Working with AI: An Analysis for Rational Integration[*]

David Lagziel[†] and Yevgeny Tsodikovich[‡]

March 21, 2024

Abstract

A decision maker uses an AI agent to estimate an unknown state, for which both possess informative private signals. Conditional on the state and the decision maker's final assessment, he prefers the AI's recommendations to be incorrect, thus affirming his own superiority or sharing the blame. Our analysis indicates that the correctness of the process is not a monotone function of participants' expertise levels: (i) a less accurate AI may lead to improved outcomes by reducing the agent's reliance on it, and (ii) a less accurate agent can enhance information aggregation leading to a superior result.

Journal classification numbers: C72, D82, D83

Keywords: AI-enhanced decisions; Bayesian learning; Guided strategies.

[*] For their valuable comments, the authors thank Nicolas Vieille, Eran Shmaya, Amnon Schreiber, Eilon Solan, Xavier Venel, Ehud Lehrer, as well as the participants of the 7th Workshop on Stochastic Methods in Game Theory in Erice, SING17 – 17th European Meeting on Game Theory, and the 12th annual conference of the Israeli Chapter of the Game Theory Society. Lagziel acknowledges the support of the Israel Science Foundation, Grant #2074/23. Tsodikovich acknowledges the support of the Israel Science Foundation Grants #2566/20, #1626/18 & #448/22.

[†] Department of Economics, Ben-Gurion University of the Negev, Israel. E-mail: Davidlag@bgu.ac.il.

[‡] Department of Economics, Bar Ilan University, Israel. E-mail: yevgets@gmail.com.

1 Introduction

Writing the first paragraph of a paper is not easy, so we decided to consult a large language model (LLM) based on our abstract. The model suggested the following sentence: 'In an era where artificial intelligence (AI) and human decision-making increasingly intersect, this paper explores the nuanced dynamics between a decision maker (DM) and an AI agent, each equipped with private insights into an uncertain situation.' Although the LLM's suggestion articulately encapsulates the key elements of our study, we decided to proceed with the current format — it does so even better.

The paper analyzes the strategic intersection between an AI agent (henceforth, AI) and a DM, where each possesses a private, informative signal about an unknown state. The AI offers a preliminary truthful assessment, and the DM provides the final decision based on his private information and the AI's recommendation. The DM strives to provide an accurate decision, but *conditional on the state and his final assessment*, he also prefers the AI to be incorrect. The reason is clear: if the DM is also incorrect, this allows him to share any liability with the AI, and in case the DM is correct, it allows him to establish superiority.[1] This payoff structure is corroborated by recent empirical evidence from Almog et al. (2024) who estimate the significant psychological costs of being overruled by AI (in the context of tennis umpires).

Our analysis shows that the DM has three optimal strategies: (i) a DM-led strategy that relies solely on his private signal; (ii) an AI-led strategy that strictly aligns with the AI's assessment; and (iii) a Guided strategy involving some degree of learning, conditional on both signals. To study the nature of these equilibria, we map the parameter space, encompassing expertise levels, and payoffs, to three disjoint sets – one for each optimal strategy. We then use the notion of *correctness*, defined by the probability that the DM arrives at a correct decision, to study the transitions between the different regimes.

The study presents three (somewhat surprising) phenomena. The first, referred to as *the dependency threshold*, states that the correctness of the process is not a monotonic function of the AI's expertise level. Specifically, an infinitesimal increase in the AI's accuracy may yield a

[1] As later discussed in Section 3.2.1, the analysis requires only one of these externalities to hold.

stark drop in the correctness. This occurs due to a transition from an optimal Guided strategy to an optimal AI-led one, triggered by state-dependent payoff externalities. Namely, if the expertise levels of both participants are rather close, the DM (optimally) “hedges” his expected payoff by completely relying on the AI’s advice. Figure 1 illustrates this effect by depicting the correctness as a function of the AI’s accuracy.

The second phenomenon, named *the humility threshold*, mirrors the first by varying the DM’s expertise level. It states that an infinitesimal increase in the DM’s expertise level can also decrease the probability of reaching a correct decision due to a shift from an optimal Guided strategy to an optimal DM-led one. Again, assuming that the expertise levels are rather close, the DM can increase his expected payoff by distinguishing his decision from the AI’s preliminary assessment.

As our reliance on AI technology is projected to increase significantly, the relevance of these findings extends well beyond the theoretical aspects. For example, in February 2024, the British Columbia Civil Resolution Tribunal held Air Canada accountable for incorrect information provided by the company’s chatbot, despite the company’s argument that the chatbot was “a separate legal entity responsible for its own actions.”[2] Additionally, many recent academic papers have been published containing explicit comments from LLMs, such as: “as of my last knowledge update” and “I don’t have access to real-time data.”[3] These examples underscore the need for a rational approach to integrating AI technologies into critical decision-making processes, a necessity that becomes increasingly apparent with the growing application of AI in fields like medical imaging and assessments.[4]

Another effect uncovered in our analysis concerns the relative expertise levels of the two agents. Given the option to choose, we ask whether a better-informed DM is more beneficial, in terms of correctness, than an AI with more information. Our analysis indicates that, in some scenarios, it is preferable to have a less-informed DM paired with a better-informed AI. This phenomenon also follows from the transition between two regimes: one where a less-informed

[2] See bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know.

[3] A Google Scholar search in March 2024 showed more than 150 academic papers containing these phrases. There are even astonishing examples of published papers that cite non-existing literature.

[4] see, for example, Arabi et al. (2021); Barragán-Montero et al. (2021) and Varoquaux and Cheplygina (2022), as well as the broad literature review in Agarwal et al. (2023).

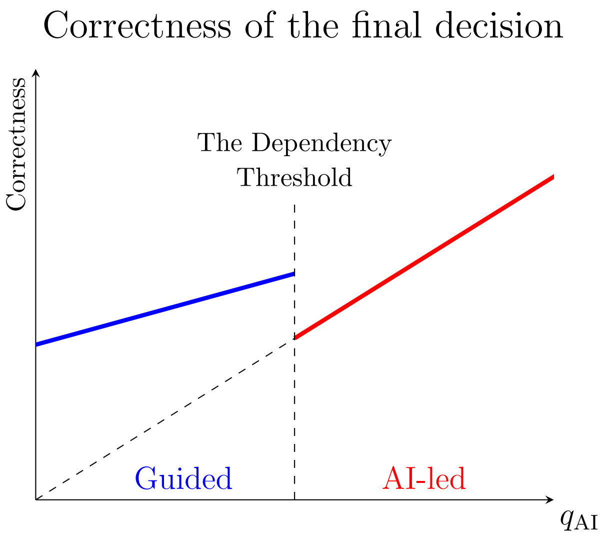

Figure 1: An infinitesimal increase in the AI's accuracy q_{AI} yields a transition from an optimal guided strategy (blue line), to an optimal AI-led strategy (red line). This leads the DM to ignore his private information, thus producing an inferior expected outcome.

