Artificial Intelligence in the Boardroom*

Daniel Ferreira Jin Li

August 27, 2026

Abstract

We analyze how the use of artificial intelligence affects the monitoring and advisory relationship between CEOs and corporate boards. AI can serve as a private advisor to CEOs, partially substituting for board advice, and thus reducing CEOs' incentives to share information with directors. In equilibrium, firms respond by reducing board independence to restore information flows. Lower monitoring intensity decreases CEO turnover, making CEOs more entrenched. Lower dismissal risk allows firms to reduce CEO compensation. While AI adoption by the board mitigates some of the negative aspects of the CEO's AI adoption, it does not restore efficient levels of board monitoring: there is either too little or too much monitoring in equilibrium. Our analysis suggests that AI adoption in the boardroom may have unintended consequences for corporate governance.

Keywords: Boards, CEOs, Artificial Intelligence, Corporate Governance

*We thank Sonny Biswas, Mike Burkart, Felix Zhiyu Feng, Luis Garicano, Enrique Ide, Dirk Jenter, Camilo Riva, Liyan Yang, seminar participants at EDHEC Business School, LSE, MBS-Southampton-Surrey seminar, Berlin Micro Theory seminar, U. of Delaware Corporate Governance Symposium, ICMA Corporate Finance and Banking Workshop, Nova SBE Entrepreneurship and Innovation Symposium, and the BSE Summer Workshop on Economics of Transformative AI. We thank Zehao Zhang for excellent research assistance. An Internet Appendix is at the end of this manuscript. A companion interactive tool that illustrates and verifies all main results is available at Boardroom IC Explorer. The authors have not received specific funding for this research, and there are no conflicts of interest to disclose. Ferreira: London School of Economics, CEPR, and ECGI; d.ferreira@lse.ac.uk. Li: HKU Business School; jli1@hku.hk.

1 Introduction

Artificial intelligence (AI) systems can solve problems and carry out complex tasks with limited human oversight. Such systems can both help experts and replace them. AI's impact is most pronounced in analytical reasoning and strategic assessment—tasks that until recently required human expertise.[1] The speed and scope of this technological change raise questions about how organizations should structure their decision-making processes and divide tasks between humans and machines.[2]

In this paper, we develop a theoretical framework for studying the impact of generative artificial intelligence on the relationship between CEOs and corporate boards. As AI increasingly serves as an analytical tool for executive decision-making,[3] it potentially disrupts the traditional information flows between executives and directors. While CEOs and other executives can now access AI-generated insights for strategic planning, idea generation, and project assessment, this capability may reduce their reliance on board expertise and alter their incentives to share information.[4] We show that these changes in communication dynamics have important implications for board composition, monitoring, and executive compensation.

The model is as follows. A CEO exerts effort to identify a problem (e.g., find a deal, generate an idea, etc.). The CEO job is a "messy job" (Garicano, Li, and Wu (2026)), that is, a bundle of tasks that must be performed by humans. Once a problem is identified, the CEO must assess it and decide whether to address it on her own or with the board's help. The CEO has access to an AI agent. We consider AI primarily as an oracle with a dual

[1] In a randomized controlled trial with 758 Boston Consulting Group consultants, Dell'Acqua et al. (2026) find that those using GPT-4 completed 12.2% more tasks, finished them 25.1% faster, and achieved over 40% higher quality output compared to controls.

[2] Babina et al. (2025) use resume and job posting data to document significant flattening of hierarchical structures with AI adoption, showing increases in junior-level workers and decreases in middle management roles.

[3] In a large international survey, Yotzov et al. (2026) find that AI adoption by top executives increased by more than 50% in 2025, and that CEOs are more frequent users of AI than other executives. Kenny, Kowalkiewicz, and Oosthuizen (2025) present cases and examples of how CEOs can use AI for strategic planning. Csaszar, Ketkar, and Kim (2024) present evidence of the use of AI as a strategic decision-making tool in firms.

[4] Larcker et al. (2025) argue that AI has the potential to alter the relationship between executives and directors significantly. Shekshnia and Yakubovich (2025) discuss how AI helps directors.

role: it is both a tool and a coworker.[5] Ransbotham et al. (2025) argue that "*this tool-coworker duality breaks down traditional management logic, which assumes that technology either substitutes or complements, automates or augments, is labor or capital, or is a tool or a worker, but not all at once*."[6] To model this duality, we assume that the AI technology has two capabilities: it enhances the CEO's problem-solving skills (a *skill-augmenting capability*), and it also solves some problems independently of the CEO's ability (a *problem-solving capability*). That is, in line with the emerging literature on the effect of AI on the future of work, we model AI technology as task automation and worker augmentation (Acemoglu, Autor, and Johnson (2026); Demirer et al. (2026)).

We consider the problem of a group of controlling shareholders who must choose the board's monitoring intensity (i.e., board independence) and the CEO's compensation contract, which we jointly refer to as a *governance structure*. An optimal governance structure must incentivize the CEO to exert effort to identify deals and to ask the board for help when the CEO cannot solve a problem on her own. Asking for help increases the likelihood that the board replaces the CEO because it is perceived as a negative signal about the CEO's ability. To induce the CEO to communicate truthfully with the board, the optimal governance structure must involve a board that is partially "CEO-friendly," that is, one that may retain the CEO even when a better replacement is available. As a consequence, the optimal structure will typically commit to some degree of ex post inefficient CEO entrenchment to enhance the CEO's ex ante incentives to seek the board's advice.

A CEO with access to an AI agent with problem-solving capabilities can use it as a substitute for the board's advice. What makes AI distinctive in this regard is that its consultation is private and nearly costless, setting it apart from traditional sources of outside advice. The more capable AI is, the less likely is the CEO to ask the board for help. Therefore, to restore information flows between the CEO and the board, the optimal board structure must be more CEO-friendly when such a capability exists (i.e., when AI is a coworker). As the CEO faces a lower risk of dismissal under a friendlier board, the optimal compensation

[5] An oracle "*serves as an advisor, offering information, predictions, or recommendations while leaving all decisions and actions to human users*" (Jiang (2025), p. 114).

[6] In a survey on the use of agentic AI in business decision-making, Ransbotham et al. (2025) find that 76% of executives view agentic AI more as a coworker than a tool.

level is lower with AI than without it. Because the CEO's incentive constraints must be binding, the decrease in dismissal risk is exactly offset by the reduction in pay, leaving the CEO no better off. Thus, the AI-driven increase in CEO entrenchment destroys firm value, without benefiting the CEO.

In contrast, if AI augments CEOs' skills (i.e., CEOs use AI primarily as a tool), CEOs become more productive, and firms find it easier to incentivize them to exert effort. Thus, CEOs can be induced to exert effort with lower pay. However, lower pay reduces CEO incentives to communicate with the board. Thus, as in the case in which AI serves as a coworker, the optimal governance structure requires low pay and low monitoring intensity. We conclude that CEO access to AI increases board friendliness and reduces CEO pay, regardless of whether CEOs use AI primarily as a tool or as a coworker. The value implications of AI adoption by CEOs are ambiguous: while AI's problem-solving capabilities are welfare-destroying, its skill-augmenting capabilities are welfare-increasing.

While CEO AI adoption unambiguously reduces both the board's monitoring intensity and CEO pay levels, a natural follow-up question is whether board AI adoption can counter the effects of CEO AI adoption. We extend the model to allow the board to directly benefit from AI adoption in its advisory and monitoring roles. We show that board AI adoption attenuates some of the effects of CEO AI adoption on monitoring and pay levels. However, these are *short-run* effects. In the long run – when AI advances significantly augment the board's advising and monitoring capabilities – AI adoption by the board converges to one of two polar cases. If AI significantly enhances the board's advice quality, the board gives up monitoring completely and specializes in advising the CEO. A fully CEO-friendly board thus obtains, leading to full CEO entrenchment and reduced pay. By contrast, if AI substantially enhances the board's monitoring capability, the board ends up "overmonitoring" and "underadvising" the CEO: it replaces some "good" CEOs and fails to advise some "bad" CEOs.

Our analysis shows that AI adoption by CEOs and boards can have unintended consequences for corporate governance because of interactions between the board's monitoring and advising roles. For evidence that boards perform an important advisory role (which at times conflicts with its monitoring role), see Coles, Daniel, and Naveen (2008), Linck,

Netter, and Yang (2008), Faleye, Hoitash, and Hoitash (2011), Field, Lowry, and Mkrtchyan (2013), Harford and Schonlau (2013), Adams, Akyol, and Verwijmeren (2018), de Haas, Ferreira, and Kirchmaier (2021), and Ewens and Malenko (2024).

Our paper relates to recent empirical work on the behavioral side effects of AI. In a series of experiments, Marcoccia, Quattrociocchi, and Capraro (2026) show that access to AI advice erodes people's willingness to admit ignorance. When an AI answer was available, participants largely stopped saying "I don't know": they attempted far more questions, answered a smaller share correctly, and reported higher confidence. They conclude that AI not only changes the answers people give, but also lowers the threshold at which they judge they know enough to answer at all. Cowgill, Hernandez-Lagos, and Wright (2024) find that generative AI lowers screening accuracy by 4–9%. Weak applicants use AI to write like strong ones, so their applications carry less signal. Chen et al. (2025) document a drop in worker motivation after firms adopt AI. Both results point to the same problem — AI makes it harder for firms to learn what they need to know about people. Our model shows that firms can partially mitigate these adverse effects by adjusting their governance structures.

A companion interactive tool solves the model for any parameter configuration, under both standard and alternative assumptions.[7] The tool uses analytical results (most of which are proven in the paper) to determine the equilibrium, before producing numerical results. The tool allows the reader to visually inspect equilibrium outcomes under almost any parameter configuration, including those that violate our maintained assumptions in the text. While the tool allows the reader to verify the accuracy of the results numerically, we note that all results — including those in the Internet Appendix — have been proved analytically.

We review the related theoretical literature in Section 2. We present our model setup in Section 3. Section 4 shows the impact of AI on board learning. Section 5 solves the model in a pre-AI economy. Section 6 then presents our main results. Section 7 extends the model to allow for AI adoption by the board. Section 8 concludes. Most proofs are in

[7]The tool can be accessed at claude.ai/code/artifact/8a42cd05-0bba-436b-a104-cbf0f48d588d. It was designed with Claude Fable 5. Anyone can access it anonymously via the link; there is no need to log in.

the Appendix. The Internet Appendix presents a more detailed discussion on contracting, assumption violations, omitted cases, extensions, and all remaining omitted proofs.

2 Related Theoretical Literature

Our paper is the first formal analysis of the impact of AI on corporate boards. Our baseline model of the CEO-board relationship relates to a theoretical literature on the dual role of boards as advisors and monitors of management,[8] which includes Adams and Ferreira (2007), Harris and Raviv (2008), Baldenius, Melumad, and Meng (2014), Baldenius, Meng, and Qiu (2019), Levit (2020), Drymiotes and Sivaramakrishnan (2021), and Jiang and Laux (2023), among others.[9] More generally, the model relates to the broader literature on information asymmetries in the boardroom, as in Warther (1998), Raheja (2005), Song and Thakor (2006), Laux (2008), Baranchuk and Dybvig (2009), Inderst and Mueller (2010), Levit (2012), Malenko (2014), Levit and Malenko (2016), Chakraborty and Yilmaz (2017), Chemmanur and Fedaseyeu (2018), and Donaldson, Malenko, and Piacentino (2020); see also Malenko (2024) for a comprehensive review of the literature on information flows in corporate governance.

In our model, the CEO must first provide information to the board before receiving advice. In this setup, CEO-friendly boards can be optimal: the board commits to low-intensity monitoring, which incentivizes the CEO to share information, as in Adams and Ferreira (2007). The logic is reminiscent of Burkart, Gromb, and Panunzi (1997) and Aghion and Tirole (1997), who show that ex-post intervention by principals undermines agents' ex-ante incentives. More generally, our model relates to the theoretical literature on delegation (e.g., Holmström (1984), Melumad and Shibano (1991), Dessein (2002), Alonso and

[8] A more recent literature has also focused on the "mediating" role of boards. Burkart, Miglietta, and Ostergaard (2023) present evidence that boards monitor management and mediate between different types of shareholders. Ewens and Malenko (2024) provide evidence that VC-backed startup boards advise and mediate between investors and executives. Also consistent with a mediating role of boards, Ferreira, Ferreira, and Mariano (2018) show evidence that changes in creditors' power relative to shareholders affect appointments to the board.

[9] To the best of our knowledge, the advisory role of boards is first mentioned in the economics and finance literature by Fama and Jensen (1983), who argue that directors "*provide an important support function to the top managers in dealing with specialized decision problems*" (p. 314).

Matouschek (2008), and Amador and Bagwell (2013), among others).

While our benchmark model shares the core trade-off in Adams and Ferreira (2007), its structure is original. In Adams and Ferreira (2007), there are no compensation contracts or CEO replacement. Unlike Adams and Ferreira's moral-hazard/cheap-talk setup, we model the optimal choice of board monitoring, CEO compensation, and CEO replacement under moral hazard, selection, and career concerns. As in Song and Thakor (2006), the CEO is responsible for generating a project idea with an unknown probability of success before deciding whether to reveal it to the board. CEOs differ in ability, enjoy private benefits, and face career concerns. The board monitors by gathering information about the CEO and occasionally replacing a low-ability CEO. CEO compensation is endogenous and depends on the CEO's performance and the board's belief about the CEO's type.

Another key innovation of our paper is to model the relationship between a CEO and a board as a knowledge hierarchy, as in Garicano (2000) and Garicano and Rossi-Hansberg (2006). Following the original contribution of Ide and Talamàs (2025), we introduce AI within a knowledge-hierarchy setup. As in Nikolowa (2015), agents learn about their knowledge over time. As in Fuchs, Garicano, and Rayo (2015), there is asymmetric information about agents' knowledge.

There is little consensus on what AI means, and no generally accepted way to introduce AI into microeconomic models. We follow recent theoretical literature modeling AI-related innovations as exogenous technological shocks that facilitate task automation and improve labor productivity. These papers typically use comparative static exercises to study the effects of different technological parameters that substitute or complement human labor. Examples of this approach include Ide and Talamàs (2025), Garicano and Rayo (2025), Ide (2026), and Bárány and Koren (2026), among others. An innovation is that our interactive tool makes these comparative static exercises *dynamic*: rather than relying on what the authors choose to show, the reader can choose which combinations of comparative statics to investigate.

There is an emerging theoretical literature on AI in organizations and governance. An early contribution is Athey et al. (2020), who introduces an AI agent to Aghion and Tirole's model of authority. Garicano and Rayo (2025), Ide (2026), and Friebel et al. (2026) study the

impact of AI-related innovations on employee training and careers. Zhong (2025) develops a framework to study the interactions between humans and technologies in a multi-layer decision-making problem. His model implies that the optimal human-AI authority allocation depends on the trade-off between correcting errors and introducing new ones, with AI typically best as the final decision-maker. Cheng, Dogan, and Yildirim (2025) study the optimal replacement of workers by AI agents in a team production problem. Huang and Ouyang (2025) develop a cheap-talk model of AI-based advising in which AI agents are unbiased (unlike human agents) but face difficulties in codifying soft information.

More closely related to our work is Chen and Han (2026). They develop a model in which biased CEOs may have access to AI technology. Because AI increases the granularity and precision of hard information available to CEOs, there is more scope for biased decisions, making agency problems more severe. Firms can mitigate some of these negative effects by changing contracts, but these fixes come at a cost. They conclude that AI adoption may reduce profits and welfare.

3 Model Setup

We present a model of the relationship between a CEO and a board of directors. Following Ide and Talamàs (2025), we build on Garicano's (2000) knowledge hierarchy framework, augmented by AI. In this framework, the CEO plays the role of a "worker" who identifies and assesses investment opportunities, while the board serves as an "advisor" who can help the CEO when needed. In addition, the board also monitors the CEO. We begin by describing the agents, the technology, and the information structure. We then present the full game in detail.

Agents, Technology, and AI. We consider a firm owned by one or more shareholders, who are residual claimants to the firm's cash flows. There are two periods. The firm must select a board at the beginning of the first period. We model the board in reduced form as a unitary agent and abstract from conflicts of interest among board members. The board must hire one CEO in each period. The CEO enjoys a nonpecuniary private benefit

$B > 0$ in each employment period. Shareholders have no specialized human capital and, thus, cannot become CEOs or board members. All agents—CEOs, board members, and shareholders—are risk-neutral, have zero discount rates, have zero outside utility, and are protected by limited liability.

The CEO's job is a "messy job:" it is an indivisible bundle of tasks that cannot be automated, involving both hard and soft skills (Garicano, Li, and Wu (2026)). Specifically, the CEO operates a technology represented by a continuum of *problems* (i.e., potential deals, projects, ideas, etc.). If the CEO exerts effort at a fixed personal cost of $c > 0$, she identifies a problem of unknown *difficulty* $x \sim U[0,1]$. One interpretation is that c is the CEO's personal cost of "deal sourcing." We initially assume that only CEOs can identify deals; this task cannot be facilitated or performed by AI agents. We relax this assumption in Subsection 6.3. From now on, we refer to "deals" and "problems" interchangeably.

The CEO has unobservable *knowledge* indexed by $z \sim U[0,1]$. After identifying a deal, the CEO has one unit of time to allocate across tasks. If the CEO spends $t_s < 1$ units of time assessing a deal, she generates an *assessment*, a. If the CEO assesses the deal without any external help, her assessment is correct if and only if $x \leq z$. A correct assessment generates output $y = 1$, while incorrect assessments have value $y = 0$. Because both x and z are unobservable to all, the probability that the CEO can produce a correct assessment is $\Pr[x \leq z] = 0.5$.

To assist with deal assessment, the CEO has access to an AI agent, whose capabilities are represented by a pair of parameters $(\alpha, \kappa) \in [1, \overline{\alpha}] \times [0,1]$, which are known to all. The firm cannot deny the CEO access to AI; the CEO's AI adoption decisions are private information. If the CEO uses AI as a skill-augmenting tool, her assessment is correct if $x \leq \alpha z$, and incorrect otherwise. That is, parameter α measures how much AI enhances the CEO's ability to solve problems. We call α AI's *skill-augmenting capability*.[10] When AI skill-augmentation is available, the unconditional probability that a randomly chosen CEO

[10]Yotzov et al. (2026) show evidence that the most common types of AI technologies used by businesses are data processing with machine learning and text generation with LLMs. These are both examples of skill-augmenting capabilities.

can successfully assess a deal is

$$\Pr(x \leq \alpha z) = \int_0^1 \int_0^1 \mathbb{1}_{x \leq \alpha z} \, dx \, dz = \int_0^1 \min(\alpha z, 1) \, dz = 1 - \frac{1}{2\alpha}. \tag{1}$$

Note that if the AI agent has greater skill-augmenting capabilities, the probability that the CEO solves the problem is higher.

Alternatively, the CEO can ask the AI agent to solve the problem autonomously, with minimal human input. That is, AI is not only a tool, but also a coworker. An AI model with *problem-solving capability* κ can autonomously solve problems with difficulty $x \leq \sqrt{\kappa}$, in no time. That is, $\sqrt{\kappa}$ is the AI agent's knowledge, which implies that κ is the probability that AI can solve a problem conditional on the CEO's not being able to solve it: $\kappa = \Pr[x \leq \sqrt{\kappa} \mid x > \alpha z]$. AI's autonomous problem-solving capability is what arguably differentiates AI models from other innovations. While α represents AI's ability to complement an agent's knowledge, κ represents AI's ability to be a substitute for human decision-making. Thus, the pair of parameters (α, κ) captures the idea that the CEO can view AI as both a tool and a coworker (Ransbotham et al. (2025)).[11]

To simplify the exposition and concentrate on the more interesting cases, we make five parametric assumptions, which we introduce gradually in this section. We explain their role in the main text where they are needed, and discuss cases in which they don't hold in the Internet Appendix; the interactive tool allows the reader to solve the model when some of these assumptions do not hold. Table 1 below collects all main parametric assumptions. Our first assumption is as follows.

Assumption 1 (CEO autonomy threshold). *The average CEO has greater problem-solving capability than the AI agent:* $1 - \frac{1}{2\alpha} > \sqrt{\kappa}$.

[11] The active use of AI agents to analyze complex business decisions independently is still rare at the time of this paper's writing, but AI agents are already being deployed in business to independently tackle more specialized tasks such as insurance claims processing, inventory management, and financial trading. Survey evidence on the use of agentic AI in business decision-making shows that even among the most advanced adopters, only 54% of respondents are pursuing scenarios in which AI decides and implements autonomously (Ransbotham et al. (2025)); yet the same respondents are two to three times more likely to expect AI systems to work independently from humans and have decision-making authority in three years compared to today.

This assumption states that the probability that a CEO of unknown type solves a problem is greater than the probability that AI solves a problem autonomously. Assumption 1 implies that the AI's problem-solving capability is below what Garicano, Li, and Wu (2026) call the "autonomy threshold": the point at which full delegation to AI dominates human performance. Without Assumption 1, the *average* CEO is better off delegating an assessment fully to AI rather than trying to assess the deal herself with AI help. We analyze the case in which Assumption 1 fails in the Internet Appendix (see Section IA.5).

Assumption 1 implies that the CEO is always weakly (and sometimes strictly) better off generating an assessment herself before asking AI to assess the deal independently. Thus, without loss of generality, we assume that the CEO always generates an assessment after identifying a deal (except, of course, when Assumption 1 fails, as we study in the Internet Appendix).

The CEO can also ask the board for help in making the assessment. That is, boards serve as advisors who specialize in providing help. Thus, firms are structured as two-layer knowledge hierarchies, as in Garicano (2000), Garicano and Rossi-Hansberg (2006), Ide and Talamàs (2025), and Bárány and Koren (2026), among others. We assume that board directors have knowledge equal to one; that is, they can solve problems of any difficulty level. This assumption is stronger than necessary; we only need the board's expected knowledge to be sufficiently high. The board's time and attention are constrained. If the CEO asks the board for help, the board allocates less time and attention to routine "compliance tasks." Specifically, diverting the board's attention away from the compliance tasks destroys one unit of output with probability $\eta \in (0,1)$. That is, if the board spends one unit of time helping the CEO assess a deal, the expected output is $1-\eta$. Thus, if the CEO can perform the assessment herself, it is best not to involve the board. Henceforth, we refer to $1-\eta$ as the *board's problem-solving capability*. Parameter η equals the "cost of help" in knowledge hierarchy models (e.g., Ide and Talamàs (2025)).[12] We make the following assumption:

[12]For most of the paper, parameters (α, κ) fully describe AI's capabilities. After presenting our core model, we also consider four natural extensions: AI may lower the cost of deal sourcing (Subsection 6.3), AI may improve the board's advice and monitoring ability (Section 7), and AI may reduce the time needed to verify assessments (Internet Appendix, Subsection IA.3.2).

Assumption 2 (**Board autonomy threshold**). *The board has greater problem-solving capability than the AI agent:* $1 - \eta > \kappa$.

This assumption states that *conditional on the CEO being unable to solve a problem*, the probability that the board solves a problem is greater than the probability that AI solves the problem autonomously. Similar to Assumption 1, Assumption 2 implies that AI does not make the board redundant.

Because problem difficulty x is unobservable, a CEO does not know whether a given assessment s is correct. The CEO can verify the correctness of an assessment by performing a verification task. If the CEO spends $t_v < 1$ units of time in the verification task, she learns whether s is correct. An assessment does not require verification. Verification is nevertheless valuable because if a CEO-generated assessment s is known to be incorrect, the CEO can ask AI or the board for a second assessment.

Lamba (2026) makes the case that, in the age of AI, verification is likely to become the main bottleneck. As in Bárány and Koren (2026), we use time constraints to model the costs of both assessment and verification. For simplicity only, we assume that the time needed for verification, t_v, is the same for any assessment, regardless of who produced it. To reduce the number of cases to consider, we now assume the following.

Assumption 3 (**Time constraints**). $t_v \leq 1 - t_s < 2t_v$.

Assumption 3 implies that when a CEO produces an assessment and, subsequently, learns it is incorrect, she must decide whether to ask *either* AI *or* the board for help. Because the CEO does not have sufficient time to verify the AI agent's solution, the sequential strategy of first asking the AI agent and then the board leads to the same expected outcome as asking only the board (given Assumption 2).

Only the first part of Assumption 3, $t_v \leq 1 - t_s$, is strictly needed for our analysis. Without it, the CEO does not learn about her ability, and the problem is trivial and uninteresting. The Internet Appendix considers the case in which $1 - t_s \geq 2t_v$ (see Section IA.3).

Given our assumptions, a CEO who identifies a deal will optimally spend time $t_s + t_v$ generating and verifying an assessment. After verification, the CEO learns that she is one of two *ex post* types: $x \leq \alpha z$ (type g) or $x > \alpha z$ (type b). After learning her type, she

Table 1: Parametric assumptions

	Name	Statement	Role
1	CEO autonomy threshold	$1 - \frac{1}{2\alpha} > \sqrt{\kappa}$	The average CEO outperforms autonomous AI; she assesses deals herself before consulting AI.
2	Board autonomy threshold	$1 - \eta > \kappa$	The board outperforms autonomous AI as a source of help; board advice remains valuable.
3	Time constraints	$t_v \leq 1 - t_s < 2t_v$	The CEO can verify one assessment but not two; AI and board advice are substitutes.
4	Efficient effort provision	$1 - \frac{2}{3}\eta \geq c$	Inducing effort from every retained CEO type is efficient.
5	Costly communication	$B > 2c$	Inducing communication costs more than inducing effort, so the help IC constraint matters.

Notes: The table lists the parametric assumptions maintained in the main text; assumption numbers refer to the text.

decides whether to ask AI for an independent assessment (action m_a), ask the board for help with the assessment (action m_b), or implement her own assessment (action m_c). To allow for mixed strategies, let $P_t(\tau)$ denote the period-t probability distribution of m over $\{m_a, m_b, m_c\}$ (i.e., the CEO's behavior strategy) conditional on the CEO's knowledge of her *ex post* type, $\tau \in \{b, g\}$. $P_t(\tau)$ determines $p_t(\tau) \in [0, 1]$, which is the probability that a CEO of *ex post* type τ in period t asks the board for advice, i.e., the probability of $m = m_b$.

For completeness, we present our two remaining parametric assumptions; we explain their role later when we use them in the analysis.

Assumption 4 (**Efficient effort provision**). *Exerting effort is always efficient:* $1 - \frac{2}{3}\eta \geq c$.

Assumption 5 (**Costly communication**). *The cost of inducing communication is no lower than the cost of inducing effort:* $B > 2c$.

Table 1 collects and summarizes all parametric assumptions we use in the main text.

Board independence and monitoring. At the beginning of the first period, shareholders choose a board for a two-period term. The board cannot be replaced in period 2. Share-

holders can choose the board's *ex ante* type $\pi \in [0,1]$, denoting its "independence" (i.e., its degree of alignment with shareholders' interests). Formally, we think of π as the result of an action that stochastically reduces the board's private cost of firing a CEO: with probability π, the board is "tough" (state of the world ω^{tough}) and faces no cost of firing the CEO, while with probability $1-\pi$, the board is "weak" (state of the world ω^{weak}) and the cost of firing the CEO is prohibitively high. In this state, the board's preferences do not align with those of shareholders only because of the firing cost.

The board observes $\hat{m} \in \{m_b, \overline{m}_b\}$, where $\overline{m}_b := \{m_a, m_c\}$. That is, the board knows only whether the CEO has asked for its help. If the CEO does not ask the board for help (i.e., $\hat{m} = \overline{m}_b$), the board is unaware of the deal and, thus, cannot assess it.[13] If $\hat{m} = m_b$, the board produces an assessment, which delivers $y_t = 1$ with probability $1-\eta$.

At the beginning of period 2, the board must decide whether to replace the incumbent CEO or appoint an external hire. If the state of the world is ω^{weak}, the board chooses to retain the CEO regardless of $\hat{m}$. In state of the world ω^{tough}, the board replaces the CEO with probability $\rho(\hat{m}) \in [0,1]$, where $\rho(\cdot)$ is optimally chosen by the board to maximize shareholders' expected continuation profit.

We interpret π as determined by the board's structure and composition. A less independent board faces a greater cost of monitoring the CEO, and is thus less likely to replace an underperforming CEO.[14] As in Adams and Ferreira (2007), we call π the board's *monitoring intensity*. The interpretation is that CEOs regularly take actions to entrench themselves, such as co-opting board members. This entrenchment is a byproduct of the collaboration between the CEO and the board. By working closely with the CEO, the board becomes CEO-friendly. A friendlier board has lower monitoring intensity.

Timing. Early in period 1, shareholders appoint a board of (observable) ex ante type $\pi \in [0,1]$ for a two-period term. All further decisions are delegated to the board, which aims to maximize shareholder value, net of its private costs of firing the CEO. Within a period,

[13] Even if the board knows that the CEO has identified a deal, it is unaware of its characteristics unless the CEO shares this information with the board.

[14] For simplicity, we assume that shareholders face no cost of choosing π; we could easily add a direct cost of independence without changing the nature of the problem.

actions occur in the following order.

(i) **Contracting**: The board hires a new CEO with (unobservable) knowledge $z \sim U[0,1]$ or retains the incumbent CEO from the previous period.

(ii) **Payoffs**: Output from the previous period's actions, y_{t-1}, is realized. Output-contingent compensation is paid. The CEO enjoys a private benefit $B > 0$ from being employed at this stage.

(iii) **CEO's actions**: The CEO decides whether to exert effort at a private cost c. If the CEO exerts effort, she finds a "deal" (i.e., a problem) with (unknown) difficulty $x \sim U[0,1]$. The CEO produces an assessment s and verifies it. After learning whether s is correct, the CEO chooses one of three actions : $m \in \{m_a, m_b, m_c\}$, where m_a denotes asking the AI agent to solve the problem autonomously, m_b denotes disclosing the deal to the board, and m_c denotes sticking with her original assessment.

(iv) **Board's actions**: The board observes $\hat{m} \in \{m_b, \overline{m}_b\}$ where $\overline{m}_b := \{m_a, m_c\}$. If $\hat{m} = m_b$, the board produces an assessment. The state of the world, $\omega \in \{\omega^{tough}, \omega^{weak}\}$, is realized and becomes observable to all; π denotes the probability of $\omega = \omega^{tough}$.

