Who Should Do What?
Adaptive Delegation in Human-AI Collaboration

Wei Gu[1] **Michael Lingzhi Li**[2] **Shixiang Zhu**[1]
[1]Carnegie Mellon University [2]Harvard Business School
{weigu, shixianz}@andrew.cmu.edu mili@hbs.edu

Abstract

As human-AI collaboration becomes increasingly common in real-world decision-making systems, it is essential to develop principled frameworks for deciding who should act and when: the AI, the human, or both. In this paper, we develop optimal delegation strategies for settings where human oversight adds value but comes at a cost. We propose an adaptive delegation framework in which a central coordinator assigns each task to either the AI, the human, or a human-in-the-loop review process. Importantly, we model the human as a cost-sensitive and adaptive agent, whose effort adapts based on the AI's accuracy. This interaction is formalized using a generalized Nash equilibrium framework, which allows us to characterize stable collaboration strategies under broad conditions. We provide theoretical guarantees that identify when adaptive delegation enables effective cooperation and how human-AI collaboration evolves according to the development of AI technologies. Numerical experiments confirm that our approach improves overall system performance under accuracy and cost constraints. These results offer practical guidance for designing agentic AI systems that balance efficiency with meaningful human involvement.

1 Introduction

AI systems are increasingly embedded in high-stakes domains such as healthcare, finance, energy, and security. In these settings, automation alone is rarely sufficient: while AI offers scale, speed, and consistency, human experts bring contextual judgment and domain expertise. The challenge is no longer whether to integrate humans and AI, but how to structure this collaboration effectively.

Today's systems often adopt rigid decision pipelines: either full automation with minimal oversight, or mandatory human review regardless of context. Such static delegation misses the opportunity to leverage the complementary strengths of humans and AI [2, 11]. In practice, the right decision-making structure varies with the task. A physician may want AI assistance for routine diagnoses but prefer to take control when uncertainty is high. A compliance officer may rely on AI for standard checks, but override it when unusual patterns arise. Ideally, an intelligent system would dynamically determine who should act, and when—balancing accuracy, cost, and human engagement.

In this paper, we propose an adaptive delegation framework for coordinator agents that govern human-AI collaboration at a granular level. For each task, the coordinator decides whether to defer to the AI, hand off the task to a human, or present an AI-generated recommendation for human review. Crucially, our model treats the human not as a passive oracle, but as an adaptive, cost-sensitive agent whose effort shifts depending on both the AI's performance and the structure of delegation over time. Empirical evidence underscores the importance of this modeling choice: studies of automation bias show that humans often reduce vigilance and become over-reliant on AI, particularly when it is perceived as accurate or dependable [12, 13]. Because human behavior evolves in response to the

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

AI's performance, this creates a feedback loop in the system: the human adapts their effort based on perceived AI reliability, and the coordinator agent updates its delegation policy based on the human's observed behavior.

We formalize this interaction using a generalized Nash equilibrium framework, allowing us to analyze how human and AI strategies co-evolve. We derive theoretical conditions on how the adaptive delegation evolves and identify when cooperation between human and AI is sustainable. Empirical results validate our model's ability to improve overall system utility under cost and accuracy constraints. Our findings offer insight into how to design agentic AI systems that not only work with humans, but also learn how to work through them.

Related Work The proposed research builds on and significantly extends the literature on human-AI collaboration and decision delegation. A central theme in prior work is the concept of learning to defer or learning to reject, where a predictive model not only produces decisions but also learns when to defer to a human expert [4, 7, 8, 14]. These models typically optimize a joint loss over the predictive accuracy and deferral behavior, assuming that the human response is fixed and independent of the AI's policy. More recent work has explored two-stage frameworks that separately learn an AI model and a deferral policy [5], though these remain largely theoretical. Our work differs in a fundamental way: we treat the human as a rational agent whose behavior is influenced by the coordinator agent, creating a strategic feedback loop. The most closely related conceptual model is the principal-agent framework introduced in [6], which considers adaptive human responses to algorithmic behavior. However, their work is limited to binary actions and static, single-shot interactions. In contrast, we model the coordinator-human-AI interaction as a dynamic and general decision process.

2 Problem Setup

We consider a decision-making system involving three agents: a pre-trained AI policy, a cost-sensitive human expert, and a coordinator agent that governs how decisions are delegated. In this system, the coordinator determines, for each given task, whether the decision should be generated by AI (`A`), made independently by the human without AI input (`H`), or presented to the human along with the AI's recommendation (`R`). When the coordinator suggests that the AI's recommendation should be presented to the human (i.e., the case `R`), the human agent will make a decision on whether to modify it (`M`) or not (`D`). This distinction between `H` and `R` gives the coordinator nuanced control: it can withhold the AI's influence when it is beneficial to maintain human engagement or avoid bias. As a concrete example, consider an email triage system for a busy executive. Routine messages may be delegated directly to the AI (`A`); legal or sensitive messages may be routed to the human without AI input (`H`); and ambiguous cases, such as media inquiries, may be shown with the AI's draft but left for human confirmation (`R`). This structure enables the system to adaptively balance efficiency, accuracy, and human oversight.