DM learns from the AI's recommendations, and the other which eliminates this possibility.

1.1 Related research

This paper encompasses elements from several different sub-fields, the primary being Bayesian learning with externalities, which expanded remarkably since the studies of Banerjee (1992), Bikhchandani et al. (1992), and Smith and Sorensen (2000).[5] Bayesian learning with externalities typically relates to either positive or negative congestion costs, under which agents' payoffs either increase or decrease depending on the number of agents who choose the same action (see, e.g., Veeraraghavan and Debo (2011), Debo et al. (2011), and Eyster et al. (2014), among others). Our study matches this line of research through the possibility of learning under positive/negative externalities among experts, as well as the basic information structure and actions. Specifically, Eyster et al. (2014) show that backward-looking negative externalities (i.e., players' actions are less profitable the more they are played by others) prevent action fixation and improve social learning. This resembles our notion of a Guided equilibrium, where the second agent learns from the first, through payoff (and information) considerations. Conversely, this is in contrast to the findings of Ali and Kartik (2012) that identifies a unique

[5] For a recent extensive review of this topic, see Bikhchandani et al. (2021).

optimal (AI-led) strategy in case agents' preferences are aligned.

A recent paper relevant to our work is Agarwal et al. (2023), which illustrates how radiologists fail to account for possible correlations between the AI's assessment and their private signals. Our model serves as a baseline framework for theirs, assuming a DM without bounded-rationality constraints. Similarly a study by Angelova et al. (2023) investigates related issues within the context of bail decisions, examining variations in judges' abilities to optimally utilize private information when overriding AI decisions. Our principal contributions focus on the shift between optimal-strategy regimes, specifically from a Guided strategy to either a DM-led or AI-led strategy. These transitions have been empirically validated in a recent study by Kanazawa et al. (2022), which assesses the effect of AI assistance on taxi drivers in Japan. This study demonstrates that AI assistance significantly boosts the productivity of lower-skilled drivers (aligning with a Guided or AI-led strategy) while having negligible impact on higher-skilled drivers (consistent with a DM-led strategy).

Structure of the paper. In Section 2 we describe the model and the key definitions. In Section 3 we present the main results. Concluding remarks are given in Section 4. To facilitate readability, proofs are relegated to the appendix.

2 The model

A decision-maker (hereafter, DM) wishes to identify a binary unknown state. To achieve this, he utilizes an AI advisor, trained specifically for such tasks, and based on the AI's recommendation and his own observation, the DM provides an assessment. While the DM's goal is to provide an accurate assessment, he faces a conflict of interest concerning the AI's accuracy. On the one hand, he prefers the AI to provide an accurate assessment, because the information conveyed by the AI aids him in making the correct decision. On the other hand, he perceives the AI as a potential rival. So, *conditional on his own decision and state*, he prefers the AI to provide an incorrect assessment. This reflects his ambition either to assert his superiority, or to share the blame (in case both assessments are false).

Formally, consider the following decision problem G. There are two states denoted by

$\theta \in \Theta = \{0, 1\}$, and a prior probability $p = \Pr(\theta = 0) > \frac{1}{2}$. Given θ, every agent $i \in \{\mathrm{AI}, \mathrm{DM}\}$ receives an independent, noisy and informative signal $s_i \in S = \{0, 1\}$, such that $\Pr(s_i = \theta|\theta) = q_i$. One can think of q_i as *the expertise level* of agent i, a measure of agent i's ability to correctly identify the state. The action set of every agent i is $A = \{0, 1\}$.

The decision problem evolves as follows. First, nature chooses a state θ according to a common, publicly known, prior p. Then, every agent i receives a private signal s_i based on the previously defined information structure. The AI is the first to publicly provide his true assessment $a_{\mathrm{AI}}(s_{\mathrm{AI}}) = s_{\mathrm{AI}}$. After observing a_{AI}, the DM makes a decision $a_{\mathrm{DM}} \in A$.

The utility of the DM is characterized by four possible payoffs: $U_1 \geq U_2 > U_3 \geq U_4$, where U_1 is the payoff in case the DM is the only one to choose the correct action, U_2 is the payoff when both are correct, U_3 is the payoff in case both are incorrect and U_4 is the DM's payoff if he is the only one to choose the wrong action. Since games are strategically equivalent under an affine transformation of payoffs, it is without loss of generality that we normalize $U_2 = 1$ and $U_3 = -1$. The realized payoffs are summarized as follows:

$$U(a_{\mathrm{DM}}, a_{\mathrm{AI}}|\theta) = \begin{cases} U_1, & \text{if } a_{\mathrm{DM}} = \theta, a_{\mathrm{AI}} \neq \theta \\ 1, & \text{if } a_{\mathrm{DM}} = \theta, a_{\mathrm{AI}} = \theta \\ -1, & \text{if } a_{\mathrm{DM}} \neq \theta, a_{\mathrm{AI}} \neq \theta \\ U_4, & \text{if } a_{\mathrm{DM}} \neq \theta, a_{\mathrm{AI}} = \theta \end{cases}$$

To ensure that the agents' signals are informative independently of the state, we assume that $\min\{q_{\mathrm{AI}}, q_{\mathrm{DM}}\} \geq p$, which is equivalent to $\Pr(\theta = x|s_i = x) > \frac{1}{2}$ for every (i, x). Otherwise, in a single-player set-up where $q_i < p$, the state $\theta = 0$ is more likely than $\theta = 1$, regardless of the agent's information.[6]

Denote the strategy of the DM by $\sigma_{\mathrm{DM}} : S \times A \to \Delta(A)$. Our first goal is to identify the optimal strategy for every composition of the parameters $(p, q_{\mathrm{AI}}, q_{\mathrm{DM}}, U_1, U_4)$. To achieve this goal, we define the three following strategies. Our analysis shows that, for every choice of the

[6] One can find some resemblance between our model and the example provided in Section 2 of Smith et al. (2021).

mentioned parameters, exactly one of these strategies is optimal:[7]

- A strategy σ_{DM} is a *DM-led strategy* if the DM's action matches his private signal, i.e., if $\sigma_{\text{DM}} = s_{\text{DM}}$ for every s_{DM}. This typically occurs when the DM is much more informed than the AI, so he ignores it.

- A strategy σ_{DM} is an *AI-led strategy* if the DM's action matches the recommendation of the AI in every realization, i.e., if $\sigma_{\text{DM}} = s_{\text{AI}}$ for every s_{AI}. This typically occurs when the AI is much more informed than the DM, so the DM ignores his private signal.

- A strategy σ_{DM} is a *Guided strategy* if $\sigma_{\text{DM}} = s_{\text{DM}}$, with the exception of $\sigma_{\text{DM}}(a_{\text{AI}} = 0, s_{\text{DM}} = 1) = 0$. That is, the DM follows his signal, unless $(a_{\text{AI}}, s_{\text{DM}}) = (0, 1)$ where he takes the decision $a_{\text{DM}} = 0$. This typically occurs when the two have close expertise levels, so that the DM learns from the AI, and weighs in both signals into his decision. Alternatively, this can be viewed as a two-step screening process, in which the DM intervenes only when the AI detects an anomaly (the less-likely state).

