\right)}.$$

Separating Equilibrium. In the separating equilibrium, the IP officer maximizes the expected social welfare generated by the low-cost firm, subject to the constraint that the low-cost firm invests in innovation. The officer's problem is:

$$\max_{q,\Delta q} \mathbb{E}[W] = pW(\psi_L)$$
$$\text{s.t.} \quad \Pi(\psi_L) \geq 0.$$

Where the participation constraint for the firm with cost ψ_L is:

$$\Pi(\psi_L) = \pi_D + q\Delta\pi + \frac{(\Delta q\Delta\pi)^2}{2\psi_L} - \overline{\psi} \geq 0.$$

Substituting the firm's optimal human-capital use $h^*_L = \frac{\Delta q\Delta\pi}{\psi_L}$, the expected welfare simplifies to

$$\mathbb{E}[W] = p\left(\pi_D + u_D - \overline{\psi} + q(\Delta\pi+\Delta u) + \frac{\Delta\pi\gamma}{\psi_L}\Delta q + \frac{\Delta\pi}{\psi_L}\left(\frac{\Delta\pi}{2}+\Delta u\right)(\Delta q)^2\right).$$

We solve the firm's participation constraint for q:

$$q^* = \frac{\overline{\psi}-\pi_D}{\Delta\pi} - \frac{(\Delta q)^2\Delta\pi}{2\psi_L}.$$

Substituting q^* into the expected welfare, the officer's maximization problem becomes:

$$\max_{\Delta q} \mathbb{E}[W] = p\Bigg(\pi_D + u_D - \overline{\psi} + \left(\frac{\overline{\psi}-\pi_D}{\Delta\pi} - \frac{(\Delta q)^2\Delta\pi}{2\psi_L}\right)(\Delta\pi+\Delta u) + \frac{\Delta\pi\gamma}{\psi_L}\Delta q + \frac{\Delta\pi}{\psi_L}\left(\frac{\Delta\pi}{2}+\Delta u\right)(\Delta q)^2\Bigg).$$

The first-order condition with respect to Δq is:

$$\frac{d\mathbb{E}[W]}{d\Delta q} = p\frac{\Delta\pi\gamma}{\psi_L} + p\Delta q\frac{\Delta\pi\Delta u}{\psi_L} = 0$$

The optimal Δq^* is:

$$\Delta q^* = \frac{\gamma}{-\Delta u}$$

Therefore, the expected welfare under separating equilibrium is:

$$\mathbb{E}[W]^*_{sep} = p\left(u_D + \frac{\Delta u}{\Delta\pi}(\overline{\psi} - \pi_D)\right) + \frac{p\Delta\pi\gamma^2}{-2\Delta u\psi_L}.$$

Comparing Pooling to Separating Equilibrium. Comparing the two equilibria, the pooling equilibrium dominates the separating equilibrium if $\mathbb{E}[W]^*_{pool} \geq \mathbb{E}[W]^*_{sep}$. The difference is:

$$\begin{aligned}
&\mathbb{E}[W]^*_{pool} - \mathbb{E}[W]^*_{sep} \\
&= (1-p)\left(u_D + \frac{\Delta u}{\Delta\pi}(\overline{\psi} - \pi_D)\right) + \left(\frac{\Delta\pi\gamma^2(\mathbb{E}[1/\psi])^2}{2\left[-\mathbb{E}[1/\psi](\Delta\pi + 2\Delta u) + \frac{\Delta\pi+\Delta u}{\psi_H}\right]} - \frac{p\Delta\pi\gamma^2}{-2\Delta u\psi_L}\right).
\end{aligned}$$

Rearranging the terms, we obtain the following condition for the pooling equilibrium to yield higher welfare:

$$\begin{aligned}
&(1-p)\left(u_D + \frac{\Delta u}{\Delta\pi}(\overline{\psi} - \pi_D)\right) \geq \\
&\Delta\pi\gamma^2\left(\frac{p}{-2\Delta u\psi_L} - \frac{(\mathbb{E}[1/\psi])^2}{2\left(-\mathbb{E}[1/\psi](\Delta\pi + 2\Delta u) + \frac{\Delta\pi+\Delta u}{\psi_H}\right)}\right)
\end{aligned}$$

In this case, the optimal $(q^*, \Delta q^*)$ provided in Proposition 2 apply. Otherwise, the $(q^*, \Delta q^*)$ for the separating equilibrium that are derived above apply. □

A.4 Proof of Corollary 1

Proof. We examine the conditions under which the optimal policy is $q^* = 0$ and $\Delta q^* = 0$.

First, consider Δq^*. From Proposition 2, assuming an interior solution exists, for Δq^* to be zero, it must be that $\gamma = 0$ (since $-(\Delta\pi + 2\Delta u) > 0$). If $\gamma > 0$, then $\Delta q^* > 0$. Thus, $\gamma = 0$ is a necessary condition.

Second, consider q^*. From Proposition 2, the optimal baseline probability is $q^* = \frac{\overline{\psi}-\pi_D}{\Delta\pi} - \frac{(\Delta q^*)^2\Delta\pi}{2\psi_H}$. If $\gamma = 0$, then $\Delta q^* = 0$. Substituting this into the expression for q^*:

$$q^* = \frac{\overline{\psi} - \pi_D}{\Delta\pi}. \tag{13}$$

For q^* to be zero (or the corner solution 0 to be optimal), we require $\frac{\overline{\psi}-\pi_D}{\Delta\pi} \leq 0$. Since $\Delta\pi > 0$, this inequality holds if and only if $\overline{\psi} \leq \pi_D$.

Therefore, the optimal IP system is dismantled ($q^* = \Delta q^* = 0$) if and only if $\gamma = 0$ and $\pi_D \geq \overline{\psi}$. □

A.5 Proof of Lemma 2

Proof. The difference in human-capital use is

$$h^* - h^{**} = \frac{\Delta q \Delta\pi(\mathbb{E}[h^*])}{\psi} - \frac{\Delta q \Delta\pi(h^{**})}{\psi} = \frac{\Delta q}{\psi}\left(\Delta\pi(\mathbb{E}[h^*]) - \Delta\pi(h^{**})\right)$$

To prove that $h_L^* < h_L^{**}$ and $h_H^* > h_H^{**}$, it suffices to show that $h_H^{**} < \mathbb{E}[h^*] < h_L^{**}$.

In the benchmark case, the firm's first-order condition for human-capital use is

$$\psi h - \Delta q \Delta\pi(h) = 0.$$

The unique solution guarantees that the left-hand side is strictly increasing in h for both ψ_H and ψ_L.

We first show that $\mathbb{E}[h^*] < h_L^{**}$. Recall that $\mathbb{E}[h^*] = ph_L^* + (1-p)h_H^*$, which is equivalent to

$$\mathbb{E}[h^*] = \Delta q \Delta\pi(\mathbb{E}[h^*])\left(\frac{p}{\psi_L} + \frac{1-p}{\psi_H}\right).$$

Evaluate the low-cost firm's first order condition for the benchmark case at $h = \mathbb{E}[h^*]$,

$$\begin{aligned}\psi_L\,\mathbb{E}[h^*] - \Delta q \Delta\pi(\mathbb{E}[h^*]) &= \psi_L \Delta q \Delta\pi(\mathbb{E}[h^*])\left(\frac{p}{\psi_L} + \frac{1-p}{\psi_H}\right) - \Delta q \Delta\pi(\mathbb{E}[h^*]) \\ &= \Delta q \Delta\pi(\mathbb{E}[h^*])\left(p + (1-p)\frac{\psi_L}{\psi_H} - 1\right) < 0,\end{aligned}$$

where the inequality follows from $\psi_L < \psi_H$. Since

$$\psi_L h_L^{**} - \Delta q \Delta\pi(h_L^{**}) = 0,$$

and the left-hand side is strictly increasing in h, it follows that

$$\mathbb{E}[h^*] < h_L^{**}.$$

Similarly, we show that $\mathbb{E}[h^*] > h_H^{**}$ by evaluating the high-cost firm's first order

condition for the benchmark case at $h = \mathbb{E}[h^*]$,

$$\begin{aligned}\psi_H \mathbb{E}[h^*] - \Delta q \Delta\pi(\mathbb{E}[h^*]) &= \psi_H \Delta q \Delta\pi(\mathbb{E}[h^*]) \left(\frac{p}{\psi_L} + \frac{1-p}{\psi_H}\right) - \Delta q \Delta\pi(\mathbb{E}[h^*]) \\ &= \Delta q \Delta\pi(\mathbb{E}[h^*]) \left(p\frac{\psi_H}{\psi_L} + (1-p) - 1\right) > 0,\end{aligned}$$

where the inequality follows from $\psi_L < \psi_H$. Since

$$\psi_H h_H^{**} - \Delta q \Delta\pi(h_H^{**}) = 0,$$

and the left-hand side is strictly increasing in h, it follows that

$$\mathbb{E}[h^*] > h_H^{**}.$$

Therefore,

$$h_H^{**} < \mathbb{E}[h^*] < h_L^{**}.$$

It follows that $h_L^* < h_L^{**}$ and $h_H^* > h_H^{**}$.

□

A.6 Proof of Proposition 3

Proof. The feedback loop is captured by the term labeled "Indirect feedback loop" in Eq. (9). If $\pi'_G(h) > \pi'_D(h)$, then $\Delta\pi'(h) > 0$ and therefore $\frac{d\Delta\pi(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]} > 0$. Conversely, in the counterfactual $\pi'_G(h) = \pi'_D(h)$, $\frac{d\Delta\pi(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]} = 0$ and the feedback loop term is equal to zero. Since $\frac{\Delta q}{\psi} > 0$, this implies that the feedback loop term leads to an increase in $\frac{dh^*}{d\Delta q}$ for both types.

To demonstrate that the feedback loop mitigates the adverse selection discount, first note that by Eq. (9),

$$\frac{dh_L^*}{d\Delta q} - \frac{dh_H^*}{d\Delta q} = \underbrace{\left(\frac{1}{\psi_L} - \frac{1}{\psi_H}\right)\Delta\pi(\mathbb{E}[h^*])}_{\text{Direct incentive}} + \underbrace{\left(\frac{1}{\psi_L} - \frac{1}{\psi_H}\right)\Delta q \frac{d\Delta\pi(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}}_{\text{Indirect feedback loop}}.$$

The second term shows that the feedback loop effect of Δq on h_L^* is larger compared to on h_H^* because $\frac{1}{\psi_L} > \frac{1}{\psi_H}$. Therefore, compared to the counterfactual without the feedback loop ($\pi'_G(h) = \pi'_D(h)$), the feedback loop increases the difference between $\frac{dh_L^*}{d\Delta q}$ and $\frac{dh_H^*}{d\Delta q}$.

□

A.7 Proof of Corollary 2

Proof. See Proof of Proposition 6 in Online Appendix B.3.1. The expression for Δq^* shows that $\Delta q^* > 0$ even if $\gamma = 0$ given that the term $\frac{\partial \mathbb{E}[W]}{\partial \mathbb{E}[h^*]} \frac{d\,\mathbb{E}[h^*]}{d\Delta q} \frac{1}{\Delta \pi}$ is positive. □

A.8 Proof of Proposition 4

Proof. In this proof, we derive the equilibrium firm investment decisions as a function of q, holding Δq fixed. To proceed, we use the following lemma to formally establish the impact of the IP policy on the firm's expected profit.

Lemma 3. *The expected profit of a firm of type ψ, denoted by $\Pi^*(h^*, \psi)$, is increasing in q and Δq.*

Proof. The following derivative based on Eq. (7) shows the effect of q and Δq on the firm's expected profit:

$$\frac{d\Pi^*}{dq} = \Delta\pi(\mathbb{E}[h^*]) > 0.$$

$$\frac{d\Pi^*}{d\Delta q} = h^* \Delta\pi(\mathbb{E}[h^*]) + \left((q + \Delta q h^*) \frac{d\pi_G(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]} + (1 - (q + \Delta q h^*)) \frac{d\pi_D(\mathbb{E}[h^*])}{d\,\mathbb{E}[h^*]} \right) \frac{d\,\mathbb{E}[h^*]}{d\Delta q} > 0.$$

Given that the profit premium $\Delta\pi(\mathbb{E}[h^*])$ is positive, both $\frac{d\Pi^*}{dq}$ and the first term of $\frac{d\Pi^*}{d\Delta q}$ are positive. The second term of $\frac{d\Pi^*}{d\Delta q}$ is also positive due to the feedback loop detailed in Section 4.2.2, where an increase in Δq induces higher consumer beliefs about the firm's human-capital use, i.e., $\frac{d\,\mathbb{E}[h^*]}{d\Delta q} > 0$. □

This monotonicity of the firm's expected profit allows us to characterize the equilibrium investment using threshold values. To isolate the firm's investment decision from its human-capital use decision, we analyze the equilibrium by varying the base grant probability q while holding the slope Δq constant. Because the expected profit of a low-cost firm $\Pi^*(\psi_L)$ is weakly higher than that of the high-cost firm $\Pi^*(\psi_H)$, the low-cost firm invests at a lower threshold q than the high-cost firm. Consequently, the possible equilibrium investment scenarios are:

1. **No investment**: If the low-cost firm cannot profitably invest under the consumer belief $\mathbb{E}[h] = 0$, the market collapses. We define q_N as the threshold value of q that makes the low-cost type break even under the belief $\mathbb{E}[h] = 0$:

$$\Pi^*(\psi_L, \mathbb{E}[h] = 0, q_N, \Delta q) = q_N \Delta\pi(0) + \pi_D(0) + \frac{(\Delta q \Delta\pi(0))^2}{2\psi_L} - \overline{\psi} = 0. \tag{14}$$

   For any $q < q_N$, no firm invests given the belief $\mathbb{E}[h] = 0$.

2. **Only the low-cost type invests**: In this equilibrium, the consumer correctly infers that only the low-cost type invests. The consumer belief is $\mathbb{E}[h] = h_L^*$, where h_L^* solves the fixed-point equation $\mathbb{E}[h] = h_L^*(\mathbb{E}[h])$:

$$h_L^* = \frac{\Delta q \Delta\pi(h_L^*)}{\psi_L}.$$

We define q_L and q_H as the thresholds that satisfy the binding participation constraints for the low-cost and high-cost types, respectively, under the belief $E[h] = h_L^*$:

$$\Pi^*(\psi_L, \mathbb{E}[h] = h_L^*, q_L, \Delta q) = q_L \Delta\pi(h_L^*) + \pi_D(h_L^*) + \frac{(\Delta q \Delta\pi(h_L^*))^2}{2\psi_L} - \overline{\psi} = 0, \quad (15)$$

$$\Pi^*(\psi_H, \mathbb{E}[h] = h_L^*, q_H, \Delta q) = q_H \Delta\pi(h_L^*) + \pi_D(h_L^*) + \frac{(\Delta q \Delta\pi(h_L^*))^2}{2\psi_H} - \overline{\psi} = 0. \quad (16)$$

Thus, an equilibrium where only the low-cost type invests exists if and only if $q \in [q_L, q_H)$.

3. **Both types invest**: In this equilibrium, the consumer correctly infers that both types invest, so the belief $\mathbb{E}[h] = \mathbb{E}[h^*]$ is the pooled average that solves the fixed-point equation

$$\mathbb{E}[h^*] = p\left(\frac{\Delta q \Delta\pi(\mathbb{E}[h^*])}{\psi_L}\right) + (1-p)\left(\frac{\Delta q \Delta\pi(\mathbb{E}[h^*])}{\psi_H}\right). \quad (17)$$

We define q_B as the threshold that satisfies the binding participation constraint for the high-cost firm under the belief $\mathbb{E}[h] = \mathbb{E}[h^*]$,

$$\Pi^*(\psi_H, \mathbb{E}[h] = \mathbb{E}[h^*], q_B, \Delta q) = q_B \Delta\pi(\mathbb{E}[h^*]) + \pi_D(\mathbb{E}[h^*]) + \frac{(\Delta q \Delta\pi(\mathbb{E}[h^*]))^2}{2\psi_H} - \overline{\psi} = 0. \quad (18)$$

Consequently, an equilibrium where both types invest exists if and only if $q \geq q_B$.

In certain regions of q, multiple equilibria may arise. To resolve this multiplicity, we apply the standard Pareto-dominance criterion. We select the equilibrium that yields strictly higher welfare. This criterion implies that whenever investment is profitable for at least one firm type under rational beliefs, the economy selects the investment equilibrium over market collapse. In addition, we restrict our analysis to the case where the firm's profit under the most pessimistic beliefs is sufficiently low such that $q_N > q_B$. This condition implies that the base grant probability q required to induce investment under the most pessimistic belief ($\mathbb{E}[h] = 0$) is strictly higher than that required to sustain the equilibrium where both types invest, streamlining the analysis by limiting the number of cases we need to consider.