3 Static Model

Consider that the coordinator implements a mixed strategy (randomized policy) with probabilities $p_{\mathtt{A}}$, $p_{\mathtt{H}}$, and $p_{\mathtt{R}}$ for decisions `A`, `H`, and `R`, respectively. Denote $\boldsymbol{p} \triangleq \{p_{\mathtt{A}}, p_{\mathtt{H}}, p_{\mathtt{R}}\}$ as the vector of probabilities representing this strategy. With the coordinator's recommendation `R`, the human agent implements a mixed strategy with a probability $q_{\mathtt{M}}$ to modify AI's decision, and a probability $q_{\mathtt{D}}$ not to modify it. Denote $\boldsymbol{q} \triangleq \{q_{\mathtt{M}}, q_{\mathtt{D}}\}$. Let $r_{\mathtt{A}}$ and $r_{\mathtt{H}}$ be the system reward for AI and human expert to solely handle a given task, respectively. Notate $c_{\mathtt{H}}$ as the cost for the human expert to deal with a task without AI input. Denote $r_{\mathtt{M}}$ and $c_{\mathtt{M}}$ as the system reward and the cost for the human expert to modify the solution generated by AI, respectively. Let b represent the cost budget of the human expert, and notate δ as the minimal target level of system reward. Notations used in the static model are summarized in Table 1 in Appendix A.

Given the human expert's best strategy $\boldsymbol{q}$, the **coordinator** aims to decide the best mixed strategy to maximize the system reward, subject to the human expert's cost budget constraint:

$$\begin{aligned} \max_{\boldsymbol{p}\geq 0} \quad & r_{\mathtt{H}}p_{\mathtt{H}} + r_{\mathtt{A}}p_{\mathtt{A}} + (r_{\mathtt{M}}q_{\mathtt{M}} + r_{\mathtt{A}}q_{\mathtt{D}})\,p_{\mathtt{R}} \\ \text{s.t.} \quad & c_{\mathtt{H}}p_{\mathtt{H}} + c_{\mathtt{M}}q_{\mathtt{M}}p_{\mathtt{R}} \leq b \\ & p_{\mathtt{A}} + p_{\mathtt{H}} + p_{\mathtt{R}} = 1 \end{aligned} \tag{1}$$

Given the coordinator's policy $\boldsymbol{p}$, the **human expert** wants to decide the best mixed strategy for minimizing his or her own effort cost, subject to maintaining a target level of system reward:

$$\begin{aligned} \min_{\boldsymbol{q} \geq 0} \quad & c_{\mathtt{M}} p_{\mathtt{R}} q_{\mathtt{M}} \\ \text{s.t.} \quad & r_{\mathtt{M}} p_{\mathtt{R}} q_{\mathtt{M}} + r_{\mathtt{A}} p_{\mathtt{R}} q_{\mathtt{D}} + r_{\mathtt{H}} p_{\mathtt{H}} + r_{\mathtt{A}} p_{\mathtt{A}} \geq \delta \\ & q_{\mathtt{M}} + q_{\mathtt{D}} = 1 \end{aligned} \tag{2}$$

The coordinator's problem (1) and the human expert's problem (2) together form a generalized Nash equilibrium [1, 3]. The main difference from the traditional Nash equilibrium [9, 10] is that the decision of each player (coordinator or human expert) will influence not only the other player's objective function, but also the other player's constraint set. As a generalized Nash equilibrium, the proposed model captures the strategic coupling between the human's cost-sensitive actions and the coordinator's reward-maximizing delegation policy. The existence of the generalized Nash equilibrium between the coordinator (1) and the human expert (2) is established in the following proposition, which illustrates when human and AI can cooperate.

Proposition 1 *Under the conditions that: (i) For each* **q** *feasible to the human expert, the constraints in (1) for the coordinator are feasible, and (ii) For every* **p** *feasible to the coordinator, the feasible region in (2) is nonempty, then the generalized Nash game between the coordinator (1) and the human expert (2) must have an equilibrium solution.*

Proposition 1 means that under the mild condition that constraints of both the coordinator (1) and the human expert (2) are feasible, thus, human and AI can cooperate. We can derive the closed-form solution for the static model, as shown in Appendix B.

4 Dynamic Model

We consider T-stage dynamic model, namely the game will be played for T rounds. Since AI could learn from the human expert, different from Section 3, we now consider the reward for AI to handle a task at time $t \in \{0, 1, 2, \cdots, T\}$ as a function, denoted as $r_{\mathtt{A}}^{t}[\bullet]$. As a supplementary of Table 1 in Appendix A, extra notations used in the dynamic model are listed in Table 2 in Appendix C.

The coordinator aims to maximize the total reward over T stages, by deciding the optimal mixed strategy at each time stage $t \in \{0, 1, 2, \cdots, T\}$ subject to the budget constraint:

$$\begin{aligned} \max_{\bar{\boldsymbol{p}} \geq 0} \quad & \sum_{t=0}^{T} \left\{ r_{\mathtt{H}} p_{\mathtt{H}}^{t} + r_{\mathtt{A}}^{t}[\bullet]\, p_{\mathtt{A}}^{t} + \left(r_{\mathtt{M}} q_{\mathtt{M}}^{t} + r_{\mathtt{A}}^{t}[\bullet]\, q_{\mathtt{D}}^{t} \right) p_{\mathtt{R}}^{t} \right\} \\ \text{s.t.} \quad & c_{\mathtt{H}} p_{\mathtt{H}}^{t} + c_{\mathtt{M}} q_{\mathtt{M}}^{t} p_{\mathtt{R}}^{t} \leq b \quad \forall t = 0, 1, 2, \cdots, T \\ & p_{\mathtt{A}}^{t} + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t} = 1 \quad \forall t = 0, 1, 2, \cdots, T \end{aligned} \tag{3}$$