Using this classification, we proceed to our main goal – to study the strategic interaction between the AI and the DM on the latter's final decision. For this purpose, we use the notion of "correctness", which measures the probability of reaching a correct decision.[8] Formally, define the *correctness of the process* to be the probability that the DM's decision matches the true state of the world, so $C(q_{\text{AI}}, q_{\text{DM}}) = \Pr(\sigma_{\text{DM}} = \theta)$. In general, the correctness depends on the expertise levels, as well as on the prior p and the payoffs U_1 and U_4. In what follows, we will assume p, U_1, and U_4 are fixed and study how the correctness varies with the agents' expertise levels.

Finally, we introduce a logit notation to represent probabilities, so

$$\tilde{p} := \ln\left(\frac{p}{1-p}\right), \qquad \tilde{q_i} := \ln\left(\frac{q_i}{1-q_i}\right).$$

[7] Other than on the boundaries between different areas in this parameter space, where different strategies coincide.

[8] See Arieli et al. (2018) for more details.

This notation allows us to simplify some of the equations and conditions resulting from Bayesian updating and present them as linear functions.

3 Main results

Our main results comprise several parts. In Theorem 1 we characterize the conditions for each of the three strategies to be optimal, as a function of $p, q_{\text{AI}}, q_{\text{DM}}, U_1$ and U_4. Figure 2 illustrates this characterization. It shows the three disjoint optimal-strategy regions in the $(q_{\text{AI}}, q_{\text{DM}})$-plane given $U_1 = -U_4$ and p. The two regions of the AI-led and DM-led strategies are rather straightforward — a significantly higher expertise level of one agent over the other. On the other hand, the region of the Guided strategy is more intriguing. It arises when neither agent has a clear dominance. This leads the DM to rely on the AI's assessment in borderline situations when the DM's signal is $s_{\text{DM}} = 1$, which contradicts an AI's recommendation of $a_{\text{AI}} = 0$ along with the biased prior (towards $\theta = 0$).

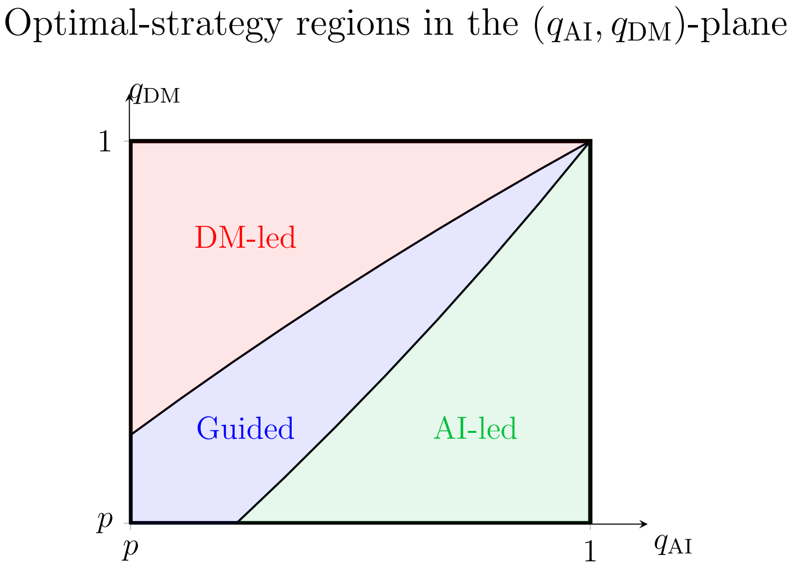
Optimal-strategy regions in the $(q_{\text{AI}}, q_{\text{DM}})$-plane
q_{DM}
1
DM-led
Guided
AI-led
p
p
1
q_{AI}

Figure 2: The different optimal-strategy regions in the $(q_{\text{AI}}, q_{\text{DM}})$-plane for $U_1 = -U_4$. Each region corresponds to exactly one of the three possible strategies: DM-led (red), Guided (blue), and AI-led (green).

We use the characterization in Theorem 1 to prove three results concerning the non-monotonic nature of the correctness. In Proposition 1, we show that correctness is not a monotone function of the AI's expertise level; thus, reducing the AI's accuracy can increase the probability of reaching the correct decision. More precisely, there exists a threshold expertise level, referred to as *the dependency threshold*, such that the DM completely ignores his private information if and only if the AI's expertise level is above this threshold. Consequently, given a less-informed AI, the problem transitions from an AI-led scenario to a Guided one, in which all relevant information is taken into account.

Similarly, in Proposition 2, we prove that an infinitesimal increase in the DM's expertise level may trigger a stark drop in correctness. That is, there exists a threshold expertise level, namely *the humility threshold*, such that the DM ignores the information conveyed by the AI if and only if the DM's expertise level is above this threshold. Thus, only a less-informed DM would consider both signals to ideally reach a more accurate decision.

In addition, Proposition 3 investigates two contrasting scenarios: one where the DM is more informed than the AI, and the reversed situation. We show that assigning the less precise signal to the DM, rather than to the AI, can, paradoxically, increase the likelihood of reaching a correct decision. This result implies that the DM, who could exploit errors made by the AI, might actually benefit from being assigned the less accurate signal. This phenomenon underscores the ability to enhance learning processes when starting with a more accurate initial assessment.

3.1 Optimal-strategy characterization

When the DM decides against the AI's assessment he weighs in two possible gains: $U_1 + 1$, which is the potential gain from non-conformity (being correct alone versus being incorrect with the AI), and $1 - U_4$, which is the potential gain from conformity (being correct with the AI versus being incorrect alone). We define the *non-conformity gain ratio* as the ratio between these two numbers, and denote it by $\gamma = \frac{U_1-(-1)}{1-U_4}$. To be in line with the logit representation of probabilities, we denote $\tilde{\gamma} = \ln \gamma$.

In general, we use the non-conformity gain ratio γ to study how the optimal strategy of the DM evolves, as a function of the payoffs. Though one can perform an analysis for every $\tilde{\gamma}$ (and for every U_1 and U_4), we limit the discussion to the range $\tilde{\gamma} \leq \tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} - \tilde{p}$. The reason is that when $\tilde{\gamma}$ is too large, it becomes the sole driving force behind the optimal strategy, irrespective of private and public information. Specifically, suppose that both signals are $s_{\mathrm{AI}} = s_{\mathrm{DM}} = 1$. This yields the highest possible posterior on the event $\{\theta = 1\}$. Still, if the previous inequality is violated, then the DM benefits from "gambling" on the low-probability event $\{\theta = 0 | s_{\mathrm{AI}} = s_{\mathrm{DM}} = 1\}$, simply because it opposes the AI's assessment.