Given the result in Lemma 3 and the assumption $\psi_H > \psi_L$, the high-cost firm requires a higher q to induce investment; therefore, it holds that $q_H > q_L$. In addition, because the firm's expected profit is strictly increasing in consumer beliefs and $h_L^* > \mathbb{E}[h^*]$, it holds that $q_H < q_B$. This ordering reflects the cost of adverse selection: the firm requires a larger investment incentive q from the IP system to compensate for the reduction in consumer beliefs.

We use NI, LI and BI to represent no investment, only low-cost type invests, and both types invest, respectively. The equilibrium is

1. if $q < q_L$, NI is the unique equilibrium.
2. if $q_L \leq q < q_H$, both NI and LI can be the equilibrium.
3. if $q_H \leq q < q_B$, NI is the unique equilibrium.
4. if $q_B \leq q < q_N$, both NI and BI can be the equilibrium.
5. if $q \geq q_N$, BI is the unique equilibrium.

In regions where multiple equilibria may arise, we select among them using the Pareto-dominance criterion. Because investment by either type yields higher welfare than the no-investment outcome, the resulting equilibrium is the one stated in the Proposition. □

A.9 Proof of Proposition 5

Proof. In this proof, we derive the expected welfare for both equilibria and analyze the impact of $\Delta\psi$. To this end, we hold ψ_L fixed and perform comparative statics based on ψ_H, implying variation in $\Delta\psi$.

We first establish the optimal q for each equilibrium regime with a fixed Δq. Under Assumption 1, the expected welfare is strictly decreasing in q. To maximize welfare within each regime, the officer sets the lowest possible q that satisfies the respective binding participation constraints. Thus, we compare the LI equilibrium (only low-cost type invests) implemented with q_L, where $\Pi(\psi_L, h_L^*, q_L) = 0$, against the BI (both type invests) equilibrium implemented with q_B, where $\Pi(\psi_H, \mathbb{E}[h^*], q_B) = 0$.

In the LI equilibrium, the expected welfare is

$$W_{LI} = p\Big(u_D(h_L^*) + (q_L + \Delta q h_L^*)\Delta u(h_L^*) + \gamma h_L^*\Big) \tag{19}$$

Where $h_L^* = \frac{\Delta q \Delta\pi(h_L^*)}{\psi_L}$. Note that the expected profit of the low-cost firm, $\Pi(\psi_L, h_L^*, q_L)$, is zero, and hence, it does not appear in W_{LI}. In addition, h_L^* and q_L are determined solely by ψ_L and Δq. Since ψ_L is fixed, W_{LI} is independent of $\Delta\psi$, implying $dW_{LI}/d\Delta\psi = 0$.

In the BI equilibrium, the officer sets q_B such that the high-cost firm makes zero profit given the belief $\mathbb{E}[h^*]$. The low-cost firm earns a positive expected profit $\Pi(\psi_L, \mathbb{E}[h^*], q_B) > 0$. The expected welfare W_B is :

$$W_B(\Delta\psi) = u_D(\mathbb{E}[h^*]) + (pP_L + (1-p)P_H)\Delta u(\mathbb{E}[h^*]) + p\Pi(\psi_L, \mathbb{E}[h^*], q_B) + \gamma(ph_L^* + (1-p)h_H^*), \tag{20}$$

where $P_\psi = q_B + \Delta q h^*$ is the total probability of granting IP protection for type ψ, $h^* = \frac{\Delta q \Delta\pi(\mathbb{E}[h^*])}{\psi}$, and $\mathbb{E}[h^*]$ is the consumer's belief given in Eq. (17). Note that the expected profit of the high-cost firm is zero and therefore does not appear in W_B. Further note that $\Delta\psi$ enters the expected welfare W_B through three channels: the human-capital use of the high-cost firm h_H^*, the consumer belief $\mathbb{E}[h^*]$, and the base grant probability q_B.

We analyze the total derivative of W_B with respect to $\Delta\psi$. This derivative is :

$$\frac{dW_B}{d\Delta\psi} = \frac{\partial W_B}{\partial h_H^*}\frac{dh_H^*}{d\Delta\psi} + \frac{\partial W_B}{\partial \mathbb{E}[h^*]}\frac{d\mathbb{E}[h^*]}{d\Delta\psi} + \frac{\partial W_B}{\partial q_B}\frac{dq_B}{d\Delta\psi} \tag{21}$$

The first term captures the direct impact of rising costs on the high-type's human-capital choice. This term changes the social value γh_H^* and the monopoly probability P_H. As $\Delta\psi$ increases, the firm reduces human-capital use h_H^*. The welfare impact of a change in h_H^* is:

$$\frac{\partial W_B}{\partial h_H^*} = (1-p)\Big(\gamma + \Delta q \Delta u(\mathbb{E}[h^*])\Big).$$

The sign of this term is ambiguous because both the social value and the social loss from monopoly rights increase in h_H^*.

The second term represents how $\Delta\psi$ affects the consumer belief $\mathbb{E}[h^*]$. As $\Delta\psi$ increases, the consumer understands that the high-cost firm uses less human capital, causing the belief to fall, that is, $d\mathbb{E}[h^*]/d\Delta\psi < 0$. Since welfare is strictly increasing in the belief, $\partial W_B/\partial\mathbb{E}[h^*] > 0$, this term is strictly negative, i.e., $\frac{\partial W_B}{\partial \mathbb{E}[h^*]}\frac{d\mathbb{E}[h^*]}{d\Delta\psi} < 0$.

The third term represents how $\Delta\psi$ affects the participation constraint of the high-cost firm. The high-cost firm's participation constraint becomes more difficult to satisfy as its marginal cost rises and the consumer belief falls. Therefore, the officer must increase the base granting probability ($dq_B/d\Delta\psi > 0$). The welfare impact of this increase is:

$$\frac{\partial W_B}{\partial q_B} = (p + (1-p))\Delta u(\mathbb{E}[h^*]) + p\Delta\pi(\mathbb{E}[h^*]). \tag{22}$$

The first part is the reduction in consumer surplus due to the increased probability of monopoly for both types of firm. The second part is the increase in the low-cost firm's expected profit. Summing these yields $\Delta u + p\Delta\pi$. Assumption 1 states that $\Delta u + \Delta\pi < 0$. Since $p < 1$ and $\Delta\pi > 0$, it follows that $\Delta u + p\Delta\pi < \Delta u + \Delta\pi < 0$. Thus, the marginal

welfare effect of increasing q_B is strictly negative. The social loss from consumer surplus outweighs the profit gain for the low-cost firm.

The first term of the total derivative $\frac{dW_B}{d\Delta\psi}$ has ambiguous sign, while the second and the third term are strictly negative. To determine the sign of the total derivative, we verify that the sum of the first term and the third term is negative. To see this, we sum up the two terms.

$$
\begin{aligned}
&\frac{\partial W_B}{\partial h_H^*}\frac{dh_H^*}{d\Delta\psi} + \frac{\partial W_B}{\partial q_B}\frac{dq_B}{d\Delta\psi} \\
&= (1-p)\Delta q \Delta u(\mathbb{E}[h^*])\frac{dh_H^*}{d\Delta\psi} + (p+(1-p))\Delta u(\mathbb{E}[h^*])\frac{dq_B}{d\Delta\psi} + (1-p)\gamma\frac{dh_H^*}{d\Delta\psi} + p\Delta\pi(\mathbb{E}[h^*])\frac{dq_B}{d\Delta\psi} \\
&= p\Delta u(\mathbb{E}[h^*])\frac{dq_B}{d\Delta\psi} + (1-p)\Delta u(\mathbb{E}[h^*])\left(\Delta q\frac{dh_H^*}{d\Delta\psi} + \frac{dq_B}{d\Delta\psi}\right) + (1-p)\gamma\frac{dh_H^*}{d\Delta\psi} + p\Delta\pi(\mathbb{E}[h^*])\frac{dq_B}{d\Delta\psi} \\
&= p\big(\Delta u(\mathbb{E}[h^*]) + \Delta\pi(\mathbb{E}[h^*])\big)\frac{dq_B}{d\Delta\psi} + (1-p)\Delta u(\mathbb{E}[h^*])\frac{dP_H}{d\Delta\psi} + (1-p)\gamma\frac{dh_H^*}{d\Delta\psi} < 0.
\end{aligned}
$$

Differentiating the participation constraint of the high-cost firm reveals that the total probability P_H must strictly increase to cover rising costs and falling consumer beliefs for the high-cost firm to be indifferent. Since P_H increases despite the fall in h_H^*, the increase in q_B must be larger than the decrease in the incentive term $\Delta q h_H^*$, which implies that $\frac{dq_B}{d\Delta\psi} > \frac{dP_H}{d\Delta\psi} > 0$. Given Assumption 1, the term $\Delta u(\mathbb{E}[h^*]) + \Delta\pi(\mathbb{E}[h^*]) < 0$ and $\Delta u(\mathbb{E}[h^*]) < 0$. Consequently, the net effect of the first and third terms is strictly negative.

Combining these results, we conclude that $dW_B/d\Delta\psi < 0$ while W_{LI} is constant over $\Delta\psi$. By the Intermediate Value Theorem, there exists a unique cutoff $\widetilde{\Delta\psi}$ such that for all $\Delta\psi > \widetilde{\Delta\psi}$, the LI equilibrium strictly dominates the BI equilibrium. □

A.10 Proof of Corollary 3

Proof. The officer maximizes expected welfare $\mathbb{E}[W]$ across firm types. We define $\Lambda \equiv -(\Delta\pi + \Delta u) > 0$ and $K \equiv -(\Delta\pi + 2\Delta u) > 0$. Substituting the firm's optimal response $h^* = \frac{\rho\Delta q\Delta\pi}{\psi}$ into the welfare function and taking the expectation over cost types yields:

$$
\mathbb{E}[W(\Delta q)] = \pi_D + u_D - \overline{\psi} + \mathbb{E}\left[\gamma h^* - \Lambda\left[q + \Delta q\left(\rho h^* + \frac{1-\rho}{2}\right)\right] - \frac{\psi}{2}(h^*)^2\right].
$$

Denote $\mathbb{E}[1/\psi] = \frac{p}{\psi_L} + \frac{1-p}{\psi_H}$, this reduces to:

$$
\mathbb{E}[W(\Delta q)] = \pi_D + u_D - \overline{\psi} - \Lambda q + \left(\gamma\rho\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right] - \frac{\Lambda(1-\rho)}{2}\right)\Delta q - \frac{\rho^2\Delta\pi K}{2}\mathbb{E}\left[\frac{1}{\psi}\right](\Delta q)^2.
$$

Differentiating $\mathbb{E}[W(\Delta q)]$ with respect to Δq yields the marginal welfare condition:

$$\frac{d\mathbb{E}[W]}{d\Delta q} = \gamma\rho\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right] - \left[\frac{\Lambda(1-\rho)}{2} + \rho^2\Delta\pi K\mathbb{E}\left[\frac{1}{\psi}\right]\Delta q\right] = 0.$$

The optimal Δq^* is positive if and only if the marginal benefit at zero exceeds the marginal cost. We evaluate the FOC at $\Delta q = 0$:

$$\left.\frac{d\mathbb{E}[W]}{d\Delta q}\right|_{\Delta q=0} = \gamma\rho\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right] - \frac{\Lambda(1-\rho)}{2}.$$

Setting this expression to be less than or equal to zero and rearranging for ρ yields the condition $\rho \leq \bar{\rho}$. □

References

Acemoglu, D. and U. Akcigit (2012). Intellectual Property Rights Policy, Competition and Innovation. *Journal of the European Economic Association 10*(1), 1–42.

Acemoglu, D., U. Akcigit, N. Bloom, and W. Kerr (2018). Innovation, Reallocation, and Growth. *American Economic Review 108*(11), 3450—3491.

Acemoglu, D. and D. Autor (2011). Skills, tasks and yechnologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier.

Acemoglu, D. and P. Restrepo (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review 108*(6), 1488—1542.

Acemoglu, D. and P. Restrepo (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives 33*(2), 3—30.

Acemoglu, D. and P. Restrepo (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy 128*(6), 2188—2244.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: An Inverted-U Relationship. *Quarterly Journal of Economics 120*(2), 701—728.

Aghion, P., C. Harris, P. Howitt, and J. Vickers (2001). Competition, Imitation and Growth with Step-by-Step Innovation. *The Review of Economic Studies 68*(3), 467–492.

Aghion, P., B. F. Jones, and C. I. Jones (2019, 05). Artificial intelligence and economic growth. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Aghion, P., J. Van Reenen, and L. Zingales (2013). Innovation and Institutional Ownership. *American Economic Review 103*(1), 277—304.

Agrawal, A., J. McHale, and A. Oettl (2018). Finding Needles in Haystacks: Artificial Intelligence and Recombinant Growth. Working Paper.

Akcigit, U. and W. R. Kerr (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy 126*(4), 1374—1443.

Akcigit, U. and Q. Liu (2016). The Role of Information in Innovation and Competition. *Journal of the European Economic Association 14*(4), 828–870.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, pp. 609–626. Princeton University Press.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics 151*(1), 103745.

Black, B. and R. J. Gilson (1998). Venture capital and the structure of capital markets: Banks versus stock markets. *Journal of Financial Economics 47*(3), 243—277.

Bleistein v. Donaldson Lithographing Co. (1903). 188 U.S. 239, 251.

Brown, J. R., S. M. Fazzari, and B. C. Petersen (2009). Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom. *Journal of Finance 64*(1), 151—185.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative AI at Work. *Quarterly Journal of Economics 140*(2), 889—942.

Budish, E., B. N. Roin, and H. Williams (2015). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review 105*(7),

2044—2085.

Castelo, N., M. W. Bos, and D. R. Lehmann (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research 56*(5), 809–825.

Cockburn, I. M., S. Kortum, and S. Stern (2002). Are All Patent Examiners Equal? The Impact of Examiner Characteristics. NBER Working Paper No. 8980.

Coles, J. L., N. D. Daniel, and L. Naveen (2006). Managerial incentives and risk-taking. *Journal of Financial Economics 79*(2), 431–468.

Colliard, J.-E. and J. Zhao (2025). Artificial intelligence and the rents of finance workers. Available at SSRN 5339402.

Concord Music Group, Inc. v. Anthropic PBC (2024). 3:23-cv-01092. (M.D. Tenn.).

Cong, L. W., Y. Lu, H. Shi, and W. Zhu (2025). Automation-Induced Innovation Shift. Working Paper.

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022). The Economics of Intangible Capital. *Journal of Economic Perspectives 36*(3), 29–52.

Cuntz, A., C. Fink, and H. Stamm (2024). Artificial Intelligence and Intellectual Property: An Economic Perspective. World Intellectual Property Organization (WIPO) Economic Research Working Paper Series No. 77.

Dargnies, M.-P., R. Hakimov, and D. Kübler (2024). Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *Management Science 72*(1), 285–301.

de Rassenfosse, G., A. B. Jaffe, and J. Waldfogel (2025). Intellectual Property and Creative Machines. *Entrepreneurship and Innovation Policy and the Economy 4*(1), 47–79.

de Rassenfosse, G., A. B. Jaffe, and M. Wasserman (2025). AI-Generated Inventions: Implications for the Patent System. *Southern California Law Review 96*(6), 1453–1478.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General 144*(1), 114.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science 384*(6702), 1306—1308.

Farre-Mensa, J., D. Hegde, and A. Ljungqvist (2020). What Is a Patent Worth? Evidence from the U.S. Patent "Lottery". *Journal of Finance 75*(2), 639—682.

Ferreira, D., G. Manso, and A. Silva (2014). Incentives to Innovate and the Decision to Go Public or Private. *Review of Financial Studies 27*(1), 256—300.

Galasso, A. and M. Schankerman (2015). Patents and Cumulative Innovation: Causal Evidence from the Courts. *Quarterly Journal of Economics 130*(1), 317—369.

Gans, J. S. (2026). Copyright Policy Options for Generative Artificial Intelligence. *Journal of Law and Economics 69*(1), 1—19.

Garen, J. E. (1994). Executive Compensation and Principal-Agent Theory. *Journal of Political Economy 102*(6), 1175—1199.

Geiger, C. (2024). Elaborating a Human Rights-Friendly Copyright Framework for Generative AI. *International Review for Intellectual Property and Competition Law 55*(7), 1129—1165.

Getty Images (US), Inc. v. Stability AI, Ltd. (2025). 3:25-cv-06891. (N.D. Cal.).

Gilbert, R. and C. Shapiro (1990). Optimal patent length and breadth. *RAND Journal of Economics 21*(1), 106—112.

He, J. and X. Tian (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics 109*(3), 856—878.

Henry, N. L. (1974). Copyright, Public Policy, and Information Technology. *Science 183*(4123), 384–391.

Holmström, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics 10*(1), 74–91.

Hopenhayn, H., G. Llobet, and M. Mitchell (2006). Rewarding Sequential Innovators: Prizes, Patents, and Buyouts. *Journal of Political Economy 114*(6), 1041–1068.

Hopenhayn, H. and F. Squintani (2016). Patent Rights and Innovation Disclosure. *The Review of Economic Studies 83*(1), 199–230.