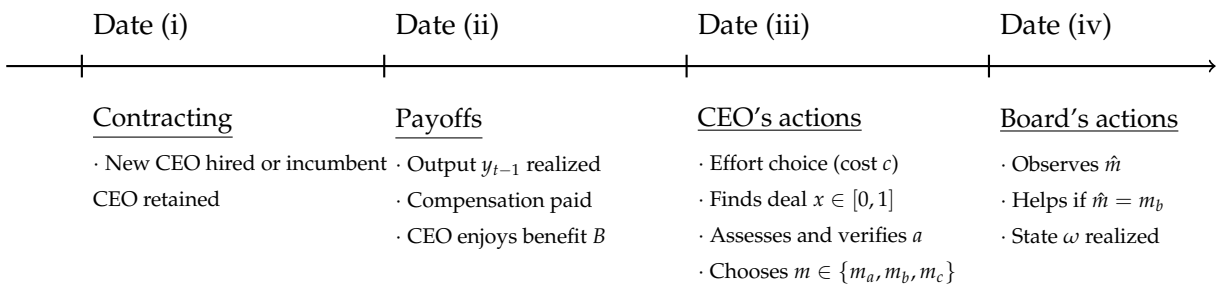


Figure 1: Timing of the game within a period.

Figure 1 provides a summary of the timing. There are two important timing assumptions. First, each period starts with a contracting stage. Since a tough board can choose not to reappoint the CEO early in period 2, and the CEO can quit at any time, any optimal long-term contract must be renegotiation-proof.[15]

[15]See Laux (2008) for a model of CEO turnover with renegotiation-proof contracts.

Second, there is "maturity mismatch" in that contracts must be (re)negotiated before the output from the previous period is realized. The interpretation is that corporate policies may have long-term consequences, and decisions to retain a CEO may be made before the full impact of these actions is known. Formally, we need only that the CEO's decision to ask the board for help is incrementally informative for the board's decision to fire or retain the CEO. Our timing assumptions retain this property in the simplest possible manner.[16]

Contracts. In date (i) of period $t = 1$, the board offers a take-it-or-leave-it contract to a CEO with unknown knowledge $z \sim U[0,1]$. In date (i) of period $t = 2$, the board offers a take-it-or-leave-it contract to either a new CEO (again with knowledge $z \sim U[0,1]$) or the incumbent CEO (i.e., the same CEO as in $t = 1$). CEO compensation contracts may be contingent on y_1, y_2, and $\theta \in \{r, f\}$—the CEO's job status in period 2—where r denotes a CEO who is retained and f a CEO who is fired. The CEO's message to the board, $\hat{m}$, is private information to the board and the CEO, and thus not verifiable. CEOs' outside options remain fixed at zero.[17] CEOs have no external wealth and can "credibly commit" to consuming any compensation received at the end of a period, for example, by borrowing against that amount.

Given the model setup, we can restrict attention to one-period contracts of the following type. At $t = 1$, the CEO is offered an output-contingent bonus w_1 conditional on $y_1 = 1$ and a severance bonus s conditional on $\theta = f$. At $t = 2$, the board offers an output-contingent bonus w_2 conditional on $y_2 = 1$ to either the incumbent CEO or a new CEO; severance pay is irrelevant in $t = 2$ because all CEOs step down at the end of the period. Assuming such one-period contracts is without loss of generality; in the Internet Appendix, we show that no other contract format can deliver higher shareholder value (see Proposition IA.1 and

[16] Flipping the order of dates (ii) (payoffs) and (iii) (CEO's actions) does not change the analysis or conclusions.

[17] Under a different information structure where competing firms could observe the CEO's decision to ask the board for help, in a separating equilibrium, CEOs who didn't ask for help would have better outside options. This would make asking for help a more negative signal and would require an even lower monitoring intensity to induce a CEO to ask the board for help. This variation of the model is more complex but yields no significant further insights.

Corollary IA.1).[18]

We can show that optimal one-period contracts implement the first-best solution with $\pi = 1$ whenever the benefit of monitoring exceeds the CEO's private benefit (see Proposition IA.2). Therefore, to make the choice of board structure meaningful under such conditions, we assume the following contracting friction:

Assumption 6 (**Severance bonus cap**). *The severance bonus is capped: $s \leq \bar{s}$.*

Note that, unlike the parametric assumptions in Table 1, Assumption 6 is a restriction on the space of contracts, and thus a friction. Because the specific reason for the existence of a cap is not important for the analysis, we assume for simplicity that $\bar{s} \geq 0$ is set exogenously. A severance pay cap may exist because of institutional frictions. While compensation for without-cause dismissal is common in CEO contracts, it is subject to both hard and soft constraints. The US tax code heavily penalizes both executives and corporations for excessive severance payments, effectively limiting severance payments to no more than three times the base salary. In addition, proxy advisors often have explicit guidelines flagging excessively generous separation contracts.[19]

Although institutional frictions that constrain severance pay are realistic, it is natural to ask which deeper frictions may also serve as a microfoundation for the cap $\bar{s}$. Rather than adding to model complexity, in the main text we take the cap as given, whereas in the Internet Appendix we consider a simple extension of the model—inspired by Berk and van Binsbergen (2022)—in which the board optimally chooses contracts to screen different CEO types (see Section IA.2). A severance pay cap then arises endogenously as an optimal contractual feature to screen out low-skill CEOs. This model extension retains all of our main results.

[18] In reality, CEO contracts are either fixed-term or "at will" (i.e., with no explicit termination date; see Gonzalez-Uribe and Groen-Xu (2017)). Fixed-term contracts usually last no longer than five years, but renewal is common. Most companies opt for at-will contracts, which do not provide protection against dismissal and are, in effect, a series of short-term contracts that are continuously renegotiated. We can interpret our one-period contracts as either fixed-term contracts with the possibility of renewal or at-will contracts.

[19] The theoretical literature on delegation typically differs from optimal contracting and mechanism design approaches by restricting the space of feasible contingent transfers, often ruling out such contracts entirely. A notable exception is the recent work by Chen (2026), who provides a dynamic model of authority delegation under optimal compensation contracts. Similarly, our model combines features from both approaches.

Given an exogenous severance pay cap, for notational simplicity, we set $\bar{s} = 0$, which is an inconsequential normalization. Because it is never optimal to choose $s < 0$, any optimal contract must choose $s = 0$ (again, this is shown formally in the Internet Appendix). Thus, the optimal one-period contracts can be described simply by w_1 and w_2.

4 AI and Board Learning

In this section, we begin our analysis by focusing on the board's learning problem. In period $t = 1$, the CEO learns more about her own knowledge by trying to assess a deal at date (iii). If the CEO learns she cannot assess the deal (i.e., her *ex post* type is b), she decides whether to seek help from the AI agent ($m = m_a$) or the board ($m = m_b$).

Under Assumption 2, a CEO who identifies a deal she cannot assess faces a trade-off. On the one hand, if she asks the board for help, she will receive better advice than the AI agent's. On the other hand, the board may learn that she was unable to assess the deal. In that case, the board may decide to replace the CEO at the next round of renegotiations. If Assumption 2 does not hold, no such trade-off exists, and the CEO does not ask the board for help. In contrast, under Assumption 2, the board's advice remains valuable.

Let S_2 denote the event in which a retained CEO can assess a deal in $t = 2$. Let $\Pr(S_2 \mid \tau)$ denote the probability of S_2 conditional on the CEO being "good" or "bad," $\tau \in \{g, b\}$. We have the following result.

Lemma 1 (**Learning**). *If the CEO from $t = 1$ is retained in $t = 2$, the probability that she can solve a newly drawn problem is $\Pr(S_2|g) = \frac{2(3\alpha-2)}{3(2\alpha-1)}$ if she is good and $\Pr(S_2|b) = \frac{1}{3}$ if she is bad.*

Note that the AI agent's skill-augmenting capability makes the "good" CEO better, but has no effect on the expected output of a "bad" CEO in period 2. The latter result arises because when AI is more skill-augmenting (larger α), the CEO who fails to solve the problem is, in expectation, of even lower type. This negative selection effect exactly cancels out the positive skill-augmenting effect in period 2.

When facing a bad CEO, if the board is "tough," it has the option to replace the CEO with a new one of unknown type. A newly-hired CEO has an average type of $E[z] = \frac{1}{2}$; call

this type $\tau = n$. Note that at $t = 2$ there are no career concerns, so CEOs always ask the board for help if they cannot assess a deal. With a newly-hired CEO, the expected output is $E[y_2 \mid n] = (1 - \frac{1}{2\alpha}) + \frac{1}{2\alpha}(1 - \eta) = 1 - \frac{\eta}{2\alpha}$. If a bad incumbent CEO is retained, the expected output is $E[y_2 \mid b] = 1 - \frac{2\eta}{3}$.[20] We define the expected output gain from replacing a bad CEO with a CEO of average type as

$$M(\alpha, \eta) := 1 - \frac{\eta}{2\alpha} - 1 + \frac{2\eta}{3} = \eta\left(\frac{2}{3} - \frac{1}{2\alpha}\right) > 0. \tag{2}$$

We can interpret $M(\alpha, \eta)$ as the value of the option to replace a bad CEO, or the *(marginal) benefit of monitoring*. Note that this value increases with the AI agent's skill-augmenting capability, α, and with the cost of board advising, η. By contrast, AI's problem-solving capability does not affect the value of board learning.

Under our assumptions, an equilibrium without board learning does not exist, as we show next.

Lemma 2 (**No uninformative equilibrium**). *There is no equilibrium with positive effort provision in both periods in which $p_t(b) = p_t(g)$ for $t \in \{1, 2\}$.*

Recall that $p_t(\tau)$ is the probability that a CEO with *ex post* type $\tau \in \{b, g\}$ asks the board for advice (i.e., $m = m_b$). Lemma 2 implies that CEOs of different *ex post* types must communicate with the board with different probabilities. It then follows that there is no full pooling of types in equilibrium; with some positive probability, the board must learn something about τ from the equilibrium play. In equilibrium, there could be either full separation of types or partial separation, in which case both types may choose the same action, but with different probabilities.

[20] Assumption 4 implies that the board finds it optimal to offer a bonus that induces all CEO types to exert effort in $t = 2$. Nothing important changes if Assumption 4 does not hold, as long as it is efficient to induce the average type (i.e., $z = 0.5$) to exert effort. We impose Assumption 4 simply to reduce the number of cases to consider. If Assumption 4 fails (while effort remains efficient for the average type), the optimal governance structure coincides with Case 1 of Proposition 1; see Section IA.4 of the Internet Appendix for a complete analysis of this case.

5 The Pre-AI Boardroom

In this section, we consider the case of a "pre-AI economy," where $\alpha = 1$ and $\kappa = 0$. This case serves as a benchmark. Solving this simpler version reveals the mechanisms more clearly and helps build intuition. We extend the benchmark model to the general case in the next section.

At the beginning of the first period, shareholders choose (w_1, π). For each (w_1, π), players play a sequential game of incomplete information. The solution concept is Perfect Bayesian Equilibrium. To solve for an equilibrium, note first that because $\kappa = 0$, at date (iii) of either period action m_c weakly dominates m_a, and thus only two choices are relevant: $m \in \{m_b, m_c\}$. Because there are no career concerns in $t = 2$, the CEO's decision whether to ask the board for help in this period is trivial. If the CEO knows how to assess the deal, she will choose m_c because it delivers w_2 with probability 1, while choosing m_b delivers w_2 with probability $1 - \eta$. Similarly, if the CEO does not know how to assess the deal, it is strictly better to ask for help and gain w_2 with probability $1 - \eta$.

In the first period, at date (iv), the board updates its beliefs about the CEO after observing $\hat{m} = m \in \{m_b, m_c\}$. Lemma 2 implies that, if effort is provided in the first period (which is the only equilibrium type we need to consider), both nodes at this date (m_b, or "ask for help," and m_c, or "do not ask for help") must be reached with positive probability in equilibrium. Thus, the board's beliefs about the CEO's type are fully pinned down by Bayesian updating, and no off-path beliefs remain to be determined. Because the game is sequential and beliefs are uniquely determined, for each (w_1, π) there is a (generically) unique equilibrium of the continuation game that follows. Since shareholders move first, they select their preferred equilibria by choosing (w_1, π). We refer to the shareholders' preferred equilibrium as the *optimal equilibrium*.[21]

We have the following result.

Lemma 3 (Good CEOs don't ask for help). *In equilibrium,* $p_1(g) = 0$.

[21] Generic uniqueness allows for a measure zero of pairs (w_1, π) to induce multiple equilibria in the continuation game. We will show in the proof of Proposition 1 that this multiplicity does not create difficulties for finding the optimal equilibrium, which is always unique.

Intuitively, because ex post "good" CEOs are better at solving problems than the board, not asking the board for help must signal something positive to the board. By contrast, ex post "bad" CEOs face a trade-off: while asking for help increases the likelihood of solving the problem, it also increases the probability of being fired. Thus, only type-b CEOs can choose mixed strategies in equilibrium: while $p_1(g) = 0$, we must have $p_1(b) > 0$ (due to Lemma 2). It follows immediately from Lemma 3 that after observing m_b, the board knows that the CEO must be of the bad type. Thus, a tough board should always replace a CEO who asks for help: $\rho(m_b) = 1$.

To streamline the exposition, we will henceforth restrict the analysis to pure-strategy equilibria. As we will formally show in the proof of Proposition 1, restricting the analysis to pure strategies is without loss of generality because an optimal equilibrium must be in pure strategies.

Under pure strategies, Lemmas 2 and 3 jointly imply $p_1(b) = 1$, that is, an *ex post* bad CEO always asks the board for help. That is, in equilibrium, bad and good CEOs must make different choices at date (iii), and these choices fully reveal the CEO's type. Accordingly, we solve the model assuming shareholders expect a bad CEO to request the board's help at $t = 1$. Then, we show that this belief is consistent with equilibrium play, that is, $p_1(b) = 1$.

Solving the model requires checking six incentive compatibility constraints. Working backward, we begin at date (iii) in $t = 2$ (note that date (iv) is redundant in $t = 2$ because the CEO cannot stay for another period). At this point, the CEO has no career concerns and thus does not need incentives to seek help from the board when she is unable to assess a deal. However, the CEO still needs to be incentivized to exert effort c to identify a deal.

Suppose first that the CEO from $t = 1$ is retained in $t = 2$. If the CEO is of type b, from Lemma 1 we know that $E[z \mid b] = \Pr(S_2 \mid b) = \frac{1}{3}$. Let w_{2b} denote the bonus offered to a CEO of type b conditional on $y_2 = 1$.[22] The probability that the output is $y_2 = 1$ if the CEO is of type b is $\frac{1}{3} + \frac{2}{3}(1 - \eta) = 1 - \frac{2}{3}\eta$. A bad CEO's "effort incentive constraint" at date (iii) of $t = 2$ is

$$\left(1 - \frac{2}{3}\eta\right) w_{2b} - c + B \geq B. \tag{3}$$

[22]Technically, y_2 realizes only in $t = 3$, after the firm shuts down; this creates no issues for the analysis.

The left-hand side of (3) is the bad CEO's payoff in $t = 2$ if she exerts effort. By exerting effort at a personal cost c, the CEO produces $y_2 = 1$ with probability $1 - \frac{2}{3}\eta$, in which case she earns w_{2b}, on top of her private benefit, B. If she does not exert effort, she collects her private benefit, which is the right-hand side of (3). At Date (i) of $t = 2$, if the board knows that the CEO is of type b, it will offer the lowest possible bonus such that (3) holds, implying $w^*_{2b} = \frac{3c}{3-2\eta}$.

Similarly, if a retained CEO is of type g, from Lemma 1 we know that $E[z \mid g] = \Pr(S_2 \mid g) = \frac{2}{3}$, and the effort incentive constraint is

$$\Big(1 - \frac{\eta}{3}\Big) w_{2g} - c + B \geq B, \tag{4}$$

implying $w^*_{2g} = \frac{3c}{3-\eta}$.

In the case of a newly appointed CEO, which we denote by type n, we have that $E[z \mid n] = \Pr(S_2 \mid n) = \frac{1}{2}$, and the effort incentive constraint is

$$\Big(1 - \frac{\eta}{2}\Big) w_{2n} - c + B \geq B, \tag{5}$$

implying $w^*_{2n} = \frac{2c}{2-\eta}$.

At date (i) of $t = 2$, a "weak" board never replaces the CEO. Because the marginal benefit of monitoring is positive ($M(1, \eta) > 0$), a "tough" board always replaces a CEO of type b, but not a CEO of type g. Intuitively, there is no reason for a firm to retain an *ex post* bad CEO because an outside replacement is expected to be more qualified. Similarly, an *ex post* good CEO is always retained because she is expected to be more qualified than outsiders.[23] That is, in equilibrium we must have $\rho(m_b) = 1$ and $\rho(m_c) = 0$.

Consider now date (iii) of $t = 1$. First, we need to ensure that a "good" CEO does not want to pretend she is "bad":

$$w_1 + B \geq (1 - \eta) w_1 + (1 - \pi) \Big[\Big(1 - \frac{\eta}{3}\Big) w^*_{2b} - c + B \Big]. \tag{6}$$

[23] An implicit assumption here is that only the board observes m and the decision of whether to retain the CEO, implying that an *ex post* good CEO does not have better outside opportunities. Allowing CEO retention decisions to signal quality to competing employers (as in Waldman (1984), Greenwald (1986), Li (2013), and Ferreira and Nikolowa (2023), among others) complicates the analysis but adds little insight.

The left-hand side of (6) is the (incremental) payoff a good CEO expects by solving the problem herself, without involving the board. In that case, the CEO guarantees w_1 and knows that she will be retained in $t = 2$. She will then enjoy her private benefit B and exert effort c in return for a second-period bonus of w_{2g}. In that case, her $t = 2$ utility is $\left(1 - \frac{\eta}{3}\right) w^*_{2g} - c + B = B$. The right-hand side of (6) is the payoff a good CEO expects by involving the board. If a good CEO asks the board for help, the board solves the problem with probability $1 - \eta$ and, with probability $1 - \pi$, it retains the CEO (whom the board now thinks is bad) and offers her a second-period bonus of w^*_{2b}.[24]

Second, still on date (iii), a bad CEO must be incentivized to ask the board for help:

$$(1 - \eta) w_1 + (1 - \pi) B \geq B. \tag{7}$$

Condition (7) is the "(asking for) help incentive constraint." If a CEO who cannot assess a deal asks the board for help, the board will solve the problem with probability $1 - \eta$, and the CEO's expected bonus is $(1 - \eta) w_1$. By revealing her type, the CEO is retained with probability $1 - \pi$, in which case she enjoys B.[25] If the CEO chooses not to ask for help, she will not solve the problem, but will be retained with probability 1, because $\rho(m_c) = 0$. If retained, the CEO is guaranteed to enjoy her private benefit B at $t = 2$. Because the board thinks the CEO is good, it will offer her a bonus of w^*_{2g}. The CEO will choose not to exert effort because her IC in (3) is violated at w^*_{2g}.

The help IC constraint can also be written as

$$\pi \leq \frac{(1 - \eta) w_1}{B}. \tag{8}$$

Intuitively, for the CEO to share information with the board, the board must be sufficiently "friendly," that is, its monitoring intensity π should not be too high. The right-hand side of (8) is the ratio between the CEO's benefit from the board's help and the cost of sharing

[24] We assume that when a good CEO asks the board for help, the assessment of the deal is delegated to the board. The board succeeds at assessing the deal with probability $1 - \eta$; the CEO's knowledge here is irrelevant.

[25] The CEO is also promised bonus w^*_{2b} in $t = 2$ in exchange for effort c, but because the effort constraint in (3) binds, the CEO's exactly breaks even.

information with the board. A "bad" CEO benefits from asking the board for help because the board can solve a problem with probability $1-\eta$, in which case the CEO receives w_1. The cost of sharing information is that, if the board is "tough," the CEO will be fired and lose her private benefit, B. This logic is reminiscent of Adams and Ferreira's (2007) model.

The sixth and final constraint is the first-period "effort incentive compatibility" constraint:

$$\underbrace{B-c}_{t=1 \text{ utility}} + \frac{1}{2}\underbrace{(w_1+B)}_{\text{g's utility}} + \frac{1}{2}\underbrace{((1-\eta)w_1+(1-\pi)B)}_{\text{b's utility}} \geq 2B. \tag{9}$$

The left-hand side of (9) is the CEO's expected lifetime payoff if she exerts effort at $t=1$ and behaves optimally thereafter. By exerting effort, the CEO identifies a deal and eventually learns her type (g or b), each with equal probabilities. The right-hand side of (9) is the CEO's payoff if she chooses instead not to exert effort at $t=1$. In that case, the first-period output is $y_1=0$, implying that she does not earn w_1. In addition, she does not learn her type; she believes her knowledge is average ($E[z]=\frac{1}{2}$). Because she does not ask the board for help, she is retained in $t=2$ and offered a bonus of w_{2g}^*. Since w_{2g}^* violates the second-period IC constraint for type n (see (5)), she does not exert effort at $t=2$. Thus, her lifetime expected utility is $2B$.

The shareholders' problem at $t=1$ is to choose a board monitoring intensity (i.e., independence) $\pi \in [0,1]$. Then, at date (i), the board chooses a first-period bonus $w_1 \geq 0$ to maximize the firm's expected profit, subject to constraints (6), (7) and (9), taking w_{2b}^*, w_{2g}^*, and w_{2n}^* as given. As preferences are aligned at this date, we may think of (w_1,π) as being chosen simultaneously to maximize profit. This is a linear programming problem. To solve it, we proceed by initially assuming that (6) is slack. After solving for the optimal values, we then check (and confirm) that (6) is indeed slack at these values. With a slight abuse of terminology, we call a pair (w_1,π) a *governance structure*.[26]

The interesting case is, of course, when the effort IC constraint (9) does not imply the help IC constraint (7). Assumption 5 is a convenient sufficient condition that ensures that this happens for any value of η. Intuitively, for the help IC constraint to matter, we need

[26]Strictly speaking, the governance structure should also include the second-period bonuses: (w_{2b}, w_{2g}, w_{2n}).

the "cost of inducing communication" in $t = 1$, B, to be larger than the "cost of inducing effort" in both periods, $2c$.

It is instructive to consider briefly the case in which Assumption 5 fails. If $B \in ((1-\eta)2c, 2c)$, the effort IC constraint still does not imply the help IC constraint. Suppose now that $B \leq (1-\eta)2c$. It can be readily checked that, for all $\pi \in [0,1]$, the effort IC constraint (9) implies the help IC constraint (7). Thus, the optimal board monitoring intensity π^* must be a corner solution that binds constraint (9) only. From (10), we see that the marginal benefit of monitoring a bad CEO is $M(1,\eta) = \frac{\eta}{6}$. Thus, the firm wants to monitor more intensively as long as the marginal benefit exceeds the bad CEO's expected loss from monitoring, which is B. If $M(1,\eta) > B$, the optimal board monitoring intensity is $\pi^* = 1$; if $M(1,\eta) < B$, the optimal board structure is $\pi^* = 0$. Thus, although a friendly board might be optimal in this case, the governance structure is determined not by the CEO's incentives to communicate but rather by the incentives for effort provision, unlike Adams and Ferreira (2007).

Figure 2 illustrates constraints (7) and (9) on the (w_1, π) plane, for a set of parameters that satisfy all parametric assumptions. Note that the help IC constraint passes through the origin and has a smaller slope than the effort IC constraint. The incentive-compatible region is the gray area to the right of both boundaries.

Given the optimal $t = 2$ bonuses, and assuming that constraints (7) and (9) bind, the firm's expected profit at the beginning of $t = 1$ is

$$\begin{aligned} V(w_1,\pi) &= \frac{1}{2}\Big(\underbrace{2-\frac{\eta}{3}-c-w_1}_{\text{Profit from good CEO}}\Big) + \frac{1}{2}\Big(\underbrace{2-\frac{5}{3}\eta-c-(1-\eta)w_1+\pi\frac{\eta}{6}}_{\text{Profit from bad CEO}}\Big) \\ &= 2-\eta-c-(1-\frac{\eta}{2})w_1+\frac{\pi}{2}M(1,\eta). \end{aligned} \tag{10}$$

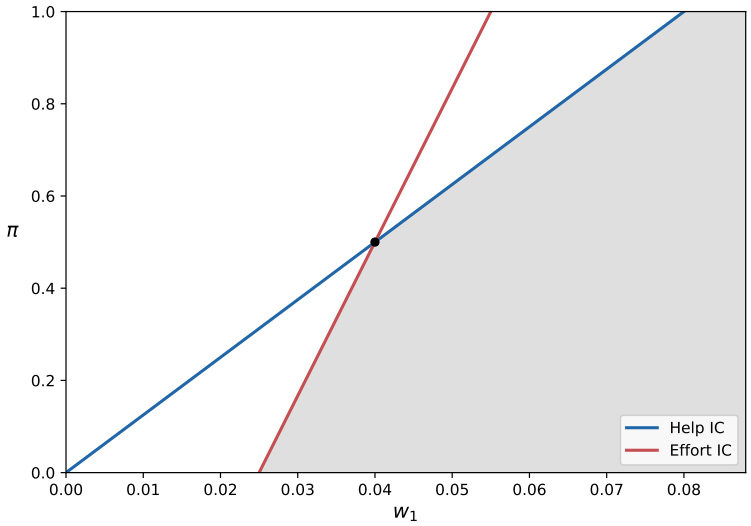


Figure 2: Incentive-compatible governance structures

(Parameters: $\eta = 0.4$, $B = 0.048$, $c = 0.02$)

From (10) we can write the *isoprofit* for a given level of profit V as

$$\pi(w_1) = \frac{1}{M(1,\eta)}[(2-\eta)w_1 + 2(\eta + c) - 4 + 2V]. \tag{11}$$

The profit increases as we move towards the Northwest in Figure 2. Shareholders must thus choose a governance structure on the highest isoprofit that intersects the incentive-compatible region.

Figure 3 shows the unique governance structure (w_1^*, π^*) when the solution is interior. This case occurs when the isoprofit slope lies between the help IC slope and the effort IC slope. This condition can be written as

$$\frac{M(1,\eta)}{B} \in \left(1, \frac{2-\eta}{1-\eta}\right). \tag{12}$$

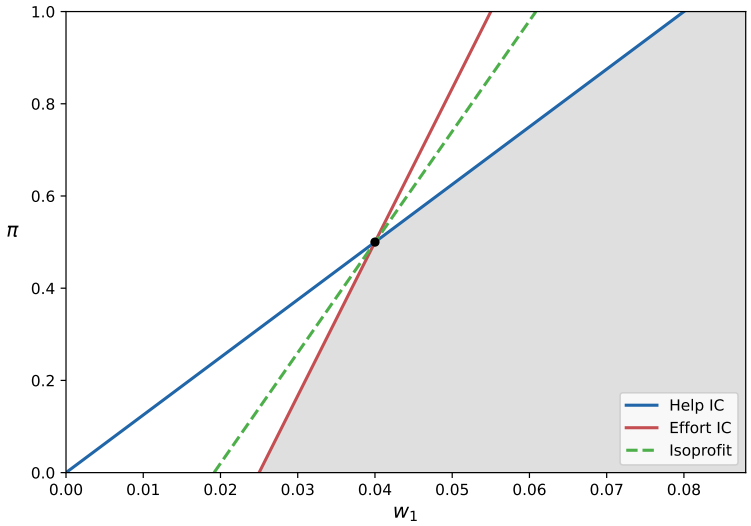


Figure 3: Optimal governance structure when the solution is interior

(Parameters: $\eta = 0.4$, $B = 0.048$, $c = 0.02$)

To understand (12), notice that if the benefit of monitoring $M(1,\eta)$ is less than the cost of monitoring B, then we must be at a corner solution where the optimal monitoring is zero ($\pi^* = 0$). If, instead, $M(1,\eta) > B$, an interior solution obtains unless the benefit of monitoring is "too high" (i.e., $M(1,\eta) > \frac{2-\eta}{1-\eta}B$), in which case the optimal solution requires $\pi^* = 1$.

The following proposition formally characterizes the optimal governance structure.

Proposition 1 (Optimal governance structure). *The (generically unique)*[27] *optimal governance structure is*

$$(w_1^*, \pi^*) = \begin{cases} (\frac{2c}{2-\eta}, 0) & \text{if } M(1,\eta) < B \\ (2c, \frac{(1-\eta)2c}{B}) & \text{if } \frac{M(1,\eta)}{B} \in (1, \frac{2-\eta}{1-\eta}) \\ (\frac{B}{1-\eta}, 1) & \text{if } M(1,\eta) > (\frac{2-\eta}{1-\eta})B \end{cases} \tag{13}$$

Proposition 1 shows that a CEO-friendly board ($\pi^* < 1$) is optimal unless the benefit of monitoring is sufficiently high ($M(1,\eta) > \frac{2-\eta}{1-\eta}B$). Under an interior solution, (w_1^*, π^*)

[27] In the knife-edge cases of $M(1,\eta) = B$ or $M(1,\eta) = (\frac{2-\eta}{1-\eta})B$, there are multiple values of π that maximize profit.

solves (7) and (9):

$$w_1^* = 2c \text{ and } \pi^* = \frac{(1-\eta)2c}{B}. \tag{14}$$

Replacing (w_1^*, π^*) in (6) confirms that this constraint is slack (see the Appendix for the proof). The optimal board "friendliness" is $1 - \pi^* > 0$.

The interactive tool replicates Figure 3 as one of the presets. By changing parameter values, the tool regenerates this figure for all cases in Proposition 1.

6 AI Adoption by the CEO

In this section, we modify the benchmark model to allow the CEO to have access to an AI agent. We start with the simpler case in which the AI agent has problem-solving capabilities ($\kappa > 0$) but no skill-augmenting capabilities ($\alpha = 1$) (Subsection 6.1), before considering the case in which AI has both capabilities (Subsection 6.2). Subsection 6.3 briefly considers an extension in which AI has deal-sourcing capabilities. Since the analysis follows essentially the same steps as in the benchmark case, for brevity, we present only the steps necessary to understand the intuition behind the results.

6.1 AI as a Coworker

Let $\kappa > 0$ and $\alpha = 1$. A CEO who verifies that her assessment is incorrect can now ask the AI agent for help. Given Assumption 3, a bad CEO cannot verify the AI's assessment; thus she must ask either the board or the AI agent for help. The only difference from the benchmark model is that the help IC constraint is now

$$(1-\eta)w_1 + (1-\pi)B \geq \kappa w_1 + B. \tag{15}$$

Intuitively, a bad CEO who chooses not to ask the board for help must choose m_a as it strictly dominates m_c. Conditional on the CEO being bad, the probability that the AI agent solves the problem is κ. This option delivers expected payoff $\kappa w_1 + B$, making the help IC constraint harder to satisfy. In turn, shareholders now need a friendlier board to induce the

CEO to communicate truthfully with the board.