We start with an optimal-strategy characterization. The following theorem depicts the DM's optimal strategy, as a function of the $p, q_{\mathrm{AI}}, q_{\mathrm{DM}}$, and γ.

Theorem 1. *Consider the previously defined decision problem G with fixed parameters $p, q_{\mathrm{AI}}, q_{\mathrm{DM}}$, and γ.*

- *The* AI-led strategy *is optimal if and only if $\tilde{\gamma} \leq \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} - \tilde{p}$.*
- *The* Guided strategy *is optimal if and only if $\tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} - \tilde{p} \leq \tilde{\gamma} \leq \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} + \tilde{p}$.*
- *The* DM-led strategy *is optimal if and only if $\tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} + \tilde{p} \leq \tilde{\gamma} \leq \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} - \tilde{p}$.*

Let us provide some intuition to the Guided outcome. This strategy is optimal whenever the expertise levels of both are rather close, as evident from the condition $\tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} \in (\tilde{\gamma} - \tilde{p}, \tilde{\gamma} + \tilde{p})$. Thus, when the DM receives a signal that matches the recommendation of the AI, it is optimal for him to follow it accordingly. Otherwise, he resorts to the prior which is biased towards an $a_{\mathrm{DM}} = 0$ recommendation.

The ability to vary the payoffs of the game, through the non-conformity gain ratio γ, shows how the transition from an optimal AI-led strategy to a DM-led one, passes through an optimal Guided strategy. Figure 3 illustrates this transition. From a mechanism design perspective, an outside stakeholder can implement any of the three strategy profiles by calibrating the payoffs according to the required γ. As each strategy profile induces a different correctness level, this allows the stakeholder to maximize the probability of reaching a correct decision.

We now turn to calculate the correctness of each strategy. Clearly, $C_{\mathrm{AI}}(q_{\mathrm{AI}}, q_{\mathrm{DM}}) = q_{\mathrm{AI}}$ for the AI-led strategy and $C_{\mathrm{DM}}(q_{\mathrm{AI}}, q_{\mathrm{DM}}) = q_{\mathrm{DM}}$ for the DM-led one. For the Guided strategy, it is straightforward to verify that

$$C_{\mathrm{G}}(q_{\mathrm{AI}}, q_{\mathrm{DM}}) = q_{\mathrm{DM}} + p q_{\mathrm{AI}}(1 - q_{\mathrm{DM}}) - (1 - p)(1 - q_{\mathrm{AI}}) q_{\mathrm{DM}}. \tag{1}$$

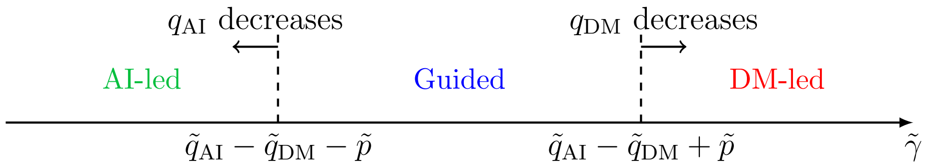
q_{AI} decreases
q_{DM} decreases
AI-led
Guided
DM-led
$\tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} - \tilde{p}$
$\tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} + \tilde{p}$
$\tilde{\gamma}$

Figure 3: The transitions between optimal strategy regimes as a function of γ. The two arrows indicate how the thresholds shift, thus expanding the Guided regime, given that the expertise levels decrease.

Two things are noteworthy. First, the correctness is symmetric in q_{AI} and q_{DM}. Second, while the Guided strategy uses two signals rather than one, it is not evident that $C_{\mathrm{G}} > \max\{C_{\mathrm{DM}}, C_{\mathrm{AI}}\}$ in general. For example, when $q_{\mathrm{AI}} \ll q_{\mathrm{DM}}$, the information of the AI is relatively inaccurate, and the correctness is improved when ignoring it and adopting the DM-led strategy. For $\tilde{\gamma} \neq 0$, this results in a discontinuity of the correctness at the transitions between the regions of optimality, which is the driving force of our next two results. We note that for $\tilde{\gamma} = 0$ such discontinuity does not exist, and that $C_{\mathrm{G}} > \max\{C_{\mathrm{DM}}, C_{\mathrm{AI}}\}$ given an optimal Guided strategy, as depicted in Figure 4.

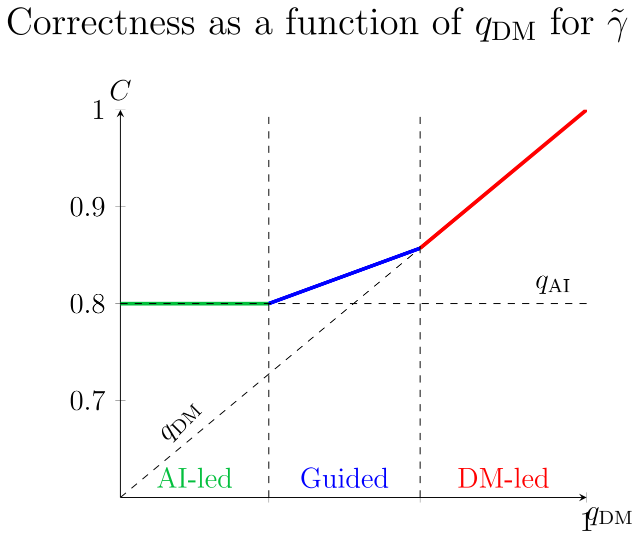

Figure 4: The correctness of the process as a function of the DM's expertise level, and a fixed level of $q_{\mathrm{AI}} = 0.8$ for the AI, and $p = 0.6$. The dashed vertical lines divide the axis into the different optimal strategy regions (from left to right): AI-led, Guided, and DM-led. Notably, the Guided regime supports a correctness level that individually supersedes the expertise levels of both agents.

3.2 Key insights

The following Proposition 1, which builds on Theorem 1, presents our first main insight: a better-informed AI does not necessarily improve the correctness of the process. In fact, one can find two disjoint intervals such that *every* expertise level of the AI in the lower interval generates strictly higher correctness than *every* expertise level in the higher interval. The limit value that separates these two intervals is referred to as *the dependency threshold* – the highest level that allows for a strategic learning process. Above this level, the AI is sufficiently good that the DM fully relies on it and completely ignores his private signal.