Hopenhayn, H. and F. Squintani (2021). On the Direction of Innovation. *Journal of Political Economy 129*(7), 1991–2022.

Horton, Jr., C. B., M. W. White, and S. S. Iyengar (2023). Bias against AI art can enhance perceptions of human creativity. *Scientific Reports 13*(1), 19001.

Johnson, W. R. (1985). The Economics of Copying. *Journal of Political Economy 93*(1), 158–174.

Kennedy, B. and E. Yam (2025). From political speeches to songs, how would Americans react if they found out AI was involved? Pew Research Center. Accessed on January 7, 2026, from https://www.pewresearch.org/short-reads/2025/09/17/from-political-speeches-to-songs-how-would-americans-react-if-they-found-out-ai-was-involved/.

Kerr, W. R., J. Lerner, and A. Schoar (2014). The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. *Review of Financial Studies 27*(1), 20—55.

Kortum, T. J. K. S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy 112*(5), 986—1018.

Lemley, M. A. (2024). How Generative AI turns copyright law on its head. *Science and Technology Law Review 25*(2), 190—212.

Lemley, M. A. and B. Sampat (2012). Examiner Characteristics and Patent Office Outcomes. *Review of Economics and Statistics 94*(3), 817–827.

Liang, A. and J. Lu (2026). Creative Ownership in the Age of AI. Working Paper.

Manewitsch, V. and A. Kalb (2025). Explaining consumer preferences for AI- and human-authored books: An explanatory experimental study. In *Proceedings of the 5th International Conference on AI Research (ICAIR 2025)*, Volume 5. Academic Conferences and Publishing International Limited.

Manso, G. (2011). Motivating Innovation. *Journal of Finance 66*(5), 1823—1860.

Naruto v. Slater (2018). 888 F.3d 418, 426. (9th Cir.).

Nordhaus, W. D. (1969). An Economic Theory of Technological Change. *American Economic Review 59*(2), 18—28.

Osborne, M. R. and E. R. Bailey (2025). Me vs. the machine? Subjective evaluations of human-and AI-generated advice. *Scientific Reports 15*(1), 3980.

Peukert, C., F. Abeillon, J. Haese, F. Kaiser, and A. Staub (2025). AI and the Dynamic Supply of Training Data. Working Paper.

Peukert, C. and M. Windisch (2025). The economics of copyright in the digital age. *Journal of Economic Surveys 39*(3), 877–903.

Phillips, G. M. and A. Zhdanov (2013). R&D and the Incentives from Merger and Acquisition Activity. *Review of Financial Studies 26*(1), 34—78.

Sampat, B. and H. L. Williams (2019). How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome. *American Economic Review 109*(1), 203—236.

Samuelson, P. (2024). Generative AI meets copyright. *Science 371*(6654), 158–161.

Scotchmer, S. (1991). Standing on the Shoulders of Giants: Cumulative Research and the Patent Law. *Journal of Economic Perspectives 5*(1), 29–41.

Thaler v. Perlmutter (2023). 687 F. Supp. 3d 140,. 142. (D.D.C.).

Thaler v. Vidal (2022). 43 F.4th 1207, 1210. (Fed. Cir.).

The New York Times Co v. Microsoft Corp (2025). 1:23-cv-11195. (S.D.N.Y.).

Tubadji, A., H. Huang, and D. J. Webber (2021, None). Cultural proximity bias in AI-acceptability: The importance of being human. *Technological Forecasting and Social Change 173*(C), None.

U.S. Copyright Office (2025a). Copyright and artificial intelligence. Technical report, U.S. Copyright Office.

U.S. Copyright Office (2025b). Identifying the Economic Implications of Artificial Intelligence for Copyright Policy. Technical report, U.S. Copyright Office.

U.S. Patent & Trademark Office (2025). Revised Inventorship Guidance for AI-Assisted Inventions. Technical report, U.S. Patent & Trademark Office.

Wang, J. T., Z. Deng, H. Chiba-Okabe, B. Barak, and W. J. Su (2024). An Economic Solution to Copyright Challenges of Generative AI. Working Paper.

Williams, H. L. (2013). Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy 121*(1), 1—27.

Wu, L., B. Lou, and L. M. Hitt (2025). Innovation Strategy After IPO: How AI Analytics Spurs Innovation After IPO. *Management Science 71*(3), 1865—1888.

Yang, S. A. and A. H. Zhang (2025). Generative AI and Copyright: A Dynamic Perspective. Working Paper.

Zhong, H. (2026). Optimal Integration: Human, Machine, and Generative AI. *Management Science 72*(5), 3629—4567.

Online Appendix for "Intellectual Property Protection for AI-Generated Output"

B Additional Results and Derivations

B.1 Microfoundation Based on Cournot Competition

This appendix develops a microfoundation for the profit and consumer surplus assumptions used in the main text (see Assumption 1 and its surrounding text). This microfoundation captures our main interpretation of the innovation investment in our model, namely that it represents the development of a new intangible asset that can be commercialized as a new product.

Suppose the representative consumer has a utility function

$$\begin{aligned} U(Q_n, Q) &= Q_n + b_1(1+\beta h)^\alpha Q - b_2 Q^2/2 \\ &= \bar{Q}_n - pQ + b_1(1+\beta h)^\alpha Q - b_2 Q^2/2, \end{aligned}$$

where Q_n is the consumption of a numeraire good (cash) and Q is the quantity consumed of the new product which firm 1 can produce if it makes the initial innovation investment. b_1 and b_2 are constants. β parameterizes the strength of the consumer's preference for human capital and α parameterizes its curvature, with $\beta = 0$ capturing the baseline model in Section 3 and $\beta > 0$ the extended model in Section 4. The consumer has some initial endowment $\bar{Q}_n$ of the numeraire good. Denoting by p the price of the new product, the budget constraint is given by $Q_n + pQ = \bar{Q}_n$.

The consumer cannot perfectly observe h. Therefore, the consumer forms an expectation based on the information set, and the optimization problem is

$$\max_{Q_n, Q} \mathbb{E}[U(Q_n, Q)] = \mathbb{E}[\bar{Q}_n - pQ + b_1(1+\beta h)^\alpha Q - b_2 Q^2/2],$$

which implies an optimal consumption $Q(p) = \frac{b_1 \mathbb{E}[(1+\beta h)^\alpha] - p}{b_2}$ and an inverse demand function

$$p(Q) = b_1 \mathbb{E}[(1+\beta h)^\alpha] - b_2 Q. \tag{A1}$$

If firm 1 did not invest in innovation, no firm produces. If firm 1 invested in innovation, then N identical firms enter the market in the production stage and firm i maximizes

$$\max_{q_i} q_i p(Q).$$

The first order condition is given by

$$\begin{aligned} & p(Q) + q_i p'(Q) = 0 \\ \Leftrightarrow\ & q_i = \frac{b_1 \mathbb{E}[(1+\beta h)^\alpha]}{(N+1) b_2}, \end{aligned}$$

where aggregate output is given by $Q = Nq_i$. This implies the price is

$$\begin{aligned} p(Q) &= b_1 \mathbb{E}[(1+\beta h)^\alpha] - b_2 \frac{N b_1 \mathbb{E}[(1+\beta h)^\alpha]}{(N+1)b_2} \\ &= \frac{b_1 \mathbb{E}[(1+\beta h)^\alpha]}{(N+1)}, \end{aligned}$$

and the profit for firm i is

$$\begin{aligned} \pi(\mathbb{E}[h]) &= q_i p(Q) \\ &= \frac{b_1^2 \mathbb{E}[(1+\beta h)^\alpha]^2}{(N+1)^2 b_2}. \end{aligned} \tag{A2}$$

In our model, $h(\psi)$ is inversely proportional to ψ, and the human-capital use h_L and h_H are linear in $\mathbb{E}[h]$. Thus, $\mathbb{E}[(1+\beta h)^\alpha]$ is strictly increasing in $\mathbb{E}[h]$, which implies $\pi'(\mathbb{E}[h]) \geq 0$.

Consumer utility is given by

$$\begin{aligned} U(Q_n, Q) &= \bar{Q}_n - pQ + b_1(1+\beta h)^\alpha Q - b_2 Q^2/2 \\ &= \bar{Q}_n - N\pi(\mathbb{E}[h]) + b_1(1+\beta h)^\alpha \left(\frac{N b_1 \mathbb{E}[(1+\beta h)^\alpha]}{(N+1)b_2} \right) - b_2 \left(\frac{N b_1 \mathbb{E}[(1+\beta h)^\alpha]}{(N+1)b_2} \right)^2 /2 \\ &= \bar{Q}_n + \pi(\mathbb{E}[h]) \left[N(N+1) \frac{(1+\beta h)^\alpha}{\mathbb{E}[(1+\beta h)^\alpha]} - N - \frac{N^2}{2} \right] \end{aligned}$$

We define u_a as the sum of consumer surplus and the profit of the $N-1$ other firms. Therefore, we add $(N-1)\pi$ to $U(Q_n, Q)$:

$$u_a = U(Q_n, Q) + (N-1)\pi.$$

B.1.1 Mapping to Baseline Model

In the baseline model in Section 3, the consumer has no preferences for human-developed products, represented by the case $\beta = 0$. Moreover, the monopoly outcome ($a = G$) is represented by the case $N = 1$, while the competitive case ($a = D$) is represented by $N > 1$. Consequently, in the Cournot model the terms in the main text are given by

$$\begin{aligned} \pi_G &= \frac{b_1^2}{4b_2}, \\ \pi_D &= \frac{b_1^2}{(N+1)^2 b_2}, \\ u_G &= \bar{Q}_n + \frac{1}{2}\pi_G, \\ u_D &= \bar{Q}_n + \left(\frac{1}{2}N^2 + N - 1 \right) \pi_D, \\ W_G(h, \psi) &= \bar{Q}_n + \frac{3}{2}\pi_G - C(h, \psi) + v(h), \\ W_D(h, \psi) &= \bar{Q}_n + \left[N + \frac{N^2}{2} \right] \pi_D - C(h, \psi) + v(h). \end{aligned}$$

As stated in Section 2, these terms are consistent with Assumption 1.

B.1.2 Mapping to Consumer Preference Model

In the extended model in Section 4 consumers have a preference for human-developed products, represented by $\beta > 0$. In this case, the consumer belief about the firm's human-capital use $\mathbb{E}[h]$ affects profits and consumer surplus. Consumer surplus and the profit of the $N-1$ other firms are given by

$$u_a(h, \mathbb{E}[h]) = U(Q_n, Q) + (N-1)\pi(\mathbb{E}[h]).$$

To analyze the ex-ante expected consumer surplus, we evaluate the $t = 0$ unconditional ex-ante expectation

$$u_a(\mathbb{E}[h]) \equiv \mathbb{E}_0[u_a(h, \mathbb{E}[h])] = \mathbb{E}_0[U(Q_n, Q)] + (N-1)\pi(\mathbb{E}[h]) = \bar{Q}_n + \left(\frac{1}{2}N^2 + N - 1\right)\pi(\mathbb{E}[h]).$$

where the ex-ante expected utility is given by

$$\begin{aligned}
\mathbb{E}_0[U(Q_n, Q)] &= \bar{Q}_n + \mathbb{E}_0\left\{\pi(\mathbb{E}[h])\left[N(N+1)\frac{(1+\beta h)^\alpha}{\mathbb{E}[(1+\beta h)^\alpha]} - N - \frac{N^2}{2}\right]\right\} \\
&= \bar{Q}_n + \mathbb{E}_0\left\{\pi(\mathbb{E}[h])\left[N(N+1)\frac{\mathbb{E}\left[(1+\beta h)^\alpha\right]}{\mathbb{E}[(1+\beta h)^\alpha]} - N - \frac{N^2}{2}\right]\right\} \\
&= \bar{Q}_n + \pi(\mathbb{E}[h])\frac{N^2}{2},
\end{aligned}$$

where the second step uses the law of iterated expectations. The expression $u_a(\mathbb{E}[h])$ shows that the ex-ante consumer surplus depends solely on the expected human capital use given the consumer's information $\mathbb{E}[h]$ instead of the actual human capital use h.

In contrast, ex-ante expected welfare is defined as the sum of firm profits and consumer utility, making it a function of both h and $\mathbb{E}[h]$:

$$\begin{aligned}
W(\psi, h, \mathbb{E}[h]) &= \pi(\mathbb{E}[h]) + u_a(\mathbb{E}[h]) - C(h, \psi) + v(h) \\
&= \bar{Q}_n + \left[N + \frac{N^2}{2}\right]\pi(\mathbb{E}[h]) - C(h, \psi) + v(h)
\end{aligned}$$

Therefore, in the Cournot model the terms in Section 4 are given by

$$\begin{aligned}
\pi_G(\mathbb{E}[h]) &= \frac{b_1^2\,\mathbb{E}[(1+\beta h)^\alpha]^2}{4b_2},\\
\pi_D(\mathbb{E}[h]) &= \frac{b_1^2\,\mathbb{E}[(1+\beta h)^\alpha]^2}{(N+1)^2 b_2},\\
u_G(\mathbb{E}[h]) &= \bar{Q}_n + \frac{1}{2}\pi_G(\mathbb{E}[h]),\\
u_D(\mathbb{E}[h]) &= \bar{Q}_n + \left(\frac{1}{2}N^2 + N - 1\right)\pi_D(\mathbb{E}[h]),\\
W_G(\psi, h, \mathbb{E}[h]) &= \bar{Q}_n + \frac{3}{2}\pi_G(\mathbb{E}[h]) - C(h,\psi) + v(h),\\
W_D(\psi, h, \mathbb{E}[h]) &= \bar{Q}_n + \left[N + \frac{N^2}{2}\right]\pi_D(\mathbb{E}[h]) - C(h,\psi) + v(h).
\end{aligned}$$

It can be verified that $\partial\pi_G(\mathbb{E}[h])/\partial\,\mathbb{E}[h] \geq \partial\pi_D(\mathbb{E}[h])/\partial\,\mathbb{E}[h]$, as required for Section 4.

B.1.3 Endogenous Entry

We now extend the Cournot model to allow for endogenous entry. Suppose each of the $N-1$ competing firms incurs a sunk entry cost $K > 0$ before competing in the production stage. We assume K is a function of technological improvements, such as improved AI models, that can reduce both the cost of copying the product K and $\overline{\psi}$. Moreover, K is product specific: some products, such as digital music, may be easily copied and thus have a low K, while other products, such as brand-specific advertising campaigns, might be difficult to copy and thus have a high K.

Firm 1 has already borne its own sunk cost $C(h,\psi)$ at the innovation stage, so it does not pay K. Free entry drives each competitor's net profit to zero, such that

$$\pi(\mathbb{E}[h]) = K, \tag{A3}$$

where $\pi(\mathbb{E}[h])$ is given by (A2). Substituting, the equilibrium total number of firms N^* satisfies

$$\frac{b_1^2\,\mathbb{E}[(1+\beta h)^\alpha]^2}{(N^*+1)^2\,b_2} = K \quad\Longrightarrow\quad N^*(\mathbb{E}[h]) = \frac{b_1\,\mathbb{E}[(1+\beta h)^\alpha]}{\sqrt{Kb_2}} - 1. \tag{A4}$$

For an interior competitive equilibrium with at least one entrant beyond firm 1, we require $b_1\,\mathbb{E}[(1+\beta h)^\alpha] > 2\sqrt{Kb_2}$.

The effect of advancements in AI technology. As discussed in the main text following Proposition 2, the same technology that reduces the fixed cost of implementing AI may also reduce the cost of copying a product if the underlying intangible asset did not receive protection. This extension formalizes this discussion. From Corollary 1, a necessary condition for the AI to "kill" the IP system is that $\pi_D \geq \overline{\psi}$. With endogenous entry, this condition becomes $K \geq \overline{\psi}$, demonstrating formally that whether AI "kills" the IP system boils down to a horse race between how fast AI drives down the cost of innovation versus the cost of copying products.

Equilibrium outcomes. Substituting (A4) into the expression for $p(Q)$:

$$p^* = \frac{b_1 \, \mathbb{E}[(1+\beta h)^\alpha]}{N^* + 1} = \sqrt{K b_2}, \tag{A5}$$

which is independent of human-capital use. Each firm produces $q^* = \sqrt{K/b_2}$, so aggregate output is

$$Q^* = N^* q^* = \left(\frac{b_1 \, \mathbb{E}[(1+\beta h)^\alpha]}{\sqrt{K b_2}} - 1 \right) \sqrt{\frac{K}{b_2}}. \tag{A6}$$

Industry output is therefore increasing in $\mathbb{E}[h]$ when $\beta > 0$.

At the endogenous equilibrium, every competitor earns gross profit $\pi^* = K$ and net profit equals zero. Firm 1 also earns gross profit K in the competitive stage, but its net payoff includes the sunk innovation cost $C(h, \psi)$. From (A4),

$$\frac{\partial N^*}{\partial \mathbb{E}[h]} = \frac{b_1 \, \frac{\partial}{\partial \mathbb{E}[h]} \mathbb{E}[(1+\beta h)^\alpha]}{\sqrt{K b_2}} \geq 0, \tag{A7}$$

with strict inequality when $\beta > 0$. Higher expected human-capital use raises demand, thereby attracting more entrants until net profits are zero. Thus, while entry is increasing in human-capital use, competitive profits π_D are independent of human-capital use.