Figure 4 illustrates the introduction of an AI agent with problem-solving capabilities. The help IC constraint becomes flatter due to $\kappa > 0$, while the effort IC remains the same. The intersection now implies a lower monitoring intensity and a lower first-period bonus in an interior solution. Solving (9) and (15) simultaneously leads to

$$w_{1AI}^* = \frac{2c}{1+\kappa} < w_1^*. \tag{16}$$

and

$$\pi_{AI}^* = \frac{(1-\eta-\kappa)2c}{B(1+\kappa)} < \pi^*. \tag{17}$$

That is, when AI has problem-solving capabilities, in an interior solution, firms must choose a lower monitoring intensity. Intuitively, by offering an imperfect alternative to the board's advice, AI increases the CEO's payoff for not seeking the board's help. This makes the CEO less likely to reveal information that can lead to her dismissal. To encourage the CEO to communicate with the board, the board must become even friendlier. As a consequence, CEOs become more entrenched and are replaced less often. More CEO entrenchment is inefficient because CEOs are kept even when a better replacement is available. The next proposition formally states these results.[28]

[28] As κ increases, eventually Assumption 1 must fail, in which case in equilibrium the board learns nothing and CEO knowledge becomes obsolete. See Section IA.6 for an analysis of this case.

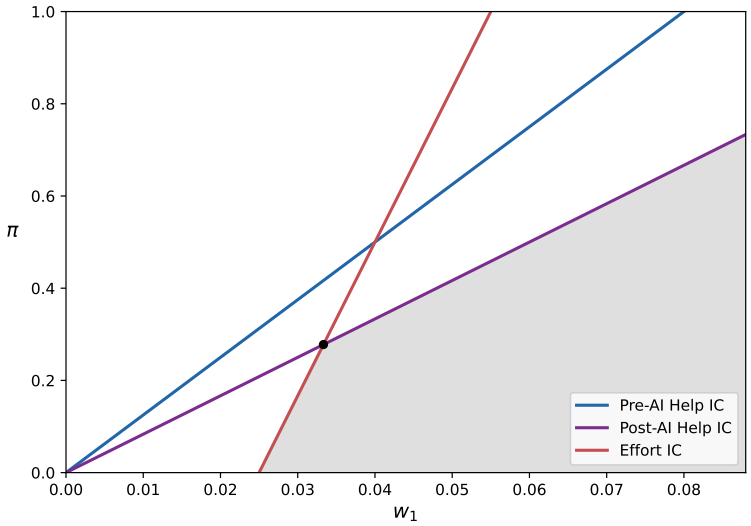


Figure 4: Incentive-compatible region under AI with problem-solving capabilities

(Parameters: $\eta = 0.4$, $B = 0.048$, $c = 0.02$, $\kappa = 0.2$, $\alpha = 1$)

Proposition 2 (Governance and AI's problem-solving capability). *Let $\kappa \in [0, 1-\eta]$ and $\alpha = 1$. The optimal governance structure is interior and given by (16) and (17) if and only if*

$$\frac{M(1,\eta)}{B} \in \left(1, \frac{2-\eta}{1-\eta-\kappa}\right). \tag{18}$$

Under an interior solution, an increase in κ decreases the first-period bonus, the board's monitoring intensity, and the firm's profit. The CEO's expected lifetime utility is independent of κ.

Proposition 2 implies that improvements in AI's problem-solving capabilities can be inefficient: an increase in κ reduces firm value without affecting the CEO's utility. The firm is worse off with a larger κ because it needs to monitor less intensively to encourage the CEO to communicate with the board. Lower monitoring intensity increases CEO entrenchment, which is inefficient. Because the CEO's lifetime effort IC constraint is binding, the firm fully bears the entrenchment cost.

Because the first-period bonus is lower under $\kappa > 0$, while the second-period bonuses are the same, incentive pay falls with κ. We can write the first-period CEO's expected

lifetime compensation as

$$E(w) = (1 - \frac{\eta}{2})w^*_{1AI} + \frac{1}{2}(1 - \frac{\eta}{3})w^*_{2g} + \frac{1}{2}(1 - \pi^*_{AI})(1 - \frac{2\eta}{3})w^*_{2b}. \tag{19}$$

The next corollary shows that expected total compensation also falls with κ.

Corollary 1 (CEO compensation and AI's problem-solving capability). *Expected total compensation is decreasing in κ.*

Thus, an increase in κ reduces both the average bonus and expected total compensation. That is, while the CEO benefits from increased job stability, her compensation falls as AI's problem-solving capabilities improve. Because the first-period CEO's utility is $2B$ regardless of κ, the CEO is indifferent to the value of κ.

6.2 AI as Both a Tool and a Coworker

We now extend the model to consider the case in which the AI agent possesses skill-augmenting capabilities (i.e., $\alpha \in [1, \overline{\alpha}]$), in addition to its problem-solving capabilities. Note that the help IC constraint in (15) does not change. The second-period bonus for a bad CEO, w^*_{2b}, also does not change. The second-period bonus for a good CEO becomes

$$w^*_{2gAI} = \frac{3c}{3 - \frac{\eta}{2\alpha - 1}} < w^*_{2g} \tag{20}$$

and the second-period bonus for a newly appointed CEO becomes

$$w^*_{2nAI} = \frac{2c}{2 - \frac{\eta}{\alpha}} < w^*_{2n}. \tag{21}$$

Intuitively, because the CEO is more competent with AI's help, it is easier to incentivize her to exert effort, requiring smaller bonuses overall.

The first-period effort incentive compatibility constraint becomes

$$B - c + (1 - \frac{1}{2\alpha})(w_1 + B) + \frac{1}{2\alpha}\left((1 - \eta)w_1 + (1 - \pi)B\right) \geq 2B. \tag{22}$$

Note that a larger α relaxes the effort IC because it increases the probability that the CEO can solve the problem on her own.

The firm's expected profit at the beginning of $t = 1$ is

$$V(w_1, \pi) = \frac{2\alpha - 1}{2\alpha}\Big(\underbrace{2 - \frac{\eta}{3(2\alpha - 1)} - c - w_1}_{\text{Profit from good CEO}}\Big) + \frac{1}{2\alpha}\Big(\underbrace{2 - \frac{5}{3}\eta - c - (1 - \eta)w_1 + \pi M(\alpha, \eta)}_{\text{Profit from bad CEO}}\Big). \tag{23}$$

Note that—all else constant—a larger α increases the firm's profit in both states (g and b), and also increases the probability that the CEO is perceived as "good." Because the profit is larger for type g than b for any governance structure (w_1, π), we conclude that α is a positive profit shifter.

The next proposition shows how α affects governance.

Proposition 3 (Governance and AI's skill-augmenting capability). *Let $\kappa \in [0, 1 - \eta]$ and $\alpha \geq 1$. The optimal governance structure is interior and given by*

$$w_{1AI}^* = \frac{2c\alpha}{2\alpha - 1 + \kappa} < w_1^*. \tag{24}$$

and

$$\pi_{AI}^* = \frac{(1 - \eta - \kappa)2c\alpha}{B(2\alpha - 1 + \kappa)} < \pi^*, \tag{25}$$

if and only if

$$\frac{M(\alpha, \eta)}{B} \in \Big(1, \frac{2\alpha - \eta}{1 - \eta - \kappa}\Big). \tag{26}$$

Under an interior solution, an increase in α decreases the first-period bonus and the board's monitoring intensity, and increases the firm's profit. The CEO's expected lifetime utility is independent of α.

Proposition 3 shows that the monitoring intensity also decreases with the AI agent's skill-augmenting capability, α. An increase in α makes the CEO more productive, increasing the benefit of exerting effort and loosening the CEO's effort IC constraint. This allows the firm to incentivize effort provision with a lower first-period bonus, w_1. But a lower first-

period bonus makes it harder to meet the help IC constraint, forcing the firm to decrease board independence, π. Figure 5 illustrates the effect of α on the optimal governance structure.

Overall, as in the case of an increase in κ, both board independence and CEO compensation fall when α increases, as the following corollary shows.

Corollary 2 (CEO compensation and AI's skill-augmenting capability). *Expected total compensation decreases with α.*

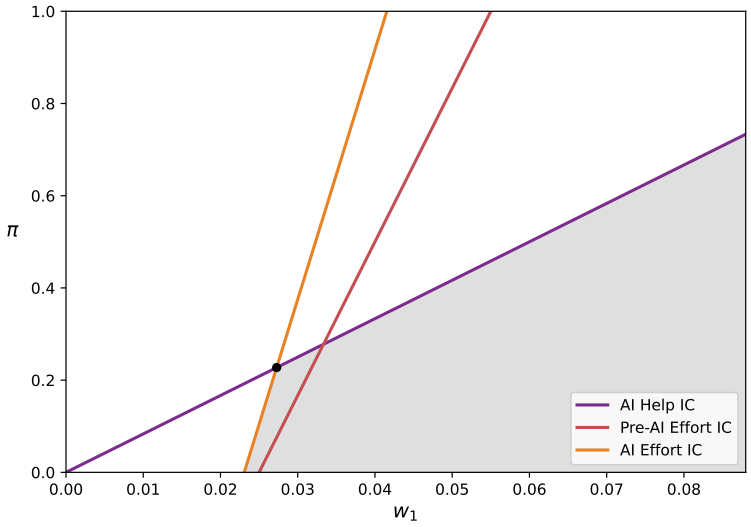


Figure 5: Incentive-compatible region under AI with both capabilities

(Parameters: $\eta = 0.4$, $B = 0.048$, $c = 0.02$, $\kappa = 0.2$, $\alpha = 1.5$)

Despite the equilibrium displaying lower board independence, the firm is better off with a higher α because the increase in CEO productivity, coupled with the fall in compensation (in both periods), more than compensates for the (inefficient) lower CEO turnover rate. This result is robust and not driven by specific parameter assumptions. This occurs because an increase in α both expands the incentive-compatibility region (See Figure 5) and shifts the profit function upward. Intuitively, α expands the technological frontier without (directly) affecting the CEO's incentives to share information. Thus, increasing α must be welfare-increasing. Because the effort IC constraint binds, the CEO's utility is unaffected by α, and the firm captures the entire welfare increase as profits.

6.3 AI and Deal Sourcing

In our model, the CEO must first exert effort at a cost c to identify a deal. While we assume that deal sourcing cannot be fully automated, it is still reasonable to expect that the CEO's adoption of AI will reduce her deal-sourcing cost.

Suppose that AI's skill-augmenting capability can also be used to reduce the deal-sourcing cost, c. As we can see directly from (16) and (17), a decrease in c decreases both w_{1AI}^* and π_{1AI}^*. Intuitively, a decrease in c has an effect similar to an increase in α: it slackens all effort incentive constraints without affecting the help IC constraint.[29] We conclude that our predictions do not depend on the channel through which AI augments the CEO's skills: if AI enhances the CEO's ability to solve problems or reduces her cost of finding deals, both CEO compensation and board monitoring decrease.

The analysis is more nuanced if we allow an AI agent to substitute for the CEO by independently identifying deals. Suppose an AI agent can autonomously identify a deal with probability $\phi \in [0, 1]$ at no cost. This case is arguably less realistic, as it conflicts with our assumption that deal sourcing is a "messy job" that needs to be performed by humans. Still, for completeness, we provide a brief analysis of this case.

The immediate effect of allowing the CEO to (covertly) delegate deal sourcing to an AI agent with deal-sourcing capability ϕ is to tighten all incentive constraints. Thus, if the optimal contract requires the CEO to exert effort instead of delegating deal sourcing to the AI agent, an increase in ϕ must increase all bonuses. If the effect of ϕ on w_1 is sufficiently strong, π will also increase with ϕ through the help IC constraint. However, as ϕ increases, inducing the CEO to exert effort must eventually become suboptimal. Intuitively, as ϕ increases, the AI agent becomes an ever cheaper alternative to the CEO for finding deals. At some point, the optimal contract must ignore the effort IC constraints. In that case, both CEO compensation and board monitoring must fall to their minimum possible values. The following proposition formalizes these results.

Proposition 4 (Governance and AI's deal-sourcing capability). *Suppose the AI agent can*

[29] Unlike c, α also affects the probabilities of the CEO being of types b and g, as well as the value of monitoring.

identify a deal with probability $\phi \in [0,1]$. Assume that (12) holds, so that an interior solution exists if $\phi = 0$. The following holds:

1. *If ϕ is sufficiently small, both w_1^* and π^* increase with ϕ.*
2. *There exists $\overline{\phi} < 1$ such that for $\phi \geq \overline{\phi}$, $w_1^* = 0$ and $\pi^* = 0$.*

Taken together, the results in this subsection show that AI's impact on governance through the deal-sourcing channel depends on whether AI complements or substitutes for the CEO's effort. When AI reduces the CEO's cost of sourcing deals, the effect mirrors that of skill augmentation: both compensation and monitoring decrease, reinforcing our earlier findings. When AI can autonomously identify deals, however, the relationship is non-monotonic. For low levels of autonomous deal-sourcing capability, the need to prevent the CEO from covertly delegating to the AI agent drives both compensation and monitoring upward. Beyond the "autonomy threshold," however, the firm optimally stops incentivizing the CEO to source deals altogether, and both compensation and monitoring collapse to zero (or their minimum possible values). Thus, the "long-run effect" of AI through deal sourcing (i.e., when AI improves so much that $\phi \geq \overline{\phi}$) is to reduce both CEO compensation and board monitoring.

6.4 Summary of Empirical Predictions

To summarize, our analysis indicates that improvements in AI (both κ and α) are associated with lower board independence and CEO turnover. AI improvements also reduce CEO incentive compensation, regardless of which capability is enhanced (κ or α). These predictions are reinforced when AI reduces the CEO's cost of sourcing deals (lower c), and they also hold in the long run when AI can autonomously identify deals with probability ϕ—though in the latter case, the short-run effects are non-monotonic, with both compensation and monitoring initially increasing before collapsing.

The welfare implications of AI adoption and improvements are ambiguous, as they depend on the relative strength of each AI capability (κ or α) and how technological advances

affect these capabilities. If CEOs use AI primarily as a substitute for the board's advice, total welfare is reduced.

Empirically, these predictions relate to the CEO's adoption of AI, all else held constant. If a shock increases AI usage across the firm as a whole, e.g., knowledge workers use AI more intensively, our predictions may not hold because the resulting improvements in firm productivity may indirectly affect the determination of optimal governance structures. Thus, empirical tests of our predictions must identify shocks to CEOs' AI use that are orthogonal to other shocks affecting firm productivity and governance.

7 AI Adoption by the Board

So far, we have assumed that AI adoption only benefits the CEO. But what if board members can also use AI in their work? In this section, we extend the model to consider how AI affects the board's performance in its advisory and monitoring roles. In Subsection 7.1, we consider technological advances that marginally improve the board's advising and monitoring capabilities. We call this case the "short-run effects of board AI adoption." Subsection 7.2 then considers discrete technological advances; we call these the "long-run effects."

7.1 Board AI adoption: Short-run effects

We first consider the case in which AI can help the board in its advisory role. Suppose the board has access to an AI agent with problem-solving capability κ. If the CEO asks the board for advice, the board can either spend one unit of time assessing the deal x, in which case output is $y_t = 1$ with probability $1 - \eta$, or ask the AI agent for an assessment, in which case output is $y_t = 1$ with probability κ. Assumption 2 ($1 - \eta > \kappa$) implies that the board always prefers to assess the deal rather than delegate this task to the AI agent. Thus, AI's problem-solving capabilities are irrelevant to the board.

Things are more interesting when the board can benefit from AI's skill-augmenting capabilities. Although we assume the board's knowledge is 1, AI can still reduce the cost

of board advice by reducing η. For example, AI may reduce the time the board spends on routine compliance tasks. Alternatively, we may assume that the board's knowledge is less than 1, and that the AI agent's skill-augmenting capability enables the board to solve harder problems *and* perform more routine compliance tasks per unit of time.

For simplicity, assume that the board's problem-solving capability is $\beta(1-\eta)$, where $\beta \in [1, \overline{\beta}]$ is the AI's *board skill-augmenting capability*. To find an interior equilibrium, we replace $1-\eta$ with $\beta(1-\eta)$ in the help IC constraint (15) and the effort IC constraint (22). The optimal first-period bonus is unchanged and given by (24). The optimal monitoring intensity becomes

$$\pi^*_{AI} = \frac{(\beta(1-\eta)-\kappa)2c\alpha}{B(2\alpha-1+\kappa)}. \tag{27}$$

It is immediate that increasing β increases the optimal monitoring intensity. Intuitively, higher β makes the board a better advisor, providing further incentives for the CEO to ask for help. Thus, all else constant, shareholders can choose a more independent board while still meeting the CEO's help IC constraint. The total CEO pay is

$$E(w) \;=\; 2c + \frac{\pi^* B}{2\alpha} \;=\; 2c + \frac{\big(\beta(1-\eta)-\kappa\big)c}{2\alpha-1+\kappa}, \tag{28}$$

which is increasing in β: pay has to increase with monitoring intensity to compensate for the expected loss of B. We conclude that AI's board skill-augmenting capability can partly counter the effect of CEO AI adoption on the monitoring intensity, has no effect on the first-period bonus, and reduces total CEO pay.

We now consider the case in which AI can also assist the board in performing its monitoring task. Suppose that, with the help of AI, the board may independently learn some of the information the CEO possesses. For example, with AI, the board may find it easier to process market information about the firm.[30] We maintain the assumption that the CEO must exert effort to identify a deal. That is, the board cannot identify a deal unknown to the CEO. For concreteness, suppose that at the beggining of date (iv), with probability $\lambda \in [0,1]$ the board learns both about the existence of a deal (with unknown difficulty x)

[30]See Ferreira, Ferreira, and Raposo (2011) for evidence that the board's monitoring function is helped by information embedded in stock prices.

and whether $z \geq x$ or $z < x$, that is, whether the CEO can assess the deal.[31] At first blush, it may seem that a bad CEO would now be more willing to ask the board for help, because the board may learn that the CEO is bad even without the CEO asking for help. However, this reasoning is incomplete because when the board learns about a problem the CEO cannot solve, the board can solve it with probability $\beta(1-\eta)$, which benefits the CEO through her first-period bonus w_1.

The help IC constraint is

$$\beta(1-\eta)w_1 + (1-\pi)B \geq (1-\lambda)(\kappa w_1 + B) + \lambda\left(\beta(1-\eta)w_1 + (1-\pi)B\right), \tag{29}$$

which implies

$$\beta(1-\eta)w_1 + (1-\pi)B \geq \kappa w_1 + B. \tag{30}$$

That is, the board's ability to independently obtain the CEO's information does not affect the help IC constraint. In an equilibrium in which the help IC constraint is met, the CEO always shares information with the board, thus λ also does not affect the effort IC constraints or the isoprofits, implying that the equilibrium is unaffected by λ.

Because advances in AI are likely to shift AI capabilities in the same direction (but perhaps by different magnitudes), the comparative static exercises we have performed thus far have some apparent limitations. In the Internet Appendix (Section IA.6), we consider shocks that improve (α, κ, β) simultaneously (small changes in λ have no impact on the equilibrium). We show that to reverse our conclusions in Propositions, shocks to AI technology must be significantly biased toward the board, meaning AI must augment directors' skills more than CEOs'. If we instead expect AI to benefit CEOs and boards in similar ways (or to benefit the CEO more), AI adoption must decrease board monitoring and CEO pay levels.

[31] It is less reasonable to assume that the board can learn about the CEO's ability to assess a particular deal without learning also what the deal is.

7.2 AI board adoption: Long-run effects

The comparative statics above are local: they describe small changes in β and λ. Here we consider large increases in such parameters.

We first set $\lambda = 0$. With the board's problem-solving capability equal to $\beta(1-\eta)$, the marginal benefit of monitoring becomes

$$M(\alpha, \eta; \beta) := \big(1 - \beta(1-\eta)\big)\Big(\frac{2}{3} - \frac{1}{2\alpha}\Big), \tag{31}$$

which is *decreasing* in β: a bad CEO advised by a better board produces almost as much as a replacement advised by the same board, so the quality gap that monitoring exploits shrinks as advice improves. Thus, β pushes the two sides of the governance trade-off in opposite directions: it relaxes the help IC constraint—raising π^* in the interior, as shown above—while eroding the value of the replacement option.

For β sufficiently large, the second effect must eventually win. The interior solution requires $M(\alpha, \eta; \beta) > B$, but $M(\alpha, \eta; \beta)$ converges to zero as $\beta(1-\eta) \to 1$. Whenever the interior solution exists at $\beta = 1$ (i.e., $M(\alpha, \eta) > B$), there is therefore a threshold $\overline{\beta} \in (1, \frac{1}{1-\eta})$ at which the optimal governance structure jumps equals that of Case 1 of Proposition 1: the monitoring intensity drops discontinuously to $\pi^*_{AI} = 0$. The effect of β on board monitoring is thus nonmonotonic. In the short run, better advice makes communication more attractive to the CEO, and the board can afford to be more independent. In the long run (i.e., when $\beta > \overline{\beta}$), advice becomes so good that replacing an underperforming CEO is no longer worth the cost, and entrenchment dominates: the board is maximally friendly, and a bad CEO remains employed.

At $\beta > \overline{\beta}$, the first-period bonus jumps down to

$$w^*_{1AI} = \frac{2\alpha c}{2\alpha - 1 + \beta(1-\eta)}, \tag{32}$$

which is now decreasing in β. The total pay jumps down to $E(w) = 2c$, and is invariant to β for $\beta > \overline{\beta}$. In this equilibrium, AI adoption by the board eventually leads to full specialization: the board advises but does not monitor the CEO, who is now fully entrenched.

The board

Now, we set $\beta = 1$ and let λ vary. We have the following result (see Section IA.7 in the Internet Appendix for the proof).

Proposition 5 (**Vet-or-out governance**). *There exists $\overline{\lambda} < 1$ such that, for all $\lambda > \overline{\lambda}$, the optimal governance structure sets $\pi^*_{AI} = 1$ and $w^*_{1AI} = 0$. The equilibrium features partial pooling: a good CEO remains silent with probability one, a bad CEO remains silent with probability $\mu^* \in (0,1)$, and the board replaces every CEO whose deal it has not inspected. Expected total CEO pay equals c, and CEO turnover equals $1 - \lambda\left(1 - \frac{1}{2\alpha}\right)$, of which $\left(1 - \frac{1}{2\alpha}\right)(1-\lambda)$ is the replacement of good CEOs.*

Above $\overline{\lambda}$, the firm switches to what we call *vet-or-out governance*: it sets $\pi^* = 1$, cuts the bonus to zero, and stops satisfying the help IC constraint. The board retains only CEOs it has positively vetted—those whose deal it inspected and found them able to handle—and replaces everyone else: CEOs who asked for help, CEOs it caught concealing, and, crucially, CEOs about whom it learned nothing. Effort no longer needs to be paid for: a CEO who sources a deal has a chance $\lambda\left(1 - \frac{1}{2\alpha}\right)$ of being inspected, vetted as good, and retained, while a CEO who shirks sources no deal, generates nothing to inspect, and is replaced along with the unvetted. The retention motive alone sustains effort whenever $\lambda\left(1 - \frac{1}{2\alpha}\right)B \geq c$, which Assumption 5 guarantees.

Finally, we let both board AI capabilities become large. Since $\overline{\beta} < \frac{1}{1-\eta}$, taking $\beta(1-\eta)$ to 1 places the communication regime in the entrenchment case above. The question is thus how good inspection must be for the firm to prefer vet-or-out governance instead, which is our next result (see Section IA.7 in the Internet Appendix for the proof).

Proposition 6 (**Undermonitoring or overmonitoring**). *There exists $\hat{\lambda} \in (0,1)$ such that, as $\beta(1-\eta) \to 1$, the optimal governance structure implies $\pi^*_{AI} = 0$ (full entrenchment, undermonitoring) for $\lambda < \hat{\lambda}$, and $\pi^*_{AI} = 1$ in a vet-or-out governance regime (overmonitoring), for $\lambda > \hat{\lambda}$.*

The two long runs distort governance in opposite directions relative to the first best. For $\lambda < \hat{\lambda}$ the board *undermonitors*: no CEO is ever replaced, advising is its only function, and every bad CEO is advised. For $\lambda > \hat{\lambda}$ it *overmonitors and underadvises*: it replaces all bad

CEOs—which is efficient—and also the good CEOs it failed to vet, while a fraction of bad CEOs stays silent and unadvised.

8 Discussion and conclusions

This article develops a framework for understanding how AI transforms the relationship between CEOs and corporate boards. This framework illuminates the corporate governance consequences of CEO AI adoption, which is still in its early stages but growing rapidly (cf. Yotzov et al. (2026)).

We model two AI capabilities: problem-solving, which substitutes for human judgment, and skill-augmenting, which complements human decision-making. Both capabilities reduce board monitoring intensity and CEO compensation, but through different channels. While AI's problem-solving capability destroys value by making it more difficult for the board to identify and replace bad CEOs, its skill-augmenting capability increases firm value despite reducing board monitoring. When boards also adopt AI, some negative effects are mitigated in the short run, but if AI greatly augments the board's advisory capabilities, board monitoring adds little value, and the board specializes in advising. In this case, the board becomes completely CEO-friendly, and CEOs are maximally entrenched. If, instead, the board uses AI mostly to improve its monitoring capability, board-CEO communication quality eventually deteriorates, and the board "underadvises" and "overmonitors" the CEO, inefficiently replacing even some capable CEOs.

Our model reveals a tension in AI adoption: Technology that boosts CEOs' capabilities also reduces their dependence on board advice, disrupting information flows between CEOs and directors. Firms respond by reducing board independence to restore communication incentives, which entrenches the CEO. This result qualifies the view that technological development improves organizational efficiency. Our model shows that AI adoption requires redesigning governance structures by adjusting both monitoring intensity and compensation packages.

Our framework extends to any setting where principals need agents to both exert effort and communicate truthfully. Consider a sales manager overseeing field agents who must

both pursue leads (effort) and report difficulties in dealing with potential customers (communication). If AI tools enable agents to handle customer interactions more effectively, agents may stop sharing information with managers. Similarly, procurement officers need to both identify suppliers (effort) and evaluate them. Better AI might prompt the procurement officer to bypass marginal suppliers and go directly to the manager, thereby creating operational risks. In each case, our model predicts that organizations must reduce monitoring intensity and adjust compensation to maintain information flows.

A broad theoretical insight from our analysis is that AI can paradoxically increase transaction costs by tightening incentive compatibility constraints. This occurs because AI affects not only the agent's equilibrium payoff, but also her off-equilibrium payoff: the value of not asking for help in our setting. Under our maintained assumption that the board's advice remains superior, the CEO goes to the board on the equilibrium path, so AI functions as an off-equilibrium threat. The governance distortion arises from AI's availability, not its use. Organizations that fail to recognize this and adapt their governance structures accordingly may find that AI adoption leads to worse outcomes than the pre-AI era.

Appendix: Proofs

Proof of Lemma 1: Using the law of total probability, we have

$$\Pr(S_2|g) = \int_0^1 \Pr(S_2|z,g) \cdot f(z|g)\, dz.$$

Since problems are independent given z, then

$$\Pr(S_2|z,g) = \Pr(S_2|z) = \min(\alpha z, 1).$$

From Bayes' theorem, the posterior distribution of z given g is

$$f(z|g) = \frac{\Pr(g|z) \cdot f(z)}{\Pr(g)} = \frac{\min(\alpha z, 1)}{1 - \frac{1}{2\alpha}}.$$

Therefore

$$\Pr(S_2|g) = \frac{1}{1-\frac{1}{2\alpha}} \int_0^1 [\min(\alpha z, 1)]^2 \, dz = \frac{1}{1-\frac{1}{2\alpha}} \int_0^{1/\alpha} (\alpha z)^2 \, dz + \int_{1/\alpha}^1 1 \, dz = \frac{2(3\alpha-2)}{3(2\alpha-1)}.$$

Similarly, we have

$$\Pr(S_2|b) = \int_0^1 \Pr(S_2|z,b) \cdot f(z|b) \, dz.$$

Since problems are independent given z, we have

$$\Pr(S_2|z,b) = \Pr(S_2|z) = \min(\alpha z, 1).$$

If $z > 1/\alpha$, the CEO can solve any problem. Therefore, $f(z|b) = 0$ for $z > 1/\alpha$. For $z \leq 1/\alpha$, we have $f(z|b) = 2\alpha(1-\alpha z)$. Then

$$\Pr(S_2|b) = \int_0^{1/\alpha} \alpha z \cdot 2\alpha(1-\alpha z) \, dz = 2\alpha^2 \left[\frac{z^2}{2} - \frac{\alpha z^3}{3} \right]_0^{1/\alpha} = \frac{1}{3}.$$

□

Proof of Lemma 2: Suppose the board is tough. At Date (iii) of $t = 1$, it must then decide whether to retain or fire the CEO. Suppose $p_1(\tau)$ is uninformative about τ, i.e., $p_1(\tau) = p$ is a constant. Thus, the board expects the incumbent CEO to be good with probability $1 - \frac{1}{2\alpha}$ regardless of whether the board observes m_b or $\{m_a, m_c\}$. Note that, in period 2, a CEO has no career concerns and thus does not need to be incentivized to ask the board for help when she cannot assess a deal. However, the CEO still needs to be incentivized to exert effort c to identify a deal. If a CEO of type τ is to exert effort, her second-period *effort incentive constraint* must hold:

$$\Big(\Pr(S_2 \mid \tau) + (1 - \Pr(S_2 \mid \tau))(1-\eta) \Big) w_2 - c + B \geq B. \tag{A.1}$$

Suppose first that the board fires the CEO and hires a replacement. We denote the replacement's type by n. We have $\Pr(S_2 \mid n) = 1 - \frac{1}{2\alpha}$. The board will make the new CEO's

incentive constraint bind, which implies a second-period expected profit of

$$v_{2n} = 1 - \frac{\eta}{2\alpha} - c. \tag{A.2}$$

Suppose instead that the firm retains the incumbent CEO, and offers a wage that is sufficient to incentivize the bad CEO, $w_{2b} = \frac{3c}{3-2\eta}$ (here we consider the case in which the CEO exerts effort in the first period, so she privately learns about her type). Under this contract, both CEO types, b and g, will exert effort in period 2. Let v_{2gb} denote the second-period profit when the firm retains an incumbent of unknown type and both types are induced to exert effort. After algebra, we get

$$v_{2gb} = \frac{(3 - 2\eta - 3c)(2\alpha - \eta)}{2\alpha(3 - 2\eta)}$$

and thus

$$v_{2n} - v_{2gb} = \frac{c\eta(4\alpha - 3)}{2\alpha(3 - 2\eta)} > 0,$$

implying that the board strictly prefers hiring a new CEO to retaining a CEO of unknown type with a contract that induces both types to exert effort.