At each time stage $t \in \{0, 1, 2, \cdots, T\}$, the human expert aims to minimize the cost subject to the constraint regarding system performance:

$$\begin{aligned} \min_{\boldsymbol{q}_t \geq 0} \quad & c_{\mathtt{M}} p_{\mathtt{R}}^{t} q_{\mathtt{M}}^{t} \\ \text{s.t.} \quad & r_{\mathtt{M}} p_{\mathtt{R}}^{t} q_{\mathtt{M}}^{t} + r_{\mathtt{A}}^{t}[\bullet]\, p_{\mathtt{R}}^{t} q_{\mathtt{D}}^{t} + r_{\mathtt{H}} p_{\mathtt{H}}^{t} + r_{\mathtt{A}}^{t}[\bullet]\, p_{\mathtt{A}}^{t} \geq \delta \\ & q_{\mathtt{M}}^{t} + q_{\mathtt{D}}^{t} = 1 \end{aligned} \tag{4}$$

Suppose that the reward for AI to handle a task at time $t \in \{1, 2, \cdots, T\}$ as an increasing (or non-decreasing) function of the number of times that the task are dealt with by the human expert before stage t, i.e., $\sum_{s=0}^{t-1} (p_{\mathtt{H}}^{s} + p_{\mathtt{R}}^{s} q_{\mathtt{M}}^{s})$.

Furthermore, assume that the performance of AI will be better than that of the human expert and converge to that of the artificial general intelligence (AGI). We assume AGI performs better than the human expert, i.e., $r_{\mathtt{AGI}} > r_{\mathtt{H}}$, then a nonlinear AI reward function can be defined as follows

$$r_{\mathtt{A}}^{t}\left[\bullet\right] = r_{\mathtt{AGI}} - \frac{\alpha}{1 + \left[\sum_{s=0}^{t-1}\left(p_{\mathtt{H}}^{s} + p_{\mathtt{R}}^{s} q_{\mathtt{M}}^{s}\right)\right]^{\beta}}, \quad \forall t = 1, 2, \cdots, T \tag{5}$$

where α and β are positive parameters. Define $r_{\mathtt{A}}^{0}\left[\bullet\right] \triangleq r_{\mathtt{A}} = r_{\mathtt{AGI}} - \alpha$. The dynamic model can be solved using dynamic programming approach. For more details, please refer to Appendix D.

5 Numerical Experiments

In this section, we validate the proposed models through numerical examples. Figures of the numerical experiments are summarized in Appendix E. Fig. 2 shows the results of the static model, which includes Case 3, Case 4, and Case 5 analyzed in Appendix B. As we can observe from Fig. 2, with the development of AI technologies (i.e., the increase in AI reward), the coordinator will first recommend the human expert to use AI to general an initial solution and then modify it (Case 3 in Appendix B). When the performance of AI reaches the minimal requirement of system performance, the coordinator will recommend fully automated AI decision-making or mandatory human decision (Case 4 in Appendix B). After AI performs no worse than the human expert, the coordinator will always recommend AI to make the decision (Case 5 in Appendix B). At the same time, the human expert will have a nonlinear decrease in the probability of modifying the AI-generated decision as the performance of AI technology improves.

The numerical results of the proposed dynamic model in Section 4 and the corresponding AI reward are illustrated in Fig. 3 and Fig. 4, respectively. These solutions are derived using the dynamic programming approach (Appendix D). Fig. 3 shows that the coordinator will first recommend the human expert to handle the task directly, which is different from the pattern of the static model in Fig. 2. One possible explanation is that, in this case, the optimal policy of the dynamic model is to involve the human expert to improve the performance of AI rapidly in order to receive larger system reward in the long run. As a result, we can observe in Fig. 4 that the AI reward first increases rapidly, then slowly.

6 Conclusion

In high-stakes decision-making environments, human experts and AI systems often possess complementary strengths: AI systems can process large-scale data efficiently with minimal cost, while human experts offer domain intuition and experience, albeit at higher cognitive or operational cost. We study a human-AI collaborative decision-making framework, where a human expert can either delegate a decision to an AI system or make the decision independently. Our goal is to design an adaptive delegation policy that governs when the human is permitted to make this choice, with the aim of maximizing the overall system reward while satisfying a constraint on the total human decision-making cost. In this paper, we formulate the human-AI collaboration as a generalized Nash equilibrium between a human expert and a coordinator. We answer the question that when human and AI can cooperate, and provide the conditions on how the adaptive delegation strategy would evolve. Finally, we validate our proposed methods using numerical experiments.

References

[1] Kenneth J Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. In *The Foundations of Price Theory Vol 5*, pages 289–316. Routledge, 1954.

[2] Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and Daniel S Weld. Is the most accurate ai the best teammate? optimizing ai for teamwork. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11405–11414, 2021.

[3] Gerard Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38(10):886–893, 1952.

[4] Michael Lingzhi Li and Shixiang Zhu. Balancing optimality and diversity: Human-centered decision making through generative curation. *arXiv preprint arXiv:2409.11535*, 2024.

[5] Anqi Mao, Christopher Mohri, Mehryar Mohri, and Yutao Zhong. Two-stage learning to defer with multiple experts. *Advances in Neural Information Processing Systems*, 36:3578–3606, 2023.

[6] Bryce McLaughlin and Jann Spiess. Designing algorithmic recommendations to achieve human-ai complementarity. *arXiv preprint arXiv:2405.01484*, 2024.

[7] Hussein Mozannar, Hunter Lang, Dennis Wei, Prasanna Sattigeri, Subhro Das, and David Sontag. Who should predict? exact algorithms for learning to defer to humans. In *International Conference on Artificial Intelligence and Statistics*, pages 10520–10545. PMLR, 2023.

[8] Hussein Mozannar and David Sontag. Consistent estimators for learning to defer to an expert. In *International Conference on Machine Learning*, pages 7076–7087. PMLR, 2020.

[9] John F Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[10] John F Nash. Non-cooperative games. In *The Foundations of Price Theory Vol 4*, pages 329–340. Routledge, 1951.

[11] Agni Orfanoudaki, Soroush Saghafian, Karen Song, Harini A Chakkera, and Curtiss Cook. Algorithm, human, or the centaur: How to enhance clinical care? 2022.