Consumer surplus and welfare. Because each competitor earns zero net profit, u_a simplifies: the $(N^* - 1)$ competitors contribute nothing to aggregate surplus beyond their own entry costs, which are exactly offset by their gross profits. Hence

$$u_a = U(Q_n, Q). \tag{A8}$$

To compute $U(Q_n, Q)$ in ex-ante terms, we substitute Q^* and use $b_1 \mathbb{E}[(1+\beta h)^\alpha] = (N^* + 1)\sqrt{K b_2}$:

$$\begin{aligned} \mathbb{E}_0[U(Q_n, Q)] &= \bar{Q}_n + Q^*(b_1 \mathbb{E}[(1+\beta h)^\alpha] - p^*) - \frac{b_2 Q^{*2}}{2} \\ &= \bar{Q}_n + N^* \sqrt{\tfrac{K}{b_2}} \cdot N^* \sqrt{K b_2} - \frac{b_2}{2} (N^*)^2 \frac{K}{b_2} \\ &= \bar{Q}_n + \frac{(N^*)^2 \, K}{2}. \end{aligned} \tag{A9}$$

Since N^* is increasing in $\mathbb{E}[h]$ when $\beta > 0$, ex-ante consumer surplus $(N^*)^2 K/2$ is also increasing in $\mathbb{E}[h]$, providing an additional channel through which human-capital use affects social welfare.

Ex-ante expected welfare under endogenous entry is

$$W(\psi, h, \mathbb{E}[h]) = \bar{Q}_n + K + \frac{(N^*)^2 \, K}{2} - C(h, \psi) + v(h), \tag{A10}$$

where $N^* = N^*(\mathbb{E}[h])$ from (A4).

Comparison with exogenous entry. Under exogenous N, firm 1's gross profit $\pi_D(\mathbb{E}[h])$ is increasing in $\mathbb{E}[h]$. Under endogenous entry, free entry dissipates this rent: firm 1's

gross profit is pinned at K regardless of $\mathbb{E}[h]$. Therefore, $\frac{\partial \Delta\pi}{\partial h}$ is greater under endogenous entry, amplifying our results that depend on this quantity being positive.

For consumers, the gains from higher $\mathbb{E}[h]$ accrue through greater entry in the case of endogenous entry compared to higher quantity produced per firm in the case of exogenous entry. While in the former case this increases total welfare due to the increased competition, it also reduces welfare due to more firms paying the entry cost. To assess whether the net effect makes welfare more or less sensitive to human-capital use in the case of endogenous entry, we take the following derivatives:

$$\frac{\partial W_{endo}}{\partial \mathbb{E}[(1+\beta h)^\alpha]} = \frac{b_1^2 \mathbb{E}[(1+\beta h)^\alpha]}{b_2} \cdot \frac{N^*}{N^*+1},$$
$$\frac{\partial W_{exo}}{\partial \mathbb{E}[(1+\beta h)^\alpha]} = \frac{b_1^2 \mathbb{E}[(1+\beta h)^\alpha]}{b_2} \cdot \frac{N^*(N^*+2)}{(N^*+1)^2}$$

Based on the above equations, it can be seen that welfare is marginally less sensitive to human-capital use under endogenous entry. For large N^*, however, this difference is small.

B.2 Microfoundation Based on Step-by-Step Innovation

In this section, we provide an alternative microfoundation based on the step-by-step innovation framework of Aghion et al. (2001), adapted to Cournot competition to align with the microfoundation in Online Appendix B.1. This model captures the "escape competition" effect, where firms innovate to move from a state of symmetric competition to a state with a competitive cost advantage. This complements the setup in Online Appendix B.1, where a firm can innovate to develop a new product, whereas this extension models gaining a competitive edge in an existing product market.

Consider a market with N firms: the innovating firm (firm 1) and $N-1$ potential rivals. The firms compete in quantities (Cournot competition). We assume that consumers have the same utility function $U(Q_n, Q_D)$ as in Online Appendix B.1, leading to the same inverse demand function as in Eq. (A1):

$$p(Q) = b_1 \mathbb{E}[(1+\beta h)^\alpha] - b_2 Q,$$

where $Q = \sum_{i=1}^{N} q_i$ is aggregate output, $b_2 > 0$ is a price sensitivity parameter, and the intercept $b_1 \mathbb{E}[(1+\beta h)^\alpha]$ captures the consumer's willingness to pay based on expected human-capital use.

Initially, all firms possess the same technology with constant marginal cost c. In this microfoundation, the innovating firm's investment decision represents a process innovation that reduces its marginal cost to $c' < c$, representing efficiency gains.

Case 1: IP Protection Denied ($a = D$). If IP protection is denied, the rivals imitate the technology. Consequently, the market remains "neck-and-neck" where all N firms operate with the lower marginal cost c'. In a symmetric Cournot equilibrium with identical costs c', each firm produces $q_D = \frac{b_1 \mathbb{E}[(1+\beta h)^\alpha] - c'}{b_2(N+1)}$ and the aggregate output is

$Q_D = Nq_D$. The price is $p_D = \frac{b_1\mathbb{E}[(1+\beta h)^\alpha]+Nc'}{N+1}$. The innovating firm's profit is:

$$\pi_D = \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c')^2}{b_2(N+1)^2}.$$

The ex-ante expected consumer surplus, defined here to include the rivals' profits to capture the total surplus not appropriated by the innovating firm, is:

$$u_D = \mathbb{E}_0[U(Q_n, Q_D)] + (N-1)\pi_D = \bar{Q}_n + \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c')^2}{b_2(N+1)^2}\left(\frac{1}{2}N^2 + N - 1\right).$$

Case 2: IP Protection Granted ($a = G$). If IP protection is granted, imitation is prevented. The market enters a "leader-follower" state. The innovating firm becomes the leader and produces with cost c', while the $N-1$ rivals are followers and retain the old cost c. In this asymmetric Cournot equilibrium, the innovating firm gains a competitive edge. The price is $p_G = \frac{b_1\mathbb{E}[(1+\beta h)^\alpha]+c'+(N-1)c}{N+1}$. The innovating firm's equilibrium quantity is

$$q_G = \frac{b_1\mathbb{E}[(1+\beta h)^\alpha] - c' + (N-1)(c-c')}{b_2(N+1)},$$

and its profit is

$$\pi_G = \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c' + (N-1)(c-c'))^2}{b_2(N+1)^2}.$$

The rivals with marginal cost c produce the following quantities:

$$q_j = \frac{b_1\mathbb{E}[(1+\beta h)^\alpha] + c' - 2c}{b_2(N+1)}, \quad \text{for } j = 2, ..., N,$$

and the rivals' profit is

$$\pi_j = \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] + c' - 2c)^2}{b_2(N+1)^2}, \quad \text{for } j = 2, ..., N.$$

The aggregate output is $Q_G = Q_D - \frac{(N-1)(c-c')}{b_2(N+1)}$, which is strictly lower than Q_D since $c > c'$. The ex-ante expected consumer surplus (plus rivals' profits) is:

$$\begin{aligned} u_G &= \mathbb{E}_0[U(Q_n, Q_G)] + \sum_{j=2}^{N} \pi_j \\ &= \bar{Q}_n + \frac{b_2}{2}\left(\frac{Nb_1\mathbb{E}[(1+\beta h)^\alpha] - c' - (N-1)c}{b_2(N+1)}\right)^2 + (N-1)b_2\left(\frac{b_1\mathbb{E}[(1+\beta h)^\alpha] + c' - 2c}{b_2(N+1)}\right)^2. \end{aligned}$$

Verification of Assumption 1. We verify that these payoffs satisfy the conditions in the main text.

(i) $\pi_G > \pi_D$: Since $c > c'$ and $N \geq 2$, it follows directly that:

$$\pi_G = \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c' + (N-1)(c-c'))^2}{b_2(N+1)^2} > \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c')^2}{b_2(N+1)^2} = \pi_D.$$

Thus, IP protection provides the firm with a competitive edge and higher profits.

(ii) $\pi_D + u_D > \pi_G + u_G$: If protection is denied ($a = D$), all N firms produce at low cost c', and aggregate output is Q_D. Therefore, $\pi_D + u_D$ is given by

$$\pi_D + u_D = \bar{Q}_n + \frac{N(N+2)(b_1\mathbb{E}[(1+\beta h)^\alpha] - c')^2}{2b_2(N+1)^2}.$$

If protection is granted ($a = G$), $N-1$ firms produce at high cost c, and aggregate output is $Q_G < Q_D$ (since $c > c'$). Therefore, $\pi_G + u_G$ is given by

$$\begin{aligned}
&\pi_G + u_G \\
&= \bar{Q}_n + \frac{b_2}{2}\left(\frac{Nb_1\mathbb{E}[(1+\beta h)^\alpha] - c' - (N-1)c}{b_2(N+1)}\right)^2 + (N-1)b_2\left(\frac{b_1\mathbb{E}[(1+\beta h)^\alpha] + c' - 2c}{b_2(N+1)}\right)^2 \\
&\quad + \frac{(b_1\mathbb{E}[(1+\beta h)^\alpha] - c' + (N-1)(c-c'))^2}{b_2(N+1)^2}.
\end{aligned}$$

Therefore, the difference is

$$(\pi_D+u_D)-(\pi_G+u_G) = \frac{(N-1)(c-c')}{2b_2(N+1)^2}\Big(2(N+2)(b_1\mathbb{E}[(1+\beta h)^\alpha]-c')-(3N+5)(c-c')\Big) > 0.$$

The difference is positive, since $q_j \geq 0$ implies that $b_1\mathbb{E}[(1+\beta h)^\alpha] - c' \geq 2(c-c')$.

This structure is consistent with Assumption 1. The innovating firm makes an investment in process innovation to escape the lower profits of symmetric competition (π_D) and achieve the higher profits of a cost leader (π_G), even though social welfare is maximized when the innovation diffuses (W_D).

B.3 Optimal IP Policy with Consumer Preference

In this section, we derive the optimal IP policy when the consumer has a preference for human-developed products. This involves comparing the maximum welfare achievable in the cases where one or both types of firms invest in innovation. The latter strategy incentivizes participation from both types of firms but suffers from the distorted belief $\mathbb{E}[h^*]$. Alternatively, the officer may choose a policy that screens out the high-cost type from investing. This results in an equilibrium where consumer beliefs are based solely on the low-cost type, consequently inducing a higher human-capital use h_L^* from the low-cost firm, but at the cost of sacrificing the potential welfare contribution from the high-cost type. The possibility of market collapse (i.e., neither invests) occurs if even the most favorable policy cannot make investment profitable for the low-cost type under the prevailing beliefs.

Welfare for a given type ψ that invests is:

$$\begin{aligned}
W(\psi, q, \Delta q) =& (q + \Delta q h^*)\big(\pi_G(\mathbb{E}[h^*]) + u_G(\mathbb{E}[h^*])\big) \\
&+ (1 - (q + \Delta q h^*))\big(\pi_D(\mathbb{E}[h^*]) + u_D(\mathbb{E}[h^*])\big) + v(h^*) - C(h^*, \psi),
\end{aligned}$$

where h^* is the firm's human-capital choice. Notice that the consumer surplus and the firm profit depend on $\mathbb{E}[h^*]$. This is because consumer decisions are made based on the perceived human-capital use $\mathbb{E}[h^*]$ (which sets the market price).

The IP officer's problem is to establish an IP system that maximizes welfare. This

involves choosing q and Δq to balance the social benefits of innovation and human capital with the social cost of monopoly power. The officer's maximization problem is

$$\max_{0\leq q,0\leq \Delta q,q+\Delta q\leq 1} \quad p\mathbf{1}_{\Pi_L\geq 0}W(\psi_L,q,\Delta q)+(1-p)\mathbf{1}_{\Pi_H\geq 0}W(\psi_H,q,\Delta q) \tag{A11}$$

where the participation constraint for a firm with cost ψ is:

$$\Pi(\psi,\mathbb{E}[h^*])=\Big(q+\Delta q h^*\Big)\pi_G(\mathbb{E}[h^*])+\Big(1-\big(q+\Delta q h^*\big)\Big)\pi_D(\mathbb{E}[h^*])-C(h^*,\psi)\geq 0.$$

Introducing a consumer preference for h adds a social benefit for the use of human capital, which then affects the officer's incentives when setting the optimal IP policy. The following proposition characterizes the optimal IP policy.

Proposition 6. *The welfare-maximizing IP policy is characterized as follows:*

(i) If it is optimal to induce both firm types to invest in innovation, the optimal policy is

$$q^*=\frac{\overline{\psi}-\pi_D}{\Delta\pi}-\frac{(\Delta q^*)^2\Delta\pi}{2\psi_H},$$
$$\Delta q^*=\frac{\gamma\left(\frac{p}{\psi_L}+\frac{1-p}{\psi_H}\right)+\frac{\partial\mathbb{E}[W]}{\partial\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}\frac{1}{\Delta\pi}}{-\left(\left(\Delta\pi+2\Delta u\right)\left(\frac{p}{\psi_L}+\frac{1-p}{\psi_H}\right)-\frac{\Delta\pi+\Delta u}{\psi_H}\right)},$$

where π_D, $\Delta\pi$ and Δu are functions of $\mathbb{E}[h^*]$.

(ii) If it is optimal to induce only the low-cost firm to invest in innovation, the optimal policy is

$$q^*=\frac{\overline{\psi}-\pi_D}{\Delta\pi}-\frac{(\Delta q^*)^2\Delta\pi}{2\psi_L},$$
$$\Delta q^*=\frac{\gamma+\frac{\partial\mathbb{E}[W]}{\partial h_L^*}\frac{dh_L^*}{d\Delta q}\frac{\psi_L}{p\Delta\pi}}{-\Delta u},$$

where π_D, $\Delta\pi$, and Δu are functions of the rational equilibrium investment h_L^*.

Proof. See Online Appendix B.3.1. □

The base grant probability q remains structurally analogous to that in the baseline model. It functions as an innovation subsidy to satisfy the firm's participation constraint. The key difference is that the constraint is now endogenous to the consumer's equilibrium beliefs.

The slope Δq, however, diverges from that in the baseline model. While it still balances the social benefit $v(h)$ with the social costs of monopoly and costly human-capital use, consumer preferences introduce a positive "belief wedge," captured by the term containing $\frac{\partial\mathbb{E}[W]}{\partial\mathbb{E}[h^*]}$ and $\frac{\partial\mathbb{E}[W]}{\partial h_L^*}$, which drives the officer to set steeper incentives than in the baseline model. This additional term captures the officer's ability to utilize the IP policy as a signal of human-capital use for the consumer. By committing to a steeper slope Δq, the officer induces higher equilibrium human-capital use, which the rational consumer correctly infers. This shift in beliefs triggers the feedback loop: it increases the

consumer's willingness to pay and the firm's expected profits (π_D and $\Delta\pi$), relaxing the firm's participation constraint. Crucially, this relaxation allows the officer to reduce the socially costly base grant probability q^*. The optimal policy "over-incentivizes" human-capital use relative to its social value γ to trigger this feedback loop and to minimize the overall welfare loss required to ensure firm investment in innovation.

B.3.1 Proof of Proposition 6

Proof. In this proof, we derive the optimal IP policy and expected welfare for two cases: (i) the pooling equilibrium, where the officer induces investment by both firm types, and (ii) the separating equilibrium, where the officer induces investment only by the low-cost type.