Alternatively, the board may choose w_{2g} to incentivize only the good CEO. In this case, we have expected profit

$$v_{2g} = \frac{2\alpha - 1}{2\alpha}(1 - \frac{\eta}{3(2\alpha - 1)} - c),$$

and thus

$$v_{2n} - v_{2g} = \frac{1}{2\alpha}(1 - \frac{2\eta}{3} - c) > 0,$$

which follows from Assumption 4. Thus, if the board learns nothing, it always prefers to replace the CEO. But if a bad CEO knows she will be replaced, she will ask for advice, because her expected payoff will be $(1 - \eta)w_1 > 0$ (because w_1 must be positive if the CEO is to exert effort in period 1, knowing she will be replaced in period 2). This implies that a bad CEO always shares information with the board, i.e., $p_1(b) = 1$. By contrast, a good CEO prefers not to ask the board for help, because she can solve the problem with probability 1, while the board solves it only with probability $1 - \eta$, implying $p_1(g) = 0$,

contradicting $p_1(b) = p_1(g) = p$. We conclude that $p_1(g) \neq p_1(b)$.

Because there are no career concerns in $t = 2$, trivially $p_2(b) = 1$ and $p_2(g) = 0$. □

Proof of Lemma 3. Conditional on knowing that her own ex post type is g at $t = 1$, a CEO who chooses m_c expects to receive w_1 with probability 1. By choosing m_b instead, the CEO expects $(1 - \eta)w_1$. Thus, a good CEO would only choose m_b with strictly positive probability if asking the board for help increases her probability of retention, i.e., if $\rho(m_b) < \rho(m_c)$. But if $\rho(m_b) < \rho(m_c)$, action $m = m_b$ is strictly dominant for a CEO of type b, implying $p_1(b) = 1$. Lemma 2 then implies $p_1(g) < 1$. This means that when the board observes m_b, it must revise down its expectation of the CEO's knowledge, as the CEO is more likely to be bad than good. But then choosing $\rho(m_b) < \rho(m_c)$ is not rational for shareholders, because a replacement CEO is equally likely to be good or bad, and the proof of Lemma 2 shows that replacement always dominates retention when the incumbent is equally likely to be good or bad. We conclude that $p_1(g) = 0$. □

Proof of Proposition 1: **Case 1:** $M(1, \eta) < B$. In this case, the isoprofit is steeper than the effort IC constraint. As argued in the text, the optimal π must be zero in this case. This implies that the help IC constraint is slack; the "bad" CEO always asks for help. The optimal first-period bonus is determined by the effort IC constraint alone:

$$B - c + \frac{1}{2}(w_1^* + B) + \frac{1}{2}\left((1 - \eta)w_1^* + B\right) = 2B, \tag{A.3}$$

which implies $w_1^* = \frac{2c}{2-\eta}$. Replacing it in (10) yields $V(w_1^*, 0) = 2 - \eta - 2c$, which is positive given Assumption 4. Because violating the effort IC constraint leads to zero profit, choosing $w_1^* = \frac{2c}{2-\eta}$ and $\pi^* = 0$ must be the unique optimal solution for this set of parameters.

Case 2: $\frac{M(1,\eta)}{B} \in (1, \frac{2-\eta}{1-\eta})$. As illustrated in Figure 3, the tangency between the highest isoprofit and the IC region happens when both (7) and (9) bind. Solving for (w_1^*, π^*) yields:

$$w_1^* = 2c \text{ and } \pi^* = \frac{(1 - \eta)2c}{B}. \tag{A.4}$$

To confirm that constraint (6) is slack, we replace (w_1^*, π^*) in (6) to get slack

$$\frac{c\big(6B - 5B\eta + 2c\eta(1-\eta)\big)}{B(3-2\eta)} > 0 \quad \text{for all } \eta \in (0,1),\ B > 0,\ c > 0.$$

Thus, the IC constraint in (6) is slack at the optimal solution.

To confirm that (A.4) is the unique optimal governance structure in this case, we also need to consider values of (w_1, π) that violate (7) under pure strategies. If any equilibrium also violates the first-period effort IC constraint, its overall profit is bounded above by v_{2n}, which is less than the profit under the optimal pure-strategy equilibrium. Thus, we do not need to consider these "shirking equilibria," as they are strictly dominated.

If (7) is violated, there is no pure-strategy equilibrium (without shirking): with $\rho(m_c) = 0$ the bad type strictly prefers m_c, leading to pooling (ruled out by Lemma 2). With $\rho(m_c) = 1$, m_b is a dominant action for type b, so we must have $p_1(g) < 1$. Then, m_c would reveal type g and $\rho(m_c) = 1$ cannot be sequentially rational. Thus, for such values, equilibria with first-period effort must involve mixed strategies. We now show that such mixed-strategy equilibria always deliver lower profits than some pure-strategy equilibria associated with governance structure (w_1^*, π^*). That is, mixed-strategy equilibria are never optimal equilibria.

Suppose (w_1, π) is such that only a strictly mixed-strategy equilibrium exists where the first-period effort constraint is satisfied. Let $p_1(b) = p \in (0,1)$ denote the equilibrium probability that a CEO of type b asks the board for help. Let $\rho(m_c) = \rho \in (0,1)$ be the equilibrium probability that a CEO who does not ask for help is replaced, conditional on the board being tough ($\omega = \omega^{tough}$). The bad CEO must be indifferent between m_b and m_c:

$$(1-\eta)w_1 + (1-\pi)B = \pi(1-\rho)B + (1-\pi)B, \tag{A.5}$$

which implies $\pi = \frac{(1-\eta)w_1}{(1-\rho)B}$. We now consider two cases.

Suppose first that $w_1 \geq 2c$. If a strict mixed-strategy equilibrium exists for governance structure (w_1, π), a tough board must be indifferent between firing or retaining a CEO who did not ask for help, implying the expected profit from retaining the incumbent must equal

$v_{2n} = 1 - \frac{\eta}{2} - c$, which is the expected second-period profit when a new CEO is hired. The weak board after m_c always retains the incumbent, faces the same posterior, and offers the same contract. So its retention profit is also v_{2n}. Thus, the unconditional expected second-period profit must be

$$v_{2n} - \frac{p}{2}(1-\pi)M(1,\eta).$$

We obtain this expression by noting that the profit after m_c must be v_{2n}, while after m_b it is v_{2n} if the CEO is replaced and v_{2b} if retained. The unconditional probability of retaining a bad CEO is $\frac{p}{2}(1-\pi)$, in which case the firm "loses" the marginal benefit of monitoring, $M(1,\eta)$.

By contrast, under (w_1^*, π^*), the optimal pure-strategy equilibrium implies an expected period 2 profit of

$$\frac{1}{2}v_g + \frac{1}{2}(\pi^* v_{2n} + (1-\pi^*)v_{2b}) = \frac{1}{2}(v_{2n} + M(1,\eta) + \pi^* v_{2n} + (1-\pi^*)(v_{2n} - M(1,\eta))$$

$$= v_{2n} + \frac{\pi^*}{2}M(1,\eta).$$

We conclude that the second-period profit is strictly larger under the pure-strategy equilibrium for the optimal governance structure, (w_1^*, π^*). Now we need to compare first-period profits. For a mixed-strategy equilibrium under (w_1, π), we have the first-period profit

$$\frac{1}{2}(1-w_1) + \frac{1}{2}p(1-\eta)(1-w_1),$$

while for the optimal pure strategy equilibrium, we have

$$\frac{1}{2}(1-w_1^*) + \frac{1}{2}(1-\eta)(1-w_1^*).$$

Since $w_1^* = 2c$ and $w_1 \geq 2c$, it follows that the first-period profit is also larger under the pure-strategy equilibrium for the optimal governance structure. We conclude that, if $w_1 \geq 2c$, any mixed-strategy equilibrium is dominated by the optimal pure-strategy equilibrium under (w_1^*, π^*).

Suppose now that $w_1 < 2c$. To prove our result, we first find an upper bound to the

profit under (w_1, π) and then show that this upper bound is strictly smaller than the profit under the pure-strategy equilibrium induced by (w_1^*, π^*).

Define the *first-best surplus*, S^{FB}, as the total surplus to all players if the game is played cooperatively: (i) the CEO exerts effort, (ii) a bad CEO always asks for help, and (iii) the board always replaces a bad CEO and retains a good CEO. S^{FB} denotes social surplus—output net of effort costs plus CEOs' private benefits. Since the firm employs one CEO each period and the employed CEO enjoys B, at least $2B$ of social surplus is non-extractable by shareholders. Moreover, each time the CEO is replaced, one additional individual collects the private benefit, and this benefit is also non-extractable: the dismissed CEO must be compensated for the benefit she loses—her utility cannot fall below her shirking payoff, and a shirker, being indistinguishable from a good CEO, keeps her job and thereby secures $2B$—while limited liability prevents shareholders from charging her successor for the benefit he gains. Each replacement thus adds B to the CEOs' total take. Because first-best play replaces the CEO with probability $\frac{1}{2}$, we define the *extractable first-best surplus* as

$$\overline{V}^{FB} := S^{FB} - 2B - \frac{1}{2}B.$$

In a pure-strategy equilibrium where both (7) and (9) bind, the $t = 1$ expected utility of an employed CEO is $2B$, which is the minimum that a CEO must get to exert effort at $t = 1$. Because (w_1^*, π^*) implies a pure strategy equilibrium where both (7) and (9) bind, the profit is

$$V(w_1^*, \pi^*) = \overline{V}^{FB} - \frac{1 - \pi^*}{2}\big(M(1, \eta) - B\big),$$

which can be verified directly from (10). The intuition is that the only source of inefficiency under a pure-strategy optimal equilibrium is that with probability $\frac{1}{2}(1 - \pi^*)$ a bad CEO is retained, which implies a surplus loss of $M(1, \eta)$, partly offset by the successor's benefit B that is not forfeited when the incumbent stays: the net cost of under-monitoring is $M(1, \eta) - B$ per unit of monitoring intensity, the same monitoring margin that determines the interior optimum. In addition, in this equilibrium, shareholders capture all of the surplus that the equilibrium allocation makes extractable.

Define $\overline{V}^m(w_1, \pi)$ as the extractable surplus under a mixed-strategy equilibrium asso-

ciated with (w_1, π). By definition, the extractable surplus is an upper bound for the profit under this equilibrium: $\overline{V}^m(w_1, \pi) \geq V(w_1, \pi)$.

We have

$$\overline{V}^m(w_1, \pi) = \overline{V}^{FB} - \frac{1 - \pi\big(p + (1-p)\rho\big)}{2}\big(M(1,\eta) - B\big) - \frac{\pi\rho}{2}\big(M(1,\eta) + B\big)$$

$$-\frac{1-p}{2}(1-\eta) + \pi\rho B.$$

We derive this expression by subtracting from the extractable first-best surplus the expected net losses from inefficient actions. There are four terms: (i) sometimes a bad CEO is retained (the second term on the right-hand side), which loses $M(1,\eta)$ but saves the successor's benefit B; (ii) sometimes a good CEO is replaced (the third term), which loses $M(1,\eta)$ and adds a successor's benefit B; (iii) sometimes a bad CEO does not ask for help (the fourth term); and (iv) the incumbent's shirking payoff is no longer $2B$ but $2B - \pi\rho B$, because a shirker is now caught in the randomized replacement of CEOs who do not ask for help, which makes $\pi\rho B$ of her rent extractable (the last term). This expression simplifies to

$$\overline{V}^m(w_1, \pi) = \overline{V}^{FB} - \frac{1 - p\pi(1-\rho)}{2}\big(M(1,\eta) - B\big) - \frac{1-p}{2}(1-\eta).$$

Now, use the fact that $\pi = \frac{(1-\eta)w_1}{(1-\rho)B}$ and $w_1 < 2c = w_1^*$ to conclude that $p\pi(1-\rho) < \pi^*$; since $M(1,\eta) > B$ in this case,

$$\frac{1 - p\pi(1-\rho)}{2}\big(M(1,\eta) - B\big) > \frac{1-\pi^*}{2}\big(M(1,\eta) - B\big),$$

implying $V(w_1^*, \pi^*) > \overline{V}^m \geq V(w_1, \pi)$. We conclude that the pure-strategy equilibrium under (w_1^*, π^*) has strictly higher profit than any mixed-strategy equilibrium.

As a final point, note that at (w_1^*, π^*), mixed strategy equilibria may also exist, with $p \in (0,1)$ and $\rho = 0$. However, any such equilibrium trivially has a lower profit than the pure-strategy equilibrium. Furthermore, such equilibria are not robust in the following sense. If under (w_1^*, π^*) the CEO is expected to play a strict mixed strategy, then shareholders

can choose instead contract $(w_1^* + \varepsilon, \pi^*)$ with $\varepsilon > 0$ infinitesimally small, which breaks the bad CEO's indifference between asking or not asking for help, leading to a unique equilibrium in pure strategies where the help IC constraint is (slightly) slack. The profit in this equilibrium is arbitrarily close to $V(w_1^*, \pi^*)$. We thus conclude that such mixed-strategy equilibria cannot be optimal and would never be chosen by shareholders.

Case 3: $M(1,\eta) > \frac{2-\eta}{1-\eta}B$. In this case, the isoprofit is flatter than the help IC constraint and, as argued in the text, the optimal board structure must be $\pi^* = 1$. The effort IC constraint is not binding; thus the help IC constraint determines the optimal first-period bonus: $w_1^* = \frac{B}{1-\eta}$. Replacing it in (10) yields $V(\frac{B}{1-\eta}, 1) = 2 - \eta - c - \frac{1-\frac{\eta}{2}}{1-\eta}B + \frac{\eta}{12}$. Because $M(1,\eta) > \frac{2-\eta}{1-\eta}B \Rightarrow \frac{\eta}{6} > \frac{2-\eta}{1-\eta}B \Rightarrow \frac{\eta}{12} > \frac{1-\frac{\eta}{2}}{1-\eta}B$, we have $V(\frac{B}{1-\eta}, 1) > 0$. □

Proof of Proposition 2: The effort IC constraint is independent of κ, and its slope is $\frac{2-\eta}{B}$. The isoprofit is also independent of κ, and its slope is $\frac{2-\eta}{M(1,\eta)}$. As we can see from Figure 3, an interior solution requires $M(1,\eta) > B$.

The help IC constraint's slope is now $\frac{1-\eta-\kappa}{B}$. The slope must be smaller than that of the isoprofit, implying

$$\frac{M(1,\eta)}{B} < \frac{2-\eta}{1-\eta-\kappa}.$$

All other steps are similar to those in the proof of Proposition 1 and are omitted for brevity. From (16) we have

$$\frac{\partial w_{1AI}^*}{\partial \kappa} = -\frac{2c}{(1+\kappa)^2} < 0,$$

and from (17) we have

$$\frac{\partial \pi_{AI}^*}{\partial \kappa} = \frac{2c(\eta-2)}{B(1+\kappa)^2} < 0.$$

From (10) we have

$$\frac{\partial V_{AI}^*}{\partial \kappa} = (1-\frac{\eta}{2})\frac{2c}{(1+\kappa)^2} + \frac{c(\eta-2)M(1,\eta)}{B(1+\kappa)^2} = \frac{c(2-\eta)}{(1+\kappa)^2}\Big(1 - \frac{M(1,\eta)}{B}\Big) < 0,$$

which follows because $M(1,\eta) > B$ in an interior solution. Because the effort IC constraint in (9) binds, the CEO's lifetime utility is $2B$. □

Proof of Corollary 1: Because the first-period effort IC constraint binds, we must have

$$E(w) + B - c + \frac{1}{2}(B - c) + \frac{1}{2}(1 - \pi^*_{AI})(B - c) = 2B,$$

which implies

$$E(w) = 2c + \frac{\pi^*_{AI}}{2}(B - c).$$

Since $B > c$ and π^*_{AI} is decreasing in κ, then $E(w)$ is decreasing in κ. □

Proof of Proposition 3: The effort IC constraint can be written as

$$\pi \leq \frac{(2\alpha - \eta)w_1 - 2\alpha c}{B}. \tag{A.6}$$

The isoprofit now becomes

$$\pi(w_1) = \frac{1}{M(\alpha, \eta)}[(2\alpha - \eta)w_1 + 2(\eta + c\alpha) - 4\alpha + 2\alpha V]. \tag{A.7}$$

As we can see from Figure 3, an interior solution requires the effort IC constraint to be steeper than the isoprofit, which implies $M(\alpha, \eta) > B$.

The help IC constraint is independent of α, and its slope is $\frac{1-\eta-\kappa}{B}$. The slope must be smaller than that of the isoprofit, implying

$$\frac{M(\alpha, \eta)}{B} < \frac{2\alpha - \eta}{1 - \eta - \kappa}.$$

All other steps are similar to those in the proof of Proposition 1 and are omitted for brevity. From (24) we have

$$\frac{\partial w^*_{1AI}}{\partial \alpha} = \frac{2c(\kappa - 1)}{(2\alpha - 1 + \kappa)^2} < 0,$$

and from (25) we have

$$\frac{\partial \pi^*_{AI}}{\partial \alpha} = \frac{(1 - \eta - \kappa)2c(\kappa - 1)}{B(2\alpha - 1 + \kappa)^2} < 0.$$

It can be shown that $\frac{\partial V^*_{AI}}{\partial \alpha} > 0$ by brute force, but that route is long and not illuminating.

A simpler and more intuitive (and elegant) proof is to note that increasing α (i) relaxes the effort IC constraint, (ii) does not affect the help IC constraint, and (iii) (strictly) increases profit for any given (w_1, π) (the arguments are presented in text, and follow directly from inspection of these expressions). It thus follows from (i)+(ii)+(iii) that the firm must do strictly better when maximizing profit under a larger α.

Because the effort IC constraint in (22) binds, the CEO's lifetime utility is $2B$. □

Proof of Corollary 2: Because the first-period effort IC constraint binds, we must have

$$E(w) + B - c + (1 - \frac{1}{2\alpha})(B - c) + \frac{1}{2\alpha}(1 - \pi^*_{AI})(B - c) = 2B,$$

which implies

$$E(w) = 2c + \frac{\pi^*_{AI}}{2\alpha}(B - c) = 2c + \frac{(1 - \eta - \kappa)c}{B(2\alpha - 1 + \kappa)}(B - c),$$

which is decreasing in α. □

Proof of Proposition 4. For notational simplicity, we work with the case in which $\alpha = 1$ and $\kappa = 0$; the results are identical in the general case. Suppose the CEO can ask an AI agent to identify a deal. If asked, the agent identifies a deal with probability $\phi \in (0, 1)$. Suppose that (12) holds, so we have an interior solution if $\phi = 0$. Suppose it is optimal to induce the CEO to exert effort under $\phi > 0$, regardless of type. A bad CEO's effort incentive constraint at Date (ii) of $t = 2$ is now

$$\Big(1 - \frac{2}{3}\eta\Big)w_{2b} - c + B \geq \phi\Big(1 - \frac{2}{3}\eta\Big)w_{2b} + B, \tag{A.8}$$

implying $w^*_{2b} = \frac{3c}{(1-\phi)(3-2\eta)}$. Similarly, we also have $w^*_{2g} = \frac{3c}{(1-\phi)(3-\eta)}$ and $w^*_{2n} = \frac{2c}{(1-\phi)(2-\eta)}$.

The help IC constraint is now

$$(1 - \eta)w_1 + (1 - \pi)\Big(\phi\Big(1 - \frac{2}{3}\eta\Big)w^*_{2b} + B\Big) \geq \phi\Big(1 - \frac{2}{3}\eta\Big)w^*_{2g} + B, \tag{A.9}$$

which simplifies to

$$\pi \leq \frac{(1-\eta)w_1 + \frac{\eta}{3-\eta}\frac{\phi}{1-\phi}c}{\frac{\phi}{1-\phi}c + B}. \tag{A.10}$$

The first-period effort IC constraint (when the help IC constraint binds) is

$$-c + \frac{1}{2}\Big(w_1 + \frac{\phi}{1-\phi}c + B\Big) + \frac{1}{2}\Big(\frac{1-\frac{2\eta}{3}}{1-\frac{\eta}{3}}\frac{\phi}{1-\phi}c + B\Big) = -c + \frac{w_1}{2} + \frac{\phi}{1-\phi}\frac{1-\frac{\eta}{2}}{1-\frac{\eta}{3}}c + B \geq$$

$$\phi(\frac{w_1}{2} + \frac{\phi}{1-\phi}\frac{1-\frac{\eta}{2}}{1-\frac{\eta}{3}}c) + (1-\phi)\Big(\frac{\phi}{1-\phi}\frac{1-\frac{\eta}{2}}{1-\frac{\eta}{3}}c\Big) + B = \phi\frac{w_1}{2} + \frac{\phi}{1-\phi}\frac{1-\frac{\eta}{2}}{1-\frac{\eta}{3}}c + B.$$

To understand the last term, note that if the CEO does not exert effort in period 1, she will not know her type in period 2 with probability $1-\phi$, in which case she will not exert effort in that period either. In that state, she will find a deal with probability ϕ (through AI), which she can assess with probability $1-\frac{\eta}{2}$, in which case she receives a bonus $w^*_{2g} = \frac{3c}{(1-\phi)(3-\eta)}$ (because the board now thinks the CEO is of type g given the lack of communication).

Solving for w_1^* yields

$$w_1^* = \frac{2c}{1-\phi}.$$

Notice that the bonus is increasing in ϕ; when $\phi = 0$, we have $w_1^* = 2c$ as in (14). Replacing this value in the help IC gives

$$\pi^* = \frac{(1-\eta)2c + \frac{\eta\phi c}{3-\eta}}{\phi c + (1-\phi)B}. \tag{A.11}$$

When $\phi = 0$, we have π^* as in (14). Notice that π^* is increasing in ϕ given Assumption 5 ($B \geq 2c$):

$$\frac{d\pi^*}{d\phi} = c\frac{\frac{\eta B}{3-\eta} - 2(1-\eta)(c-B)}{[\phi c + (1-\phi)B]^2} > 0.$$

If $\phi = 1$, trivially the board does not need to pay for the CEO to exert effort, because the AI agent always finds a deal. Thus, the effort IC constraints are irrelevant in this case. By continuity, there exists $\overline{\phi} < 1$ such that it is optimal to ignore all effort IC constraints for

$\phi \geq \overline{\phi}$. If all second-period bonuses are zero, we have that the help IC constraint implies $\pi^* = \frac{(1-\eta)w_1}{B}$. Then, we have that the derivative of V with respect to w_1 is

$$\frac{dV}{dw_1} = \phi\Big[\frac{M(1,\eta)}{B}\frac{1-\eta}{2} - \frac{2-\eta}{2}\Big],$$

which is negative because (12) holds. Thus, the optimal first period bonus is $w_1^* = 0$, implying that the optimal monitoring intensity is also $\pi^* = 0$ for all $\phi \geq \overline{\phi}$. □

References

ACEMOGLU, D., D. AUTOR, AND S. JOHNSON (2026): "Building Pro-Worker Artificial Intelligence," Tech. Rep. 34854, National Bureau of Economic Research.

ADAMS, R. B., A. C. AKYOL, AND P. VERWIJMEREN (2018): "Director Skill Sets," *Journal of Financial Economics*, 130, 641–662.

ADAMS, R. B. AND D. FERREIRA (2007): "A Theory of Friendly Boards," *Journal of Finance*, 62, 217–250.

AGHION, P. AND J. TIROLE (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1–29.

ALONSO, R. AND N. MATOUSCHEK (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–293.

AMADOR, M. AND K. BAGWELL (2013): "The Theory of Optimal Delegation with an Application to Tariff Caps," *Econometrica*, 81, 1541–1599.

ATHEY, S. C., K. A. BRYAN, AND J. S. GANS (2020): "The Allocation of Decision Authority to Human and Artificial Intelligence," *AEA Papers and Proceedings*, 110, 80–84.

BABINA, T., A. FEDYK, A. HE, AND J. HODSON (2025): "Firm Investments in Artificial Intelligence Technologies and Changes in Workforce Composition," in *Technology, Productivity, and Economic Growth*, ed. by S. Basu, L. Eldridge, J. Haltiwanger, and E. Strassner, University of Chicago Press, vol. 83 of *Studies in Income and Wealth*.

BALDENIUS, T., N. MELUMAD, AND X. MENG (2014): "Board Composition and CEO Power," *Journal of Financial Economics*, 112, 53–68.

BALDENIUS, T., X. MENG, AND L. QIU (2019): "Biased Boards," *The Accounting Review*, 94, 1–27.

BARANCHUK, N. AND P. H. DYBVIG (2009): "Consensus in Diverse Corporate Boards," *The Review of Financial Studies*, 22, 715–747.

BÁRÁNY, Z. L. AND M. KOREN (2026): "Broken Ladders: AI, Teamwork, and the Dynamics of Skill Formation in the Workplace," Working paper. First version: May 2025.

BERK, J. B. AND J. H. VAN BINSBERGEN (2022): "Regulation of Charlatans in High-Skill Professions," *Journal of Finance*, 77, 1219–1258.

BURKART, M., D. GROMB, AND F. PANUNZI (1997): "Large Shareholders, Monitoring, and the Value of the Firm," *The Quarterly Journal of Economics*, 112, 693–728.

BURKART, M., S. MIGLIETTA, AND C. OSTERGAARD (2023): "Why Do Boards Exist? Governance Design in the Absence of Corporate Law," *The Review of Financial Studies*, 36, 1788–1836.

CHAKRABORTY, A. AND B. YILMAZ (2017): "Authority, Consensus, and Governance," *The Review of Financial Studies*, 30, 4267–4316.

CHEMMANUR, T. J. AND V. FEDASEYEU (2018): "A Theory of Corporate Boards and Forced CEO Turnover," *Management Science*, 64, 4798–4817.

CHEN, J. J. (2026): "Optimal Managerial Authority," *Journal of Finance*, forthcoming.

CHEN, J. J. AND B. Y. HAN (2026): "When Corporate AI Adoption Backfires," Working Paper, Boston University and University of Maryland.

CHEN, Y., J. GONG, J. LI, AND Z. ZHAO (2025): "Better Technology, Worse Motivation: GenAI's Mediocrity Trap," Available at SSRN: `https://ssrn.com/abstract=5208163`.

CHENG, X., M. DOGAN, AND P. YILDIRIM (2025): "Artificial Intelligence in Team Dynamics: Who Gets Replaced and Why?" Working paper.

COLES, J. L., N. D. DANIEL, AND L. NAVEEN (2008): "Boards: Does One Size Fit All?" *Journal of Financial Economics*, 87, 329–356.

COWGILL, B., P. HERNANDEZ-LAGOS, AND N. L. WRIGHT (2024): "Does AI cheapen talk? Theory and evidence from global entrepreneurship and hiring," *Theory and Evidence From Global Entrepreneurship and Hiring (July 26, 2024). Columbia Business School Research Paper*.

CSASZAR, F. A., H. KETKAR, AND H. KIM (2024): "Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors," *Strategy Science*, 9, 322–345.

DE HAAS, R., D. FERREIRA, AND T. KIRCHMAIER (2021): "The Inner Workings of the Board: Evidence from Emerging Markets," *Emerging Markets Review*, 48, 100777.

DELL'ACQUA, F., E. MCFOWLAND, III, E. MOLLICK, H. LIFSHITZ, K. C. KELLOGG, S. RAJENDRAN, L. KRAYER, F. CANDELON, AND K. R. LAKHANI (2026): "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality," *Organization Science*, 37, 403–423.

DEMIRER, M., J. J. HORTON, N. IMMORLICA, B. LUCIER, AND P. SHAHIDI (2026): "Chaining Tasks, Redefining Work: A Theory of AI Automation," Tech. Rep. 34859, National Bureau of Economic Research.

DESSEIN, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69, 811–838.

DONALDSON, J. R., N. MALENKO, AND G. PIACENTINO (2020): "Deadlock on the Board," *The Review of Financial Studies*, 33, 4445–4488.

DRYMIOTES, G. AND K. SIVARAMAKRISHNAN (2021): "Strategic Director Appointments," *Journal of Accounting Research*, 59, 1303–1347.

EWENS, M. AND N. MALENKO (2024): "Board Dynamics over the Startup Life Cycle," *Journal of Finance*, forthcoming.

FALEYE, O., R. HOITASH, AND U. HOITASH (2011): "The Costs of Intense Board Monitoring," *Journal of Financial Economics*, 101, 160–181.

FAMA, E. F. AND M. C. JENSEN (1983): "Separation of Ownership and Control," *The Journal of Law and Economics*, 26, 301–325.

FERREIRA, D., M. A. FERREIRA, AND B. MARIANO (2018): "Creditor Control Rights and Board Independence," *Journal of Finance*, 73, 2385–2422.

FERREIRA, D., M. A. FERREIRA, AND C. RAPOSO (2011): "Board Structure and Price Informativeness," *Journal of Financial Economics*, 99, 523–545.

FERREIRA, D. AND R. NIKOLOWA (2023): "Talent Discovery and Poaching Under Asymmetric Information," *Economic Journal*, 133, 201–234.

FIELD, L., M. LOWRY, AND A. MKRTCHYAN (2013): "Are Busy Boards Detrimental?" *Journal of Financial Economics*, 109, 63–82.

FRIEBEL, G., Y. HUANG, J. LI, S. SHUKLA, AND A. ZHANG (2026): "Pyramids, Diamonds, and Oscillations: AI and the Structure of Internal Labor Markets," Working paper, available at SSRN 6570519.

FUCHS, W., L. GARICANO, AND L. RAYO (2015): "Optimal Contracting and the Organization of Knowledge," *The Review of Economic Studies*, 82, 632–658.

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.

GARICANO, L., J. LI, AND Y. WU (2026): *Messy Jobs: The Work That AI Cannot Reach*, Upriver Press.

GARICANO, L. AND L. RAYO (2025): "Training in the Age of AI: A Theory of Career Viability," CEPR Discussion Paper 20634, Centre for Economic Policy Research, Paris & London, revised March 2026.

GARICANO, L. AND E. ROSSI-HANSBERG (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121, 1383–1435.

GONZALEZ-URIBE, J. AND M. GROEN-XU (2017): "CEO Contract Horizon and Innovation," Available at SSRN: `https://ssrn.com/abstract=2633763`.

GREENWALD, B. C. (1986): "Adverse Selection in the Labour Market," *Review of Economic Studies*, 53, 325–347.