Proposition 1 (The dependency threshold). *Fix $(q_{\mathrm{DM}}, \gamma, p)$ such that $\tilde{q}_{\mathrm{DM}} > -\tilde{\gamma} > \tilde{p}$. There exists an interval $(\underline{q}_{\mathrm{AI}}, \overline{q}_{\mathrm{AI}})$ and an interior value q^*, such that for every $q^-_{\mathrm{AI}} \in (\underline{q}_{\mathrm{AI}}, q^*)$ and every $q^+_{\mathrm{AI}} \in (q^*, \overline{q}_{\mathrm{AI}})$, the correctness is higher when the* AI *has the lower quality signal, i.e., $C(q^-_{\mathrm{AI}}, q_{\mathrm{DM}}) > C(q^+_{\mathrm{AI}}, q_{\mathrm{DM}})$.*

Three clarifications are in order. First, the preliminary assumption that $\tilde{q}_{\mathrm{DM}} > -\tilde{\gamma} > \tilde{p}$ enables us to shift from a Guided equilibrium to an AI-led one, by varying the AI's expertise level q_{AI}. Otherwise, e.g., in case $\tilde{\gamma}$ is significantly smaller than $-\tilde{q}_{\mathrm{DM}} - \tilde{p}$, we are left only with an AI-led optimal strategy, independently of q_{AI}. Second, the transition between regimes hinges on the fact $\tilde{\gamma} < 0$, which implies that $U_1 < |U_4|$. In other words, the positive externalities are small compared to the negative ones, so the DM is inclined to ignore his private information and blindly follow the AI, further along the guided equilibrium. Third, note that the potential loss from crossing the dependency threshold q^*_{DM} is not necessarily small. Figure 5 illustrates the potential magnitude of this non-monotone effect, where an infinitesimal increase in q_{DM} triggers a drop from 0.81 to 0.75 in the overall correctness value.

The following Proposition 2 mirrors the previous result, so that a better-informed DM completely ignores the AI's assessment only to produce an inferior result, in terms of correctness. The limit value where the regime shifts from an optimal Guided strategy to an optimal DM-led one is referred to as *the humility threshold* – the highest level that enables some strategic learning.

Proposition 2 (The humility threshold). *Fix $(q_{\mathrm{AI}}, \gamma, p)$ such that $\tilde{q}_{\mathrm{AI}} > \tilde{\gamma} > \tilde{p}$. There exists an interval $(\underline{q}_{\mathrm{DM}}, \overline{q}_{\mathrm{DM}})$ and an interior value q^*, such that for every $q^-_{\mathrm{DM}} \in (\underline{q}_{\mathrm{DM}}, q^*)$ and every $q^+_{\mathrm{DM}} \in (q^*, \overline{q}_{\mathrm{DM}})$, the correctness is higher when the* DM *has the lower quality signal, i.e., $C(q_{\mathrm{AI}}, q^-_{\mathrm{DM}}) > C(q_{\mathrm{AI}}, q^+_{\mathrm{DM}})$.*

Similarly to Proposition 1, the condition $\tilde{q}_{\mathrm{AI}} > \tilde{\gamma} > \tilde{p}$ enables us to shift from a Guided equilibrium to a DM-led one, by varying the DM's accuracy. Moreover, the fact that $\tilde{\gamma} > 0$ suggests that $U_1 > |U_4|$, such that positive externalities overtake the negative ones, and the DM is inclined to strictly follow his private signal.

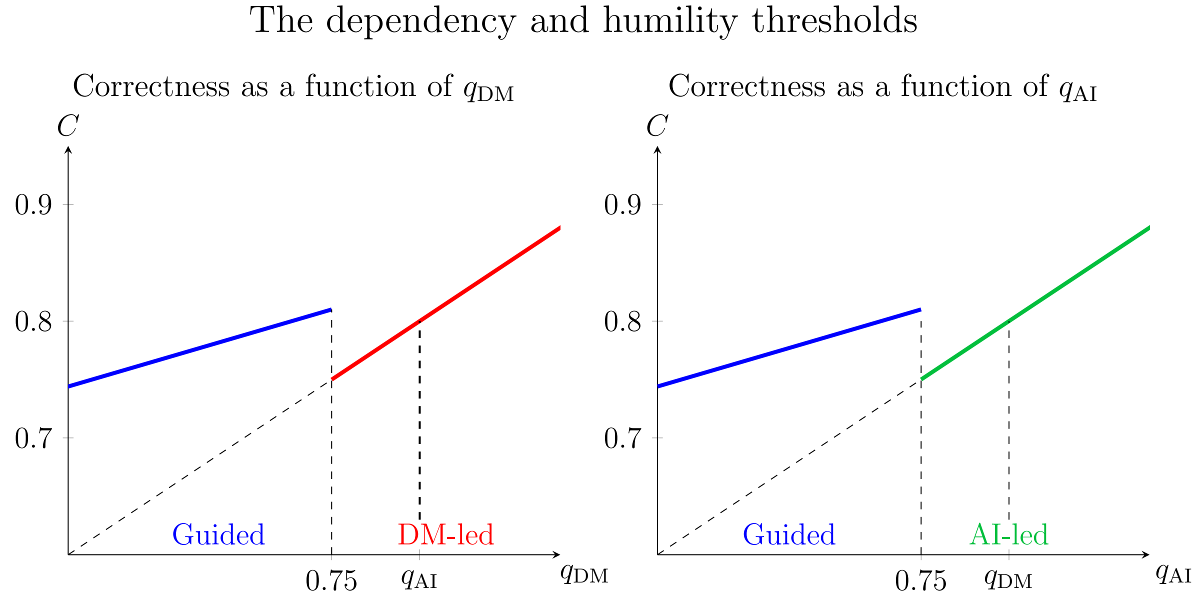

Figure 5: An illustration of the humility (left) and dependency (right) thresholds. Left: The correctness of the process as a function of the DM's expertise level, given $p = 0.6, q_{\mathrm{AI}} = 0.8$, and $\gamma = 2$. The dotted (blue) line describes the correctness under a Guided strategy, and the solid (red) line describes the correctness under a DM-led strategy. An increase in the DM's expertise level above the threshold causes the correctness to drop. Note that the Humility threshold is below q_{AI}. Right: The correctness of the process as a function of the AI's expertise level, given $p = 0.6, q_{\mathrm{DM}} = 0.8$, and $\gamma = 0.5$. The dotted (blue) line describes the correctness under a Guided strategy, and the solid (green) line describes the correctness under an AI-led strategy. An increase in the AI's expertise level above the threshold causes the correctness to drop. Note that the dependency threshold is below q_{DM}.

Our final proposition addresses the hypothetical decision between opting for a better-informed DM versus a better-informed AI. Proposition 3 demonstrates that, in terms of correctness, there are instances where it is advantageous to pair a less-informed DM with a well-informed AI. Conversely, a setup where the DM is more informed than the AI can restrict the learning process inherent in a DM-led regime.