In the pooling equilibrium, the IP officer maximizes expected social welfare subject to the constraint that both firm types invest in innovation. Let $\mathbb{E}[h^*]$ denote the consumer's beliefs about human-capital use that is consistent with the equilibrium outcome. The officer's problem is given by:

$$\begin{aligned} &\max_{q,\Delta q} \mathbb{E}[W] \\ &\text{s.t.} \quad \Pi(\psi_H, \mathbb{E}[h^*]) \geq 0. \end{aligned} \tag{A12}$$

Where the participation constraint for the firm with cost ψ_H is:

$$\Pi(\psi_H, \mathbb{E}[h^*]) = \pi_D(\mathbb{E}[h^*]) + q\Delta\pi(\mathbb{E}[h^*]) + \frac{\Big(\Delta q \Delta\pi(\mathbb{E}[h^*])\Big)^2}{2\psi_H} - \overline{\psi} \geq 0. \tag{A13}$$

We first solve the firm's participation constraint for q. Given that the expected welfare decreases with the base granting probability q, the optimal q is obtained by solving the binding participation constraint for the firm with cost ψ_H,

$$q^* = \frac{\overline{\psi} - \pi_D(\mathbb{E}[h^*])}{\Delta\pi(\mathbb{E}[h^*])} - \frac{(\Delta q)^2 \Delta\pi(\mathbb{E}[h^*])}{2\psi_H}. \tag{A14}$$

We substitute q^* into the objective function, reducing the dimensionality of the maximization problem to a single choice variable Δq. The unconstrained maximization problem becomes:

$$\begin{aligned} \max_{\Delta q} \mathbb{E}[W] = {} & \pi_D(\mathbb{E}[h^*]) + u_D(\mathbb{E}[h^*]) - \overline{\psi} \\ & + \left(\frac{\overline{\psi} - \pi_D(\mathbb{E}[h^*])}{\Delta\pi(\mathbb{E}[h^*])} - \frac{(\Delta q)^2 \Delta\pi(\mathbb{E}[h^*])}{2\psi_H} \right) \Big(\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*]) \Big) \\ & + (\Delta q)^2 \Delta\pi(\mathbb{E}[h^*]) \Big(\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*]) \Big) \mathbb{E}\left[\frac{1}{\psi}\right] \\ & + \gamma \Delta q \Delta\pi(\mathbb{E}[h^*])\, \mathbb{E}\left[\frac{1}{\psi}\right] - \frac{\Big(\Delta q \Delta\pi(\mathbb{E}[h^*])\Big)^2}{2} \mathbb{E}\left[\frac{1}{\psi}\right] \end{aligned} \tag{A15}$$

where $\mathbb{E}\left[\frac{1}{\psi}\right] = \frac{p}{\psi_L} + \frac{1-p}{\psi_H}$. The first-order condition of Δq is:

$$\frac{d\,\mathbb{E}[W]}{d\Delta q} = \frac{\partial\,\mathbb{E}[W]}{\partial \Delta q} + \frac{\partial\,\mathbb{E}[W]}{\partial\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q} = 0. \tag{A16}$$

We derive the partial derivative $\frac{\partial\,\mathbb{E}[W]}{\partial \Delta q}$ by holding $\mathbb{E}[h^*]$ constant. This partial derivative simplifies to:

$$\begin{aligned}\frac{\partial\,\mathbb{E}[W]}{\partial \Delta q} =& \gamma \Delta\pi(\mathbb{E}[h^*])\,\mathbb{E}\left[\frac{1}{\psi}\right] + \Delta q \Delta\pi(\mathbb{E}[h^*])\,\mathbb{E}\left[\frac{1}{\psi}\right]\Big(\Delta\pi(\mathbb{E}[h^*]) + 2\Delta u(\mathbb{E}[h^*])\Big) \\ &- \Delta q \Delta\pi(\mathbb{E}[h^*])\frac{\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*])}{\psi_H}.\end{aligned} \tag{A17}$$

Substituting this expression back into the total first-order condition yields:

$$\begin{aligned}&\gamma \Delta\pi(\mathbb{E}[h^*])\,\mathbb{E}\left[\frac{1}{\psi}\right] + \Delta q \Delta\pi(\mathbb{E}[h^*])\left(\mathbb{E}\left[\frac{1}{\psi}\right]\Big(\Delta\pi(\mathbb{E}[h^*]) + 2\Delta u(\mathbb{E}[h^*])\Big) - \frac{\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*])}{\psi_H}\right) \\ &+ \frac{\partial\,\mathbb{E}[W]}{\partial\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q} = 0.\end{aligned} \tag{A18}$$

Rearranging the terms to isolate Δq^* on the left-hand side, we obtain:

$$\begin{aligned}&- \Delta q^* \Delta\pi(\mathbb{E}[h^*])\left(\mathbb{E}\left[\frac{1}{\psi}\right]\Big(\Delta\pi(\mathbb{E}[h^*]) + 2\Delta u(\mathbb{E}[h^*])\Big) - \frac{\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*])}{\psi_H}\right) \\ &= \gamma \Delta\pi(\mathbb{E}[h^*])\,\mathbb{E}\left[\frac{1}{\psi}\right] + \frac{\partial\,\mathbb{E}[W]}{\partial\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}.\end{aligned} \tag{A19}$$

Dividing both sides by the term multiplying Δq^* and subsequently dividing the numerator and denominator by $\Delta\pi(\mathbb{E}[h^*])$, we arrive at the optimal slope characterized in the lemma:

$$\Delta q^* = \frac{\gamma\,\mathbb{E}\left[\frac{1}{\psi}\right] + \frac{\partial\,\mathbb{E}[W]}{\partial\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}\frac{1}{\Delta\pi(\mathbb{E}[h^*])}}{-\left(\Big(\Delta\pi(\mathbb{E}[h^*]) + 2\Delta u(\mathbb{E}[h^*])\Big)\,\mathbb{E}\left[\frac{1}{\psi}\right] - \frac{\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*])}{\psi_H}\right)}. \tag{A20}$$

The derivative $\frac{\partial\,\mathbb{E}[W]}{\partial\,\mathbb{E}[h^*]}$ is derived from taking derivative of $\mathbb{E}[W]$ with respect to the con-

sumer beliefs $\mathbb{E}[h^*]$:

$$
\begin{aligned}
&\frac{\partial \mathbb{E}[W]}{\partial\, \mathbb{E}[h^*]} \\
&= \Big(\pi_D'(\mathbb{E}[h^*]) + u_D'(\mathbb{E}[h^*])\Big) \\
&\quad + \Big(\Delta\pi'(\mathbb{E}[h^*]) + \Delta u'(\mathbb{E}[h^*])\Big)\left(\frac{\overline{\psi} - \pi_D(\mathbb{E}[h^*])}{\Delta\pi(\mathbb{E}[h^*])} - \frac{(\Delta q)^2 \Delta\pi(\mathbb{E}[h^*])}{2\psi_H} + \Delta q \mathbb{E}[h^*]\right) \\
&\quad + \Big(\Delta\pi(\mathbb{E}[h^*]) + \Delta u(\mathbb{E}[h^*])\Big)\left(\frac{-\pi_D'(\mathbb{E}[h^*])}{\Delta\pi(\mathbb{E}[h^*])} - \frac{\Big(\overline{\psi} - \pi_D(\mathbb{E}[h^*])\Big)\Delta\pi'(\mathbb{E}[h^*])}{(\Delta\pi(\mathbb{E}[h^*]))^2} - \frac{(\Delta q)^2 \Delta\pi'(\mathbb{E}[h^*])}{2\psi_H}\right) \\
&\quad + \Delta\pi'(\mathbb{E}[h^*]) \Delta q \mathbb{E}\left[\frac{1}{\psi}\right]\Big(\gamma + \Delta q \Delta u(\mathbb{E}[h^*])\Big).
\end{aligned}
$$

The derivative $\frac{d\,\mathbb{E}[h^*]}{d\Delta q}$ is derived as follows. In equilibrium, the consumer beliefs must be consistent with the firm's human-capital use. Therefore,

$$
\mathbb{E}[h^*] = \Delta q \Delta\pi(\mathbb{E}[h^*]) \mathbb{E}\left[\frac{1}{\psi}\right].
$$

Applying the implicit function theorem, the derivative of $\mathbb{E}[h^*]$ with respect to Δq is

$$
\frac{d\,\mathbb{E}[h^*]}{d\Delta q} = \frac{\Delta\pi(\mathbb{E}[h^*]) \mathbb{E}\left[\frac{1}{\psi}\right]}{1 - \Delta q \Delta\pi'(\mathbb{E}[h^*]) \mathbb{E}\left[\frac{1}{\psi}\right]}.
$$

In the separating equilibrium, the IP officer maximizes the expected social welfare generated by the low-cost firm, subject to the constraint that the low-cost firm invests in innovation. In a separating equilibrium, the consumer correctly infers that the entrant is the low-cost type, so their belief is exactly the firm's optimal human-capital choice: $\mathbb{E}[h] = h_L^*$. The officer's problem is:

$$
\begin{aligned}
&\max_{q, \Delta q} \ \mathbb{E}[W] = pW(\psi_L) \\
&\text{s.t.} \quad \Pi(\psi_L, h_L^*) \geq 0.
\end{aligned}
\tag{A21}
$$

Where the participation constraint for the firm with cost ψ_L is:

$$
\Pi(\psi_L, h_L^*) = \pi_D(h_L^*) + q\Delta\pi(h_L^*) + \frac{(\Delta q \Delta\pi(h_L^*))^2}{2\psi_L} - \overline{\psi} \geq 0. \tag{A22}
$$

We solve the firm's participation constraint for q:

$$
q^* = \frac{\overline{\psi} - \pi_D(h_L^*)}{\Delta\pi(h_L^*)} - \frac{(\Delta q)^2 \Delta\pi(h_L^*)}{2\psi_L}. \tag{A23}
$$

We substitute q^* into the objective function. The unconstrained maximization problem

becomes:

$$\begin{aligned}\max_{\Delta q} \mathbb{E}[W] = p\Bigg(& \pi_D(h_L^*) + u_D(h_L^*) - \overline{\psi} \\ & + \left(\frac{\overline{\psi} - \pi_D(h_L^*)}{\Delta\pi(h_L^*)} - \frac{(\Delta q)^2 \Delta\pi(h_L^*)}{2\psi_L} \right) \Big(\Delta\pi(h_L^*) + \Delta u(h_L^*) \Big) \\ & + \frac{(\Delta q)^2 \Delta\pi(h_L^*)}{\psi_L} \Big(\Delta\pi(h_L^*) + \Delta u(h_L^*) \Big) \\ & + \frac{\gamma \Delta q \Delta\pi(h_L^*)}{\psi_L} - \frac{\Big(\Delta q \Delta\pi(h_L^*) \Big)^2}{2\psi_L} \Bigg)\end{aligned} \tag{A24}$$

The first-order condition with respect to Δq is:

$$\frac{d\mathbb{E}[W]}{d\Delta q} = \frac{\partial \mathbb{E}[W]}{\partial \Delta q} + \frac{\partial \mathbb{E}[W]}{\partial h_L^*} \frac{dh_L^*}{d\Delta q} = 0. \tag{A25}$$

We derive the partial derivative $\frac{\partial \mathbb{E}[W]}{\partial \Delta q}$ holding h_L^* constant:

$$\begin{aligned}\frac{\partial \mathbb{E}[W]}{\partial \Delta q} &= p \left(\frac{\gamma \Delta\pi(h_L^*)}{\psi_L} + \frac{\Delta q \Delta\pi(h_L^*)}{\psi_L} \Big(\Delta\pi(h_L^*) + 2\Delta u(h_L^*) \Big) - \frac{\Delta q \Delta\pi(h_L^*)}{\psi_L} \Big(\Delta\pi(h_L^*) + \Delta u(h_L^*) \Big) \right) \\ &= p \left(\frac{\gamma \Delta\pi(h_L^*)}{\psi_L} + \frac{\Delta q \Delta\pi(h_L^*) \Delta u(h_L^*)}{\psi_L} \right).\end{aligned} \tag{A26}$$

Substituting this back into the total first-order condition:

$$p \frac{\Delta\pi(h_L^*)}{\psi_L} \Big(\gamma + \Delta q \Delta u(h_L^*) \Big) + \frac{\partial \mathbb{E}[W]}{\partial h_L^*} \frac{dh_L^*}{d\Delta q} = 0. \tag{A27}$$

Rearranging to isolate Δq^*:

$$-\Delta q^* \left(p \frac{\Delta\pi(h_L^*) \Delta u(h_L^*)}{\psi_L} \right) = p \frac{\gamma \Delta\pi(h_L^*)}{\psi_L} + \frac{\partial \mathbb{E}[W]}{\partial h_L^*} \frac{dh_L^*}{d\Delta q}. \tag{A28}$$

Dividing both sides by the term multiplying Δq^* and simplifying the fraction yields the result in the lemma:

$$\Delta q^* = \frac{\gamma + \frac{\partial \mathbb{E}[W]}{\partial h_L^*} \frac{dh_L^*}{d\Delta q} \frac{\psi_L}{p \Delta\pi}}{-\Delta u}. \tag{A29}$$

The derivative $\frac{\partial \mathbb{E}[W]}{\partial h_L^*}$ is derived by taking the derivative of the substituted $\mathbb{E}[W]$ with

respect to h_L^*:

$$\begin{aligned}\frac{\partial \mathbb{E}[W]}{\partial h_L^*} = p\Bigg(& (\pi_D'(h_L^*) + u_D'(h_L^*)) \\ & + \Big(\Delta\pi'(h_L^*) + \Delta u'(h_L^*)\Big)\left(\frac{\overline{\psi} - \pi_D(h_L^*)}{\Delta\pi(h_L^*)} - \frac{(\Delta q)^2\Delta\pi(h_L^*)}{2\psi_L} + \Delta q h_L^*\right) \\ & + \Big(\Delta\pi(h_L^*) + \Delta u(h_L^*)\Big)\left(\frac{-\pi_D'(h_L^*)}{\Delta\pi(h_L^*)} - \frac{\Big(\overline{\psi} - \pi_D(h_L^*)\Big)\Delta\pi'(h_L^*)}{(\Delta\pi(h_L^*))^2} - \frac{(\Delta q)^2\Delta\pi'(h_L^*)}{2\psi_L}\right) \\ & + \frac{\Delta q\Delta\pi'(h_L^*)}{\psi_L}\Big(\gamma + \Delta q\Delta u(h_L^*)\Big)\Bigg).\end{aligned}$$

Finally, $\frac{dh_L^*}{d\Delta q}$ is derived from the firm's optimal choice condition in equilibrium, $h_L^* = \frac{\Delta q\Delta\pi(h_L^*)}{\psi_L}$:

$$\frac{dh_L^*}{d\Delta q} = \frac{\Delta\pi(h_L^*)/\psi_L}{1 - \frac{\Delta q\Delta\pi'(h_L^*)}{\psi_L}}.$$

□

B.4 Extended Adverse Selection Analysis

This appendix extends Section 4.2.1 along two dimensions. Section B.4.1 analyzes how the slope Δq reshapes the equilibrium thresholds in Proposition 4, complementing the main text, which fixes Δq and varies q. Section B.4.2 then introduces an alternative specification in which the high-ψ firm is "AI focused," i.e., less efficient at using human capital ($\psi_H > \psi_L$) but has greater potential to bring down innovation cost ($\overline{\psi}_H < \overline{\psi}_L$) by substituting AI for human input. We show that in this alternative specification, the base grant probability q may lose its ability to selectively deter the high-ψ type, while the slope Δq retains this role.

B.4.1 The Role of Δq in Investment

We start by revisiting the three thresholds (q_L, q_H, q_B) defined in Appendix A.8. Differentiating the defining equations (15), (16), and (18) implicitly with respect to Δq yields

$$\frac{dq_L}{d\Delta q} = -\frac{1}{\Delta\pi(h_L^*)}\left\{\frac{\partial\Pi^*(\psi_L, h_L^*, q_L, \Delta q)}{\partial\Delta q} + \frac{\partial\Pi^*(\psi_L, h_L^*, q_L, \Delta q)}{\partial h_L^*}\frac{dh_L^*}{d\Delta q}\right\}, \tag{A30}$$

$$\frac{dq_H}{d\Delta q} = -\frac{1}{\Delta\pi(h_L^*)}\left\{\frac{\partial\Pi^*(\psi_H, h_L^*, q_H, \Delta q)}{\partial\Delta q} + \frac{\partial\Pi^*(\psi_H, h_L^*, q_H, \Delta q)}{\partial h_L^*}\frac{dh_L^*}{d\Delta q}\right\}, \tag{A31}$$

$$\frac{dq_B}{d\Delta q} = -\frac{1}{\Delta\pi(\mathbb{E}[h^*])}\left\{\frac{\partial\Pi^*(\psi_H, \mathbb{E}[h^*], q_B, \Delta q)}{\partial\Delta q} + \frac{\partial\Pi^*(\psi_H, \mathbb{E}[h^*], q_B, \Delta q)}{\partial\,\mathbb{E}[h^*]}\frac{d\,\mathbb{E}[h^*]}{d\Delta q}\right\}. \tag{A32}$$

By Lemma 3, the expected profit Π^* is strictly increasing in Δq and in the relevant equilibrium consumer belief, and the direct effect and feedback loop in Section 4.2.2 gives

$dh_L^*/d\Delta q > 0$ and $d\,\mathbb{E}[h^*]/d\Delta q > 0$ by Eq. (9). Therefore every bracketed term above is strictly positive, so

$$\frac{dq_L}{d\Delta q} < 0, \qquad \frac{dq_H}{d\Delta q} < 0, \qquad \frac{dq_B}{d\Delta q} < 0. \tag{A33}$$

This implies that raising Δq uniformly relaxes every participation constraint and reduces the q needed to incentivize investment. Economically, Δq strengthens the IP incentive $\frac{(\Delta q \Delta \pi)^2}{2\psi}$ and, via the feedback loop, raises consumer willingness to pay. Both channels lead to increased profits and lower the required innovation subsidy.