HARFORD, J. AND R. J. SCHONLAU (2013): "Does the Director Labor Market Offer Ex Post Settling-up for CEOs? The Case of Acquisitions," *Journal of Financial Economics*, 110, 18–36.

HARRIS, M. AND A. RAVIV (2008): "A Theory of Board Control and Size," *The Review of Financial Studies*, 21, 1797–1832.

HOLMSTRÖM, B. (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, ed. by M. Boyer and R. E. Kihlstrom, Amsterdam: North-Holland, 115–141.

HUANG, J. AND S. OUYANG (2025): "Soft Information, Hard Decisions: AI Advising," Working Paper.

IDE, E. (2026): "Automation, AI, and the Intergenerational Transmission of Knowledge," Working paper, IESE Business School.

IDE, E. AND E. TALAMÀS (2025): "Artificial Intelligence in the Knowledge Economy," *Journal of Political Economy*, 133, 3762–3800.

INDERST, R. AND H. M. MUELLER (2010): "CEO Replacement Under Private Information," *The Review of Financial Studies*, 23, 2935–2969.

JIANG, W. (2025): "Corporate Finance and Governance with Artificial Intelligence: Old and New," in *Artificial Intelligence in Finance*, ed. by T. Foucault, L. Gambacorta, W. Jiang, and X. Vives, CEPR Press, chap. 4.

JIANG, X. AND V. LAUX (2023): "What Role Do Boards Play in Companies with Visionary CEOs?" *Journal of Accounting Research*, 61, 1757–1796.

KENNY, G., M. KOWALKIEWICZ, AND K. OOSTHUIZEN (2025): "How CEOs Are Using Gen AI for Strategic Planning," *Harvard Business Review Digital Article*, online.

LAMBA, R. (2026): "The Verification Hill," Working paper, Cornell University.

LARCKER, D. F., A. SERU, B. TAYAN, AND L. YOLER (2025): "The Artificially Intelligent Boardroom," Tech. Rep. 1, Stanford Graduate School of Business.

LAUX, V. (2008): "Board Independence and CEO Turnover," *Journal of Accounting Research*, 46, 137–171.

LEVIT, D. (2012): "Expertise, Structure, and Reputation of Corporate Boards," SSRN Working Paper.

——— (2020): "Words Speak Louder Without Actions," *Journal of Finance*, 75, 91–131.

LEVIT, D. AND N. MALENKO (2016): "The Labor Market for Directors and Externalities in Corporate Governance," *Journal of Finance*, 71, 775–808.

LI, J. (2013): "Job Mobility, Wage Dispersion, and Technological Change: An Asymmetric Information Perspective," *European Economic Review*, 60, 105–126.

LINCK, J. S., J. M. NETTER, AND T. YANG (2008): "The Determinants of Board Structure," *Journal of Financial Economics*, 87, 308–328.

MALENKO, N. (2014): "Communication and Decision-Making in Corporate Boards," *The Review of Financial Studies*, 27, 1486–1532.

——— (2024): "Information Flows, Organizational Structure, and Corporate Governance," in *Handbook of Corporate Finance*, ed. by D. Denis, Edward Elgar Publishing, chap. 14, 511–547.

MARCOCCIA, C., W. QUATTROCIOCCHI, AND V. CAPRARO (2026): "AI advice suppresses people's willingness to say "I don't know", even when the advice is wrong and accuracy is incentivized," .

MELUMAD, N. D. AND T. SHIBANO (1991): "Communication in Settings with No Transfers," *RAND Journal of Economics*, 22, 173–198.

NIKOLOWA, R. (2015): "Career Dynamics and Span of Control," *Economics Letters*, 128, 6–8.

RAHEJA, C. G. (2005): "Determinants of Board Size and Composition: A Theory of Corporate Boards," *Journal of Financial and Quantitative Analysis*, 40, 283–306.

RANSBOTHAM, S., D. KIRON, S. KHODABANDEH, S. IYER, AND A. DAS (2025): "The Emerging Agentic Enterprise: How Leaders Must Navigate a New Age of AI," *MIT Sloan Management Review and Boston Consulting Group*.

SHEKSHNIA, S. AND V. YAKUBOVICH (2025): "How Pioneering Boards Are Using AI," *Harvard Business Review*, 103, 56–64.

SONG, F. AND A. V. THAKOR (2006): "Information Control, Career Concerns, and Corporate Governance," *Journal of Finance*, 61, 1845–1896.

WALDMAN, M. (1984): "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics*, 15, 255–267.

WARTHER, V. A. (1998): "Board Effectiveness and Board Dissent: A Model of the Board's Relationship to Management and Shareholders," *Journal of Corporate Finance*, 4, 53–70.

YOTZOV, I., J. M. BARRERO, N. BLOOM, P. BUNN, S. J. DAVIS, K. M. FOSTER, A. JALCA, B. H. MEYER, P. MIZEN, M. A. NAVARRETE, P. SMIETANKA, G. THWAITES, AND B. Z. WANG (2026): "Firm Data on AI," Working Paper 34836, National Bureau of Economic Research, Cambridge, MA.

ZHONG, H. (2025): "Optimal Integration: Human, Machine, and Generative AI," *Management Science*, forthcoming.

Internet Appendix to "Artificial Intelligence in the Boardroom"

Daniel Ferreira Jin Li

This draft: August 2026

This Internet Appendix contains supplementary material for "Artificial Intelligence in the Boardroom." Section IA.1 presents a more detailed discussion of contracting: it shows that the one-period contracts assumed in the main text are without loss of generality, that in the absence of contracting frictions optimal one-period contracts implement the first-best allocation with $\pi = 1$, and that negative severance is never optimal. Section IA.2 provides a microfoundation for the severance bonus cap (Assumption 6 in the main text). Section IA.3 considers the case in which $1 - t_s \geq 2t_v$, so that the CEO has enough time to verify a second assessment, and uses it to analyze the governance effects of AI-driven reductions in verification time. Section IA.4 characterizes the pre-AI benchmark when Assumption 4 fails, showing that the model collapses to Case 1 of Proposition 1. Section IA.5 characterizes the model when Assumption 1 fails: the CEO then devotes her single verification slot to the AI agent's assessment rather than her own, board communication loses its informational content, and governance becomes irrelevant.

Unless otherwise noted, notation and assumptions are as in the main text. Sections IA.1 and IA.2 consider the benchmark (pre-AI) economy with $\alpha = 1$ and $\kappa = 0$ (Remark IA.3 discusses the general case); Section IA.3 sets $\alpha = 1$ and $\kappa > 0$; Section IA.5 considers the general case $\alpha \geq 1$, $\kappa \in [0, 1]$. Throughout, $M := M(1, \eta) = \frac{\eta}{6}$ denotes the marginal benefit of monitoring.

IA.1 Contracting

The main text restricts attention to one-period contracts: a first-period bonus w_1 conditional on $y_1 = 1$, a severance bonus s conditional on $\theta = f$, and a second-period bonus w_2 conditional on $y_2 = 1$. This section establishes the two claims made about this restriction in Section 3 of the main text: no other contract format can deliver higher shareholder value (Proposition IA.1), and, in the absence of contracting frictions, optimal one-period contracts implement the first-best allocation with $\pi = 1$ (Proposition IA.2). Along the way we also verify that negative severance is never optimal, so that under Assumption 6 (with $\bar{s} = 0$) any optimal contract sets $s = 0$.

IA.1.1 Verifiable Events and General Contracts

The verifiable variables are y_1, y_2, and the period-1 CEO's job status $\theta \in \{r, f\}$. The CEO's message $\hat{m}$ is private information to the board and the CEO and hence not contractible; the state ω, while observable, is not verifiable, which is immaterial on the equilibrium path because dismissals occur only in state ω^{tough}, so conditioning on θ subsumes conditioning on ω. By limited liability, all payments to the CEO are nonnegative, and because the CEO consumes all compensation at the end of each period, payments cannot be recovered ex post: there is no bonding, no deferral beyond the period, and no clawback.

A general (possibly long-term, possibly renegotiated) arrangement for the period-1 CEO is then a collection of nonnegative payments contingent on the realized verifiable history (y_1, θ, y_2). Note that y_2 is verifiable regardless of the period-1 CEO's job status: if $\theta = f$, it is the output produced by the replacement. Thus, a fired CEO's severance may, in principle, be indexed to her successor's performance; Step 1 of the proof of Proposition IA.1 shows that such contingencies are permitted but useless. Because all players are risk neutral and do not discount, only *expected* payments conditional on each verifiable event matter. Accordingly, define

$$\begin{aligned} w_{1r} &:= E[\text{pay} \mid y_1 = 1, \theta = r], & w_{1s} &:= E[\text{pay} \mid y_1 = 1, \theta = f], \\ s_1 &:= E[\text{pay} \mid y_1 = 0, \theta = r], & s_0 &:= E[\text{pay} \mid y_1 = 0, \theta = f], \end{aligned} \tag{IA.1.1}$$

where the expectations are taken over y_2 and, for retained-CEO events, include all period-2 payments in excess of the period-2 contract offered at Date (i) of $t = 2$.

Proposition IA.1 (One-period contracts are without loss of generality). *Every renegotiation-proof contract is weakly dominated, in terms of shareholder value, by the combination of (i) a period-1 contract paying (w_{1r}, w_{1s}, s_0) as in (IA.1.1) with $s_1 = 0$, and (ii) the incentive-compatible one-period bonus contract offered at Date (i) of $t = 2$.*

Proof. Step 1: payments to a fired CEO need not depend on y_2. If $\theta = f$, second-period output is generated by a replacement whose knowledge and problem are independent draws; conditional on $(y_1, \theta = f)$, y_2 is therefore independent of the incumbent's type, actions, and information. Indexing severance to y_2 thus attaches a pure lottery to the payment: replacing any y_2-contingent payment with its conditional expectation leaves the CEO's expected payoff at every information set, and the firm's expected cost, unchanged. Hence fired-state pay is summarized by (w_{1s}, s_0).

Step 2: pay of a retained CEO beyond the fresh period-2 contract is a retention rent. At Date (i) of $t = 2$ the board makes a take-it-or-leave-it offer, taking as given any claims inherited from the period-1 contract. Renegotiation-proofness implies that total period-2 y_2-contingent pay is just sufficient for the second-period effort IC constraint of the CEO type the board wishes to employ; any inherited claims in excess of this level constitute a non-negative rent that is paid *because* the CEO is retained and can depend, through the inherited claims, at most on y_1. By risk neutrality, such rents are equivalent to increases in the event payments w_{1r} (if $y_1 = 1$) or s_1 (if $y_1 = 0$). Hence, without loss of generality, the period-2 contract is the IC-binding one-period bonus contract, and all remaining payments are summarized by $(w_{1r}, w_{1s}, s_0, s_1)$.

Step 3: $s_1 = 0$. The event $(y_1 = 0, \theta = r)$ is reached on the equilibrium path only when a type-b CEO asks for help, the board fails, and the board is weak—probability $\eta(1 - \pi)$ conditional on type b. Off the path, it is the *certain* outcome of shirking (no effort implies $y_1 = 0$ and, absent a request for help, retention), and it is also reached with probability one by a type-b CEO who deviates by not asking for help. Raising s_1 by $\varepsilon > 0$ therefore raises the right-hand sides of the effort IC constraint and of the help IC constraint by ε, while

raising the corresponding equilibrium payoffs by at most $\eta(1-\pi)\varepsilon < \varepsilon$: both constraints tighten, and expected pay rises. Setting $s_1 = 0$ is thus optimal.

Step 4: nothing else is feasible. Limited liability rules out negative payments (dismissal penalties or bonds), and consumption at the end of each period rules out recovering period-1 pay at $t = 2$. The feasible set therefore collapses to (w_{1r}, w_{1s}, s_0) plus the fresh period-2 contract, as claimed. □

Proposition IA.1 reduces the space of contracts to three period-1 instruments, but it leaves open whether the *two-instrument* main-text format—a bonus w_1 paid if $y_1 = 1$ and a single severance payment s paid if $\theta = f$, so that $w_{1r} = w_1$, $w_{1s} = w_1 + s$, and $s_0 = s$—can span this space. The answer turns on whether the two severance instruments, the success-state premium $w_{1s} - w_{1r}$ and the failure-state payment s_0, are payoff-equivalent. Define the *expected severance premium*

$$\delta := (1-\eta)(w_{1s} - w_{1r}) + \eta s_0, \tag{IA.1.2}$$

the additional expected pay of a dismissed CEO relative to a retained one, and note that the main-text contract delivers $\delta = (1-\eta)s + \eta s = s$.

Corollary IA.1 (The main-text format is without loss of generality). *Restrict attention to pure-strategy equilibria with communication.*[1] *Every period-1 contract (w_{1r}, w_{1s}, s_0) is payoff-equivalent to the main-text contract $(w_1, s) = (w_{1r}, \delta)$: each player's expected payoff, on the equilibrium path and after every deviation, and the firm's expected cost, are unchanged. Restricting attention to contracts (w_1, s) is therefore without loss of generality, and Assumption 6's cap on s is equivalent to a cap on δ.*

Proof. Dismissal occurs only after m_b ($\rho(m_b) = 1$ and $\rho(m_c) = 0$ in the equilibria under consideration), and after m_b the assessment is produced by the board, which succeeds with probability $1 - \eta$ regardless of the CEO's type or prior play. Hence *every* type and strategy that reaches the event $\theta = f$ faces the same conditional distribution of y_1 (a CEO who

[1] In mixed-strategy equilibria—which the proof of Proposition 1 shows are never optimal—the CEO may be dismissed after m_c, in which case the conditional distribution of y_1 given dismissal is type-dependent and the equivalence need not hold. This does not affect the characterization of optimal governance.

does not exert effort identifies no deal and thus cannot ask for help), so that expected pay conditional on dismissal equals

$$(1-\eta)w_{1s} + \eta s_0 = (1-\eta)w_{1r} + \delta$$

in every case; the firm's expected outlay per dismissal is the same expression. Replacing (w_{1s}, s_0) with $(w_{1r}+\delta,\, \delta)$ preserves δ, and hence preserves every one of these expected payments as well as all payments in retained states. □

The equivalence rests on a single property of the model: every dismissed CEO faces the same output distribution, because dismissal follows a request for help and it is then the board that produces the assessment. Section IA.2 introduces a class of applicants for whom this property fails; there, the distinction between the two severance instruments becomes payoff-relevant.

IA.1.2 The Frictionless Benchmark

We now drop Assumption 6 and characterize optimal governance when severance pay is unrestricted. By Corollary IA.1, we may work directly with the main-text contract (w_1, s), identifying the severance payment with the expected premium of (IA.1.2), $s = \delta$; for continuity with Section IA.2, we write w_{1r} for w_1. Severance enters the main text's pre-AI benchmark in three places. A type-b CEO who asks for help now collects the expected premium δ when dismissed, so her help IC constraint becomes

$$(1-\eta)w_{1r} + \pi\delta + (1-\pi)B \;\geq\; B \quad\Longleftrightarrow\quad \pi(B-\delta) \;\leq\; (1-\eta)w_{1r}. \tag{IA.1.3}$$

Substituting the same continuation payoff into the first-period effort IC constraint gives

$$B - c + \frac{1}{2}(w_{1r}+B) + \frac{1}{2}\big[(1-\eta)w_{1r} + \pi\delta + (1-\pi)B\big] \;\geq\; 2B, \tag{IA.1.4}$$

which simplifies to $(2-\eta)w_{1r} - \pi(B-\delta) \geq 2c$. Finally, because a dismissal occurs with probability $\frac{1}{2}\pi$ and then costs the firm the expected premium δ, expected profit is the main-

text expression net of expected severance:

$$V \;=\; 2-\eta-c-\Big(1-\frac{\eta}{2}\Big)w_{1r}-\frac{\pi}{2}\delta+\frac{\pi}{2}M. \tag{IA.1.5}$$

Proposition IA.2 (The frictionless benchmark). *Absent a severance cap, the minimal expected compensation cost of implementing effort, communication, and monitoring intensity π is $c+\frac{\pi B}{2}$, for every $\pi\in[0,1]$. Hence shareholder value is linear in π,*

$$V(\pi)=2-\eta-2c+\frac{\pi}{2}\big(M-B\big), \tag{IA.1.6}$$

and the optimal monitoring intensity is $\pi^*=1$ if $M>B$ and $\pi^*=0$ if $M<B$. If $M>B$, optimal one-period contracts implement the first-best allocation—effort in both periods, full communication, and certain replacement of a type-b CEO—and form a continuum

$$\Big\{\, w_{1r}=\frac{2c+u}{2-\eta},\;\; \delta=B-u \;:\; u\in[0,\,2c(1-\eta)]\,\Big\}, \tag{IA.1.7}$$

ranging from full insurance $(w_{1r},\delta)=(\frac{2c}{2-\eta},B)$ to $(w_{1r},\delta)=(2c,\,B-2c(1-\eta))$. A negative expected premium ($\delta<0$) is weakly dominated for every π and strictly suboptimal when $M>B$.

Proof. Let $u:=\pi(B-\delta)$ denote the CEO's expected net dismissal exposure. The effort IC constraint (IA.1.4) reads $(2-\eta)w_{1r}\geq 2c+u$ and binds at a cost minimum, so expected compensation is

$$\Big(1-\frac{\eta}{2}\Big)w_{1r}+\frac{\pi}{2}\delta=\frac{2c+u}{2}+\frac{\pi B-u}{2}=c+\frac{\pi B}{2},$$

independent of u. Feasibility requires the help IC constraint (IA.1.3), $u\leq(1-\eta)w_{1r}$, which at $w_{1r}=\frac{2c+u}{2-\eta}$ is equivalent to $u\leq 2c(1-\eta)$; since severance is uncapped, any $u\in[0,\min\{\pi B,2c(1-\eta)\}]$ is attainable with $\delta=B-\frac{u}{\pi}\in[0,B]$, so the feasible set is nonempty for every $\pi\in[0,1]$. Substituting the cost into (IA.1.5) yields (IA.1.6); linearity delivers the corner solutions, and at $\pi^*=1$ the set of cost-minimizing contracts is (IA.1.7). The allocation at $\pi^*=1$ is first best: both IC constraints hold, the board is tough with probability one, and a type-b CEO is always replaced. Finally, $\delta<0$ leaves the cost $c+\frac{\pi B}{2}$

unchanged but requires $u > \pi B$, shrinking the set of implementable π (in particular, $\pi = 1$ requires $u = B - \delta > B > 2c(1-\eta)$ by Assumption 5, violating feasibility); it is therefore weakly dominated, and strictly so when $M > B$, since it precludes the optimum $\pi^* = 1$. With the cap $\bar{s} = 0$, feasibility requires $\delta \leq 0$, and the dominance of $\delta < 0$ implies that any optimal contract sets $s = 0$, as claimed in the main text. □

Proposition IA.2 verifies the claim made in Section 3 of the main text: whenever monitoring has positive net private value ($M > B$, which holds under the interior condition of Proposition 1), the absence of contracting frictions makes board friendliness unnecessary—severance pay is a frictionless instrument that compensates the CEO for dismissal at a flat cost of $\frac{B}{2}$ per unit of monitoring intensity. In main-text notation, the continuum (IA.1.7) is the set of contracts (w_1, s) with $s = B + 2c - (2-\eta)w_1$: for example, $(w_1, s) = \big(2c,\, B - 2c(1-\eta)\big)$ or, with full insurance against dismissal, $(w_1, s) = \big(\frac{2c}{2-\eta},\, B\big)$. The geometry of Proposition 1 follows directly: once the cap forces $\delta = 0$, the flat-cost implementation is feasible only up to $\pi = \frac{(1-\eta)2c}{B}$ (the point at which the help IC constraint binds at the effort-IC-binding bonus), and any further monitoring requires raising w_{1r} above $2c$, which is strictly more expensive. This is precisely the interior solution (15) of the main text and the corner logic of its Case 3.

Two remarks are in order. First, the private and social values of turnover differ. Total surplus is $S(\pi) = V(\pi) + 2B + \frac{\pi}{2}B = 2 - \eta - 2c + 2B + \frac{\pi}{2}M$, which is strictly increasing in π for *all* parameter values, because dismissal merely transfers the private benefit B from the incumbent to the replacement. Shareholders, however, must compensate the dismissed incumbent for her lost B (through severance) while being unable to extract the replacement's B (by limited liability); the wedge is exactly B. Thus, when $M < B$, even the frictionless private optimum $\pi^* = 0$ features socially insufficient CEO turnover. The wedge rests on two deeper frictions that removing the severance cap does not touch: an *information* friction—the firm must buy the bad CEO's self-revelation and hence pay her the continuation rent B she loses upon dismissal, which is the source of the $\frac{B}{2}$ marginal cost of monitoring—and a *pledgeability* friction—limited liability and the absence of CEO wealth prevent the firm from charging the incoming CEO for the identical rent B she gains (the

firm cannot "sell the job"). Eliminating either friction would align private and social incentives: with observable types, no dismissal compensation is owed; with pledgeable private benefits, the replacement's B would be recouped at hiring; in both cases the private value of monitoring rises to $M > 0$, and $\pi = 1$ becomes privately optimal for all parameter values. Accordingly, the statement that frictionless one-period contracts implement the first best is conditional on monitoring being privately valuable, $M \geq B$—which holds in the region of interest of the main text, where the interior condition of Proposition 1 is satisfied. Second, the frictionless benchmark assumes away entry by unqualified applicants; Section IA.2 shows that the very generosity of severance in (IA.1.7) is what attracts such applicants, and that screening them out endogenously restores a binding cap.

IA.2 A Microfoundation for the Severance Bonus Cap

In the main text, we impose an exogenous cap on severance pay (Assumption 6), which we normalize to zero. The cap matters because severance pay insures the CEO against dismissal: absent any cap, the firm could compensate a dismissed CEO for her lost private benefit, thereby making truthful communication with the board costless to the CEO. In this section, we show that a cap arises endogenously when the firm must design contracts that do not attract unqualified applicants—which we call *charlatans*—to the CEO position. Our modeling of charlatans follows Berk and van Binsbergen (2022), who study agents who know they lack skill but can pose as skilled professionals. In their setting, as in ours, entry by charlatans is disciplined by the structure of compensation rather than by direct screening. The exogenous cap $\bar{s}$ of the main text is then interpreted as the reduced form of this screening problem, and the main-text solution corresponds to the limiting case in which the cap binds at zero.

IA.2.1 Setup: Charlatans and Contracts

Charlatans. We extend the model by introducing a new type of agent. In addition to the unit measure of regular CEO candidates with unknown knowledge $z \sim U[0,1]$, there is an

unbounded pool of *charlatans*. A charlatan differs from a regular candidate in two ways. First, he knows that he has no knowledge: $z = 0$. Second, he is not a natural candidate for the job: masquerading as one requires effort at a personal cost $K > 0$. Applying is costless—a plausible-looking application is cheap to produce—but a charlatan who is selected must incur K to see the masquerade through: fabricating a record that survives the board's due diligence, rehearsing the part, and sustaining the pretense in office. A masquerading charlatan is observationally identical to a regular candidate, so the board cannot screen applicants directly and hires at random from the pool. Like all agents, charlatans are risk neutral and protected by limited liability.

A masquerade, however, does not survive a full period of scrutiny:

Assumption IA.1 (Detection). *By the end of Date (iv) of $t = 1$, the board learns whether its CEO is a charlatan. This information is not verifiable, so contracts cannot be made contingent on it. The board's private cost of firing applies only to regular CEOs: in state ω^{weak}, firing a detected charlatan is costless.*

Assumption IA.1 embodies three natural ideas. First, a fabricated candidacy does not withstand a period of close interaction: the board learns whether it is dealing with a genuine draw from the candidate pool, even though it never observes a regular CEO's knowledge z—detection is *category* learning, not ability learning. It is the close interaction of the first period—including acting on the CEO's request for help—that unmasks a charlatan, so detection arrives too late to affect the board's Date (iv) actions. Second, because the board's information is unverifiable, the separation is contractually a without-cause termination and severance remains owed. Third, the board's private cost of firing reflects its alignment with, and attachment to, a legitimate executive. Taken together, detection makes retention worthless to a charlatan: he is dismissed at Date (i) of $t = 2$ with probability one, by either board type.

Free entry of charlatans disciplines contract design. Let U^{ch} denote a charlatan's expected payoff conditional on being hired, net of the masquerading cost K. If the contract offered by the firm implies $U^{ch} > 0$, taking the job is profitable for a charlatan; because applications are costless, an unbounded mass of charlatans then applies, the probability of

hiring a regular candidate is zero, and the firm is inundated with charlatans.[2] We therefore require the governance structure to be *charlatan-proof*:

$$U^{ch} \leq 0. \tag{IA.2.1}$$

Restricting attention to charlatan-proof contracts is therefore without loss of optimality: the only alternative is to hire a charlatan with probability one, which earns less than the charlatan-proof optimum under our maintained assumptions.[3]

Contracts. By Proposition IA.1, we may describe the period-1 contract by the event payments (w_{1r}, w_{1s}, s_0) of (IA.1.1), with $s_1 = 0$.[4] Unlike in Section IA.1, however, we cannot invoke Corollary IA.1 to collapse the two severance instruments into a single payment: charlatan entry breaks the payoff equivalence on which that result rests, as this section makes precise. We refer to w_{1s} as the *severance bonus* and to

$$\Delta := w_{1s} - w_{1r} \tag{IA.2.2}$$

as the *severance premium*. Period-2 contracts are as in the main text; because all CEOs step down at the end of $t = 2$, severance is a period-1 instrument only.

Charlatan strategies and the no-entry condition. Consider a charlatan who is hired at $t = 1$. He collects the private benefit B at Date (ii). At Date (iii) he can exert effort at cost c, identify a deal, and—knowing that he cannot solve it ($z = 0$)—ask the board for help, which succeeds with probability $1 - \eta$. By Lemma 3 in the main text, on the equilibrium

[2] Costless applications make inundation an all-or-nothing affair. If instead K had to be sunk at the application stage, entry would be self-limiting—charlatans would apply only until the chance of being selected justifies K—and the firm would face a diluted, but not degenerate, applicant pool. Charlatan-proofness would then be a choice rather than a necessity, and the cap derived below would be soft rather than hard.

[3] If the firm hires a charlatan, its expected profit is at most $(1 - \eta) + v_{2n}$ minus the charlatan's expected pay, where $v_{2n} = 1 - \frac{\eta}{2} - c$. A sufficient condition for this to fall short of the charlatan-proof optimum characterized in Proposition IA.3 is $c < \frac{1}{2}$, which is implied by Assumption 4 whenever $\eta > \frac{3}{4}$ and holds for all our numerical illustrations.

[4] Proposition IA.1 applies unchanged: its proof concerns regular CEOs only, and no charlatan ever reaches the event $(y_1 = 0, \theta = r)$, since an idle charlatan is detected and dismissed (Assumption IA.1).

path asking for help identifies the CEO as type b, so the request itself raises no suspicion. His dismissal at Date (i) of $t = 2$ is nonetheless certain (Assumption IA.1): a tough board replaces any CEO who asks for help, and a weak board, having detected him, fires him at no cost.[5] This "working" strategy yields

$$U_w^{ch} = B - c + (1-\eta)w_{1s} + \eta s_0 - K. \tag{IA.2.3}$$

Alternatively, the charlatan can stay idle: he exerts no effort, output is $y_1 = 0$, and, once detected and dismissed, he collects s_0. This "lazy" strategy yields

$$U_\ell^{ch} = B + s_0 - K. \tag{IA.2.4}$$

The no-entry condition (IA.2.1) is thus $\max\{U_w^{ch}, U_\ell^{ch}\} \leq 0$, that is,

$$K \geq B + \max\left\{ s_0,\ (1-\eta)w_{1s} + \eta s_0 - c \right\}. \tag{IA.2.5}$$

Our first result shows that severance pay optimally takes the form of a *success-contingent* bonus: the firm pays fired CEOs only in the state $y_1 = 1$.

Lemma IA.1 (Severance is paid as a success-contingent bonus). *Fix any charlatan-proof governance structure in which a type-b CEO asks the board for help. Replacing (w_{1s}, s_0) with $(\hat{w}_{1s}, 0)$, where $(1-\eta)\hat{w}_{1s} = (1-\eta)w_{1s} + \eta s_0$, leaves the firm's profit, all of the regular CEO's incentive constraints, and the working charlatan's payoff (IA.2.3) unchanged, and weakly relaxes the lazy charlatan's no-entry constraint, strictly so if $s_0 > 0$. Hence it is without loss of generality to set $s_0 = 0$.*

Proof. On the equilibrium path, the only CEO who is fired is a type-b CEO who asked for help; her output succeeds with probability $1-\eta$. Thus both her expected severance receipts and the firm's expected severance payments depend on (w_{1s}, s_0) only through the combination $(1-\eta)w_{1s} + \eta s_0$, and the same is true of the working charlatan's payoff (IA.2.3),

[5] Firing is also strictly optimal for the board: a retained charlatan delivers at most $1 - \eta - c$ in period 2 (he can identify a deal but never solve one himself), while a fresh draw from the candidate pool delivers $v_{2n} = 1 - \frac{\eta}{2} - c > 1 - \eta - c$.

because a working charlatan succeeds with the same probability $1-\eta$. The substitution therefore changes no payoff on the equilibrium path or after any deviation that involves exerting effort and asking for help. The only object it affects is U_ℓ^{ch} in (IA.2.4), which falls by s_0 (a regular CEO cannot exploit s_0: she is dismissed only after asking for help, in which case output succeeds with probability $1-\eta$). □

The intuition is that a working charlatan mimics a genuine type-b CEO perfectly, so conditioning severance on success cannot separate the two, but paying severance only upon success at least forces a charlatan to work for it and starves the lazy entry strategy. The result echoes the central insight of Berk and van Binsbergen (2022) that charlatans are deterred by compensation that is sensitive to performance: here, it is precisely the performance-*insensitive* component of severance that must be eliminated. Lemma IA.1 is thus the converse of Corollary IA.1: a lazy charlatan reaches dismissal with $y_1 = 0$ with probability one, so—unlike every agent in the charlatan-free model—his payoff depends on s_0 separately from the expected premium δ of (IA.1.2), and the two severance instruments are no longer payoff-equivalent. In particular, the main-text two-instrument contract (w_1, s), which sets $s_0 = s$, is weakly dominated in this section whenever severance is used. In light of Lemma IA.1, we henceforth set $s_0 = 0$, so that the expected severance premium (IA.1.2) reduces to $\delta = (1-\eta)\Delta$. We also assume from now on that $K \geq B$, so that charlatans do not enter merely to collect the private benefit; Proposition IA.3 imposes a parameter range under which this holds. The no-entry condition then reduces to a cap on the severance bonus:

$$(1-\eta)\, w_{1s} \;\leq\; K - (B-c). \tag{IA.2.6}$$

Condition (IA.2.6) is the microfounded counterpart of Assumption 6 in the main text: the maximum payment a fired CEO can expect is pinned down by the cost of masquerading, net of the rents $(B-c)$ that any hired charlatan pockets. Note that the instrument being capped is not a separation payment in the everyday sense: by Lemma IA.1, severance is optimally delivered as a bonus on verifiable first-period output that happens to accrue to a departing executive. A regulatory or contractual ceiling on explicit separation pay (s_0) would therefore not bind at the optimum; the relevant friction must cap the success-

contingent bonus w_{1s} itself, and charlatan-proofness does exactly that.