[12] Raja Parasuraman and Dietrich H Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010.

[13] Linda J Skitka, Kathleen Mosier, and Mark D Burdick. Accountability and automation bias. *International Journal of Human-Computer Studies*, 52(4):701–717, 2000.

[14] Bryan Wilder, Eric Horvitz, and Ece Kamar. Learning to complement humans. *arXiv preprint arXiv:2005.00582*, 2020.

A Notations Used in the Static Model in Section 3

Table 1. Notations Used in the Static Model.

Model inputs	
$r_{\mathtt{A}}$	System reward for a decision to be generated by AI in the static model
$r_{\mathtt{H}}$	System reward for a decision to be generated by the human expert
$r_{\mathtt{M}}$	System reward for the human expert to modify the solution generated by AI
$c_{\mathtt{H}}$	Cost for the human expert to derive a decision
$c_{\mathtt{M}}$	Cost for the human expert to modify the solution generated by AI
b	Cost budget of the human expert
δ	The minimal target level of system reward
$f(r_{\mathtt{A}})$	A function of AI reward that captures the development of AI technology
Decision variables	
$\boldsymbol{p} \triangleq \{p_{\mathtt{A}}, p_{\mathtt{H}}, p_{\mathtt{R}}\}$	Mixed strategy of the coordinator in the static model
$\boldsymbol{q} \triangleq \{q_{\mathtt{M}}, q_{\mathtt{D}}\}$	Mixed strategy of the human expert in the static model

B Closed-form Solution for the Static Model in Section 3

In this subsection, we will show that closed-form solutions for both the coordinator and the human expert can be derived under the following (mild) assumptions:

(a) $r_{\mathtt{H}} = r_{\mathtt{M}} > \delta$;

(b) $c_{\mathtt{M}} \times \max\{\frac{\delta - r_{\mathtt{A}}}{r_{\mathtt{H}} - r_{\mathtt{A}}}, 0\} \leq b$;

(c) $c_{\mathtt{M}} < c_{\mathtt{H}}$.

Assumptions (a) and (b) are to guarantee the feasibility of constraints for the human expert (2) and the coordinator (1), respectively. These two assumptions ensure that the conditions of Proposition 1 are satisfied. To be specific, assumption (a) means that the minimal target level of system reward, δ, is reasonable and can be achieved; assumption (b) represents that the cost budget b is enough so that it is possible to hire human expert to maintain the minimal system reward requirement. For simplicity, we need assumption (c) that means the cost for human expert to modify the solution generated by AI is smaller than the cost for human expert to generate the solution directly. Note that the situation that $c_{\mathtt{M}} \geq c_{\mathtt{H}}$ (e.g., AI makes mistakes) can be analyzed in a similar way.

Define the function $f(r_{\mathtt{A}}) \triangleq \frac{\delta - r_{\mathtt{A}}}{r_{\mathtt{H}} - r_{\mathtt{A}}}$, which captures the development of AI technology. When the performance of AI cannot satisfy the minimal target level of system reward, namely $\delta > r_{\mathtt{A}}$, it is not hard to see that $f(r_{\mathtt{A}})$ decreases as $r_{\mathtt{A}}$ increases since

$$f'(r_{\mathtt{A}}) = \frac{\delta - r_{\mathtt{H}}}{(r_{\mathtt{H}} - r_{\mathtt{A}})^2} \leq 0 \text{ under assumption (a) } \delta < r_{\mathtt{H}}$$

The closed-form solution for the static model between the coordinator (1) and the human expert (2) can be derived as follows. The results of the static model are summarized in Fig. 1.

• Case 1: $r_{\mathtt{H}} > \delta > r_{\mathtt{A}}$ and $f(r_{\mathtt{A}}) > \frac{b}{c_{\mathtt{M}}}$

When AI is under-developed, namely the reward of AI $r_{\mathtt{A}}$ is so small that $f(r_{\mathtt{A}}) > \frac{b}{c_{\mathtt{M}}}$, the equilibrium between coordinator and human expert does not exist. As a result, human and AI cannot cooperate. In this situation, the assumption (b) for the feasibility of the coordinator's problem (1) is violated. It means that the performance of AI cannot satisfy the minimal target level of system reward. However,