We next explore the relative effect Δq has on the high- and low-type firms' participation constraints. Subtracting (15) from (16), the $\pi_D(h_L^*)$ and $\overline{\psi}$ terms cancel, and dividing through by $\Delta\pi(h_L^*) > 0$ yields a closed form for the gap:

$$q_H - q_L \;=\; \frac{(\Delta q)^2\,\Delta\pi(h_L^*)}{2}\left(\frac{1}{\psi_L} - \frac{1}{\psi_H}\right). \tag{A34}$$

This quantity is strictly positive since $\psi_L < \psi_H$. Differentiating (A34) with respect to Δq, and using $\Delta\pi'(\cdot) > 0$ together with $dh_L^*/d\Delta q > 0$ from Eq. (9),

$$\frac{d(q_H - q_L)}{d\Delta q} \;=\; \left(\frac{1}{\psi_L} - \frac{1}{\psi_H}\right)\Delta q\left[\Delta\pi(h_L^*) + \frac{\Delta q}{2}\,\Delta\pi'(h_L^*)\,\frac{dh_L^*}{d\Delta q}\right] > 0. \tag{A35}$$

Economically, the IP-incentive component $(\Delta q \Delta\pi)^2/(2\psi)$ is scaled by $1/\psi$, so the low-cost firm captures a larger marginal benefit from a higher Δq than the high-cost firm. Raising Δq therefore slackens the low-cost participation constraint faster, and the feedback loop reinforces this through the induced rise in h_L^*. That is, the "low-cost-only" (LI) region $[q_L, q_H)$ widens as Δq increases.

The following proposition summarizes the results from this section:

Proposition 7. *Consider the thresholds (q_L, q_H, q_B) from Proposition 4. Then*

(i) q_L, q_H, and q_B are all strictly decreasing in Δq;

(ii) the LI region $[q_L, q_H)$ strictly widens in Δq.

Proof. Part (i) follows from (A33). Part (ii) follows from (A35). □

Proposition 7 highlights that Δq plays a role in screening firms to avoid inefficient investment decisions resulting from adverse selection. Because Δq rewards human capital at the margin and the low-cost firm uses more human capital, a higher Δq disproportionately slackens the low-cost firm's participation constraint. This widens the window of q over which only the low-cost firm invests, mitigating the adverse selection problem. This role is complementary to Δq's role in mitigating distortion in human-capital use, discussed in Section 4.2.2.

B.4.2 Heterogeneity in $\overline{\psi}$ Firm

We now depart from the main-text assumption of a common fixed cost and allow $\overline{\psi}$ to be type-dependent. This alternative specification is particularly relevant to the setting with adverse selection because it can generate a different ordering of investment thresholds.

Write $\overline{\psi}_i$ for $i \in \{L, H\}$ and assume

$$\psi_L < \psi_H \qquad \text{and} \qquad \overline{\psi}_H < \overline{\psi}_L. \tag{A36}$$

We interpret the high-ψ firm as "AI focused" in the sense that it is less efficient in using human capital ($\psi_H > \psi_L$) but has greater potential to bring down innovation costs ($\overline{\psi}_H < \overline{\psi}_L$) by substituting AI for human capital. This interpretation captures differential access to or specialization in AI and human capital. The firm's expected profit under belief $\mathbb{E}[h]$ and policy $(q, \Delta q)$ becomes

$$\Pi^*(\psi, \mathbb{E}[h], q, \Delta q) = q\Delta\pi(\mathbb{E}[h]) + \pi_D(\mathbb{E}[h]) + \frac{\left(\Delta q\, \Delta\pi(\mathbb{E}[h])\right)^2}{2\psi} - \overline{\psi}. \tag{A37}$$

All other elements of the model are unchanged.

In the model of Section 4, the low-ψ firm is more profitable than the high-ψ firm ($\Pi^*(\psi_L) > \Pi^*(\psi_H)$), and so the high-$\psi$ firm's participation constraint is more difficult to satisfy. Under (A36), however, the ordering of expected profits across types at a common belief $\mathbb{E}[h]$ is no longer unambiguous:

$$\Pi^*(\psi_H, \mathbb{E}[h], q, \Delta q) - \Pi^*(\psi_L, \mathbb{E}[h], q, \Delta q) = (\overline{\psi}_L - \overline{\psi}_H) - \frac{(\Delta q\, \Delta\pi(\mathbb{E}[h]))^2}{2}\left(\frac{1}{\psi_L} - \frac{1}{\psi_H}\right). \tag{A38}$$

The first term captures the high-ψ firm's fixed-cost advantage from AI. The second captures the low-ψ firm's marginal-cost advantage, which scales in $(\Delta q)^2$. The relative size of the two cost advantages determines which firm is more profitable at the common belief, and thus determines which firm's participation constraint is most difficult to satisfy. This results in two regimes.

Lemma 4. *Define the crossing value*

$$\Delta q^\dagger(\mathbb{E}[h]) \equiv \frac{1}{\Delta\pi(\mathbb{E}[h])}\sqrt{\frac{2\,\psi_L\,\psi_H\,(\overline{\psi}_L - \overline{\psi}_H)}{\psi_H - \psi_L}}. \tag{A39}$$

For $\Delta q < \Delta q^\dagger(\mathbb{E}[h])$, the "reversed regime" holds: $\Pi^*(\psi_H) > \Pi^*(\psi_L)$. For $\Delta q > \Delta q^\dagger(\mathbb{E}[h])$, the "standard regime" holds: $\Pi^*(\psi_L) > \Pi^*(\psi_H)$, as in Proposition 4.

Proof. The crossing value is derived from setting Eq. (A38) to zero and re-arranging for Δq. □

The crossing value $\Delta q^\dagger$ is increasing in the fixed-cost gap $\overline{\psi}_L - \overline{\psi}_H$ and decreasing in the marginal-cost gap $\psi_H - \psi_L$. Intuitively, a larger AI-driven fixed-cost advantage requires a steeper IP slope to restore the standard profit ordering, while a larger marginal-cost advantage for the low-ψ firm lowers the required slope.

The standard regime replicates the structure of Proposition 4 (with thresholds that shift to reflect the type-specific fixed costs). We now analyze the "reversed regime." Focus on the case where $\Delta q < \Delta q^\dagger(h_L^*)$. Since $h_L^* > \mathbb{E}[h^*] > h_H^*$ and $\Delta q^\dagger(\cdot)$ is decreasing in the belief, this condition also implies $\Delta q < \Delta q^\dagger(\mathbb{E}[h^*])$ and $\Delta q < \Delta q^\dagger(h_H^*)$. Hence the high-$\psi$ firm is more profitable than the low-ψ firm under all relevant beliefs: h_H^*, the pooled belief $\mathbb{E}[h^*]$, and h_L^*. The construction in Appendix A.8 carries over with L and H interchanged:

- Let h_H^* solve the fixed-point $h_H^* = \Delta q\, \Delta\pi(h_H^*)/\psi_H$, the belief in the "only H invests" (HI) equilibrium.

- Define q'_H as the threshold at which $\Pi^*(\psi_H, h^*_H, q, \Delta q) = 0$, and q'_L as the threshold at which $\Pi^*(\psi_L, h^*_H, q, \Delta q) = 0$.
- Define q'_B as the threshold at which, under the pooled belief $\mathbb{E}[h^*]$, the low-ψ firm (now the marginal entrant) just breaks even.

Provided the analogue of Assumption 1 holds with $\overline{\psi}$ replaced by $\overline{\psi}_L$ (the larger of the two fixed costs), Lemma 3 applies to each type separately, since $\overline{\psi}_i$ enters Π^* additively and drops out of the derivatives in q and Δq. Each type's participation constraint is therefore strictly increasing in q for a given fixed belief.

However, when it comes to the equilibrium structure, one substantial difference arises compared to Proposition 4. The reason is that, in the reversed regime, a low-ψ entrant raises the pooled belief since $h^*_L > h^*_H$. This is in contrast to Proposition 4, where the high-ψ entrant lowers the pooled belief $\mathbb{E}[h^*]$. The marginal entrant therefore benefits from a positive rather than negative informational externality on the pooled belief. Therefore, the inequality between the two thresholds flips: $q'_B < q'_L$. Thus, there is no market-breakdown region. Instead, the high-ψ firm may invest alone over $[q'_H, q'_B)$, and both types invest above q'_B.

Proposition 8. *Assume (A36) and fix $\Delta q < \Delta q^\dagger(h^*_L)$. There exist thresholds q'_H and q'_B such that the unique Pareto-dominant equilibrium takes one of two forms:*

(a) If $q'_H < q'_B$:

(i) If $q < q'_H$, no investment occurs.

(ii) If $q'_H \le q < q'_B$, only the high-ψ firm invests.

(iii) If $q \ge q'_B$, both types of firm invest.

(b) If $q'_B \le q'_H$:

(i) If $q < q'_B$, no investment occurs.

(ii) If $q \ge q'_B$, both types of firm invest.

Proof. Eq. (A37) gives the expected profit with type-specific fixed costs $\overline{\psi}_i$. Lemma 3 continues to hold pointwise in $\overline{\psi}_i$: $\partial\Pi^*/\partial q = \Delta\pi(\mathbb{E}[h^*]) > 0$ and $\partial\Pi^*/\partial\Delta q > 0$ for each i. Monotonicity in q therefore characterizes each participation constraint via a threshold.

Since $h^*_L > h^*_H$ and $\Delta q^\dagger(\cdot)$ is decreasing in beliefs, $\Delta q < \Delta q^\dagger(h^*_L)$ implies $\Delta q < \Delta q^\dagger(h^*_H)$. Therefore, Eq. (A38) evaluated at $\mathbb{E}[h] = h^*_H$ yields $\Pi^*(\psi_H, h^*_H, q, \Delta q) > \Pi^*(\psi_L, h^*_H, q, \Delta q)$ for every q. Define q'_H and q'_L as the solutions to $\Pi^*(\psi_H, h^*_H, q, \Delta q) = 0$ and $\Pi^*(\psi_L, h^*_H, q, \Delta q) = 0$, respectively. By monotonicity and the strict profit ranking just established, $q'_H < q'_L$.

Define q'_B as the value of q at which the low-ψ firm just breaks even under the pooled belief: $\Pi^*(\psi_L, \mathbb{E}[h^*], q'_B, \Delta q) = 0$. Because $\mathbb{E}[h^*] > h^*_H$ by construction (the pooled belief is the convex combination of h^*_L and h^*_H, and $h^*_L > h^*_H$ in the reversed regime, so $\mathbb{E}[h^*]$ lies between them), the marginal entrant's constraint is strictly looser at the pooled belief than at h^*_H: $q'_B < q'_L$.

Moreover, because $\Delta q < \Delta q^\dagger(h^*_L)$, Eq. (A38) evaluated at $\mathbb{E}[h] = h^*_L$ implies

$$\Pi^*(\psi_H, h^*_L, q, \Delta q) > \Pi^*(\psi_L, h^*_L, q, \Delta q)$$

for every q. Hence, whenever the low-ψ firm is willing to invest under the belief h_L^*, the high-ψ firm is also willing to invest. Therefore, an equilibrium with only L invests (LI) cannot occur.

The possible equilibrium regimes—no investment (NI), only H invests (HI), both invest (BI)—follow by the same enumeration as in Appendix A.8. In case (a), multiplicity arises on $[q'_H, q'_B)$ (NI versus HI), on $[q'_B, q'_L)$ (HI versus BI versus NI), and on $[q'_L, q'_N)$ (NI versus BI), where q'_N is defined analogously to q_N in (14) but using the *most profitable* type at the pessimistic belief $\mathbb{E}[h] = 0$, which in the reversed regime is the high-ψ type:

$$\Pi^*(\psi_H, \mathbb{E}[h] = 0, q'_N, \Delta q) = q'_N \Delta\pi(0) + \pi_D(0) + \frac{(\Delta q\, \Delta\pi(0))^2}{2\psi_H} - \overline{\psi}_H = 0.$$

Pareto dominance selects HI over NI, and BI over HI and NI in these regions, delivering the case-(a) characterization. In case (b), $q'_B \leq q'_H$, so the HI candidate is infeasible for all $q < q'_H$; on $[q'_B, q'_H)$ BI is the unique non-trivial equilibrium and Pareto-dominates NI, and for $q \geq q'_H$ BI Pareto-dominates HI and NI as before. The HI region therefore collapses, leaving NI for $q < q'_B$ and BI for $q \geq q'_B$. We impose the analogue of the main-text condition $q_N > q_B$, namely $q'_N > q'_L$, which streamlines the case enumeration exactly as in Appendix A.8.

□

Proposition 8 has several implications for the effects of q and Δq. Most notably, it reveals that q no longer operates as a screening device for excluding the high-ψ in the reversed regime. In the baseline of Section 4, the officer deters the high-ψ firm by lowering q below q_H, resulting in a separating equilibrium where only the low-ψ firm invests and consumers form the belief h_L^*. In the reversed regime, the analogous action (lowering q below q'_B) may instead exclude the low-ψ firm, such that only the high-ψ firm invests with consumers forming the belief h_H^*. Thus, any separating equilibrium induced by q alone in the reversed regime supports the type that uses less human capital. Tightening q therefore cannot implement the LI outcome if $\Delta q < \Delta q^\dagger$.

However, this also means that Δq picks up an additional role. By Lemma 4, raising Δq above $\Delta q^\dagger(\mathbb{E}[h^*])$ flips the regime back to the standard one in which $\Pi^*(\psi_L) > \Pi^*(\psi_H)$. Consequently, if the social planner prefers the LI equilibrium over the HI or BI equilibria, the optimal policy entails setting Δq strictly above $\Delta q^\dagger$, even when the direct social value of human capital γ and the feedback-loop considerations of Section 4.2.2 would, in isolation, call for a smaller Δq. This additional role for Δq complements its role formalized in Proposition 7 as facilitating screening in the standard regime.

C Extensions

C.1 Alternative Cost Function

In the baseline model, the cost function $C(h, \psi) = \overline{\psi} + \frac{\psi}{2}h^2$ is globally increasing in h. Therefore, $h = 0$ should be interpreted as the minimum baseline level of human input, beyond which human capital is always more expensive at the margin than AI. To capture the idea that a mix of human and AI inputs can be cost-minimizing more explicitly, in

this appendix we consider an alternative specification:

$$C(h, \psi) = \overline{\psi}(1 - h) + \frac{\psi}{2}h^2. \tag{A40}$$

Note that $\overline{\psi}$ should no longer be interpreted as the lowest cost the firm can achieve by leveraging AI as much as possible. Instead, the marginal cost of human-capital use is

$$\frac{\partial C}{\partial h} = -\overline{\psi} + \psi h,$$

which is negative for low h and positive for high h. The cost-minimizing level of human capital absent any IP incentives is $h^{cm} = \overline{\psi}/\psi$, which is interior whenever $\overline{\psi} < \psi$. Intuitively, some human involvement reduces costs, for example, because human oversight improves the efficiency of AI systems. Beyond a point, the marginal cost of additional human capital dominates. This is broadly consistent in spirit with complementarity between human and AI inputs at low levels of human-capital use but substitutability at high levels.

The key properties required for the baseline analysis are preserved. First, the cost function remains convex in h: $\frac{\partial^2 C}{\partial h^2} = \psi > 0$. Second, the single-crossing property holds for $h > 0$: $\frac{\partial^2 C}{\partial h \partial \psi} = h > 0$, ensuring that a higher cost type ψ has a higher marginal cost of human capital at any positive h.

Firm's Optimal Human-Capital Use. The general first-order condition for human-capital use given in Eq. (4) continues to hold. Substituting $\frac{\partial C}{\partial h} = -\overline{\psi} + \psi h$, the condition becomes $\Delta q \cdot \Delta \pi + \overline{\psi} = \psi h$, yielding the optimal interior human-capital use

$$h^* = \min\left\{1,\ \frac{\Delta q \cdot \Delta \pi + \overline{\psi}}{\psi}\right\}. \tag{A41}$$

Compared to the baseline solution $h^* = \frac{\Delta q \cdot \Delta \pi}{\psi}$, the alternative cost function adds the term $\overline{\psi}/\psi$ to the optimal human-capital use. This reflects the cost-reduction benefit of human capital at low levels: even absent any IP incentive ($\Delta q = 0$), the firm optimally chooses $h^{cm} = \overline{\psi}/\psi > 0$ to minimize costs. The IP incentive $\Delta q \cdot \Delta \pi$ operates on top of this baseline level.

Investment Decision. Substituting h^* from Eq. (A41) into the profit function (assuming an interior solution), the firm's equilibrium expected profit is

$$\begin{aligned}\Pi(h^*, \psi) &= \pi_D + q\Delta\pi + \Delta q \cdot \frac{\Delta q \cdot \Delta \pi + \overline{\psi}}{\psi} \cdot \Delta \pi - \overline{\psi}\left(1 - \frac{\Delta q \cdot \Delta \pi + \overline{\psi}}{\psi}\right) - \frac{\psi}{2}\left(\frac{\Delta q \cdot \Delta \pi + \overline{\psi}}{\psi}\right)^2 \\ &= \pi_D + q\Delta\pi + \frac{(\Delta q \cdot \Delta \pi + \overline{\psi})^2}{2\psi} - \overline{\psi}.\end{aligned}$$

Therefore, the firm invests if and only if

$$\pi_D + q\Delta\pi + \frac{(\Delta q \cdot \Delta \pi + \overline{\psi})^2}{2\psi} \geq \overline{\psi}. \tag{A42}$$

Compared to the baseline participation constraint, $\pi_D + q\Delta\pi + \frac{(\Delta q \cdot \Delta\pi)^2}{2\psi} \geq \overline{\psi}$, the alternative cost function relaxes the participation constraint through the additional term $\frac{2\overline{\psi}\Delta q \cdot \Delta\pi + \overline{\psi}^2}{2\psi}$ on the left-hand side. This reflects the cost savings from the interior cost-minimizing input mix, which makes the innovation investment more attractive.