Proposition 3. *Assume $\tilde{\gamma} > \tilde{p}$ and fix $q^H > q^L$ such that $\tilde{q}^H - \tilde{q}^L < \tilde{p}$. In case $(q_{\text{AI}}, q_{\text{DM}}) = (q^H, q^L)$ yields the* Guided optimal strategy *with correctness $C_G(q^H, q^L)$, then reversing the expertise levels to $(q_{\text{AI}}, q_{\text{DM}}) = (q^L, q^H)$ would make the* DM-led strategy *optimal with a lower correctness of $C_{\text{DM}}(q^L, q^H) < C_G(q^H, q^L)$. Hence, higher correctness is obtained when the* DM *is given the lower quality signal.*

To gain some intuition for this result, consider the signals that both agents receive. Once the DM is better-informed, it is sub-optimal for him to deviate from his relatively high-accuracy signal toward the AI's less-accurate assessment. In practice, the ordering $(q_{\text{AI}}, q_{\text{DM}}) = (q^L, q^H)$ yields a DM-led optimal strategy, which implies that the DM's decision is based entirely on his own (superior) signal. In case the relative accuracy reverses, the less-informed DM takes into account all available information and reverts to the AI's recommendation if it is also supported by the prior.

3.2.1 One-sided externalities

This analysis presupposes the existence of both positive and negative externalities – that is, the DM prefers the AI's recommendations to be incorrect, whether the DM's own decision is correct or incorrect. Nevertheless, our findings remain valid even if one of these externalities is absent, meaning either $U_1 = U_2$ or $U_4 = U_3$. To see this, redefine the non-conformity ratio as $U_1 = \gamma(1 - U_4) - 1$, so that infinitely many pairs of (U_1, U_4) yield the same γ. In particular, setting $U_4 = U_3 = -1$ results in $U_1 = 2\gamma - 1$. Nevertheless, the presence of at least one externality is essential for our conclusions to hold. Otherwise, the strategic consideration is lost, and the DM faces a simple maximization problem given the two signals. This scenario typically arises when externalities are symmetric, i.e., $U_1 = -U_4$ which corresponds to $\gamma = 1$.

4 In conclusion

This paper provides an analysis of AI-assisted decision-making processes. Our analysis depicts a mapping from the agents' expertise levels to the DM's optimal strategy, showing that: (i) better-informed rational agents may generate worse recommendations, and (ii) the outcome is typically improved when the agent with the higher level is the AI advisor, which provides the initial assessment. We conclude that there is some benefit from regulating the integration of AI in the intermediate phase where the agents' expertise levels are rather close.

One path for such rational integration involves adjusting the liability for the DM in the event of an error, or the reward for a correct decision. By implementing this approach, it is possible to incentivize the DM to prioritize the correctness of the decision-making process. Such optimal adjustments of liability and reward are typically robust to small fluctuations in expertise levels and prior probabilities. Consequently, external stakeholders can ensure that the decision-making process is aligned with their interests.

Moreover, the Guided strategy enhances resource utilization efficiency, especially concerning human resources. According to this strategy, the DM defers to the AI's assessment when $a_{\mathrm{AI}} = 0$, regardless of their own information. This protocol positions the AI as a preliminary screening device, primarily tasked with identifying potential anomalies (the rare $\theta = 1$ state), so that the DM is consulted only when an anomaly is indeed detected, effectively serving as a secondary-level analysis. It aligns with the common practice of allocating simpler tasks to computers, while reserving human interventions for more complex scenarios. Our results indicate that this method is not only practical but efficient as well.

A Proofs

A.1 Proof of Theorem 1

Proof. Fix $q_{\mathrm{AI}}, q_{\mathrm{DM}}, p$, and γ. Our analysis is divided into four different states depending on $a_{\mathrm{AI}} = s_{\mathrm{AI}}$ and s_{DM}.

Consider first the case where $a_{\mathrm{AI}} = 0$ and $s_{\mathrm{DM}} = 0$. This state occurs with probability $q_{\mathrm{AI}} q_{\mathrm{DM}}$ when $\theta = 0$, and with probability $(1 - q_{\mathrm{AI}})(1 - q_{\mathrm{DM}})$ when $\theta = 1$. The DM's best response in this state is $a_{\mathrm{DM}} = 0$ if and only if

$$p q_{\mathrm{AI}} q_{\mathrm{DM}} \cdot 1 + (1-p)(1-q_{\mathrm{AI}})(1-q_{\mathrm{DM}}) \cdot (-1) \geq p q_{\mathrm{AI}} q_{\mathrm{DM}} \cdot U_4 + (1-p)(1-q_{\mathrm{AI}})(1-q_{\mathrm{DM}}) \cdot U_1,$$

(the best response is $a_{\mathrm{DM}} = 1$ when the inequality is reversed). The last inequality can be rearranged into

$$\frac{p}{1-p} \cdot \frac{q_{\mathrm{AI}}}{1-q_{\mathrm{AI}}} \cdot \frac{q_{\mathrm{DM}}}{1-q_{\mathrm{DM}}} \cdot \frac{1-U_4}{1+U_1} \geq 1,$$

or, equivalently,

$$\tilde{p} + \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} - \tilde{\gamma} \geq 0. \tag{2}$$

Next, consider the case where $a_{\mathrm{AI}} = 0$ and $s_{\mathrm{DM}} = 1$. This state occurs with probability $q_{\mathrm{AI}}(1 - q_{\mathrm{DM}})$ when $\theta = 0$, and with probability $(1 - q_{\mathrm{AI}}) q_{\mathrm{DM}}$ when $\theta = 1$. The DM's best response in this state is $a_{\mathrm{DM}} = 0$ if and only if

$$p q_{\mathrm{AI}}(1-q_{\mathrm{DM}}) \cdot 1 + (1-p)(1-q_{\mathrm{AI}}) q_{\mathrm{DM}} \cdot (-1) \geq p q_{\mathrm{AI}}(1-q_{\mathrm{DM}}) \cdot U_4 + (1-p)(1-q_{\mathrm{AI}}) q_{\mathrm{DM}} \cdot U_1,$$

which can be rearranged into

$$\tilde{p} + \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} - \tilde{\gamma} \geq 0. \tag{3}$$

Third, consider the case where $a_{\mathrm{AI}} = 1$ and $s_{\mathrm{DM}} = 0$. This state occurs with probability $(1 - q_{\mathrm{AI}}) q_{\mathrm{DM}}$ when $\theta = 0$, and with probability $q_{\mathrm{AI}}(1 - q_{\mathrm{DM}})$ when $\theta = 1$. The DM's best response in this state is $a_{\mathrm{DM}} = 0$ if and only if

$$p(1-q_{\mathrm{AI}})q_{\mathrm{DM}} \cdot U_1 + (1-p)q_{\mathrm{AI}}(1-q_{\mathrm{DM}}) \cdot U_4 \geq p(1-q_{\mathrm{AI}})q_{\mathrm{DM}} \cdot (-1) + (1-p)q_{\mathrm{AI}}(1-q_{\mathrm{DM}}) \cdot 1,$$

which can be rearranged into

$$\tilde{p} - \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} + \tilde{\gamma} \geq 0. \tag{4}$$