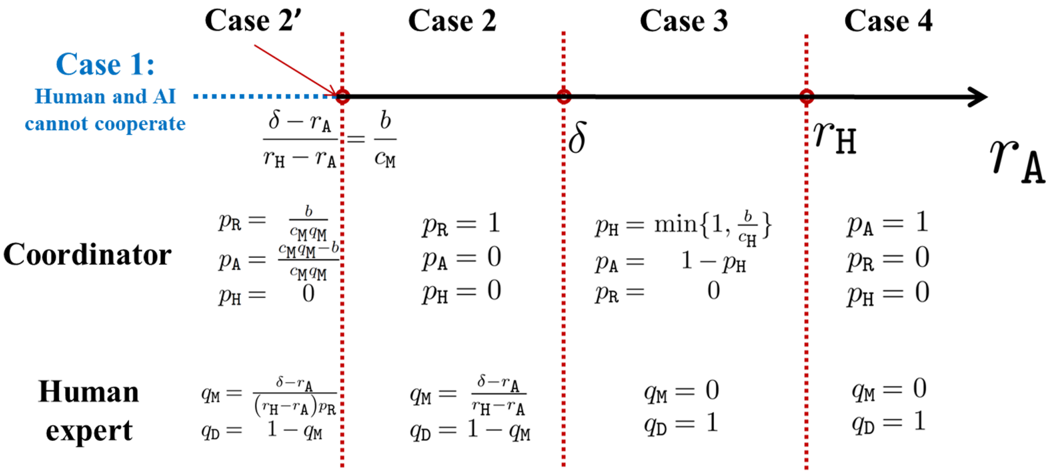
Case 2′
Case 2
Case 3
Case 4
Case 1:
Human and AI cannot cooperate
$\frac{\delta - r_{\mathtt{A}}}{r_{\mathtt{H}} - r_{\mathtt{A}}} = \frac{b}{c_{\mathtt{M}}}$
δ
$r_{\mathtt{H}}$
$r_{\mathtt{A}}$
Coordinator
$p_{\mathtt{R}} = \frac{b}{c_{\mathtt{M}} q_{\mathtt{M}}}$
$p_{\mathtt{A}} = \frac{c_{\mathtt{M}} q_{\mathtt{M}} - b}{c_{\mathtt{M}} q_{\mathtt{M}}}$
$p_{\mathtt{H}} = 0$
$p_{\mathtt{R}} = 1$
$p_{\mathtt{A}} = 0$
$p_{\mathtt{H}} = 0$
$p_{\mathtt{H}} = \min\{1, \frac{b}{c_{\mathtt{H}}}\}$
$p_{\mathtt{A}} = 1 - p_{\mathtt{H}}$
$p_{\mathtt{R}} = 0$
$p_{\mathtt{A}} = 1$
$p_{\mathtt{R}} = 0$
$p_{\mathtt{H}} = 0$
Human expert
$q_{\mathtt{M}} = \frac{\delta - r_{\mathtt{A}}}{(r_{\mathtt{H}} - r_{\mathtt{A}}) p_{\mathtt{R}}}$
$q_{\mathtt{D}} = 1 - q_{\mathtt{M}}$
$q_{\mathtt{M}} = \frac{\delta - r_{\mathtt{A}}}{r_{\mathtt{H}} - r_{\mathtt{A}}}$
$q_{\mathtt{D}} = 1 - q_{\mathtt{M}}$
$q_{\mathtt{M}} = 0$
$q_{\mathtt{D}} = 1$
$q_{\mathtt{M}} = 0$
$q_{\mathtt{D}} = 1$

Figure 1: Closed-form Solution for the Static Model.

we do not have enough budget to hire the human expert for reaching the minimal target level of system reward. Note that this case is eliminated by assumption (b).

- Case 2: $r_{\mathtt{H}} > \delta > r_{\mathtt{A}}$ and $f(r_{\mathtt{A}}) < \frac{b}{c_{\mathtt{M}}}$

In this scenario, the performance of AI, $r_{\mathtt{A}}$, reaches certain threshold, i.e., $f(r_{\mathtt{A}}) \leq \frac{b}{c_{\mathtt{M}}}$, but still cannot reach the minimal target level of system reward, namely, $r_{\mathtt{H}} > \delta > r_{\mathtt{A}}$. The coordinator will always recommend the human expert to use AI (in other words, to modify the solution derived from AI), while the human expert will modify the decision made by AI with a positive probability. Specifically, the optimal mixed strategy of the coordinator is $p_{\mathtt{R}} = 1$ and $p_{\mathtt{A}} = p_{\mathtt{H}} = 0$, and best policy of the human expert is as below:

$$\begin{array}{rl} q_{\mathtt{M}} = & \frac{\delta - r_{\mathtt{A}}}{r_{\mathtt{H}} - r_{\mathtt{A}}} \in (0, 1] \\ q_{\mathtt{D}} = & 1 - q_{\mathtt{M}} \end{array}$$

Note that the situation $f(r_{\mathtt{A}}) = \frac{b}{c_{\mathtt{M}}}$ is a degenerate case (case 2′) such that the equilibrium solution is not unique. The coordinator will recommend the human expert to use AI with a positive probability (possibly smaller than 1). In particular, the best mixed strategy of the coordinator is as follows:

$$p_{\mathtt{R}} = \frac{b}{c_{\mathtt{M}} q_{\mathtt{M}}} \in (0, 1],\ p_{\mathtt{A}} = \frac{c_{\mathtt{M}} q_{\mathtt{M}} - b}{c_{\mathtt{M}} q_{\mathtt{M}}},\ \text{and } p_{\mathtt{H}} = 0 \tag{6}$$

Accordingly, the human expert will modify the solution derived from AI with a positive probability:

$$q_{\mathtt{M}} = \frac{\delta - r_{\mathtt{A}}}{(r_{\mathtt{H}} - r_{\mathtt{A}})\, p_{\mathtt{R}}} \in (0, 1] \text{ and } q_{\mathtt{D}} = 1 - q_{\mathtt{M}} \tag{7}$$

- Case 3: $r_{\mathtt{H}} > r_{\mathtt{A}} \geq \delta$

This is the case that AI reaches the minimal requirement of system performance, but cannot perform better than the human expert. In this situation, the human expert will not modify the decision generated by AI since it already matches the minimal requirement. Thus, in order to maximize the system reward, the coordinator will use up the budget to hire the human expert, and recommend AI to handle the rest of tasks. As a result, the coordinator will recommend fully automated AI decision-making or mandatory human decision. In particular, the best policy of the coordinator is $p_{\mathtt{H}} = \min\{1, \frac{b}{c_{\mathtt{H}}}\}, p_{\mathtt{A}} = 1 - p_{\mathtt{H}}$ and $p_{\mathtt{R}} = 0$. The optimal mixed strategy for the human expert is $q_{\mathtt{M}} = 0$ and $q_{\mathtt{D}} = 1$.