IA.2.2 The Contracting Problem

We now solve for the optimal governance structure, which here consists of (w_{1r}, w_{1s}, π), taking the optimal period-2 bonuses $(w^*_{2b}, w^*_{2g}, w^*_{2n})$ from the main text as given (they are unaffected by period-1 severance). As in the main text, we look for an equilibrium in which the CEO exerts effort and a type-b CEO asks the board for help, and we then verify all remaining incentive constraints.

With severance, a type-b CEO who asks for help receives w_{1s} with probability $1-\eta$ if the board is tough (probability π), and receives w_{1r} with probability $1-\eta$ plus the continuation benefit B if the board is weak. The help IC constraint becomes

$$(1-\eta)\big[\pi w_{1s} + (1-\pi)w_{1r}\big] + (1-\pi)B \;\geq\; B \quad\Longleftrightarrow\quad \pi\big(B-\delta\big) \;\leq\; (1-\eta)\, w_{1r}. \tag{IA.2.7}$$

The first-period effort IC constraint becomes

$$B - c + \frac{1}{2}\big(w_{1r} + B\big) + \frac{1}{2}\Big[(1-\eta)\big[\pi w_{1s} + (1-\pi)w_{1r}\big] + (1-\pi)B\Big] \;\geq\; 2B, \tag{IA.2.8}$$

which simplifies to $(2-\eta)w_{1r} - \pi(B-\delta) \geq 2c$. The good CEO's truth-telling constraint becomes

$$w_{1r} + B \;\geq\; (1-\eta)\big[\pi w_{1s} + (1-\pi)w_{1r}\big] + (1-\pi)\Big[\Big(1-\frac{\eta}{3}\Big)w^*_{2b} - c + B\Big]. \tag{IA.2.9}$$

The firm's expected profit is as in the main text, minus expected severance payments: the premium Δ is paid with probability $\frac{1}{2}\pi(1-\eta)$, an expected cost of $\frac{\pi}{2}\delta$:

$$V(w_{1r}, \Delta, \pi) \;=\; 2 - \eta - c - \Big(1-\frac{\eta}{2}\Big)w_{1r} - \frac{\pi}{2}\,\delta + \frac{\pi}{2}\,M. \tag{IA.2.10}$$

An isomorphism. Note first that, written in terms of (w_{1r}, δ), constraints (IA.2.7) and (IA.2.8) and the profit function (IA.2.10) coincide with their frictionless counterparts (IA.1.3),

(IA.1.4), and (IA.1.5): severance affects regular CEOs only through the expected premium; only the charlatans' payoffs depend on the finer split. Inspection of these expressions also reveals that, for a fixed premium δ, the model is isomorphic to the benchmark model of the main text with

$$\widetilde{B}(\delta) := B - \delta \qquad \text{and} \qquad \widetilde{M}(\delta) := M - \delta. \tag{IA.2.11}$$

The severance premium is a transfer to the dismissed type-b CEO: it lowers her cost of revealing herself to the board and lowers the firm's net benefit of replacing her by exactly the same amount, so that $\widetilde{M} - \widetilde{B} = M - B$ for every δ. In particular, when both (IA.2.7) and (IA.2.8) bind, subtracting one from the other yields $w_{1r} = 2c$ *regardless of* δ—the same invariance as in Proposition 1—and

$$\pi = \frac{(1-\eta)2c}{B-\delta}, \tag{IA.2.12}$$

which is increasing in δ: severance insurance buys monitoring intensity.

Before characterizing the optimum, we record a model-free bound implied by charlatan-proofness alone.

Lemma IA.2 (**A ceiling on monitoring intensity**). *Any charlatan-proof governance structure with $\Delta \geq 0$ that induces a type-b CEO to ask the board for help satisfies*

$$\pi \;\leq\; \frac{K-(B-c)}{B}. \tag{IA.2.13}$$

In particular, if $K < 2B - c$, full monitoring ($\pi = 1$) is infeasible.

Proof. Using (IA.2.7), $\pi \leq 1$, $\delta \geq 0$, and (IA.2.6),

$$\pi B = \pi(B-\delta) + \pi\delta \leq (1-\eta)w_{1r} + \delta = (1-\eta)w_{1s} \leq K - (B-c).$$

For $\pi = 1$, (IA.2.7) directly requires $(1-\eta)w_{1r} + \delta \geq B$, which contradicts (IA.2.6) whenever $K < 2B - c$, for any sign of δ. □

Lemma IA.2 isolates the economics of the friction. A working charlatan is a perfect

mimic of a type-b CEO—he asks for help, succeeds with the same probability, and is fired—*except* that he places no value on retention. The only instrument that screens charlatans out is therefore the retention margin: the part of the bad CEO's compensation for truth-telling that is delivered as continuation value B rather than as cash. But high monitoring intensity is precisely a promise to withhold retention. As $\pi \to 1$, the bad CEO must be compensated for dismissal entirely in cash, her package converges to the charlatan's, and screening becomes impossible unless masquerading is expensive enough ($K \geq 2B - c$).

IA.2.3 Optimal Governance

Define

$$\overline{\delta} := K - B - (1 - 2\eta)c, \tag{IA.2.14}$$

which, by (IA.2.6), is the largest expected severance premium consistent with charlatan-proofness when $w_{1r} = 2c$. We characterize the optimum under the following parameter restrictions:

$$\max\{B,\, B + (1 - 2\eta)c\} \;\leq\; K \;<\; 2B - c, \tag{IA.2.15}$$

Condition (IA.2.15) ensures that charlatan-proofness permits a nonnegative severance premium ($\overline{\delta} \geq 0$), that lazy entry is deterred ($K \geq B$), and that full monitoring is infeasible ($K < 2B - c$); Assumption 5 ($B \geq 2c$) guarantees that the admissible range is nonempty. Define

$$D \;:=\; B - \overline{\delta} - 2c(1 - \eta) \;=\; 2B - K - c, \tag{IA.2.16}$$

which is strictly positive on the range (IA.2.15), and

$$\overline{M} \;:=\; B + \frac{\left(B - \overline{\delta}\right)^2}{(1 - \eta)\, D}. \tag{IA.2.17}$$

The threshold $\overline{M}$ separates two regimes of valuable monitoring: below it, the optimum is the vertex at which the help IC constraint, the effort IC constraint, and the no-entry condition all bind—three constraints pinning down three instruments; above it, the firm

pushes monitoring beyond the vertex, and the marginal instrument changes.[6]

Proposition IA.3 (Optimal governance with charlatans). *Suppose (IA.2.15) holds. The optimal governance structure is unique (up to instruments that are payoff-irrelevant in the first case) and has $s_0^* = 0$ and:*

(1) If $M < B$: $\pi^* = 0$ and $w_{1r}^* = \frac{2c}{2-\eta}$; the severance instruments are payoff-irrelevant (any $w_{1s}^* \in \left[0, \frac{K-(B-c)}{1-\eta}\right]$ is optimal), since no CEO is ever dismissed.

(2) If $B < M \leq \overline{M}$:

$$w_{1r}^* = 2c, \qquad w_{1s}^* = \frac{K-(B-c)}{1-\eta}, \qquad \pi^* = \frac{(1-\eta)2c}{2B-K+(1-2\eta)c}, \tag{IA.2.18}$$

with the help IC constraint, the effort IC constraint, and the no-entry condition all binding ($U^{ch} = 0$). Moreover: (i) π^*, the firm's profit, and the CEO's expected total compensation are strictly increasing in K; (ii) the CEO's expected lifetime utility equals $2B$, independent of K.

(3) If $M > \overline{M}$:

$$w_{1r}^* = \frac{\pi^* D}{(1-\eta)(1-\pi^*)} > 2c, \qquad w_{1s}^* = \frac{K-(B-c)}{1-\eta}, \qquad \pi^* = 1 - \sqrt{\frac{D}{(1-\eta)(M-B)}}, \tag{IA.2.19}$$

with the help IC constraint and the no-entry condition binding and the effort IC constraint slack.

In every case $\pi^* < 1$: full monitoring is infeasible.

Proof. Fix a governance structure (w_{1r}, Δ, π) inducing effort and communication, and write $w \equiv w_{1r}$; the sign of Δ is unrestricted.

Step 0: the case $M < B$. By Proposition IA.2, the minimal expected compensation cost of implementing effort, communication, and monitoring intensity π is $c + \frac{\pi B}{2}$ even without

[6] An earlier draft characterized the optimum only under the sufficient condition $\frac{M-\overline{\delta}}{B-\overline{\delta}} \in \left(1, \frac{2-\eta}{1-\eta}\right)$, which is equivalent to $M < B + \frac{B-\overline{\delta}}{1-\eta}$ and is strictly stronger than $M \leq \overline{M}$, since $\frac{(B-\overline{\delta})^2}{D} > B - \overline{\delta}$. Proposition IA.3 dispenses with it.

the no-entry condition; the cap only shrinks the feasible set, so $V(\pi) \leq 2 - \eta - 2c + \frac{\pi}{2}(M - B)$ for every feasible governance structure, with equality attainable at $\pi = 0$ under the cap (severance is unused there). For $M < B$ the bound is strictly decreasing in π, so $\pi^* = 0$, and at $\pi = 0$ the effort IC constraint alone pins down $w_{1r}^* = \frac{2c}{2-\eta}$. For the remainder of the proof assume $M > B$.

Step 1: the help IC constraint (IA.2.7) binds at the optimum. Suppose not. If (IA.2.8) is also slack, a small reduction in w raises V and preserves all constraints. If (IA.2.8) binds while (IA.2.7) is slack, increase π and w along the effort boundary, $dw = \frac{B-\delta}{2-\eta} d\pi$; if the no-entry constraint (IA.2.6) permits, then

$$dV = \Big[- \Big(1 - \frac{\eta}{2}\Big) \frac{B - \delta}{2 - \eta} + \frac{M - \delta}{2} \Big] d\pi = \frac{M - B}{2}\, d\pi > 0.$$

If instead (IA.2.6) binds, move along the effort and no-entry boundaries jointly: $d\delta = -(1 - \eta) dw$ and $d\pi = \frac{(2-\eta) - \pi(1-\eta)}{B-\delta} dw$, so that

$$dV = \frac{\big[(2 - \eta) - \pi(1 - \eta)\big](M - B)}{2(B - \delta)}\, dw > 0,$$

and w increases until (IA.2.7) binds; it must bind at some $\delta > 0$ because at $\delta = 0$ the effort and no-entry boundaries imply $\pi B = (2 - \eta)w - 2c > (1 - \eta)w$ for $w > 2c$, violating (IA.2.7), while $\pi = 1$ is excluded by Lemma IA.2.

Step 2: the no-entry condition (IA.2.6) binds at the optimum. With (IA.2.7) binding, $\pi = \frac{(1-\eta)w}{B-\delta}$, and substituting $\pi(B - \delta) = (1 - \eta)w$ into (IA.2.8) shows that the effort IC constraint reduces to $w \geq 2c$, independently of δ. Holding w fixed and raising δ,

$$\frac{\partial V}{\partial \delta}\Big|_{(IA.2.7) \text{ binds}} = \frac{(1 - \eta)w}{2} \cdot \frac{M - B}{(B - \delta)^2} > 0,$$

so the firm raises δ until (IA.2.6) binds: $\delta = K - (B - c) - (1 - \eta)w$.

Step 3: the binding path and the two regimes. By Steps 1 and 2, the help IC constraint and the no-entry condition bind at the optimum. Parametrize the resulting one-dimensional path by $w \geq 2c$: the no-entry condition gives $\delta(w) = \overline{\delta} - (1 - \eta)(w - 2c)$, and the binding

help IC constraint gives

$$\pi(w) = \frac{(1-\eta)w}{B-\delta(w)} = \frac{(1-\eta)w}{D+(1-\eta)w}, \tag{IA.2.20}$$

a strictly increasing map from $[2c, \infty)$ onto $\left[\frac{(1-\eta)2c}{B-\bar{\delta}}, 1\right)$ with inverse $w(\pi) = \frac{\pi D}{(1-\eta)(1-\pi)}$. By the computation in Step 2, the effort IC constraint reduces to $w \geq 2c$ along the path, so the entire path is feasible: both ICs bind exactly at the vertex $w = 2c$, i.e. at $\pi_v := \frac{(1-\eta)2c}{B-\bar{\delta}}$, and the effort IC constraint is slack for $w > 2c$. Because the help IC constraint binds, the wage bill is $\frac{1}{2}(w + \pi B)$ and

$$V(\pi) = 2 - \eta - c + \frac{\pi}{2}(M-B) - \frac{1}{2}w(\pi), \qquad V''(\pi) = -\frac{D}{(1-\eta)(1-\pi)^3} < 0: \tag{IA.2.21}$$

V is strictly concave on $[\pi_v, 1)$ and $w(\pi) \to \infty$ as $\pi \to 1$. For $\pi < \pi_v$, the argument of Step 0 applies with $M > B$: the compensation cost is bounded below by $c + \frac{\pi B}{2}$, and this bound is attainable under the cap on all of $[0, \pi_v]$, so V increases at rate $\frac{M-B}{2}$ there and the optimum satisfies $\pi^* \geq \pi_v$. Using $1 - \pi_v = \frac{D}{B-\bar{\delta}}$, the right derivative at the vertex is

$$V'(\pi_v^+) = \frac{1}{2}\Big[(M-B) - \frac{D}{(1-\eta)(1-\pi_v)^2}\Big] = \frac{1}{2}\Big[(M-B) - \frac{(B-\bar{\delta})^2}{(1-\eta)D}\Big] = \frac{1}{2}\big(M - \overline{M}\big). \tag{IA.2.22}$$

If $M \leq \overline{M}$, concavity implies $\pi^* = \pi_v$, which is (IA.2.18). If $M > \overline{M}$, the first-order condition $V'(\pi) = 0$ gives $(1-\pi)^2 = \frac{D}{(1-\eta)(M-B)}$, i.e. the unique interior optimum (IA.2.19), with $\pi^* < 1$ because $D > 0$ and $w_{1r}^* = w(\pi^*) > 2c$ because $\pi^* > \pi_v$.

Step 4: remaining constraints. At the optimum the help IC constraint binds, $\pi^*(B - \delta^*) = (1-\eta)w_{1r}^*$, so the slack in the truth-telling constraint (IA.2.9) equals $\eta w_{1r}^* + \pi^*(B-\delta^*) - (1-\pi^*)\frac{c\eta}{3-2\eta} = w_{1r}^* - (1-\pi^*)\frac{c\eta}{3-2\eta} \geq 2c - \frac{c\eta}{3-2\eta} > 0$ in both branches with $\pi^* > 0$, since $w_{1r}^* \geq 2c$; a regular CEO who shirks is retained (she is not a charlatan, so Assumption IA.1 is silent about her), and the right-hand side of (IA.2.8) remains $2B$; lazy entry is deterred by $K \geq B$; and charlatan entry at $t = 2$ is deterred because a $t = 2$ charlatan (hired as a replacement) obtains at most $B - c + (1-\eta)w_{2n}^* - K = B - c + \frac{2c(1-\eta)}{2-\eta} - K \leq B + (1 -$

$2\eta)c - K \leq 0$.[7]

Step 5: uniqueness and other equilibria. For a fixed δ, the continuation game is the benchmark game of the main text with private benefit $\widetilde{B}(\delta)$ in the dismissal state, since δ enters the CEO's dismissal payoff and the firm's cost of firing symmetrically. The arguments in the proof of Proposition 1 ruling out mixed-strategy equilibria and equilibria that violate the help IC constraint therefore apply after replacing B with $\widetilde{B}(\delta^*)$ and M with $\widetilde{M}(\delta^*)$, and are omitted.

Comparative statics. Profit at the optimum exceeds the main-text optimum $V(2c, \pi^*_{|\delta=0})$ by

$$c(1-\eta)\,\overline{\delta}\,\frac{M-B}{B(B-\overline{\delta})} \;>\; 0, \tag{IA.2.23}$$

and is increasing in K because $\frac{dV^*}{d\overline{\delta}} = c(1-\eta)\frac{M-B}{(B-\overline{\delta})^2} > 0$. Because (IA.2.8) binds, the CEO's lifetime utility is $2B$; the accounting identity in the proof of Corollary 1 (now inclusive of severance receipts) yields expected total compensation $E(w) = 2c + \frac{\pi^*}{2}(B-c)$, which is increasing in K through π^*. □

Proposition IA.3 answers the question posed at the start of this section: the threat of charlatan entry places an endogenous upper bound on the severance bonus, given by (IA.2.6), and whenever monitoring is privately valuable ($M > B$) the resulting optimal governance structure is interior, with the no-entry condition binding and monitoring intensity strictly between zero and one. The two regimes of valuable monitoring differ in which margin is active. Below $\overline{M}$, additional monitoring is financed by severance insurance: the firm raises the premium until charlatan-proofness binds, at no cost in the uniform bonus. Above $\overline{M}$, the premium is exhausted, and further monitoring must be bought by raising the uniform bonus w_{1r} along the help IC constraint—which is costly precisely because it is paid to good and bad types alike, and prohibitively so as $\pi \to 1$, which is why full monitoring is never reached. In this regime the severance premium $\Delta^* = w^*_{1s} - w^*_{1r}$ shrinks as M grows and eventually turns negative: when monitoring is sufficiently valuable, the firm pays a *retention* premium, and π^* then exceeds the ceiling of Lemma IA.2, which applies only to

[7] Charlatans cannot pose as the retained incumbent, so entry at $t = 2$ can occur only through the replacement pool.

$\Delta \geq 0$ (the bound $\pi^* < 1$, whose proof does not restrict the sign of Δ, is what survives). The comparative statics in K are intuitive: a higher masquerading cost relaxes the screening problem, allowing the firm to insure the CEO more generously against dismissal, monitor more intensively, and pay more in expectation; Corollary IA.2 collects the remaining signs. The two ends of the admissible range connect this section to the main text.

Corollary IA.2 (Comparative statics). *In the regime $B < M \leq \overline{M}$, the optimal monitoring intensity π^* of (IA.2.18) is strictly increasing in K and in c, and strictly decreasing in B and in η. In the regime $M > \overline{M}$, the optimal monitoring intensity π^* of (IA.2.19) is strictly increasing in K, in c, and in M, and strictly decreasing in B.*

Proof. Write $\pi_v^* = \frac{2c(1-\eta)}{B-\overline{\delta}}$ with $B - \overline{\delta} = 2B - K + (1-2\eta)c$. Then $\frac{\partial \pi_v^*}{\partial K} = \frac{(\pi_v^*)^2}{2c(1-\eta)} > 0$, $\frac{\partial \pi_v^*}{\partial c} = \frac{2(1-\eta)(2B-K)}{(B-\overline{\delta})^2} > 0$ (using $K < 2B - c < 2B$), $\frac{\partial \pi_v^*}{\partial B} = -\frac{4c(1-\eta)}{(B-\overline{\delta})^2} < 0$, and $\frac{\partial \pi_v^*}{\partial \eta} = -\frac{2cD}{(B-\overline{\delta})^2} < 0$. In the high-$M$ regime, $1 - \pi^* = \sqrt{\frac{D}{(1-\eta)(M-B)}}$ with $D = 2B - K - c$ strictly decreasing in K and in c and strictly increasing in B, while $M - B$ is decreasing in B; the stated monotonicities follow. □

Corollary IA.3 (The main-text cap and the first best as limiting cases). *(i) As $\overline{\delta} \downarrow 0$ (i.e., $K \downarrow B + (1-2\eta)c$, feasible when $\eta \leq \frac{1}{2}$), the optimal contract converges to the main-text solution under Assumption 6 with $\overline{s} = 0$: $w_{1r}^* = w_{1s}^* = 2c$ and $\pi^* = \frac{(1-\eta)2c}{B}$. (ii) If instead $K \geq 2B - c$, the first best is attained, for example by $w_{1r}^* = 2c$, $w_{1s}^* = \frac{B}{1-\eta}$, and $\pi^* = 1$, with the bad CEO fully insured against dismissal.*

Part (i) shows that the exogenous cap of the main text is the reduced form of a masquerading cost at the bottom of the admissible range; none of the main text's results rely on the cap being exactly zero, since for any K in the range (IA.2.15) the model is isomorphic to the benchmark with $(\widetilde{B}, \widetilde{M})$ in place of (B, M). Part (ii) recovers the frictionless benchmark of Proposition IA.2: when mimicking qualified candidates is sufficiently costly, the no-entry condition stops binding at the first-best contract, severance is again a frictionless instrument, and CEO-friendliness is unnecessary.

As a numerical illustration, take the parameters used in Figures 2–5 of the main text ($\eta = 0.4$, $B = 0.048$, $c = 0.02$), for which $M = \frac{\eta}{6} = 0.0\overline{6} > B$ and the benchmark solution

is $\pi^* = 0.5$. The admissible range (IA.2.15) is $K \in [0.052, 0.076)$, condition $M \leq \overline{M}$ of Proposition IA.3 holds throughout (indeed $\overline{M} \geq 0.208$ on the range), and as K rises from 0.052 to 0.076 the optimal monitoring intensity rises from 0.5 (the main-text value) toward 1 (the first best).

Remark IA.1 (**Retention premia when mimicry is cheap**)**.** The range (IA.2.15) requires $K \geq B + (1-2\eta)c$, so that $\overline{\delta} \geq 0$. For $\eta < \frac{1}{2}$, the interval $\big[B,\, B + (1-2\eta)c\big)$ is nonempty, and on it charlatan-proofness forces a *negative* premium: $\overline{\delta} < 0$. No step of the proof of Proposition IA.3 uses $\overline{\delta} \geq 0$, so the characterization extends verbatim: in the regime $B < M \leq \overline{M}$,

$$w_{1s}^* = \frac{K - (B-c)}{1-\eta} \;<\; 2c \;=\; w_{1r}^*, \qquad \pi^* = \frac{(1-\eta)2c}{B - \overline{\delta}} \;<\; \frac{(1-\eta)2c}{B},$$

the firm pays a retention premium, and the board is *friendlier* than in the main text.[8] At the bottom of the range, $K = B$, the contract is $w_{1s}^* = \frac{c}{1-\eta}$ with $\operatorname{sign}(w_{1s}^* - w_{1r}^*) = \operatorname{sign}(2\eta - 1)$: severance pay proper emerges only when the board's advice is sufficiently unreliable ($\eta > \frac{1}{2}$). The main-text cap thus sits at a knife-edge, $K = B + (1-2\eta)c$: for costlier mimicry the firm pays genuine severance and monitors more than in Proposition 1; for cheaper mimicry it pays a premium to *stay* and monitors less. Together with Remark IA.4, this sharpens the empirical content of the mechanism: where executive quality is easy to counterfeit, the model predicts not only friendlier boards but also compensation tilted toward retention rather than separation.

Remark IA.2 (**An alternative setup: short horizons and voluntary separations**)**.** An equivalent way to make retention worthless to charlatans is to replace Assumption IA.1 with two other assumptions: charlatans have a short horizon (they must retire after one period), and any separation at Date (i) of $t = 2$ is recorded as a dismissal ($\theta = f$) whether initiated by the board or by the CEO—a *separation convention* motivated by the fact that the initiating party of an amicable exit is not verifiable. A hired charlatan then asks to be fired under a weak

[8] The only step whose stated bound used $K \geq B + (1-2\eta)c$ is the $t = 2$ entry check in Step 4; the exact condition is $K \geq B - \frac{c\eta}{2-\eta}$, which holds for all $K \geq B$.

board, his payoffs (IA.2.3) and (IA.2.4) are unchanged, and Proposition IA.3 goes through verbatim—except that the convention hands *regular* CEOs a new deviation: a successful CEO can induce her own dismissal to grab the severance bonus. Staying must then dominate this "golden handshake," $w_{1r} + B \geq w_{1s}$, i.e., $\Delta \leq B$, which tightens the admissible range (IA.2.15) to $K < \min\{2B - c,\ (2-\eta)B + (1-2\eta)c\}$. If $\eta B > 2c(1-\eta)$, this internal constraint binds before charlatan-proofness does for large K: the premium is capped at $\delta = (1-\eta)B$ and monitoring intensity at $\pi = \frac{2c(1-\eta)}{\eta B} < 1$ *for every* K, so this variant features two independent microfoundations for a severance cap—outside entry by charlatans and inside handshake-grabbing by good CEOs—and the first best is unattainable no matter how costly masquerading is.

Remark IA.3 (**Charlatans and AI**)**.** The analysis extends to the AI economy of Sections 6–7 of the main text with one twist. A charlatan has $z = 0$, so his own AI-augmented assessment is never correct; but he can ask the AI agent to solve the problem autonomously, which succeeds with probability $\Pr(x \leq \sqrt{\kappa}) = \sqrt{\kappa}$—not κ, which is the corresponding probability for a regular type-b CEO. The charlatan's best route to the severance bonus therefore succeeds with probability $\rho^{ch} := \max\{1-\eta, \sqrt{\kappa}\}$, and the no-entry condition becomes $\rho^{ch} w_{1s} \leq K - (B-c)$; note that Assumption 2 ($1 - \eta > \kappa$) does not guarantee $1 - \eta > \sqrt{\kappa}$. Repeating the steps above, the interior solution has $w_{1r}^* = \frac{2c\alpha}{2\alpha - 1 + \kappa}$ as in Proposition 3 and

$$\pi^* = \frac{(1-\eta-\kappa)\, w_{1r}^*}{G + (1-\eta) w_{1r}^*}, \qquad G := B - \frac{(1-\eta)\big[K - (B-c)\big]}{\rho^{ch}} > 0,$$

where $G > 0$ follows from $K < 2B - c$ and $\rho^{ch} \geq 1 - \eta$. Because π^* is increasing in w_{1r}^* and decreasing in κ (directly, and through ρ^{ch}) and w_{1r}^* is decreasing in both κ and α, all channels reinforce the main text's comparative statics: improvements in either AI capability reduce monitoring intensity, exactly as in Propositions 2 and 3. Interestingly, the endogenous cap introduces a partial offset—a lower bonus w_{1r}^* mechanically relaxes the no-entry constraint, permitting a larger severance premium—but the offset is never strong enough to overturn the signs.

Remark IA.4 (**Interpreting** K)**.** The parameter K measures how costly it is to manufacture a credible CEO candidacy—a polished track record, fluent command of the vocabulary of expertise, convincing references. Technologies that cheapen expert mimicry map into a lower K: Cowgill et al. (2024) document that generative AI degrades screening precisely because less-qualified individuals use it to imitate expert communication. Through the lens of this section, such technologies tighten the severance cap (IA.2.6) and, by Proposition IA.3, reduce board monitoring intensity and expected CEO pay. This is a distinct channel from those in the main text: AI reduces board independence not only by improving the CEO's outside advice (κ) and productivity (α), but also by flooding the managerial labor market with potential charlatans.

IA.3 Verification: The Case $1 - t_s \geq 2t_v$

Assumption 3 in the main text ($t_v \leq 1 - t_s < 2t_v$) implies that a CEO who learns that her own assessment is incorrect does not have enough time to verify a *second* assessment: she must choose between the AI agent and the board, and AI's problem-solving capability enters the help IC constraint as an outside option. This section considers the complementary case,

$$t_v \leq 1 - t_s \qquad \text{and} \qquad 1 - t_s \geq 2t_v, \tag{IA.3.1}$$

in which the CEO can verify one additional assessment. Throughout this section we set $\alpha = 1$ and $\kappa > 0$, and maintain Assumption 2 ($1 - \eta > \kappa$). Assumption 1 plays no role here: with time for two verifications, the CEO can verify both her own assessment and the AI agent's, so the order in which she generates them is payoff-irrelevant—under either order, she involves the board exactly when both assessments fail. Verification priority matters only when verification capacity is scarce (Assumption 3); see Section IA.5. Subsection IA.3.2 uses the analysis to study AI-driven reductions in verification time, the extension announced in Section 3 of the main text.

IA.3.1 Sequential Verification

Under (IA.3.1), a CEO who verifies that her own assessment is incorrect can query the AI agent (which responds in no time), spend t_v verifying the AI's solution, and—if the AI's solution is also incorrect—still ask the board for help, since the board's assessment requires none of the CEO's time. Sequential consultation, which Assumption 3 precludes, is now feasible: AI and the board become *complements* rather than substitutes in the supply of help.

The bad CEO's optimal plan at Date (iii) is therefore: query the AI agent and verify its solution; if it is correct (probability $\kappa = \Pr[x \leq \sqrt{\kappa} \mid x > z]$), implement it and do not contact the board; if it is incorrect, decide whether to ask the board (m_b) or give up (m_c). Conditional on asking, the board solves the problem with probability $1 - \eta$. Hence a type-b CEO who follows this plan produces $y_1 = 1$ with total probability

$$\kappa + (1 - \kappa)(1 - \eta) = 1 - \hat{\eta}, \qquad \hat{\eta} := \eta(1 - \kappa), \tag{IA.3.2}$$

so sequential access reduces the effective cost of help from η to $\hat{\eta}$. The same is true in $t = 2$ (where there are no career concerns), so the period-2 bonuses and continuation values are those of the main text with η replaced by $\hat{\eta}$; in particular, the marginal benefit of monitoring becomes $\widehat{M} := M(1, \hat{\eta}) = \frac{\hat{\eta}}{6}$.