Robustness of the Main Results. The alternative cost function shifts the firm's human-capital choice upward by an additive constant $\overline{\psi}/\psi$ relative to the baseline, but does not alter the structure of the model's key trade-offs. We briefly argue that each of the main results is qualitatively preserved.

The sufficient-statistic result from Section 3.3.1 carries over immediately. The "killing AI" result (Proposition 1) is strengthened. The condition for $h^* = 1$ becomes $\Delta q \cdot \Delta\pi + \overline{\psi} \geq \psi$, which is easier to satisfy than the baseline condition $\Delta q \cdot \Delta\pi \geq \psi$. Intuitively, because the firm already has a cost-based incentive to use human capital up to h^{cm}, the IP system needs to bridge a smaller gap to push the firm to full human-capital use.

The "killing the IP system" result (Corollary 1) is qualitatively unchanged. The participation constraint (A42) is relaxed relative to Lemma 1, so sufficiently low AI costs continue to eliminate the need for the innovation-subsidy role of the IP system ($q^* = 0$). Regarding the human-capital-subsidy role, the key observation is that when $\gamma = 0$, the officer has no reason to set $\Delta q > 0$: the firm already chooses $h^{cm} = \overline{\psi}/\psi$ to minimize costs, and any positive Δq would induce human-capital use beyond the cost-minimizing level at the social cost of increased monopoly rights. Thus, $\gamma = 0$ continues to imply $\Delta q^* = 0$. Conversely, when $\gamma > 0$, the officer faces the same trade-off as in Proposition 2. The slope incentivizes human-capital use above the firm's private optimum but comes at the welfare cost of monopoly rights. Therefore, it remains optimal to set $\Delta q^* > 0$ provided the signal is sufficiently informative. The level of the optimal Δq^* is different than in the baseline, because the cost structure already provides some human-capital use.

Finally, the analysis of consumer preferences and adverse selection from Section 4 extends to this cost function. The feedback loop identified in Corollary 2 operates through the interaction of Δq with consumer beliefs $\mathbb{E}[h^*]$ and the profit premium $\Delta\pi(\mathbb{E}[h^*])$. The structure of this feedback loop is preserved under the alternative cost function: increasing Δq raises equilibrium beliefs, which widens the profit premium, which further increases human-capital use, and so on. Similarly, the adverse selection results in Propositions 4 and 5 rely on the ordering of participation thresholds across firm types and the wedge between pooled and separating beliefs. While the specific threshold values q_L, q_H, and q_B change under the alternative cost function, the strict ordering $q_L < q_H < q_B$ is preserved because it follows from $\psi_H > \psi_L$ and the fact that pooled beliefs remain strictly below separating beliefs.

C.2 Consumer's Information

In this extension, we introduce an informative signal that allows the consumer to update their beliefs regarding the firm's input. To formalize the information structure, we assume that the consumer observes a noisy signal s_c from an information set I. This signal is drawn from a conditional probability density function $f(s_c|h)$, which is differentiable in h. We impose the standard assumption that the signal distribution satisfies the Monotone Likelihood Ratio Property (MLRP). Formally, for any two levels of human capital $h' > h$, the likelihood ratio $\frac{f(s_c|h')}{f(s_c|h)}$ is non-decreasing in the signal s_c. This assumption ensures that the signal is informative: a higher realization of s_c indicates a higher likelihood that

the firm uses more human capital.

Based on conjectures about the firm's human-capital use, $\hat{h}_L$ and $\hat{h}_H$, the consumer forms a posterior belief $\mathbb{E}[h|s_c]$ using Bayes' rule:

$$\mathbb{E}[h|s_c] = \frac{p \Pr[s_c|\hat{h}_L]\hat{h}_L + (1-p)\Pr[s_c|\hat{h}_H]\hat{h}_H}{p \Pr[s_c|\hat{h}_L] + (1-p)\Pr[s_c|\hat{h}_H]}. \tag{A43}$$

An implication of the MLRP assumption is that the consumer's posterior belief $\mathbb{E}[h|s_c]$ is strictly increasing in the observed signal s_c.

The introduction of the informative signal s_c leads to a crucial departure from the analysis in Section 4. Previously, the firm could not signal its human-capital use to the consumer. In this setting, however, the firm's choice of h determines the distribution of the signal s_c, and consequently, the consumer's willingness to pay. Importantly, the introduction of the informative signal s_c provides a mechanism to mitigate the adverse selection problem identified in Section 4. In the setting without consumer information, one way to resolve this problem is for the IP policy to deter investment by the high-cost firm. In this extension, the informative signal s_c enables differentiation without requiring exclusion. Because the low-cost firm chooses a higher level of human capital than the high-cost firm, it is more likely to generate favorable signals. Since the consumer updates their beliefs $\mathbb{E}[h|s_c]$ upward upon observing a favorable signal, the low-cost firm realizes a higher expected willingness to pay than the high-cost firm. The informative signal reduces the distortion that benefits the high-cost firm at the expense of the low-cost firm in the pooling equilibrium. Consequently, the presence of consumer information relaxes the officer's need to utilize the IP policy as an instrument for resolving adverse selection.

Proposition 9. *In any pooling equilibrium where both types of firms invest in innovation, the presence of the informative signal s_c ensures that the low-cost firm achieves strictly higher consumer beliefs $\mathbb{E}[h|s_c]$ (and thus higher expected profit) than the high-cost firm.*

Proof. See Online Appendix C.2.1. □

We now characterize the firm's optimal human-capital choice. The firm anticipates that increasing human-capital use h shifts the distribution $f(s_c|h)$ toward more favorable outcomes. This mechanism alters the firm's maximization problem, leading to the first-order condition defined in Eq. (12). We repeat the condition here:

$$\psi h^* = \underbrace{\Delta q \cdot \mathbb{E}_{s_c}[\pi_G(\mathbb{E}[h|s_c]) - \pi_D(\mathbb{E}[h|s_c])]}_{\text{IP Incentive}} + \underbrace{\int_{s_c} \frac{\partial f(s_c|h)}{\partial h}\Big((q+\Delta q h)\pi_G(\mathbb{E}[h|s_c]) + (1-(q+\Delta q h))\pi_D(\mathbb{E}[h|s_c])\Big)ds_c}_{\text{Information Incentive}}.$$

The firm's optimal human-capital use equates the marginal cost ψh to the sum of the "IP Incentive" and the "Information Incentive." The following lemma shows that the "Information Incentive" motivates the firm to use more human capital relative to the case where this incentive is absent.

Lemma 5. *For any IP policy $(q, \Delta q)$, the "Information Incentive" component is strictly positive.*

Proof. See Online Appendix C.2.2. □

The intuition for a positive information incentive is as follows. For a given IP policy $(q, \Delta q)$, the firm anticipates that higher h is more likely to generate favorable signals s_c (i.e., $\frac{\partial f(s_c|h)}{\partial h} > 0$). The rational consumer, upon observing such a signal, forms higher posterior beliefs $\mathbb{E}[h|s_c]$ about the firm's human-capital use. Since the consumer values h, these higher beliefs translate directly into a higher willingness to pay, increasing the firm's profits $\pi_a(\mathbb{E}[h|s_c])$. Therefore, the "Information Incentive" provides incentives for the firm to use human capital in addition to the IP incentive.

Having characterized the firm's optimal human-capital choice, we now show that the presence of consumer information amplifies the firm's responsiveness to the slope Δq. In Section 4, the policy Δq incentivizes the use of human capital by increasing the probability of obtaining protection and triggering a feedback loop via consumer beliefs. In this extension, the information received by the consumer creates an additional channel for Δq to influence the firm's decision. A steeper slope Δq increases the marginal benefit of shifting the signal distribution toward favorable outcomes. Since a favorable consumer signal s_c is correlated with higher consumer willingness to pay, the profit premium $\Delta\pi$ associated with IP protection is largest precisely when the signal is most favorable. Consequently, a higher Δq increases the expected marginal return of generating positive signals through additional human-capital use. This complementarity implies that the firm's human-capital use h^* becomes more sensitive to the slope Δq when consumers observe informative signals. The following proposition formally characterizes this interaction.

Proposition 10. *Compared to a counterfactual in which the term labeled "Information Incentive" in Eq. (12) is absent, the presence of this term strictly increases the marginal effect of the slope Δq on the firm's human-capital use.*

Proof. See Online Appendix C.2.3. □

The presence of the "Information Incentive" also has implications for the optimal IP policy. Because the "Information Incentive" provides a strictly positive motivation for the firm to use human capital that is independent of monopoly rights, it functions as a partial substitute for the slope Δq. The officer can leverage this information-driven mechanism to maximize welfare while reducing the reliance on the IP system to incentivize human-capital use. Furthermore, because the firm is more responsive to Δq as established in Proposition 10, the officer can induce the socially optimal level of human capital with a smaller Δq than would be required if the firm were motivated solely by the IP policy incentives.

Special case: $s_c = a$. In the following we consider a special case in which the consumer's signal is the officer's granting decision itself. In this case, it is not immediate that the firm's incentive to use human capital can still be separated into the "IP Incentive" and the "Information Incentive" as in Eq. (12), because the probability of obtaining IP protection is exactly the probability that the consumer observes the favorable signal. Nevertheless, we show in the following that the same two components are still at work in determining the firm's human-capital use in this special case.

In the general case, the consumer's signal s_c and the officer's decision a are different. As a result, for a given consumer signal s_c, the firm evaluates its expected payoff by averaging over both granting outcomes, with weights given by the probability of obtaining

protection. This means that the firm's expected payoff is:

$$\Pi(h,\psi)=\int_{s_c}\Big(\Pr[a=G\mid h]\pi_G(\mathbb{E}[h|s_c])+\Pr[a=D\mid h]\pi_D(\mathbb{E}[h|s_c])\Big)f(s_c|h)ds_c-C(h,\psi).$$

By contrast, when $s_c=a$, the consumer signal and the granting decision coincide, and the firm's uncertainty is reduced to one dimension. Conditional on $a=G$, the firm only receives $\pi_G(\mathbb{E}[h\mid a=G])$, while conditional on $a=D$, it only receives $\pi_D(\mathbb{E}[h\mid a=D])$. Hence, the firm's expected payoff is

$$\Pi(h,\psi)=\Pr[a=G\mid h]\pi_G(\mathbb{E}[h\mid a=G])+\Pr[a=D\mid h]\pi_D(\mathbb{E}[h\mid a=D])-C(h,\psi).$$

The firm's first-order condition determines its human-capital use:

$$\psi h^*=\Delta q\Big(\pi_G\left(\mathbb{E}[h\mid a=G]\right)-\pi_D\left(\mathbb{E}[h\mid a=D]\right)\Big).$$

Although the right-hand side of this term collapses the "IP Incentive" and the "Information Incentive" into a single term, the firm's marginal benefit can still be decomposed in a way that isolates the same two components:

$$\psi h^*=\underbrace{\Delta q\Big(\pi_G(\mathbb{E}[h\mid a=D])-\pi_D(\mathbb{E}[h\mid a=D])\Big)}_{\text{IP Incentive}}+\underbrace{\Delta q\Big(\pi_G(\mathbb{E}[h\mid a=G])-\pi_G(\mathbb{E}[h\mid a=D])\Big)}_{\text{Information Incentive}}.$$

The first term captures the increase in the firm's profits generated by monopoly rights, evaluated at a common posterior belief $a=D$. The second term captures the additional increase in the profits generated by the more favorable consumer beliefs associated with a granting decision $a=G$. Relative to the general case, the key difference is that the two components are no longer independent, because the same realization a determines both whether protection is granted and what the consumer learns about the firm's human-capital use.

C.2.1 Proof of Proposition 9

Proof. We begin by establishing that the consumer's posterior belief $\mathbb{E}[h|s_c]$ is strictly increasing in the signal s_c. Recall that the consumer's posterior belief is:

$$\mathbb{E}[h|s_c]=\frac{pf(s_c|\hat{h}_L)\hat{h}_L+(1-p)f(s_c|\hat{h}_H)\hat{h}_H}{pf(s_c|\hat{h}_L)+(1-p)f(s_c|\hat{h}_H)}.$$

We differentiate $\mathbb{E}[h|s_c]$ with respect to s_c. The derivative is given by:

$$\frac{\partial\mathbb{E}[h|s_c]}{\partial s_c}=\frac{p(1-p)(\hat{h}_L-\hat{h}_H)\left(\frac{\partial f(s_c|\hat{h}_L)}{\partial s_c}f(s_c|\hat{h}_H)-f(s_c|\hat{h}_L)\frac{\partial f(s_c|\hat{h}_H)}{\partial s_c}\right)}{\Big(pf(s_c|\hat{h}_L)+(1-p)f(s_c|\hat{h}_H)\Big)^2}.$$

The sign of this derivative is determined by the sign of the numerator. Since the signal distribution is assumed to follow the MLRP and given that $\hat{h}_L>\hat{h}_H$, the likelihood ratio $\frac{f(s_c|\hat{h}_L)}{f(s_c|\hat{h}_H)}$ is increasing in s_c. This means that the derivative of this ratio with respect to s_c

is non-negative:

$$\frac{\partial}{\partial s_c}\left(\frac{f(s_c|\hat{h}_L)}{f(s_c|\hat{h}_H)}\right) = \frac{\frac{\partial f(s_c|\hat{h}_L)}{\partial s_c} f(s_c|\hat{h}_H) - f(s_c|\hat{h}_L)\frac{\partial f(s_c|\hat{h}_H)}{\partial s_c}}{(f(s_c|\hat{h}_H))^2} > 0.$$

This implies that $\frac{\partial f(s_c|\hat{h}_L)}{\partial s_c} f(s_c|\hat{h}_H) - f(s_c|\hat{h}_L)\frac{\partial f(s_c|\hat{h}_H)}{\partial s_c} > 0$. Since $p \in (0,1)$ and $\hat{h}_L > \hat{h}_H$, the entire numerator is strictly positive. Thus, $\frac{\partial \mathbb{E}[h|s_c]}{\partial s_c} > 0$.

We now analyze the firm's expected profit $\mathbb{E}_{s_c|h}[\pi_a(\mathbb{E}[h|s_c])]$. Recall that the profit functions $\pi_G(\cdot)$ and $\pi_D(\cdot)$ are strictly increasing in consumer beliefs $\mathbb{E}[h|s_c]$. Combining this with the result above, the firm's profit conditional on the signal, $\pi_a(\mathbb{E}[h|s_c])$ for $a \in \{G, D\}$, is strictly increasing in s_c.