- Case 4: $r_{\mathtt{A}} \geq r_{\mathtt{H}}$

When AI performs no worse than the human expert, the coordinator will always recommend AI to complete the task. Thus, the optimal policy of the coordinator is $p_{\mathtt{A}} = 1$ and $p_{\mathtt{R}} = p_{\mathtt{H}} = 0$. And the best mixed strategy of the human expert is $q_{\mathtt{M}} = 0$ and $q_{\mathtt{D}} = 1$.

C Extra Notations Used in the Dynamic Model in Section 4

Table 2. Extra Notations Used in the Dynamic Model.

Model inputs	
T	Time horizon of the dynamic model
$r_{\mathtt{A}}^t[\bullet]$	System reward function for a decision to be generated by AI at time stage $t \in \{1, 2, \cdots, T\}$ in the dynamic model
$r_{\mathtt{AGI}}$	System reward for a decision to be generated by the artificial general intelligence
α, β	Positive parameters in the nonlinear AI reward function $r_{\mathtt{A}}^t[\bullet]$
P_t	The constraint set of the coordinator at time stage $t \in \{0, 1, 2, \cdots, T\}$ in the dynamic model
$\hat{r}_{\mathtt{A}}^t[\bullet]$	Approximated system reward function for a decision to be generated by AI at time stage $t \in \{0, 1, 2, \cdots, T\}$ in the dynamic model
Decision variables	
$\bar{\boldsymbol{p}} \triangleq \{p_{\mathtt{A}}^t, p_{\mathtt{H}}^t, p_{\mathtt{R}}^t\}_{t\in\{1,2,\cdots,T\}}$	Mixed strategy of the coordinator in the dynamic model
$\boldsymbol{p}_t \triangleq \{p_{\mathtt{A}}^t, p_{\mathtt{H}}^t, p_{\mathtt{R}}^t\}$	Mixed strategy of the coordinator at time stage $t \in \{0, 1, 2, \cdots, T\}$ in the dynamic model
$\boldsymbol{q}_t \triangleq \{q_{\mathtt{M}}^t, q_{\mathtt{D}}^t\}$	Mixed strategy of the human expert at time stage $t \in \{0, 1, 2, \cdots, T\}$ in the dynamic model
x_t	State of the system in dynamic programming, which represents the cumulative probability that the human expert involves in handling the task before time stage $t \in \{0, 1, 2, \cdots, T\}$
$g_t(x_t, \boldsymbol{p}_t)$	The system reward incurred at time stage $t \in \{0, 1, 2, \cdots, T\}$ in dynamic programming
$V_t(x_t)$	The optimal system reward of the subproblem that starts at time stage $t \in \{0, 1, 2, \cdots, T\}$ with initial state x_t and ends at time stage T in dynamic programming

D Dynamic Programming to Solve the Dynamic Model in Section 4

Based on the closed-form solutions derived in Appendix B, under the assumptions of (a) $r_{\mathtt{H}} = r_{\mathtt{M}} > \delta$, (b) $c_{\mathtt{H}} \leq b$, (c) $c_{\mathtt{M}} < c_{\mathtt{H}}$, the closed-form solution for the human expert (4) at time $t \in \{0, 1, 2, \cdots, T\}$ can be derived as follows:

$$\begin{aligned} q_{\mathtt{M}}^{t,*} &= \begin{cases} 0, & \text{if } r_{\mathtt{H}} \geq \delta > r_{\mathtt{A}}^t[\bullet] \text{ and } f\left(r_{\mathtt{A}}^t[\bullet]\right) > \dfrac{b}{c_{\mathtt{M}}} \\ \dfrac{\delta - r_{\mathtt{A}}^t[\bullet]}{\left(r_{\mathtt{H}} - r_{\mathtt{A}}^t[\bullet]\right) p_{\mathtt{R}}^t}, & \text{if } r_{\mathtt{H}} \geq \delta > r_{\mathtt{A}}^t[\bullet] \text{ and } f\left(r_{\mathtt{A}}^t[\bullet]\right) = \dfrac{b}{c_{\mathtt{M}}} \\ \dfrac{\delta - r_{\mathtt{A}}^t[\bullet]}{r_{\mathtt{H}} - r_{\mathtt{A}}^t[\bullet]}, & \text{if } r_{\mathtt{H}} \geq \delta > r_{\mathtt{A}}^t[\bullet] \text{ and } f\left(r_{\mathtt{A}}^t[\bullet]\right) \leq \dfrac{b}{c_{\mathtt{M}}} \\ 0, & \text{if } r_{\mathtt{A}}^t[\bullet] \geq \delta \end{cases} \\ q_{\mathtt{D}}^{t,*} &= 1 - q_{\mathtt{M}}^{t,*} \end{aligned} \tag{8}$$

Note that we relax the assumption (b) of Appendix B to $c_{\mathtt{M}} \leq b$ for feasibility of the problem, as analyzed in Case 1 of Appendix B. Plug in the closed-form solution of the human expert (8) at each time stage $t \in \{0, 1, 2, \cdots, T\}$ to the coordinator's problem (3), we obtain the following single-level multi-stage optimization problem:

$$\begin{aligned} \max_{\bar{\boldsymbol{p}}\geq 0} \quad & \sum_{t=0}^{T}\left\{r_{\mathtt{H}}p_{\mathtt{H}}^{t} + r_{\mathtt{A}}^{t}\left[\bullet\right]p_{\mathtt{A}}^{t} + \left(r_{\mathtt{M}}q_{\mathtt{M}}^{t,*} + r_{\mathtt{A}}^{t}\left[\bullet\right]q_{\mathtt{D}}^{t,*}\right)p_{\mathtt{R}}^{t}\right\} \\ \text{s.t.} \quad & c_{\mathtt{H}}p_{\mathtt{H}}^{t} + c_{\mathtt{M}}q_{\mathtt{M}}^{t,*}p_{\mathtt{R}}^{t} \leq b \quad \forall t = 0,1,2,\cdots,T \\ & p_{\mathtt{A}}^{t} + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t} = 1 \quad \forall t = 0,1,2,\cdots,T \end{aligned} \tag{9}$$