On the equilibrium path, the no-ask pool now contains both good CEOs and lucky bad CEOs (those whose problem the AI solved), in proportions $\frac{1}{1+\kappa}$ and $\frac{\kappa}{1+\kappa}$, while asking reveals that the CEO is bad *and* that the AI failed; since $E[z \mid b,\, x > \sqrt{\kappa}] < E[z \mid b] = \frac{1}{3}$, a tough board replaces a CEO who asks, a fortiori. At Date (i) of $t = 2$, the board offers a retained no-ask CEO whichever bonus maximizes its continuation profit: for κ close to zero, the bonus targeted at type g (under which a retained lucky bad CEO shirks), and otherwise the pooling bonus $\frac{3c}{3-2\hat{\eta}}$, which induces both types in the pool to work. Which bonus the board picks turns out not to matter for governance. Under either bonus, a retained type-b CEO's continuation utility is B; under the pooling bonus, a retained good CEO additionally earns an informational rent $R > 0$, and a CEO who shirked at $t = 1$ and pools with the no-askers earns rent $\frac{R}{2}$—exactly half, because period-2 success probabilities are linear in expected knowledge and the shirker's type is the midpoint of the pool's. The rent terms

then cancel from the first-period effort IC constraint: the good CEO's rent R accrues with probability $\frac{1}{2}$ on the equilibrium path, matching the shirker's certain rent of $\frac{R}{2}$.[9] The incentive constraints derived below are therefore identical in the two regimes. We assume only that $\kappa \leq \overline{\kappa}$, where $\overline{\kappa} > 0$ is the largest κ for which the tough board prefers retaining the no-ask pool, under its optimal bonus, to replacing it. The preference is strict at $\kappa = 0$, so $\overline{\kappa} > 0$ exists, and the restriction is mild: at the parameter values used in the main text's figures, retention is optimal for all $\kappa \leq 0.85$.

Two constraints change relative to the main text. First, because the CEO consults the AI agent *before* deciding whether to involve the board, AI's capability no longer appears as an outside option: conditional on reaching the board-consultation decision, the AI option is already exhausted. The help IC constraint—now conditional on AI failure—reverts to its *pre-AI* form:

$$(1-\eta)w_1 + (1-\pi)B \;\geq\; B \quad\Longleftrightarrow\quad \pi \leq \frac{(1-\eta)w_1}{B}. \tag{IA.3.3}$$

Second, the effort IC constraint becomes

$$B - c + \frac{1}{2}(w_1 + B) + \frac{1}{2}\Big[\kappa(w_1 + B) + (1-\kappa)\big((1-\eta)w_1 + (1-\pi)B\big)\Big] \;\geq\; 2B, \tag{IA.3.4}$$

which simplifies to $(2-\hat{\eta})w_1 - (1-\kappa)\pi B \geq 2c$: the AI agent raises the return to finding a deal (a lucky bad CEO now earns the bonus and keeps her job), so effort is cheaper to incentivize.

Proposition IA.4 (Governance under sequential verification). *Let $\alpha = 1$, $\kappa \in (0, \overline{\kappa}]$, and suppose (IA.3.1) holds. The optimal governance structure is interior and given by*

$$w_1^* = \frac{2c}{1+\kappa}, \qquad \pi_{seq}^* = \frac{(1-\eta)2c}{(1+\kappa)B}, \tag{IA.3.5}$$

if and only if

$$\frac{\widehat{M}}{B} \in \Big(1,\; \frac{2-\hat{\eta}}{(1-\kappa)(1-\eta)}\Big). \tag{IA.3.6}$$

[9] Formally, with q_τ denoting a retained type-τ CEO's period-2 success probability, $q_n = \frac{q_g + q_b}{2}$ implies $q_n w_{2b} - c = \frac{1}{2}\left(q_g w_{2b} - c\right)$ at the pooling bonus $w_{2b} = c/q_b$, at which the bad type's rent is zero.

Under an interior solution: (i) the first-period bonus coincides with the main-text bonus $w_{1AI}^* = \frac{2c}{1+\kappa}$ *of Proposition 2, while monitoring intensity is strictly higher,*

$$\frac{\pi_{seq}^*}{\pi_{AI}^*} = \frac{1-\eta}{1-\eta-\kappa} > 1, \tag{IA.3.7}$$

though still lower than the pre-AI level $\pi^* = \frac{(1-\eta)2c}{B}$*; (ii) both* w_1^* *and* π_{seq}^* *are decreasing in* κ*; (iii) the CEO's expected lifetime utility is* $2B$*, independent of* κ*.*

Proof. When (IA.3.3) and (IA.3.4) both bind, substituting $\pi B = (1-\eta)w_1$ into the effort IC constraint gives $w_1\big[(2-\hat{\eta}) - (1-\kappa)(1-\eta)\big] = 2c$; since $\hat{\eta} + (1-\kappa)(1-\eta) = 1-\kappa$, the bracket equals $1+\kappa$, which yields (IA.3.5). In either continuation regime, expected profit takes the form

$$V = \Gamma - \Big(1 - \frac{\hat{\eta}}{2}\Big)w_1 + \frac{(1-\kappa)\widehat{M}}{2}\,\pi,$$

where Γ collects terms independent of (w_1, π): first-period wages are paid with probability $\frac{1}{2} + \frac{1}{2}(1-\hat{\eta}) = 1 - \frac{\hat{\eta}}{2}$, and π matters only through the replacement of a revealed-bad CEO, which occurs with probability $\frac{1-\kappa}{2}\pi$ and gains $\widehat{M} = v_{2n} - v_{2b}$, where $v_{2b} = 1 - \frac{2\hat{\eta}}{3} - c$ is the continuation profit from a retained revealed-bad CEO. The isoprofit slope in the (w_1, π) plane is thus $\frac{2-\hat{\eta}}{(1-\kappa)\widehat{M}}$, while the help and effort IC boundaries have slopes $\frac{1-\eta}{B}$ and $\frac{2-\hat{\eta}}{(1-\kappa)B}$. The tangency argument of Proposition 1 then implies that the solution is interior if and only if the isoprofit slope lies between the two, which is (IA.3.6); all remaining steps (ruling out mixed strategies and equilibria violating the help IC constraint) mirror the proof of Proposition 1 and are omitted. Part (i) follows from comparing (IA.3.5) with Proposition 2; parts (ii) and (iii) are immediate from (IA.3.5) and the binding effort IC constraint. □

Proposition IA.4 decomposes the governance effect of AI's problem-solving capability into the two channels at work in the main text. Under Assumption 3, κ operates both as an *outside option* (flattening the help IC constraint) and as a *productivity shifter* (lowering the bonus needed for effort); under sequential verification, the CEO extracts AI's value on the equilibrium path rather than off it, the first channel disappears, and only the second remains. Both channels push governance in the same direction, which is why the main

text's predictions are robust: even when AI complements the board, improvements in κ reduce CEO pay and board monitoring. But the entrenchment cost is strictly smaller—monitoring falls by less, per (IA.3.7)—because the firm no longer needs to bribe the CEO away from a private substitute for board advice.

Sequential verification introduces one new distortion worth noting: because lucky bad CEOs pool with good CEOs, the board's learning is coarser, and the unconditional probability that a bad CEO survives into $t = 2$ rises to $\kappa + (1-\kappa)(1-\pi^*_{seq})$. AI thus entrenches bad CEOs through two routes—friendlier boards and camouflage—and the second route operates even though the help IC constraint is pre-AI in form. For this reason, the overall profit comparison between the two time-constraint regimes is ambiguous in general: sequential access raises output ($1-\hat{\eta} > 1-\eta$) and monitoring, but degrades period-2 selection—retained lucky bad CEOs are less productive than the good CEOs they pool with, and inducing them to work requires leaving informational rents to good CEOs.

IA.3.2 AI and Verification Time

We now use the previous subsection to analyze the extension mentioned in Section 3 of the main text: AI may reduce the time needed to verify assessments. Suppose the verification technology is $t_v(a)$, decreasing in AI capability a (e.g., AI-assisted checking of a deal assessment), with t_s fixed. Because time enters the model as a constraint rather than as a utility cost, marginal changes in t_v are neutral *within* each regime; verification time matters only through which time-constraint regime applies:

$$\underbrace{1 - t_s < 2t_v(a)}_{\text{Assumption 3: AI is a substitute}} \qquad \text{versus} \qquad \underbrace{1 - t_s \geq 2t_v(a)}_{\text{(IA.3.1): AI is a complement}} \quad . \tag{IA.3.8}$$

As a rises and $t_v(a)$ falls below $\frac{1-t_s}{2}$, the economy switches from the first regime to the second, and by Propositions 2 and IA.4 the optimal governance structure jumps from $\left(\frac{2c}{1+\kappa}, \frac{(1-\eta-\kappa)2c}{(1+\kappa)B}\right)$ to $\left(\frac{2c}{1+\kappa}, \frac{(1-\eta)2c}{(1+\kappa)B}\right)$: the first-period bonus is unchanged, while board monitoring intensity rises discretely by the factor $\frac{1-\eta}{1-\eta-\kappa} > 1$. A type-$b$ CEO's success probability also rises, from $1-\eta$ to $1-\hat{\eta}$.

The economics is that verification speed determines *how* the CEO uses AI. When verification is slow, consulting AI forecloses consulting the board, so AI is a substitute for board advice and its capability is an off-equilibrium threat that must be bought off with board friendliness. When verification is fast, AI becomes a first-pass filter whose failures are passed on to the board, the threat disappears, and monitoring intensity recovers. Consistent with the view that verification is becoming the key bottleneck in the age of AI (Lamba (2026)), the governance implications of AI adoption thus hinge on the verification technology: AI tools that merely generate solutions erode board monitoring, whereas AI tools that also accelerate the *checking* of solutions partially restore it. This yields an empirical prediction that distinguishes our mechanism from a pure productivity story: improvements in verification-augmenting AI should be associated with *higher* board independence, holding problem-solving capability fixed.

IA.4 The Benchmark without Assumption 4

Assumption 4 in the main text ($1 - \frac{2}{3}\eta \geq c$) guarantees that inducing effort is efficient for every retained CEO type, so that in $t = 2$ the board offers effort-inducing bonuses to all types. This section characterizes the pre-AI benchmark ($\alpha = 1$, $\kappa = 0$) when Assumption 4 fails. Following the footnote to Assumption 4, we maintain that effort is efficient for the average type:

$$1 - \frac{2}{3}\eta \;<\; c \;\leq\; 1 - \frac{\eta}{2}. \tag{IA.4.1}$$

Assumptions 3 and 5 are maintained throughout, and the severance cap (Assumption 6) applies with $\bar{s} = 0$; Assumptions 1 and 2 hold automatically in the pre-AI benchmark.

Second-period contracts. The bonuses for a new CEO and for a retained good CEO are unchanged: their success probabilities, $1 - \frac{\eta}{2}$ and $1 - \frac{\eta}{3}$, weakly exceed c under (IA.4.1), so inducing their effort remains profitable and $w^*_{2n} = \frac{2c}{2-\eta}$ and $w^*_{2g} = \frac{3c}{3-\eta}$ are as in the main text. What changes is the treatment of a retained CEO known to be of type b: inducing her effort requires $w^*_{2b} = \frac{3c}{3-2\eta}$, an expected wage cost of c against expected output $1 - \frac{2}{3}\eta < c$.

The board therefore offers her no incentive pay; she shirks, enjoys B, and produces nothing, so that

$$v_{2b} = 0, \tag{IA.4.2}$$

in contrast to $v_{2b} = 1 - \frac{2}{3}\eta - c > 0$ under Assumption 4. Her continuation utility is B in either regime: with effort her IC constraint binds, without it she collects only the private benefit. The marginal benefit of monitoring becomes the *entire* surplus generated by a replacement,

$$M' := v_{2n} - v_{2b} = 1 - \frac{\eta}{2} - c \ \in \ \Big[0, \ \frac{\eta}{6}\Big), \tag{IA.4.3}$$

with $M' \to \frac{\eta}{6} = M(1, \eta)$ as $c \to 1 - \frac{2}{3}\eta$: the benefit of monitoring is continuous across the boundary of Assumption 4. That M' is *smaller* than $M(1, \eta)$—even though a retained bad CEO is now worth less—may seem puzzling, but it reflects how c moves both values. Over the whole parameter space,

$$M = v_{2n} - v_{2b} = \min\Big\{\frac{\eta}{6}, \ 1 - \frac{\eta}{2} - c\Big\}. \tag{IA.4.4}$$

Under Assumption 4, both the retained bad CEO and her replacement work, and each costs the firm exactly c in expected wages (their effort IC constraints bind); as c rises, v_{2n} and v_{2b} fall in parallel, and their difference is the pure *quality gain* $\frac{\eta}{6}$, independent of c. At $c = 1 - \frac{2}{3}\eta$, v_{2b} reaches its floor of zero—the board can always withhold incentive pay—and stops falling, while v_{2n} keeps falling one-for-one with c. Beyond this kink, replacement no longer swaps one worker for a better one at equal wage cost; it *activates production at all*, and its value is the full (and, at these effort costs, small) surplus $1 - \frac{\eta}{2} - c < \frac{\eta}{6}$.

First-period constraints. Because every retained CEO's continuation utility equals B at every relevant node—a good CEO works at a binding IC constraint, while a retained bad CEO and a shirker collect only B—the period-1 incentive constraints are *unchanged*: the help IC constraint, $\pi B \leq (1 - \eta)w_1$, and the effort IC constraint, $(2 - \eta)w_1 - \pi B \geq 2c$, are exactly those of the main text. The good CEO's truth-telling constraint becomes strictly easier: a CEO believed to be bad is no longer offered w^*_{2b}, so the deviation payoff loses the

(positive) rent term and the constraint reduces to $w_1 + B \geq (1-\eta)w_1 + (1-\pi)B$, which holds trivially. Expected profit is

$$V(w_1, \pi) = \frac{1}{2}\Big(2 - \frac{\eta}{3} - c - w_1\Big) + \frac{1}{2}\Big((1-\eta)(1-w_1) + \pi M'\Big), \tag{IA.4.5}$$

the main-text expression with M' in place of $M(1,\eta)$ and the retained-bad continuation profit set to zero.

Optimal governance.

Proposition IA.5 (Without Assumption 4, the model collapses to Case 1). *Under (IA.4.1) and Assumptions 3 and 5, the optimal governance structure coincides with Case 1 of Proposition 1:*

$$w_1^* = \frac{2c}{2-\eta}, \qquad \pi^* = 0. \tag{IA.4.6}$$

Inducing first-period effort remains optimal, and the firm's profit is strictly positive.

Proof. The constraint set and (IA.4.5) have the structure of the main-text problem with $M(1,\eta)$ replaced by M', so the characterization in Proposition 1 applies with M'. Assumption 5 and (IA.4.1) imply

$$B \;\geq\; 2c \;>\; 2\Big(1 - \frac{2}{3}\eta\Big) \;=\; 2 - \frac{4}{3}\eta \;>\; \frac{\eta}{6} \;>\; M',$$

where the third inequality holds for all $\eta \in (0,1)$; hence $M' < B$ always, Case 1 of Proposition 1 obtains, and the effort IC constraint alone pins down (IA.4.6). Profit is $V^* = \frac{1}{2}\big(3 - \frac{4}{3}\eta - c\big) - c > 0$, since (IA.4.1) implies $c \leq 1 - \frac{\eta}{2} < 1 - \frac{4}{9}\eta$. Finally, any strategy that does not induce first-period effort yields first-period output zero and second-period profit at most $v_{2n} = M'$; since $V^* - M' = \frac{1}{2} - \frac{\eta}{6} - \frac{c}{2} \geq 0$ if and only if $c \leq 1 - \frac{\eta}{3}$, which is implied by (IA.4.1), inducing effort is optimal. □

At $\pi^* = 0$ the help IC constraint is strictly slack: a bad CEO asks for help because $(1-\eta)w_1^* > 0$ and asking carries no dismissal risk. Communication is free and complete; what the firm gives up is monitoring, not information.

Which proofs change. Assumption 4 does real work in exactly one proof: that of Lemma 2. There, a tough board facing an uninformative retained pool compares replacement with retention, and the paper's own expression $v_{2n} - v_{2g} = \frac{1}{2}\left(1 - \frac{2}{3}\eta - c\right)$ shows that the sign of this comparison *is* Assumption 4. Under (IA.4.1), a tough board prefers retaining a fifty–fifty pool—offering w^*_{2g} and employing only its good half—to replacing it; more generally, it retains any pool whose share of good types exceeds $\overline{\gamma} = \frac{M'}{1-\eta/3-c} < \frac{1}{2}$. Two consequences follow. First, Lemma 2 fails as stated: uninformative continuation equilibria now exist—for instance, both types asking for help, supported by the off-path belief that a CEO who does not ask is bad—because pooling no longer triggers replacement. Such equilibria are strictly dominated for shareholders (good CEOs produce $1 - \eta$ instead of 1, and the board learns nothing), so the shareholder-preferred equilibrium remains the separating one and the characterization above is unaffected; what is lost is only the generic uniqueness of the continuation equilibrium for an arbitrary governance structure. Second, Lemma 3 survives on the optimal equilibrium path: a revealed bad CEO is still replaced by a tough board ($v_{2n} \geq 0 = v_{2b}$) and a revealed good CEO is still retained ($1 - \frac{\eta}{3} - c > v_{2n}$), so $\rho(m_b) = 1$, $\rho(m_c) = 0$, and a good CEO strictly prefers not to ask.

Interpretation: a case-reducing assumption. Proposition IA.5 gives Assumption 4 a precise status: it is a *case-reducing* assumption. Its failure creates no new outcomes—the optimal governance structure is the one already listed as Case 1 of Proposition 1, and under Assumption 5 even Proposition 1's stated condition for Case 1 classifies correctly, since $M(1, \eta) = \frac{\eta}{6} < B$ holds automatically when $B \geq 2c > 2 - \frac{4}{3}\eta$. What Assumption 4 removes is confinement to that case: under (IA.4.1), $M' < B$ for every admissible parameter configuration, so Cases 2 and 3 are empty and the board is maximally friendly ($\pi^* = 0$)—even though entrenchment is now more costly ex post, since a retained bad CEO produces nothing. These two facts have a common cause: when effort costs are high enough that a bad CEO is not worth employing, a replacement barely breaks even either ($M' \leq \frac{\eta}{6}$), while the private benefit that monitoring must overcome satisfies $B \geq 2c > 2 - \frac{4}{3}\eta$. Thus the footnote to Assumption 4 in the main text is exactly right once stated precisely: nothing new happens when the assumption fails—the model simply collapses into Case 1—but the

interior region that is the paper's object of interest exists only under Assumption 4.

IA.5 The Model without Assumption 1

Assumption 1 in the main text ($1 - \frac{1}{2\alpha} > \sqrt{\kappa}$) states that the average CEO is more likely to solve a problem than the AI agent is to solve one autonomously. Under Assumption 3, this comparison governs a decision that the main text's timing leaves implicit: the CEO has time to verify exactly *one* assessment, and she chooses which. This section first formalizes the main text's claim that, under Assumption 1, generating and verifying her own assessment first is without loss of generality (Lemma IA.3). It then characterizes optimal governance when Assumption 1 fails, in the general model of Section 6 of the main text ($\alpha \geq 1$, $\kappa \in [0, 1]$). Throughout, Assumptions 2–5 are maintained and the severance cap (Assumption 6) applies with $\bar{s} = 0$; "Assumption 1 fails" means $\sqrt{\kappa} > 1 - \frac{1}{2\alpha}$. (At the knife edge $\sqrt{\kappa} = 1 - \frac{1}{2\alpha}$ the CEO is indifferent between the two plans below, and the shareholder-preferred equilibrium is the main-text one.)

IA.5.1 The Verification-Priority Decision

Consider Date (iii) of $t = 1$, after the CEO has exerted effort and identified a deal. Under Assumption 3 ($t_v \leq 1 - t_s < 2t_v$), the CEO can generate both assessments—her own takes t_s, the AI agent's takes no time—but can verify only one of them. Her plan therefore reduces to a choice of *verification priority*:

(a) *Assess-first* (the main-text timing): she verifies her own assessment and learns her *ex post* type; if bad, she chooses between the AI agent's unverified solution (success probability $\kappa = \Pr[x \leq \sqrt{\kappa} \mid x > \alpha z]$) and the board.

(b) *AI-first*: she verifies the AI agent's solution, which is correct with unconditional probability $\Pr(x \leq \sqrt{\kappa}) = \sqrt{\kappa}$. If it is correct, she implements it—learning nothing about her own type. If it is incorrect, she knows only that $x > \sqrt{\kappa}$; because $1 - t_v \geq t_s$, she can

still generate her own (now unverifiable) assessment, which is correct with probability

$$p := \Pr[x \le \alpha z \mid x > \sqrt{\kappa}] = 1 - \frac{1+\sqrt{\kappa}}{2\alpha}, \tag{IA.5.1}$$

or she can ask the board.[10]

Write $s := \sqrt{\kappa}$, $P_g := 1 - \frac{1}{2\alpha} = \Pr(x \le \alpha z)$, and let $X := (1-\eta)w_1 + (1-\pi)B$ denote the value of asking the board, so that the main text's help IC constraint reads $X \ge \kappa w_1 + B$.

Lemma IA.3 (Verification priority). *Fix a governance structure (w_1, π), $w_1 \ge 0$, and the main-text continuation: a retained CEO's continuation utility is B in every branch, and asking triggers replacement with probability π. Let U_{own} and U_{AI} denote the CEO's Date-(iii) payoffs under the assess-first and AI-first plans.*

(i) If $s \le P_g$ (Assumption 1) and the help IC constraint holds, then $U_{own} \ge U_{AI}$: the assess-first timing is without loss of generality. The preference is strict whenever the help IC constraint is slack, $w_1 > 0$, and $s < P_g$; when the help IC constraint binds, the CEO is exactly indifferent.

(ii) If $s > P_g$ (Assumption 1 fails), then $U_{AI} \ge U_{own}$ for every *governance structure, with strict preference whenever the help IC constraint holds and $\eta w_1 + \pi B > 0$.*

Proof. The two payoffs are

$$U_{own} = P_g(w_1 + B) + (1 - P_g)\max\{\kappa w_1 + B,\ X\}, \qquad U_{AI} = s(w_1 + B) + (1-s)\max\{p\, w_1 + B,\ X\}.$$

Three identities drive the comparison. First, $\kappa = s^2$ and $(1-s)p = P_g - s + s^2(1 - P_g)$, so that

$$P_g + (1 - P_g)\kappa \ = \ s + (1-s)p \ = \ \Pr[x \le \alpha z \text{ or } x \le s]: \tag{IA.5.2}$$

[10] Since x and z are independent, $\Pr[x \le \alpha z,\ x > \sqrt{\kappa}] = \int_0^1 \max\{0, \min(\alpha z, 1) - \sqrt{\kappa}\}\, dz = \frac{(1-\sqrt{\kappa})(2\alpha - 1 - \sqrt{\kappa})}{2\alpha}$; dividing by $\Pr(x > \sqrt{\kappa}) = 1 - \sqrt{\kappa}$ gives (IA.5.1). Plans that waste the verification slot are dominated and are ignored throughout: implementing an unverified assessment without consulting anyone forgoes a free second opinion, and asking the board outright yields X below, which is weakly dominated by the AI-first plan with a board fallback.

if neither plan involves the board, both deliver the union success probability and $U_{own} = U_{AI}$—sequencing matters only through *which* assessment the board's alternative displaces. Second, if both failure branches use the board,

$$U_{own} - U_{AI} = (P_g - s)\big(w_1 + B - X\big) = (P_g - s)(\eta w_1 + \pi B). \tag{IA.5.3}$$

Third, $p - \kappa = 1 - \frac{1+s}{2\alpha} - s^2$ is strictly decreasing in s and equals zero at $s = P_g$ (there, $p = 1 - (1 + P_g)(1 - P_g) = P_g^2 = \kappa$); hence $p \geq \kappa$ for $s \leq P_g$ and $p < \kappa$ for $s > P_g$.

(i) Let $s \leq P_g$ and $X \geq \kappa w_1 + B$. If also $X \geq p\, w_1 + B$, both branches use the board and (IA.5.3) gives $U_{own} - U_{AI} \geq 0$, strict when $s < P_g$ and $\eta w_1 + \pi B > 0$. If instead $X < p\, w_1 + B$ (possible since $p \geq \kappa$), the assess-first bad type still uses the board while the AI-first plan falls back on the own assessment; substituting $(1 - s)p$ from above,

$$U_{own} - U_{AI} = (1 - P_g)\big(X - \kappa w_1 - B\big) \geq 0,$$

with equality if and only if the help IC constraint binds. (ii) Let $s > P_g$. If the help IC constraint holds, then $p < \kappa$ implies $X \geq \kappa w_1 + B > p\, w_1 + B$, so both branches use the board and (IA.5.3) gives $U_{AI} - U_{own} = (s - P_g)(\eta w_1 + \pi B) \geq 0$, strict when $\eta w_1 + \pi B > 0$. If the help IC constraint fails, the assess-first bad type prefers the AI agent and U_{own} equals the union value (IA.5.2), while $U_{AI} \geq s(w_1 + B) + (1 - s)(p\, w_1 + B)$ equals the same union value; hence $U_{AI} \geq U_{own}$. □

Two readings of Lemma IA.3 are worth making explicit. Part (i) is the formal content of the main text's claim that Assumption 1 makes the assess-first timing without loss of generality; the indifference at a binding help IC constraint is resolved, as elsewhere in the paper, in favor of the shareholders' preferred plan. Part (ii) shows that the failure of Assumption 1 cannot be contracted away: no governance structure makes assess-first strictly optimal—the best any contract achieves is indifference, and only by inducing the CEO to bypass the board entirely.

IA.5.2 Optimal Governance without Assumption 1

Suppose $s > P_g$. Assumption 2 then pins down the CEO's fallback after an AI failure: $s > P_g$ is equivalent to $1 - s < \frac{1}{2\alpha}$, so

$$1 - \kappa = (1-s)(1+s) < \frac{1+s}{2\alpha} \qquad \Longrightarrow \qquad 1 - \eta \;>\; 1 - \frac{1+s}{2\alpha} \;=\; p, \tag{IA.5.4}$$

where the implication uses $\eta < 1 - \kappa$ (Assumption 2): absent firing risk, the board is a better fallback than the CEO's own unverified assessment. Define the per-period success probability under the AI-first plan with a board fallback,

$$Q := s + (1-s)(1-\eta) = 1 - \eta(1 - \sqrt{\kappa}). \tag{IA.5.5}$$

Proposition IA.6 (Without Assumption 1, governance is irrelevant). *Let $\sqrt{\kappa} > 1 - \frac{1}{2\alpha}$, and maintain Assumptions 2–5 and the severance cap with $\bar{s} = 0$. In the shareholders' preferred equilibrium:*

(i) in both periods the CEO queries the AI agent first and verifies its solution; if it fails, she asks the board; output is $y_t = 1$ with probability Q in each period;

(ii) the CEO never learns her type, and $\hat{m}$ is uninformative about z: the board's replacement option has no value, a tough board is indifferent between retaining and replacing any CEO, and $\rho(m_b) = 0$;

(iii) expected profit is $V = 2(Q - c)$, independent of π: board independence is irrelevant;

(iv) $w_1^* = w_2^* = \frac{c}{Q}$, the CEO's expected lifetime utility is $2B$, and her expected total compensation is $2c$.

Proof. Step 1: play at Date (iii). By Lemma IA.3(ii), the CEO adopts the AI-first plan in $t = 1$; given $\rho(m_b) = 0$ (Step 3), asking carries no firing risk and (IA.5.4) makes the board the optimal fallback. In $t = 2$ there are no career concerns and the comparison is purely technological: the AI-first plan with a board fallback succeeds with probability Q, which exceeds the assess-first success probability $P_g + (1 - P_g)(1 - \eta)$ by $(s - P_g)\eta > 0$.

Step 2: no learning. Under the AI-first plan the CEO's information at the board-consultation decision is the event $\{x > s\}$, which is independent of z; her asking decision is measurable with respect to it, so the board's posterior about z equals the prior after either message: $E[z \mid \hat{m}] = \frac{1}{2}$. Moreover, with the board as fallback, output never depends on the CEO's knowledge: $y_t = 1$ if the AI solves the problem ($x \leq s$) or, failing that, if the board does.

Step 3: retention. By Step 2, a retained CEO and a fresh draw are identical to the firm—each succeeds with probability Q at expected wage cost c (Step 4)—so a tough board is indifferent for any ρ. Setting $\rho(m_b) = 0$ is weakly optimal for shareholders: with $\rho(m_b)\pi > 0$, the CEO asks after an AI failure only if $(1 - \eta - p)w_1 \geq \rho\pi B$, which can fail and (weakly) reduce output, whereas $\rho(m_b) = 0$ guarantees asking.

Step 4: wages. Assumption 4 guarantees that effort remains worth inducing: since $s > P_g \geq \frac{1}{2}$, $Q > 1 - \frac{\eta}{2} > 1 - \frac{2}{3}\eta \geq c$. By Step 2, output is independent of z, so information about one's type is worthless and there are no informational rents: in $t = 2$ every employed CEO succeeds with probability Q, and the binding effort IC constraint gives $w_2^* = \frac{c}{Q}$. In $t = 1$, a shirker finds no deal, is retained, and works in $t = 2$ at w_2^* for zero surplus; her lifetime payoff is $2B$. Exerting effort yields $Qw_1 - c + B + B$, so the effort IC constraint is $Qw_1 \geq c$ and the board offers $w_1^* = \frac{c}{Q}$.

Step 5: profit and utility. Per-period profit is $Q(1 - w_t^*) = Q - c$, so $V = 2(Q - c)$, independent of π. Both effort IC constraints bind, so the CEO earns B per period; expected compensation is $Qw_1^* + Qw_2^* = 2c$. □

The verification-priority threshold. The two regimes connect at $\sqrt{\kappa} = 1 - \frac{1}{2\alpha}$ in an instructive way. Output is continuous: at the threshold, $Q = 1 - \frac{\eta}{2\alpha}$, exactly the main text's per-period expected output (average retained-CEO output in $t = 2$ equals $\frac{2\alpha-1}{2\alpha}\left(1 - \frac{\eta}{3(2\alpha-1)}\right) + \frac{1}{2\alpha}\left(1 - \frac{2\eta}{3}\right) = 1 - \frac{\eta}{2\alpha} = Q$). What changes discretely is the governance side. At an interior optimum of Proposition 3, the binding effort IC constraint gives $w_1^* = \frac{2\alpha c + \pi^* B}{2\alpha - \eta} = \frac{c}{Q} + \frac{\pi^* B}{2\alpha - \eta}$: the main-text bonus is the pure effort-IC wage plus compensation for firing risk, an expected period-1 cost of $Qw_1^* - c = \frac{\pi^* B}{2\alpha}$, while monitoring generates the period-2 se-

lection gain $\frac{\pi^* M(\alpha,\eta)}{2\alpha}$. Netting the two,

$$V^{\text{main}} - V^{\text{AI-first}} = \frac{\pi^*\big(M(\alpha,\eta) - B\big)}{2\alpha} \qquad \text{at } \sqrt{\kappa} = 1 - \frac{1}{2\alpha}. \tag{IA.5.6}$$

Profit is therefore continuous across the threshold in Case 1 of Proposition 1 (where $\pi^* = 0$), and drops discretely by exactly the monitoring surplus under an interior solution or Case 3. Beyond the threshold, $\frac{\partial V}{\partial \sqrt{\kappa}} = 2\eta > 0$. Combined with Propositions 2 and 3, profit is non-monotonic in AI's problem-solving capability: it declines in κ under the interior solution, falls discretely by the monitoring surplus at $\kappa = \left(1 - \frac{1}{2\alpha}\right)^2$, and rises in κ thereafter. AI problem-solving capability is thus most damaging at intermediate levels: strong enough to erode the CEO's incentive to communicate, not yet strong enough to displace her as the object of verification.