Lastly, consider the case where $a_{\mathrm{AI}} = 1$ and $s_{\mathrm{DM}} = 1$. This state occurs with probability $(1-q_{\mathrm{AI}})(1-q_{\mathrm{DM}})$ when $\theta = 0$, and with probability $q_{\mathrm{AI}}q_{\mathrm{DM}}$ when $\theta = 1$. The DM's best response in this state is $a_{\mathrm{DM}} = 0$ if and only if

$$p(1-q_{\mathrm{AI}})(1-q_{\mathrm{DM}}) \cdot U_1 + (1-p)q_{\mathrm{AI}}q_{\mathrm{DM}} \cdot U_4 \geq p(1-q_{\mathrm{AI}})(1-q_{\mathrm{DM}}) \cdot (-1) + (1-p)q_{\mathrm{AI}}q_{\mathrm{DM}} \cdot 1,$$

which can be rearranged into

$$\tilde{p} - \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} + \tilde{\gamma} \geq 0. \tag{5}$$

In a **DM-led strategy**, $a_{\mathrm{DM}}(\cdot, s_{\mathrm{DM}}) = s_{\mathrm{DM}}$, so it is optimal if Ineq. (2) and Ineq. (4) hold while Ineq. (3) and Ineq. (5) are reversed:

$$\tilde{p} + \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} - \tilde{\gamma} \geq 0, \tag{6a}$$

$$\tilde{p} + \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} - \tilde{\gamma} \leq 0, \tag{6b}$$

$$\tilde{p} - \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} + \tilde{\gamma} \geq 0, \tag{6c}$$

$$\tilde{p} - \tilde{q}_{\mathrm{AI}} - \tilde{q}_{\mathrm{DM}} + \tilde{\gamma} \leq 0. \tag{6d}$$

Note that Ineq. (6d) implies that Ineq. (6a) holds, and Ineq. (6b) yields Ineq. (6c). Hence, a DM-led strategy is optimal if and only if $\tilde{q}_{AI} - \tilde{q}_{\mathrm{DM}} + \tilde{p} \leq \tilde{\gamma} \leq \tilde{q}_{\mathrm{AI}} + \tilde{q}_{\mathrm{DM}} - \tilde{p}$.

In an **AI-led strategy**, $a_{\mathrm{DM}}(a_{\mathrm{AI}}, \cdot) = a_{\mathrm{AI}}$, so Ineq. (2) and Ineq. (3) should hold while

Ineq. (4) and Ineq. (5) are reversed:

$$\tilde{p} + \tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} - \tilde{\gamma} \geq 0, \tag{7a}$$

$$\tilde{p} + \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} - \tilde{\gamma} \geq 0, \tag{7b}$$

$$\tilde{p} - \tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} + \tilde{\gamma} \leq 0, \tag{7c}$$

$$\tilde{p} - \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} + \tilde{\gamma} \leq 0. \tag{7d}$$

Note that Ineq. (7d) implies that Ineq. (7a) holds, and Ineq. (7c) yields Ineq. (7b) and Ineq. (7d). Hence, an AI-led strategy is optimal if and only if $\tilde{\gamma} \leq \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} - \tilde{p}$.

In a **Guided strategy**, $a_{\text{DM}}(a_{\text{AI}}, s_{\text{DM}}) = s_{\text{DM}}$ except for $a_{\text{DM}}(0, 1) = 0$ so Ineq. (2), Ineq. (3) and Ineq. (4) should hold while Ineq. (5) is reversed:

$$\tilde{p} + \tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} - \tilde{\gamma} \geq 0, \tag{8a}$$

$$\tilde{p} + \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} - \tilde{\gamma} \geq 0, \tag{8b}$$

$$\tilde{p} - \tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} + \tilde{\gamma} \geq 0, \tag{8c}$$

$$\tilde{p} - \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} + \tilde{\gamma} \leq 0. \tag{8d}$$

Note that Ineq. (8d) implies that Ineq. (8a) holds, and Ineqs. (8c) and (8b) are equivalent to $\tilde{\gamma} - \tilde{p} \leq \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} \leq \tilde{\gamma} + \tilde{p}$. Since $q_{\text{DM}} \geq p$, one can show that Ineq. (8b) yields Ineq. (8d) as follows,

$$\tilde{q}_{\text{AI}} + \tilde{q}_{\text{DM}} \geq \tilde{q}_{\text{AI}} + \tilde{p} \geq \tilde{\gamma} + \tilde{q}_{\text{DM}} \geq \tilde{\gamma} + \tilde{p}.$$

So, one only needs to sustain the inequalities $\tilde{\gamma} - \tilde{p} \leq \tilde{q}_{\text{AI}} - \tilde{q}_{\text{DM}} \leq \tilde{\gamma} + \tilde{p}$, as stated in the theorem. This concludes our proof. □

A.2 Proof of Proposition 1

Proof. Fix $(q_{\text{DM}}, \gamma, p)$ so that $\tilde{q}_{\text{DM}} > -\tilde{\gamma} > \tilde{p}$. Take $\tilde{q}^* = \tilde{\gamma} + \tilde{q}_{\text{DM}} + \tilde{p}$. The condition $\tilde{q}_{\text{DM}} > -\tilde{\gamma} > \tilde{p}$ assures that $\tilde{\gamma} + \tilde{q}_{\text{DM}} > 0$, which implies that $\tilde{q}^* = \tilde{\gamma} + \tilde{q}_{\text{DM}} + \tilde{p} > \tilde{p}$. In addition, we get that $-\tilde{\gamma} - \tilde{p} > 0$, so that $\tilde{q}_{\text{DM}} = \tilde{q}^* - \tilde{\gamma} - \tilde{p} > \tilde{q}^* > \tilde{p}$, as needed.

From Theorem 1, we know that the optimal strategy for every sufficiently close $q_{\mathrm{AI}} < q^*$ is the Guided strategy (note that q_{AI} cannot be too small so that the optimal strategy is not the DM-led), and the optimal strategy for every $q_{\mathrm{AI}} > q^*$ is the AI-led one. Thus, there exists an open interval $(\underline{q}_{\mathrm{AI}}, \overline{q}_{\mathrm{AI}})$ which contains q^*, such that for every $q_{\mathrm{AI}}^- \in (\underline{q}_{\mathrm{AI}}, q^*)$ the optimal strategy is a Guided one, whereas for every $q_{\mathrm{AI}}^+ \in (q^*, \overline{q}_{\mathrm{AI}})$ the optimal strategy is an AI-led strategy.