We use dynamic programming to solve problem (9). Define x_t as the *state* of the system, which represents the cumulative probability that the human expert involves in handling the task before time stage t (in particular, $x_0 = 0$), i.e.,

$$x_t = \sum_{s=0}^{t-1}\left(p_{\mathtt{H}}^{s} + p_{\mathtt{R}}^{s}q_{\mathtt{M}}^{s,*}\right) \leq t-1, \quad \forall t = 1,2,\cdots,T$$

Thus, the reward of AI in time stage t is

$$r_{\mathtt{A}}^{t}\left(x_t\right) = r_{\mathtt{AGI}} - \frac{\alpha}{1+\left(x_t\right)^{\beta}}, \quad \forall t = 0,1,2,\cdots,T$$

The *decision variable* or *action* to be selected at time stage $t \in \{0,1,2,\cdots,T\}$ is the coordinator's decision $\boldsymbol{p}_t \triangleq \{p_{\mathtt{A}}^{t}, p_{\mathtt{H}}^{t}, p_{\mathtt{R}}^{t}\}$, which belongs to the constraint set

$$P_t \triangleq \left\{\left(p_{\mathtt{A}}^{t}, p_{\mathtt{H}}^{t}, p_{\mathtt{R}}^{t}\right) \geq 0 : c_{\mathtt{H}}p_{\mathtt{H}}^{t} + c_{\mathtt{M}}q_{\mathtt{M}}^{t,*}p_{\mathtt{R}}^{t} \leq b,\, p_{\mathtt{A}}^{t} + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t} = 1\right\}$$

The *state transition* in the system can be defined as

$$x_{t+1} = x_t + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t}q_{\mathtt{M}}^{t,*}, \quad \forall t = 0,1,2,\cdots,T$$

The system reward incurred at time stage $t \in \{0,1,2,\cdots,T\}$, denoted as $g_t(x_t, \boldsymbol{p}_t)$, can be computed as

$$g_t(x_t, \boldsymbol{p}_t) = r_{\mathtt{H}}p_{\mathtt{H}}^{t} + r_{\mathtt{A}}^{t}\left(x_t\right)p_{\mathtt{A}}^{t} + \left[r_{\mathtt{M}}q_{\mathtt{M}}^{t,*} + r_{\mathtt{A}}^{t}\left(x_t\right)q_{\mathtt{D}}^{t,*}\right]p_{\mathtt{R}}^{t}$$

Notate $V_t(x_t)$ as the optimal system reward of the subproblem that starts at time stage $t \in \{0,1,2,\cdots,T\}$ with initial state x_t and ends at time stage T. Specifically, define $V_{T+1}(\bullet) = 0$. To be precise, we have

$$\begin{aligned} V_t(x_t) \triangleq & \max_{\left(\boldsymbol{p}_t, \boldsymbol{p}_{t+1}, \cdots, \boldsymbol{p}_T\right)\geq 0} && \sum_{s=t}^{T}\left\{r_{\mathtt{H}}p_{\mathtt{H}}^{s} + r_{\mathtt{A}}^{s}\left[\bullet\right]p_{\mathtt{A}}^{s} + \left(r_{\mathtt{M}}q_{\mathtt{M}}^{s,*} + r_{\mathtt{A}}^{s}\left[\bullet\right]q_{\mathtt{D}}^{s,*}\right)p_{\mathtt{R}}^{s}\right\} \\ \text{s.t.} & && c_{\mathtt{H}}p_{\mathtt{H}}^{s} + c_{\mathtt{M}}q_{\mathtt{M}}^{s,*}p_{\mathtt{R}}^{s} \leq b \quad \forall s = t, t+1, \cdots, T \\ & && p_{\mathtt{A}}^{s} + p_{\mathtt{H}}^{s} + p_{\mathtt{R}}^{s} = 1 \quad \forall s = t, t+1, \cdots, T \end{aligned}$$

where

$$\begin{aligned} r_{\mathtt{A}}^{t}\left[\bullet\right] = & \quad r_{\mathtt{AGI}} - \tfrac{\alpha}{1+\left(x_t\right)^{\beta}} \\ r_{\mathtt{A}}^{s}\left[\bullet\right] = & \quad r_{\mathtt{AGI}} - \frac{\alpha}{1+\left[x_t + \sum\limits_{r=t+1}^{s}\left(p_{\mathtt{H}}^{r} + p_{\mathtt{R}}^{r}q_{\mathtt{M}}^{r,*}\right)\right]^{\beta}}, \quad \forall s = t+1, t+2, \cdots, T \end{aligned}$$

From the definition of $V_t(x_t)$, we have at time stage $t \in \{0,1,2,\cdots,T\}$,

$$\begin{aligned} V_t(x_t) &= \max_{\boldsymbol{p}_t \in P_t}\left\{g_t(x_t, \boldsymbol{p}_t) + V_{t+1}(x_{t+1})\right\} \\ &= \max_{\boldsymbol{p}_t \in P_t}\left\{g_t(x_t, \boldsymbol{p}_t) + V_{t+1}(x_t + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t}q_{\mathtt{M}}^{t,*})\right\} \end{aligned}$$

Thus, the *Bellman equation* can be written as follows:

$$\begin{aligned} V_t(x_t) &= \max_{\boldsymbol{p}_t \in P_t}\left\{g_t(x_t, \boldsymbol{p}_t) + V_{t+1}(x_t + p_{\mathtt{H}}^{t} + p_{\mathtt{R}}^{t}q_{\mathtt{M}}^{t,*})\right\}, \, \forall t = 0,1,2,\cdots,T \\ V_{T+1}(\bullet) &= 0 \end{aligned} \tag{10}$$