In any pooling equilibrium where both types invest, the low-cost firm chooses human capital h_L^* and the high-cost firm chooses h_H^* such that $h_L^* > h_H^*$. The MLRP assumption implies that the signal distribution $f(s_c|h)$ satisfies First-Order Stochastic Dominance (FOSD) with respect to h. Therefore, the distribution $f(s_c|h_L^*)$ stochastically dominates $f(s_c|h_H^*)$. By the definition of FOSD, the expected value of any function that is strictly increasing in the signal s_c is strictly higher under the dominating distribution. Applying this property to the firm's profit conditional on the signal $\pi_a(\mathbb{E}[h|s_c])$ yields:

$$\mathbb{E}_{s_c|h_L^*}[\pi_a(\mathbb{E}[h|s_c])] > \mathbb{E}_{s_c|h_H^*}[\pi_a(\mathbb{E}[h|s_c])] \quad \text{for } a \in \{G, D\}.$$

Applying the same logic to the consumer belief function $\mathbb{E}[h|s_c]$ yields

$$\mathbb{E}_{s_c|h_L^*}[\mathbb{E}[h|s_c]] > \mathbb{E}_{s_c|h_H^*}[\mathbb{E}[h|s_c]].$$

Thus, the low-cost firm achieves strictly higher expected consumer beliefs and expected profit than the high-cost firm. □

C.2.2 Proof of Lemma 5

Proof. The firm's expected profit function is:

$$\Pi(h,\psi) = \int_{s_c} \Big((q+\Delta q h)\pi_G(\mathbb{E}[h|s_c]) + \big(1-(q+\Delta q h)\big)\pi_D(\mathbb{E}[h|s_c])\Big) f(s_c|h) ds_c - C(h,\psi). \tag{A44}$$

The firm chooses human-capital use $h \in [0,1]$ to maximize its expected profit $\Pi(h,\psi)$, taking the consumer belief $\mathbb{E}[h|s_c]$ as given. We take the derivative of $\Pi(h,\psi)$ with respect to h, which yields the first-order condition with respect to h given in Eq. (12). We repeat the condition here:

$$\begin{aligned}\psi h^* =& \Delta q\, \mathbb{E}_{s_c}[\pi_G(\mathbb{E}[h|s_c]) - \pi_D(\mathbb{E}[h|s_c])] \\ &+ \int_{s_c} \frac{\partial f(s_c|h)}{\partial h}\Big((q+\Delta q h)\pi_G(\mathbb{E}[h|s_c]) + \big(1-(q+\Delta q h)\big)\pi_D(\mathbb{E}[h|s_c])\Big) ds_c\end{aligned}$$

In the following, we prove that the "Information Incentive" term on the right-hand side is positive. We denote the term in the large parenthesis as the firm's expected profit conditional on the signal s_c. First, we observe that this conditional profit is strictly increasing in s_c. This holds because the assumed Monotone Likelihood Ratio Property (MLRP) ensures that the consumer's posterior beliefs $\mathbb{E}[h|s_c]$ are increasing in s_c, and

the firm's profits are increasing in consumer beliefs. Second, the MLRP implies that the signal distribution $f(s_c|h)$ satisfies First-Order Stochastic Dominance (FOSD) in h. By the definition of FOSD, an increase in h shifts the probability mass such that the expected value of any increasing function of s_c increases. This implies:

$$\int_{s_c} \frac{\partial f(s_c|h)}{\partial h}\Big((q+\Delta q h)\pi_G(\mathbb{E}[h|s_c]) + (1-(q+\Delta q h))\pi_D(\mathbb{E}[h|s_c])\Big)ds_c > 0.$$

Thus, the "Information Incentive" term is strictly positive. For any given level of IP incentive, the presence of the informative signal adds a strictly positive marginal benefit to human-capital use. □

C.2.3 Proof for Proposition 10

Proof. To determine the impact of the "Information Incentive" on the firm's responsiveness to Δq, we compare the cross-partial derivative of the firm's expected profit with respect to human-capital use h and the slope Δq, i.e., $\frac{\partial^2 \Pi}{\partial h \partial \Delta q}$, in this extension with a counterfactual case where the "Information Incentive" is absent. This cross-partial derivative measures the complementarity between the firm's human-capital use and the slope. A higher value implies that an increase in the slope Δq generates a stronger marginal incentive for the firm to use human capital h.

In this extension, the firm's equilibrium human-capital use h^* is determined by the first-order condition (12). We repeat the condition here:

$$\begin{aligned}\frac{\partial \Pi}{\partial h} &= \Delta q \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])] - \psi h \\ &\quad + \int_{s_c} \frac{\partial f(s_c|h)}{\partial h}\Big((q+\Delta q h)\pi_G(\mathbb{E}[h|s_c]) + \big(1-(q+\Delta q h)\big)\pi_D(\mathbb{E}[h|s_c])\Big)ds_c = 0.\end{aligned}$$

Differentiating the above with respect to Δq yields the following cross-partial derivative:

$$\frac{\partial^2 \Pi}{\partial h \partial \Delta q} = \left(\mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])] + \Delta q \frac{\partial \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])]}{\partial \Delta q}\right) + h\int_{s_c} \frac{\partial f(s_c|h)}{\partial h}\Delta\pi(\mathbb{E}[h|s_c])ds_c. \tag{A45}$$

The term in parentheses corresponds to the direct effect of Δq and its impact on consumer beliefs (i.e., the feedback loop). These two effects of Δq have been analyzed in Section 4 extensively. The second term is unique to the presence of the "Information Incentive." Since $\Delta\pi(\mathbb{E}[h|s_c])$ is strictly increasing in $\mathbb{E}[h|s_c]$, and $\mathbb{E}[h|s_c]$ is increasing in s_c by the MLRP assumption and the fact that $f(s_c|h)$ satisfies FOSD, this integral is strictly positive. Thus, the "Information Incentive" adds a positive term to the cross-partial derivative.

In the counterfactual case where the "Information Incentive" term is absent, the firm's marginal benefit of human-capital use is driven solely by the "IP Incentive" $\Delta q \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])]$ and the cross-partial derivative is given by:

$$\frac{\partial^2 \Pi}{\partial h \partial \Delta q} = \frac{\partial}{\partial \Delta q}\left(\Delta q \cdot \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h \mid s_c])]\right) = \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])] + \Delta q \frac{\partial \mathbb{E}_{s_c}[\Delta\pi(\mathbb{E}[h|s_c])]}{\partial \Delta q}.$$

Comparing the two cross-partial derivatives, we conclude that the presence of the "Information Incentive" increases the firm's responsiveness to Δq.

□

C.3 Continuous Signal

In this extension, we develop the continuous-signal version of our model discussed in Section 5.1 in the main text. This model more directly maps the IP policy to a threshold for human-capital use and allows us to implement a potential practical limitation to the IP policy. Specifically, we assume the officer can only set the threshold for human-capital use, rather than the "intercept" and "slope" of the policy separately.

Consider a model in which the IP officer sets a human-capital-use threshold for granting protection, $\bar{h} \in [0, 1]$. In contrast to our baseline model, in which the officer receives a binary signal, suppose the officer receives a continuous signal $s = h + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$, about the firm's use of human capital. The officer then forms a rational expectation about the firm's human-capital use, i.e., $\mathbb{E}[h|s]$, reflecting the officer's own assessment of the use of human capital. This is given by:

$$\mathbb{E}[h|s] = \frac{pf(s|\hat{h}_L)\hat{h}_L + (1-p)f(s|\hat{h}_H)\hat{h}_H}{pf(s|\hat{h}_L) + (1-p)f(s|\hat{h}_H)} \tag{A46}$$

where $\hat{h}_L$ and $\hat{h}_H$ are the officer's conjectured human-capital use for the low-cost and high-cost types, respectively. Given the signal structure, the signal s follows a normal distribution, that is, $s|h \sim \mathcal{N}(h, \sigma^2)$, therefore, the function $f(s|\hat{h})$ is the normal probability density function.

Granting Decision The officer grants IP protection if and only if

$$\mathbb{E}[h|s] > \bar{h}. \tag{A47}$$

Since the cost function $C(h, \psi)$ satisfies the single-crossing property, i.e., $\frac{\partial^2 C}{\partial h \partial \psi} > 0$, the conjectured human-capital use satisfies $\hat{h}_L > \hat{h}_H$. Consequently, the likelihood functions $f(s|\hat{h}_L)$ and $f(s|\hat{h}_H)$ satisfy the Monotone Likelihood Ratio Property, that is, $\frac{f(s|\hat{h}_L)}{f(s|\hat{h}_H)}$ increases in s. This implies that $\mathbb{E}[h|s]$ is monotonically increasing in s. Therefore, the granting decision reduces to a threshold rule on s. Specifically, there exists a unique signal $\bar{s}$ such that the officer grants IP protection if and only if $s > \bar{s}$. It is straightforward to see that $\bar{s}$ is a function of the officer's conjectured human-capital use $\hat{h}_L$ and $\hat{h}_H$ and the committed threshold $\bar{h}$.

Human-Capital Use. The general formula for the firm's first order condition for human-capital use given in the baseline model continues to hold in this setting. That is, the effect of the IP policy on human-capital use is entirely driven by the marginal probability of being granted IP protection:

$$\frac{\partial \Pr[s > \bar{s}|h]}{\partial h}\Delta\pi = \frac{\partial C(h, \psi)}{\partial h}.$$

As in the baseline model, the IP policy is defined as a cumulative distribution function that is a function of human-capital use, and so the optimal level of human-capital use is driven by the probability density function.

Given the assumption that the signal's error term is normally distributed, $\epsilon \sim \mathcal{N}(0, \sigma^2)$, the probability of receiving IP protection is $\Pr[s > \bar{s}|h] = 1 - \Phi\left(\frac{\bar{s}-h}{\sigma}\right)$, where Φ is the cumulative distribution function of the standard normal distribution. The derivative of this probability with respect to h is therefore $\frac{1}{\sigma}\phi\left(\frac{\bar{s}-h}{\sigma}\right)$, where ϕ is the standard normal probability density function.

From the firm's first order condition, we can derive the effect of the threshold $\bar{s}$ on the firm's human-capital use.

Lemma 6. *An increase in the threshold $\bar{h}$ can either increase or decrease the use of human capital relative to AI:*

$$\frac{dh^*}{d\bar{h}} \geq 0 \text{ if and only if } \bar{s} \leq h^*,$$

where $\bar{s}$ is defined as $\mathbb{E}[h^*|\bar{s}] = \bar{h}$. In the limit $\bar{h} \to \infty$, only AI is used:

$$\lim_{\bar{h}\to\infty} h^* = 0.$$

Proof. The firm's first order condition with respect to h is given by:

$$g(h^*, \bar{s}) \equiv \frac{\pi_G - \pi_D}{\sigma\psi}\,\phi\left(\frac{\bar{s} - h^*}{\sigma}\right) - h^*, \tag{A48}$$

so that the optimal human-capital use h^* is defined by $g(h^*, \bar{s}, \sigma) = 0$.

By the implicit function theorem,

$$\frac{dh^*}{d\bar{s}} = -\frac{\partial g/\partial \bar{s}}{\partial g/\partial h^*}.$$

Differentiating $g(h^*, \bar{s}, \sigma)$ with respect to h^* and $\bar{s}$, and using the fact that the standard normal density satisfies $\phi'(x) = -x\phi(x)$, gives

$$\frac{\partial g}{\partial h^*} = -\frac{\pi_G - \pi_D}{\sigma\psi}\frac{1}{\sigma}\phi'\left(\frac{\bar{s} - h^*}{\sigma}\right) - 1 = \frac{\pi_G - \pi_D}{\sigma\psi}\frac{\bar{s} - h^*}{\sigma}\frac{1}{\sigma}\,\phi\left(\tfrac{\bar{s}-h^*}{\sigma}\right) - 1,$$

$$\frac{\partial g}{\partial \bar{s}} = -\frac{\pi_G - \pi_D}{\sigma\psi}\frac{\bar{s} - h^*}{\sigma}\frac{1}{\sigma}\,\phi\left(\tfrac{\bar{s}-h^*}{\sigma}\right).$$

To determine the sign of the derivative $dh^*/d\bar{s}$, note that the partial derivative $\partial g/\partial h^* \leq 0$. This is because $g(h^*, \bar{s}, \sigma)$ is the first order condition of the firm's problem with respect to h, so $\partial g/\partial h$ is the second order condition which is negative evaluated at the local maximum h^*. Therefore, it suffices to determine the sign of $\partial g/\partial \bar{s}$:

$$dh^*/d\bar{s} \geq 0 \iff \partial g/\partial \bar{s} \geq 0 \iff \bar{s} \leq h^*,$$

where we have used that $\pi_G - \pi_D \geq 0$ and $\phi\left(\frac{\bar{s}-h^*}{\sigma}\right) \geq 0$. Since $\bar{s}$ is increasing in $\bar{h}$, this implies the sign of the derivative stated in the lemma.

The second part of the lemma follows from evaluating $g(h^*, \bar{s}, \sigma)$ in the respective limit:

$$\lim_{\bar{h}\to\infty} g(h^*, \bar{s}, \sigma) = -h^*,$$

which implies that $g(h^*, \bar{s}, \sigma) = 0$ can only be satisfied at $h^* = 0$. □

Lemma 6 echoes the first result discussed in Section 3.3: the effect of policy leniency (i.e., a lower $\bar{s}$) on h^* is ambiguous. This is because the effect is determined by how a change in $\bar{s}$ changes the marginal probability of receiving IP protection. In the case of the continuous signal, $\bar{s} \leq h^*$ implies that an increase in $\bar{s}$ increases this marginal probability, while $\bar{s} > h^*$ implies that an increase in $\bar{s}$ decreases this marginal probability.

C.4 Dynamic Model Training

To explore the implications of dynamic model training, we embed the baseline model in a two-period economy corresponding to two independent rounds of innovation. The IP officer sets the IP policy at the beginning of the first period and it applies to both periods. We assume that parameters are identical in both periods except for the AI cost floor, which we assume to be weakly decreasing over time, such that $\overline{\psi}_2 \leq \overline{\psi}_1$.

To model the feedback between human-capital use, IP protection, and AI model training, we assume that the period-2 AI cost floor is a function $\overline{\psi}_2 = \overline{\psi}_2(h_1, a_1)$, where h_1 is the firm's period-1 human-capital choice and $a_1 \in \{G, D\}$ is the period-1 grant decision. This function captures three channels through which period-1 decisions can affect $\overline{\psi}_2$ in a reduced form. First, more human-created content may improve training data quality, lowering $\overline{\psi}_2$. We refer to this as the "model-collapse channel" because it has often been discussed in the context of AI models being trained on AI-generated data, which can potentially undermine the quality of AI output. Second, more human and less AI-use results in lower revenue for the AI model provider, increasing $\overline{\psi}_2$. Third, IP protection ($a_1 = G$) restricts content from training use, implying higher $\overline{\psi}_2$. Consequently, the overall sign of the effect of period-1 decisions on $\overline{\psi}_2$ is ambiguous and depends on the relative strength of these channels. We carry this sign as unrestricted throughout the following analysis.

When considering the firm's choice of human-capital use, it is important to note that we model only one firm in the economy. However, in an economy with many innovative firms, the effect of an individual firm's human-capital use on future innovation costs may be small, such that firms do not internalize this effect. In that case, the firm's optimal human-capital use reduces to the static setting, as given by Eq. (5). Therefore, the incentives for human-capital use derived from our baseline model extend to the dynamic setting.

The effect of human-capital use and IP protection are, however, internalized by the IP officer when setting the IP policy. We first derive q^* under the assumption that it is optimal for the officer to incentivize investment by both types of firms in both periods. Since the AI cost floor is assumed to be weakly decreasing over time for all h_1 and a_1, this implies binding the participation constraint for the high-cost firm in the first period. Thus, q^* is the same as in the baseline model, which is given in Proposition 2 (with $\overline{\psi} = \overline{\psi}_1$).

To derive Δq^*, the officer maximizes total welfare across the two periods, with period-2 welfare discounted by a factor δ:

$$\max_{\Delta q} \mathbb{E}[\mathbf{W}] = \mathbb{E}[W_1] + \delta \mathbb{E}[W_2]$$

Substituting the social benefit function $v(h)$, the firm's optimal human-capital use h^*,

and the optimal slope q^* into expected welfare gives

$$\max_{\Delta q} \mathbb{E}[\mathbf{W}] = (1+\delta)\left[\pi_D + u_D + \left(\frac{\overline{\psi} - \pi_D}{\Delta\pi} - \frac{(\Delta q)^2 \Delta\pi}{2\psi_H}\right)(\Delta\pi + \Delta u) + \Delta\pi\gamma\mathbb{E}\left[\frac{1}{\psi}\right]\Delta q + \Delta\pi\left(\frac{\Delta\pi}{2} + \Delta u\right)\mathbb{E}\left[\frac{1}{\psi}\right](\Delta q)^2\right] - \overline{\psi}_1 - \delta\overline{\psi}_2(h^*, a_1),$$

where $\mathbb{E}\left[\frac{1}{\psi}\right] = \frac{p}{\psi_L} + \frac{1-p}{\psi_H}$. The first-order condition with respect to Δq is

$$\frac{d\mathbb{E}[\mathbf{W}]}{d\Delta q} = (1+\delta)\left[\Delta\pi\gamma\mathbb{E}\left[\frac{1}{\psi}\right] + 2\Delta q\left(\Delta\pi\mathbb{E}\left[\frac{1}{\psi}\right]\left(\frac{\Delta\pi}{2} + \Delta u\right) - \frac{\Delta\pi}{2\psi_H}(\Delta\pi + \Delta u)\right)\right] - \delta\frac{d\overline{\psi}_2}{d\Delta q} = 0.$$

Solving for the optimal slope gives

$$\Delta q^* = \frac{\gamma\mathbb{E}\left[\frac{1}{\psi}\right] - \frac{\delta}{1+\delta}\frac{d\overline{\psi}_2}{d\Delta q}}{-\left(\mathbb{E}\left[\frac{1}{\psi}\right](\Delta\pi + 2\Delta u) - \frac{\Delta\pi + \Delta u}{\psi_H}\right)}.$$

This expression is similar to the optimal slope given in Proposition 2, with the additional term in the numerator reflecting the discounted effect of Δq on $\overline{\psi}_2$. This additional term maps to the three channels through which period-1 decisions (h_1 and a_1) affect $\overline{\psi}_2$. Intuitively, if a larger Δq leads to a greater reduction in the future AI cost floor (i.e., $\frac{d\overline{\psi}_2}{d\Delta q} < 0$), then the officer chooses a larger Δq^* than in the static setting. This would be the case if, for example, the model-collapse channel dominates the other two channels discussed above. Conversely, if the opposite is true, then the officer chooses a smaller Δq^* than in the static setting.