Let us compute the correctness in both cases. Under an AI-led strategy, the correctness is q_{AI}, whereas under a Guided strategy, the correctness is

$$C(q_{\mathrm{AI}}, q_{\mathrm{DM}}) \quad = \quad q_{\mathrm{AI}} + pq_{\mathrm{DM}}(1 - q_{\mathrm{AI}}) - (1-p)(1-q_{\mathrm{DM}})q_{\mathrm{AI}}.$$

The fact that $\tilde{\gamma} < 0$ implies that $\tilde{q}^* = \tilde{\gamma} + \tilde{q}_{\mathrm{DM}} + \tilde{p} < \tilde{q}_{\mathrm{DM}} + \tilde{p}$. Thus, $\tilde{q}_{\mathrm{DM}} + \tilde{p} > \tilde{q}^*$ which suggests that

$$pq_{\mathrm{DM}}(1 - q^*) > (1-p)(1-q_{\mathrm{DM}})q^*,$$

or equivalently,

$$q^* + pq_{\mathrm{DM}}(1 - q^*) - (1-p)(1-q_{\mathrm{DM}})q^* > q^*,$$

and the LHS is indeed the correctness under a Guided strategy. Hence, $\lim\limits_{q_{\mathrm{AI}} \to q^*-} C_G(q_{\mathrm{AI}}, q_{\mathrm{DM}}) > \lim\limits_{q_{\mathrm{AI}} \to q^*+} q_{\mathrm{AI}}$, and the result follows by the continuity of the correctness function on both sides of q^*. □

A.3 Proof of Proposition 2

Proof. The proof follows similarly to the proof of Proposition 1. Fix $(q_{\mathrm{AI}}, \gamma, p)$ so that $\tilde{q}_{\mathrm{AI}} > \tilde{\gamma} > \tilde{p}$ and take q^* such that $\tilde{\gamma} - \tilde{p} = \tilde{q}_{\mathrm{AI}} - \tilde{q}^*$. Since $\tilde{q}_{\mathrm{AI}} > \tilde{\gamma} > \tilde{p}$, we can fix such q^* without violating the conditions of $\min\{\tilde{q}_{\mathrm{DM}}, \tilde{q}_{\mathrm{AI}}\} > \tilde{p}$. This also implies that $\tilde{\gamma} > 0$.

From Theorem 1, we know that the optimal strategy for $q_{\mathrm{DM}} > q^*$ is the DM-led one, and the optimal strategy for $q_{\mathrm{DM}} < q^*$ (but not too small, so that the optimal strategy is not the AI-led) is the Guided one. The correctness in the first case is simply q_{DM} and in the second case is C_G given in Eq. (1). Similarly to Proposition 1, it is straightforward to verify that

$\lim_{q_{\text{DM}} \to q^*-} C_G(q_{\text{AI}}, q_{\text{DM}}) > \lim_{q_{\text{DM}} \to q^*+} q_{\text{DM}}$. The result follows from the continuity of the correctness function in each of the regions $q_{\text{DM}} > q^*$ and $q_{\text{DM}} < q^*$. □

A.4 Proof of Proposition 3

Proof. Fix (γ, p, q^H, q^L) such that $q^H > q^L$, where $\tilde{q}^H - \tilde{q}^L < \tilde{p}$, and $(q_{\text{AI}}, q_{\text{DM}}) = (q^H, q^L)$ yields an optimal Guided strategy with correctness

$$\begin{aligned} C_G(q^H, q^L) &= q^L + pq^H(1-q^L) - (1-p)(1-q^H)q^L \\ &= q^H + pq^L(1-q^H) - (1-p)(1-q^L)q^H > q^H, \end{aligned}$$

where the first equality follows from the symmetry of the expression w.r.t. q^H and q^L, and the inequality follows from the assumption that $\tilde{p} + \tilde{q}^L > \tilde{q}^H$.

Now consider $(q_{\text{AI}}, q_{\text{DM}}) = (q^L, q^H)$. Following the proof of Theorem 1, one needs to show that Ineq. (6b) and Ineq. (6d) hold. First, note that Ineq. (8d) and Ineq. (6d) are identical, independently of the ordering of $(q_{\text{AI}}, q_{\text{DM}})$. So one only needs to show that Ineq. (6b) holds, i.e., $\tilde{q}^L - \tilde{q}^H \leq \tilde{\gamma} - \tilde{p}$. This inequality follows from the fact that $\tilde{\gamma} > \tilde{p}$ and $q^H > q^L$. Thus, we established an optimal DM-led strategy with correctness $C_{\text{DM}}(q^L, q^H) = q^H < C_G(q^H, q^L)$, as needed. □

References

Agarwal, N., A. Moehring, P. Rajpurkar, and T. Salz (2023, July). Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology.

Ali, S. N. and N. Kartik (2012, November). Herding with collective preferences. *Economic Theory 51*(3), 601–626.

Almog, D., R. Gauriot, L. Page, and D. Martin (2024, February). AI Oversight and Human Mistakes: Evidence from Centre Court. arXiv:2401.16754 [cs, econ, q-fin].

Angelova, V., W. S. Dobbie, and C. Yang (2023, September). Algorithmic Recommendations and Human Discretion.

Arabi, H., A. AkhavanAllaf, A. Sanaat, I. Shiri, and H. Zaidi (2021, March). The promise of artificial intelligence and deep learning in PET and SPECT imaging. *Physica Medica 83*, 122–137.

Arieli, I., M. Koren, and R. Smorodinsky (2018). The one-shot crowdfunding game. *ACM EC 2018 - Proceedings of the 2018 ACM Conference on Economics and Computation 18*, 213–214. arXiv: 1805.11872 Publisher: Association for Computing Machinery, Inc ISBN: 9781450358293.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics 107*(3), 797–817. Publisher: Oxford Academic.

Barragán-Montero, A., U. Javaid, G. Valdés, D. Nguyen, P. Desbordes, B. Macq, S. Willems, L. Vandewinckele, M. Holmström, F. Löfman, S. Michiels, K. Souris, E. Sterpin, and J. A. Lee (2021, March). Artificial intelligence and machine learning for medical imaging: a technology review. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB) 83*, 242–256.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy 100*(5), 992–1026. Publisher: University of Chicago Press.

Bikhchandani, S., D. A. Hirshleifer, O. Tamuz, and I. Welch (2021). Information Cascades and Social Learning. *SSRN Electronic Journal*.

Debo, L. G., C. Parlour, and U. Rajan (2011). Signaling Quality via Queues. *Management Science 58*(5), 876–891. Publisher: INFORMS.

Eyster, E., A. Galeotti, N. Kartik, and M. Rabin (2014). Congested observational learning. *Games and Economic Behavior 87*, 519–538. Publisher: Academic Press.

Kanazawa, K., D. Kawaguchi, H. Shigeoka, and Y. Watanabe (2022, October). AI, Skill, and Productivity: The Case of Taxi Drivers.

Smith, L. and P. Sorensen (2000). Pathological outcomes of observational learning. *Econometrica 68*(2), 371–398. Publisher: John Wiley & Sons, Ltd.

Smith, L., P. N. Sørensen, and J. Tian (2021, October). Informational Herding, Optimal Experimentation, and Contrarianism. *The Review of Economic Studies 88*(5), 2527–2554.

Varoquaux, G. and V. Cheplygina (2022, April). Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digital Medicine 5*(1), 1–8.

Veeraraghavan, S. K. and L. G. Debo (2011). Herding in Queues with Waiting Costs: Rationality and Regret. *Manufacturing and Service Operations Management 13*(3), 329–346. Publisher: INFORMS.