E Figures for the Numerical Experiments in Section 5

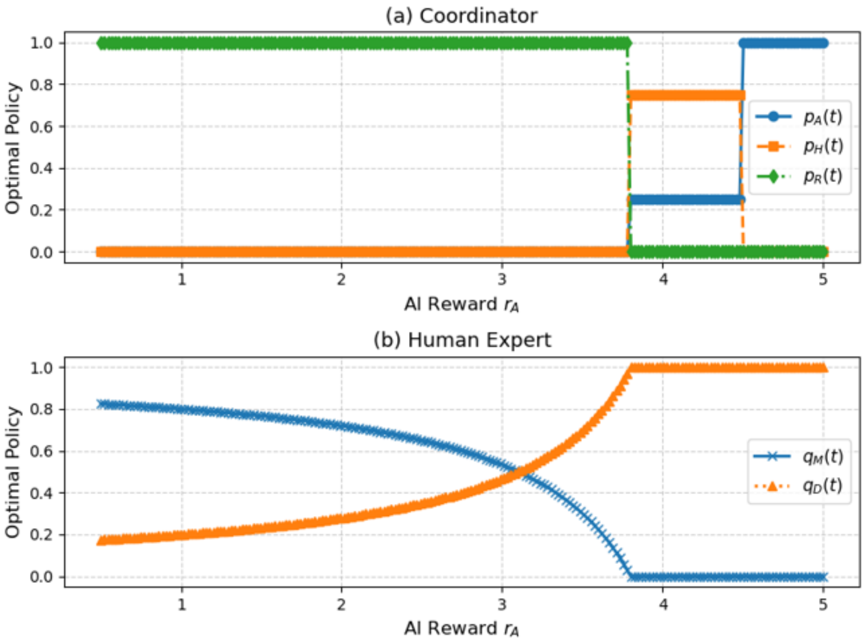

Figure 2: Numerical Results of the Static Model.

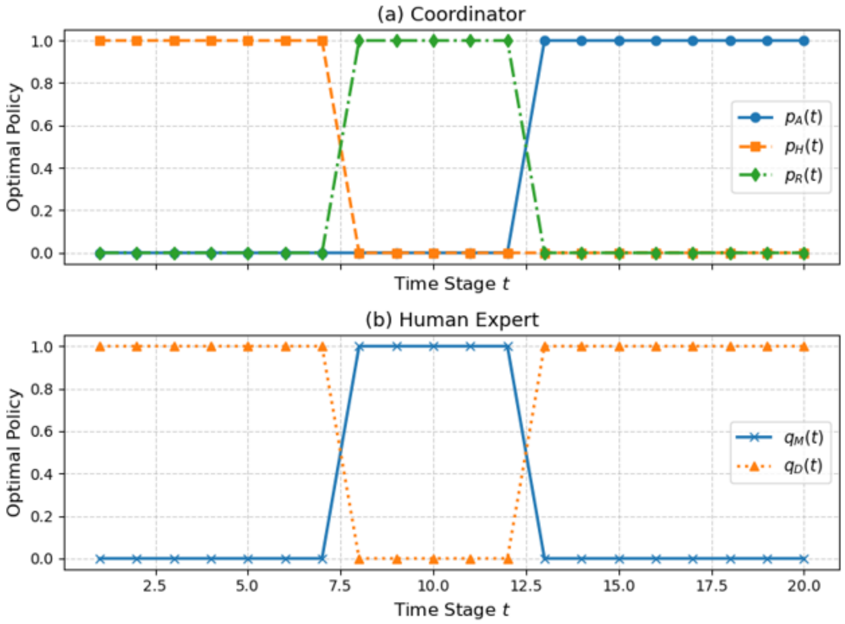

Figure 3: Numerical Results of the Dynamic Model.

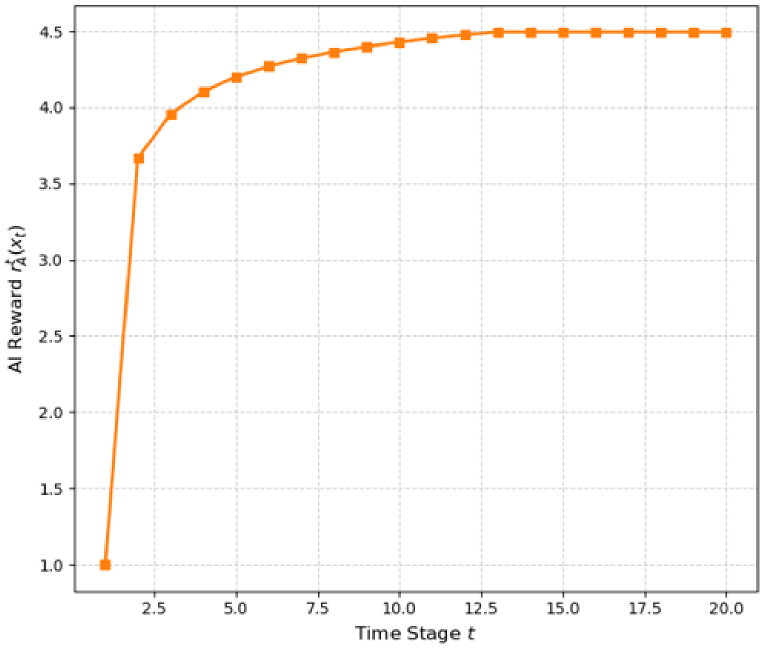

Figure 4: AI Reward of the Dynamic Model.