The model then has one autonomy threshold per assumption. At $\kappa = 1 - \eta$ (Assumption 2), the AI agent displaces the board as the supplier of advice. At $\sqrt{\kappa} = 1 - \frac{1}{2\alpha}$ (Assumption 1), it displaces the CEO as the object of verification—and with her, the informational content of board communication: beyond this threshold, board independence is irrelevant, incentive pay collapses to the pure effort-IC level ($w_t^* = \frac{c}{Q}$, total expected compensation $2c$), and further improvements in κ raise firm value.

IA.6 When AI capabilities comove

Because advances in AI are likely to shift AI capabilities in the same direction (but perhaps by different magnitudes), the comparative static exercises we have performed thus far have some apparent limitations. In this section, we consider shocks that improve (α, κ, β) simultaneously. Let a denote a latent variable measuring AI's overall capabilities. The interpretation is that a higher a increases each of AI's specific capabilities. The effect of changes in a on the monitoring intensity π^*_{AI} depends on the particular correlation structure among the elements in $(\alpha(a), \kappa(a), \beta(a))$. Define the *relative knowledge advantage* the board has over AI as an advisor as

$$R(a) = \beta(a)(1 - \eta) - \kappa(a). \tag{IA.6.1}$$

Assumption 2 and $\beta(a) \in [1, \overline{\beta}]$ imply $R(a) > 0$. We say that a shock $\epsilon_a > 0$ to the latent variable a is *biased against the board* if $R(a + \epsilon_a) - R(a) < 0$, *biased towards the board* if $R(a + \epsilon_a) - R(a) > 0$, and *neutral* if $R(a + \epsilon_a) - R(a) = 0$. Assuming differentiability, as $\epsilon_a \to 0$, the direction of the bias is determined by the sign of $R'(a)$.

We then have

$$\frac{\partial \pi^*_{AI}}{\partial a} = \frac{2c}{B(2\alpha(a) - 1 + \kappa(a))}\Big(R'(a)\alpha(a) - R(a)\frac{\alpha'(a)(1-\kappa(a)) + \kappa'(a)\alpha(a)}{2\alpha(a) - 1 + \kappa(a)}\Big). \quad \text{(IA.6.2)}$$

Note that since $\alpha'(a)(1 - \kappa(a)) + \kappa'(a)\alpha(a) > 0$, an AI improvement can increase the board's monitoring intensity only if it is *sufficiently* biased towards the board, that is, if

$$R'(a) > R(a)\frac{\alpha'(a)(1-\kappa(a)) + \kappa'(a)\alpha(a)}{\alpha(a)(2\alpha(a) - 1 + \kappa(a))}. \quad \text{(IA.6.3)}$$

To help with interpretation, we rearrange (IA.6.3) and simplify notation to get

$$\varepsilon_{Ra} \geq \frac{|\varepsilon_{w_1\alpha}|}{1-\kappa}\Big((1-\kappa)\varepsilon_{\alpha a} + \kappa\varepsilon_{\kappa a}\Big), \quad \text{(IA.6.4)}$$

where ε_{yx} denotes the elasticity of variable y with respect to parameter x. Because the right-hand side of (IA.6.4) is positive, for π^*_{AI} to increase with advances in AI, we need the board's relative knowledge advantage to be sufficiently responsive to AI technology shocks (i.e., high ε_{Ra}). This condition is harder to obtain when (i) the first-period bonus is very sensitive to AI's skill-augmenting capability (i.e., high $|\varepsilon_{w_1\alpha}|$), (ii) AI's skill-augmenting capability is very sensitive to AI technology shocks (i.e., high $\varepsilon_{\alpha a}$), and (iii) AI's problem-solving capability is very sensitive to AI technology shocks (i.e., high $\varepsilon_{\kappa a}$).

A natural benchmark is to consider changes in a that are *unbiased*, that is, changes in a such that $\varepsilon_{Ra} = 0$. In that case, we have

$$\frac{\partial \pi^*_{AI}}{\partial a} = -2cR(a)\frac{\alpha'(a)(1-\kappa(a)) + \kappa'(a)\alpha(a)}{B(2\alpha(a) - 1 + \kappa(a))^2} < 0. \quad \text{(IA.6.5)}$$

We thus conclude that technological change that improves AI capabilities in an unbiased manner will decrease the board's monitoring intensity. To put it differently, an AI improve-

ment can increase the board's monitoring intensity only if it is sufficiently biased towards the board, that is, if the board's relative advantage over AI as an advisor increases significantly.[11]

To understand the intuition, note that an increase in κ affects the optimal monitoring intensity through two routes. First, it reduces incentives to ask for help for a given first-period bonus. Second, it reduces the optimal first-period bonus, further reducing incentives to ask for help. In contrast, an increase in β increases incentives to ask for help for a given bonus, but does not affect the optimal bonus (note that the optimal bonus is independent of η as well). Therefore, a shock of similar magnitude to both κ and $\beta(1-\eta)$ would not change incentives to ask for help for a given first-period bonus, but would still lower the bonus, tightening the help IC constraint. The net effect on the monitoring intensity is thus negative. In addition, if the same shock also increases α, the first-period bonus is further reduced, further contributing to the tightening of the help IC constraint.

IA.7 Long-Run Effects of Board AI Adoption: Proofs of Propositions 5 and 6

Throughout this section, $\varphi = 0$, Assumptions 1–6 hold with $\bar{s} = 0$ (and Assumption 4 strictly, $q_{2b} > c$), and $\beta(1-\eta) > \kappa$. Write $\eta_b := 1 - \beta(1-\eta)$ for the AI-assisted board's failure probability, $P_g := 1 - \frac{1}{2\alpha}$ and $p_b := \frac{1}{2\alpha}$ for the probabilities that a CEO who exerts effort is ex post good or bad, and

$$q_{2b} := 1 - \tfrac{2\eta_b}{3}, \qquad q_{2g} := 1 - \tfrac{\eta_b}{3(2\alpha-1)}, \qquad q_{2n} := 1 - \tfrac{\eta_b}{2\alpha}$$

for the second-period success probabilities of a retained bad CEO, a retained good CEO, and a replacement, with continuation profits $v_{2\tau} := q_{2\tau} - c$ and $M := v_{2n} - v_{2b} = \eta_b\big(\frac{2}{3} - \frac{1}{2\alpha}\big)$. We use three identities repeatedly. First, $P_g q_{2g} + p_b q_{2b} = q_{2n}$: a retained CEO of

[11] A natural assumption is that technological improvements measured by a have decreasing marginal returns on skill-augmentation (for both the CEO and the board). Under that assumption, because $\beta(1-\eta) > \kappa$, an increase in a benefits the CEO more than the board, implying $R'(a) < 0$.

unknown type has the same expected second-period success probability as a fresh draw. It implies both $P_g v_{2g} + p_b v_{2b} = v_{2n}$ and $P_g(v_{2g} - v_{2n}) = p_b M$. Second, $q_{2g} - q_{2b} = \frac{2\alpha}{2\alpha - 1} M$. Finally, let $\overline{\mu} := \max\{1 - 2\alpha c,\, 0\}$.

The inspection technology is as in the main text: at the beginning of date (iv) of $t = 1$, with probability λ the board learns the deal and whether $z \geq x$, conditional on a deal existing. It can then assess the deal exactly as if the CEO had asked for help, and a tough board replaces a CEO it has learned to be bad. A CEO who exerts no effort sources no deal and generates nothing to inspect. Call a first-period CEO *vetted* if the board inspected her deal and learned $z \geq x$, and *silent* if she neither asked for help nor was inspected.

Fix a governance structure (w_1, π) with $w_1 > 0$ that violates the help IC constraint,

$$\pi B \;>\; \big(\beta(1-\eta) - \kappa\big)\, w_1, \tag{IA.7.1}$$

and induces first-period effort. (The optimal structure of Proposition 5 has $w_1 = 0$; it is treated as the limit $w_1 \downarrow 0$ of what follows, which is innocuous because all payoffs below are continuous in w_1; see the Remark at the end of this section.) As in the proof of Proposition 1, let $\mu := 1 - p_1(b)$ denote the probability that a bad CEO conceals and $\rho := \rho(m_c)$ the probability that a tough board replaces a silent CEO.

Lemma IA.4 (Equilibrium without the help IC constraint). *Under (IA.7.1) and for $\lambda < 1$, no equilibrium in pure strategies exists. In the semi-separating equilibrium:*

(i) The concealment probability $\mu^* \in (0,1)$ makes a tough board indifferent between retaining and replacing a silent CEO; it depends on second-period parameters only—not on (w_1, π, λ). When the board's preferred contract for a retained CEO of unknown type induces effort from both types (the case $q_{2b} v_{2b} \geq 2\alpha c\, v_{2n}$, where $r_g := c\,\frac{q_{2g} - q_{2b}}{q_{2b}}$ is the good type's rent under the bonus $\frac{c}{q_{2b}}$),

$$\mu^* = 1 - \frac{2\alpha c}{q_{2b}}; \tag{IA.7.2}$$

otherwise $\mu^* = \frac{M}{v_{2n}}$. Equivalently, μ^* is the larger of the two expressions.

(ii) A tough board replaces a silent CEO with probability $\rho^* = 1 - \frac{(\beta(1-\eta) - \kappa) w_1}{\pi B} \in (0,1)$.

(iii) At μ^*, the firm's second-period value of a retained silent CEO equals v_{2n}. Expected profit is therefore

$$\begin{aligned} V_R(w_1,\pi) = S_x(1-w_1) + \lambda P_g v_{2g} + (1-\lambda)\big(P_g + p_b\mu^*\big)v_{2n} \\ + p_b\big(1-\mu^*(1-\lambda)\big)\big[\pi v_{2n} + (1-\pi)v_{2b}\big], \end{aligned} \tag{IA.7.3}$$

where

$$S_x := P_g + p_b\big[\big(1-\mu^*(1-\lambda)\big)\beta(1-\eta) + \mu^*(1-\lambda)\kappa\big]$$

is the probability of first-period success. In particular, V_R does not depend on ρ, is strictly increasing in π, and is strictly decreasing in w_1.

Proof. No pure-strategy equilibrium. A bad CEO who asks obtains $\beta(1-\eta)w_1 + (1-\pi)B$. If she conceals, then with probability λ she is inspected, the board assesses the deal, and a tough board replaces her—exactly the payoff from asking—and with probability $1-\lambda$ she is silent, succeeds with probability κ, and is replaced only if the board is tough and replaces silent CEOs. Suppose $\mu = 0$. Then all silent CEOs are good; the board retains them ($\rho = 0$), since $v_{2g} > v_{2n}$; and concealing pays $\lambda\big[\beta(1-\eta)w_1 + (1-\pi)B\big] + (1-\lambda)\big[\kappa w_1 + B\big]$. Concealing beats asking by $(1-\lambda)\big[\pi B - (\beta(1-\eta)-\kappa)w_1\big] > 0$ by (IA.7.1), a contradiction. Suppose $\mu = 1$. Then the silent pool has the composition of the candidate pool, so by the first identity a retained silent CEO produces in expectation exactly as much as a replacement—but costs strictly more. If the board offers the pooling bonus $\frac{c}{q_{2b}}$, both types work and the expected value per silent CEO is $P_g(v_{2g} - r_g) + p_b v_{2b} = v_{2n} - P_g r_g < v_{2n}$; if it offers $\frac{c}{q_{2g}}$, the bad type shirks and the value is $P_g v_{2g} = v_{2n} - p_b v_{2b} < v_{2n}$ (using both parts of the first identity, and $r_g > 0$, $v_{2b} > 0$). A tough board therefore replaces every silent CEO ($\rho = 1$), silence provides no protection, and asking beats concealing by $(1-\lambda)\big(\beta(1-\eta)-\kappa\big)w_1 > 0$, a contradiction.

(ii) In a mixed equilibrium the bad CEO is indifferent. The inspected branches of the two actions coincide and cancel, so indifference requires $\kappa w_1 + (1-\pi\rho)B = \beta(1-\eta)w_1 + (1-\pi)B$, i.e., $\pi(1-\rho)B = \big(\beta(1-\eta)-\kappa\big)w_1$, which is (ii); interiority follows from (IA.7.1) and $w_1 > 0$. (A good CEO strictly prefers silence: her deal succeeds whether or not she asks, and asking raises her replacement probability from $(1-\lambda)\pi\rho$ to π, so silence is worth

at least $\pi B\big[1-(1-\lambda)\rho\big] > 0$ to her.)

(i) For the tough board to mix, it must be indifferent about silent CEOs. Inspection reaches good and bad CEOs at the same rate, so the silent pool consists of good CEOs and concealing bad CEOs in proportions P_g to $p_b\mu$, independently of λ. Let $R(\mu)$ denote the expected second-period value per retained silent CEO under the board's preferred bonus. R is continuous and strictly decreasing, $R(0) = v_{2g} > v_{2n}$, and $R(1) < v_{2n}$ (shown above), so a unique $\mu^* \in (0,1)$ solves $R(\mu^*) = v_{2n}$. If the pooling bonus is preferred at μ^*, the indifference condition is $P_g(v_{2g} - r_g) + p_b\mu v_{2b} = (P_g + p_b\mu)v_{2n}$, i.e., $P_g(v_{2g} - v_{2n}) - P_g r_g = p_b\mu(v_{2n} - v_{2b})$; using $P_g(v_{2g} - v_{2n}) = p_b M$ this reads $p_b M - P_g r_g = p_b\mu M$, so $\mu^* = 1 - \frac{P_g r_g}{p_b M}$, and by the second identity $\frac{P_g r_g}{p_b M} = (2\alpha - 1)\frac{c}{q_{2b}} \cdot \frac{q_{2g} - q_{2b}}{M} = \frac{2\alpha c}{q_{2b}}$, which gives (IA.7.2). If instead the bonus $\frac{c}{q_{2g}}$ is preferred, indifference is $P_g v_{2g} = (P_g + p_b\mu)v_{2n}$, giving $\mu^* = \frac{P_g(v_{2g} - v_{2n})}{p_b v_{2n}} = \frac{M}{v_{2n}}$. The pooling bonus is preferred at μ^* exactly when $p_b\mu^* v_{2b} \geq P_g r_g$; substituting either candidate reduces this to $q_{2b} v_{2b} \geq 2\alpha c\, v_{2n}$, so the two cases partition the parameter space and μ^* is the larger of the two expressions.

(iii) Count the contributions to expected profit. First period: output is 1 for good CEOs, $\beta(1-\eta)$ for bad CEOs who ask or are inspected, and κ for silent bad CEOs, giving success probability S_x; the bonus w_1 is paid upon success, so the first period contributes $S_x(1 - w_1)$. Second period: vetted good CEOs (mass λP_g) are retained and contribute v_{2g}; bad CEOs who asked or were inspected (mass $p_b(1 - \mu^*(1-\lambda))$) contribute $\pi v_{2n} + (1-\pi)v_{2b}$; and each silent CEO (mass $(1-\lambda)(P_g + p_b\mu^*)$) contributes exactly v_{2n}, whether replaced (a fresh draw is worth v_{2n}) or retained (the board's indifference at μ^* values retention at v_{2n}, inclusive of the rents its chosen bonus concedes). Summing gives (IA.7.3); ρ appears nowhere. Monotonicity: $\frac{\partial V_R}{\partial \pi} = p_b\big(1 - \mu^*(1-\lambda)\big)M > 0$ and $\frac{\partial V_R}{\partial w_1} = -S_x < 0$. □

Proof of Proposition 5. Let V^c denote the value of the optimal governance structure satisfying the help IC constraint. As shown in the text, V^c does not depend on λ. Equilibria without first-period effort yield at most v_{2n} and are dominated: the structure $(w_1, \pi) = \big(\frac{c}{S}, 0\big)$, where $S := 1 - \frac{\eta_b}{2\alpha}$, satisfies the help IC and effort IC constraints and yields $S - c + P_g v_{2g} + p_b v_{2b} = S - c + v_{2n} > v_{2n}$, using the first identity and $S \geq q_{2b} > c$. It therefore suffices to compare V^c with the values attainable in the region (IA.7.1).

Step 1: the best structure in the region is $(w_1, \pi) = (0, 1)$. By Lemma IA.4(iii), V_R is strictly increasing in π and strictly decreasing in w_1, so for every structure in the region,

$$V_R(w_1, \pi) \;\le\; V^\circ(\lambda) \;:=\; S_x + \lambda P_g v_{2g} + (1-\lambda) P_g v_{2n} + p_b v_{2n},$$

the value of (IA.7.3) at $(0, 1)$ (after simplifying the v_{2n} terms). The bound is attained in the limit $w_1 \downarrow 0$ at $\pi = 1$ provided effort is incentive-compatible there. At $(w_1, \pi) = (0, 1)$ with $\rho^* \to 1$, a CEO who exerts effort is retained only if vetted, which occurs with probability λP_g, while a shirker sources no deal, is silent with certainty, and is replaced; effort is therefore optimal if and only if

$$-c + B + \lambda P_g B \;\ge\; B \qquad \Longleftrightarrow \qquad \lambda \;\ge\; \lambda_0 := \frac{2\alpha c}{(2\alpha - 1)B},$$

and $\lambda_0 < 1$ because $B > 2c$ (Assumption 5) implies $(2\alpha - 1)B > (2\alpha - 1)2c \ge 2\alpha c$ for $\alpha \ge 1$.

Step 2: V° is strictly increasing, and $V^\circ(1) > V^c$. Differentiating, $\frac{dV^\circ}{d\lambda} = p_b \mu^* \big(\beta(1-\eta) - \kappa\big) + P_g(v_{2g} - v_{2n}) > 0$ (both inspection margins are valuable: a silent bad CEO gets the board's assessment instead of the AI agent's, and a good CEO is vetted rather than exposed to the purge). At $\lambda = 1$ there are no silent CEOs: $S_x = S$ and $V^\circ(1) = S + P_g v_{2g} + p_b v_{2n}$. Any structure (w_1, π) satisfying the help IC constraint and inducing effort has $w_1 > 0$ (with $w_1 = 0$ a shirker, who is silent and believed good, retains office in both periods and obtains $2B$, strictly more than the $-c + 2B - \pi p_b B$ from exerting effort) and value $S(1 - w_1) + P_g v_{2g} + p_b\big[\pi v_{2n} + (1-\pi) v_{2b}\big]$, so

$$V^\circ(1) - V(w_1, \pi) \;=\; S\, w_1 + p_b(1-\pi)M \;>\; 0.$$

Hence $V^\circ(1) > V^c$.

Step 3: the threshold. Define $\overline{\lambda} := \inf\{\lambda \in [\lambda_0, 1] : V^\circ(\lambda) \ge V^c\}$, which is well defined and smaller than 1 by Step 2 and continuity. For every $\lambda > \overline{\lambda}$, Step 2 gives $V^\circ(\lambda) > V^c$, and by Step 1 the optimal governance structure is $(w_1, \pi) = (0, 1)$ in the region (IA.7.1), with the equilibrium of Lemma IA.4: good CEOs silent, bad CEOs silent with probability μ^*, and, as $w_1 \downarrow 0$, $\rho^* \to 1$: the board replaces every CEO it has not vetted.

Step 4: pay and turnover. At the optimum the first-period bonus is zero, and every second-period bonus $w_{2\tau} = \frac{c}{q_{2\tau}}$ pays c in expectation to whoever is employed, so expected total CEO pay is c. Retained CEOs are exactly the vetted good ones, of mass λP_g; turnover is $1 - \lambda P_g = 1 - \lambda\big(1 - \frac{1}{2\alpha}\big)$, of which the bad CEOs account for $p_b = \frac{1}{2\alpha}$ (all of them: they ask, are caught, or are purged) and good CEOs for $P_g(1-\lambda)$. □

Proof of Proposition 6. All limits below are as $\eta_b \to 0$; every object is a rational function of η_b whose denominator does not vanish at $\eta_b = 0$, hence continuous there, so each strict limit inequality holds for $\beta(1-\eta)$ close enough to 1, which is the sense of the statement. In the limit, $q_{2b}, q_{2g}, q_{2n} \to 1$, $v_{2b}, v_{2g}, v_{2n} \to 1 - c$, $M \to 0$, and $r_g \to 0$. Since $q_{2b} v_{2b} \to 1 - c$ and $2\alpha c\, v_{2n} \to 2\alpha c(1-c)$, the case of (IA.7.2) applies in the limit exactly when $2\alpha c < 1$, and $\mu^* \to 1 - 2\alpha c$; if $2\alpha c \geq 1$, the other case applies and $\mu^* = \frac{M}{v_{2n}} \to 0$. In both cases $\mu^* \to \overline{\mu}$.

Step 1: the two values in the limit. For $\beta > \overline{\beta}$, the optimal structure satisfying the help IC constraint is the corner of Case 1, with $\pi^*_{AI} = 0$ and $w^*_{1AI} = \frac{2\alpha c}{2\alpha - \eta_b}$; the help IC constraint holds strictly at $\pi = 0$, so every bad CEO asks and is advised, and no CEO is replaced. Its value is $S(1 - w^*_{1AI}) + P_g v_{2g} + p_b v_{2b} = S(1 - w^*_{1AI}) + v_{2n}$, and since $S \to 1$ and $w^*_{1AI} \to c$,

$$V^c \;\to\; V^{ent} := 2 - 2c,$$

with expected CEO pay $S\, w^*_{1AI} + c \to 2c$. In the region (IA.7.1), the effort IC constraint at monitoring intensity π pins the minimal bonus. A direct computation of the effort IC constraint of Lemma IA.4 in the limit (all second-period rents vanish) gives the effort-minimal bonus

$$w_E(\pi, \lambda) \;=\; \max\Big\{0,\; \frac{c - \pi\lambda P_g B}{1 - (1-\kappa)\big(p_b + P_g\lambda\big)}\Big\},$$

whose numerator is decreasing in π (for $\lambda > 0$) and whose denominator is positive for $\lambda < 1$, exceeding $1 - (p_b + P_g\lambda) = P_g(1-\lambda) > 0$; at $\lambda = 1$ the constraint reads $-c + \pi P_g B \geq 0$ directly, so $w_E(1,1) = 0$ because $P_g B > c$ ($\lambda_0 < 1$). Because V_R is decreasing in w_1, increasing in π, and w_E is decreasing in π, the attainable optimum in the region is at $\pi = 1$, $w_1 = w_E(1,\lambda)$, which lies in the region (IA.7.1): $(1-\kappa)w_E(1,\lambda) < B$ reduces to the λ-free inequality $(1-\kappa)c < B\big[1 - (1-\kappa)p_b\big]$, which holds because $B\big[1 - (1-\kappa)p_b\big] \geq$

$B(1-p_b) = P_g B > c \geq (1-\kappa)c$. Thus

$$V^{reg}(\lambda) \;=\; S_x(\lambda)\big(1 - w_E(1,\lambda)\big) + (1-c), \qquad \text{with} \quad S_x(\lambda) = 1 - p_b\overline{\mu}(1-\lambda)(1-\kappa),$$

using that all second-period terms of (IA.7.3) converge to $1-c$. Note $w_E(1,\lambda) = 0$ exactly for $\lambda \geq \lambda_0 := \frac{2\alpha c}{(2\alpha-1)B}$, where $V^{reg} = V^{\circ}$ of Proposition 5.

Step 2: single crossing. V^{reg} is continuous and nondecreasing in λ: S_x is nondecreasing, and $w_E(1,\cdot)$ is nonincreasing, since

$$\begin{aligned}\frac{\partial}{\partial\lambda}\Big[\frac{c - \lambda P_g B}{1-(1-\kappa)(p_b + P_g\lambda)}\Big] \;&\propto\; (1-\kappa)c - B\big[1-(1-\kappa)p_b\big] \\ &<\; (1-\kappa)c - 2c(1-p_b) \;\leq\; -\kappa c \;\leq\; 0,\end{aligned}$$

where the first inequality is strict because $B > 2c$ (Assumption 5) and $1-(1-\kappa)p_b \geq 1-p_b > 0$, and the second uses $p_b \leq \frac{1}{2}$; hence $w_E(1,\cdot)$ is strictly decreasing wherever positive, and V^{reg} is strictly increasing wherever $w_E(1,\lambda) > 0$ or $\overline{\mu} > 0$. At the endpoints,

$$\begin{aligned}V^{reg}(0) - V^{ent} \;&=\; c - p_b\overline{\mu}(1-\kappa) - S_x(0)\, w_E(1,0) \\ &\leq\; c - p_b\overline{\mu}(1-\kappa) - \frac{S_x(0)}{1-(1-\kappa)p_b}\, c \\ &=\; -p_b\overline{\mu}(1-\kappa) - \Big[\frac{S_x(0)}{1-(1-\kappa)p_b} - 1\Big]c \;<\; 0,\end{aligned}$$

because $S_x(0) = 1 - p_b\overline{\mu}(1-\kappa) \geq 1-(1-\kappa)p_b$ makes the bracketed term nonnegative and, when $\overline{\mu} = 0$, strictly positive (as $S_x(0) = 1 > 1-(1-\kappa)p_b$, using $\kappa < 1$ and $p_b > 0$); and

$$V^{reg}(1) \;=\; 2 - c \;>\; V^{ent}.$$

Hence there is a unique $\hat{\lambda} \in (0,1)$ with $V^{reg}(\lambda) < V^{ent}$ for $\lambda < \hat{\lambda}$ and $V^{reg}(\lambda) > V^{ent}$ for $\lambda > \hat{\lambda}$.

Step 3: the optimal regime. Fix $\lambda < \hat{\lambda}$. Every structure in the region (IA.7.1) has value at most $V^{reg}(\lambda) < V^{ent}$ (by the monotonicities of V_R and the definition of w_E, whether or not the bounding structure is itself in the region), and equilibria without effort are dominated

as in the proof of Proposition 5; so for $\beta(1-\eta)$ close enough to 1 the optimal structure is the entrenchment corner, with $\pi^*_{AI} = 0$. Fix $\lambda > \hat{\lambda}$. Then $V^{reg}(\lambda) > V^{ent} \geq$ the value of every structure satisfying the help IC constraint (whose optimum is the entrenchment corner for $\beta > \overline{\beta}$), so for $\beta(1-\eta)$ close enough to 1 the optimal structure is in the region, at $\pi^*_{AI} = 1$ and $w^*_{1AI} = w_E(1,\lambda)$, with the equilibrium of Lemma IA.4: vet-or-out governance, with a bonus that is strictly positive for $\lambda < \lambda_0$ and zero for $\lambda \geq \lambda_0$.

Step 4: bounds, payoffs, and the distributional limit. For $\lambda \geq \lambda_0$, $V^{reg}(\lambda) = 2 - c - \frac{\overline{\mu}(1-\kappa)(1-\lambda)}{2\alpha}$, so $V^{reg}(\lambda) > V^{ent}$ if and only if $\lambda > \overline{\lambda}_\infty := 1 - \frac{2\alpha c}{\overline{\mu}(1-\kappa)}$ (with $\overline{\lambda}_\infty := -\infty$ if $\overline{\mu} = 0$); for $\lambda < \lambda_0$, $V^{reg}(\lambda) \leq V^\circ(\lambda)$, which is below V^{ent} for $\lambda < \overline{\lambda}_\infty$. Hence

$$\max\{0, \overline{\lambda}_\infty\} \;\leq\; \hat{\lambda} \;\leq\; \max\{\overline{\lambda}_\infty, \lambda_0\},$$

and $\hat{\lambda} = \overline{\lambda}_\infty$ whenever $\overline{\lambda}_\infty \geq \lambda_0$, in which case the bonus is already zero at the switch. Turnover, the unadvised fraction of bad CEOs, and expected pay for $\lambda \geq \lambda_0$ are as in Proposition 5 (with $\mu^* \to \overline{\mu}$). Just above $\hat{\lambda}$ when $\hat{\lambda} < \lambda_0$, the bonus $w_E(1,\lambda)$ is still positive and a tough board replaces silent CEOs only with probability $\rho^* < 1$ (Lemma IA.4), so turnover is lower by $(1-\lambda)\big(P_g + p_b\overline{\mu}\big)(1-\rho^*)$, expected pay is $c + S_x w_E(1,\lambda)$, and the unadvised fraction of bad CEOs is unchanged at $\overline{\mu}(1-\lambda)$. Finally, as $\lambda \to 1$: output converges to 1 per period under both regimes ($S \to 1$, $S_x \to 1$, and all second-period success probabilities converge to 1), while profits converge to $2-2c$ and $2-c$. Total CEO surplus under entrenchment is the incumbent's utility, $2B$ (her effort IC constraint binds, and a shirker retains office in both periods); under vet-or-out governance the incumbent obtains $-c + B + P_g B$ and her successors, of mass p_b, obtain B each, for a total of $2B - c$. The firm gains exactly the c that CEOs lose. □

Remark IA.5 (The zero-bonus limit). At $w_1 = 0$ exactly, the bad CEO is indifferent among her actions and the equilibrium of Lemma IA.4 is a limit rather than strictly unique: any $\mu \geq \mu^*$ together with $\rho = 1$ is also an equilibrium. As in the corresponding discussion in the proof of Proposition 1, an infinitesimal bonus $\varepsilon > 0$ makes the semi-separating equilibrium (μ^*, ρ^*) with $\rho^* = 1 - \frac{(\beta(1-\eta)-\kappa)\varepsilon}{\pi B}$ the unique outcome and changes all values by at most $O(\varepsilon)$; the statements of Propositions 5 and 6 are understood in this sense.

References

BERK, J. B. AND J. H. VAN BINSBERGEN (2022): "Regulation of Charlatans in High-Skill Professions," *Journal of Finance*, 77, 1219–1258.

COWGILL, B., P. HERNANDEZ-LAGOS, AND N. L. WRIGHT (2024): "Does AI cheapen talk? Theory and evidence from global entrepreneurship and hiring," *Theory and Evidence From Global Entrepreneurship and Hiring (July 26, 2024). Columbia Business School Research Paper*.

LAMBA, R. (2026): "The Verification Hill," Working paper, Cornell University.