

Supplementary Information

Supplementary Note 1: Details of Analytical Model and Additional Results

To ease exposition of the results, we provide a summary of the notation in Supplementary Table 1.

Supplementary Table 1: Summary of notations

Notation	Definition	Comments
p	: True health status of a <i>patient</i> , either sick (s) or healthy (h)	$p \in \{s, h\}$
r	: The <i>risk</i> assessment by an <i>algorithm</i>	
μ_i/σ_i	: Mean/standard deviation of risk assessment by an algorithm for a sick (s) or healthy (h) patient,	$i \in \{s, h\}$
$d(r)$	: The <i>delegation</i> rule for a given risk r	
η	: A human <i>expert</i> 's risk assessment (sick vs. healthy) for a given risk r	$\eta \in \{s, h\}$
$a(r)$	: An <i>algorithm</i> 's risk assessment (sick vs. healthy) for a given risk r	$a(r) \in \{s, h\}$
λ	: The prevalence of disease in the patient population	$\lambda \in (0, 1)$
o	: Diagnostic decision <i>outcomes</i> : False negative (FN) /false positive (FP) /true positive (TP) /true negative (TN)	$o \in \{FN, FP, TP, TN\}$
$P_E(o)$	: Probability of various outcomes (o) for human expert	
t_j	: Thresholds under different strategies	$j \in \{D, A\}$
I/I_k	: Information content of algorithm/algorithm k	$k \in \{1, 2\}$
c_j	: The cost of algorithm ($.a$), human expert ($.e$), follow-up ($.f$), and liability for misdiagnosing a sick patient ($.l$).	$j \in \{a, e, f, l\}$
$C_j(t_j)$	: Expected costs under delegation or automation with threshold t_j	$j \in \{A, D\}$
C_j^*	: Optimal costs of different strategies	$j \in \{A, D\}$
A_k	: Algorithm k	$k \in \{1, 2\}$
ρ_i	: The correlation between risk assessed by an algorithm for sick (s) and healthy (h) patients	$i \in \{s, h\}$

For brevity of presentation, we let

$$t_1 := \frac{\lambda}{1-\lambda} \times \frac{P_E(TP)(c_l - c_f) - c_e}{c_e + P_E(FP)c_f} \text{ and } t_2 := \frac{\lambda}{1-\lambda} \times \frac{c_l - c_f}{c_f}. \quad (1)$$

Variables t_1 and t_2 represent design parameters for the delegation and automation strategies, respectively. They indicate the relative harm (measured in costs) due to missing disease versus that from subjecting a healthy patient to follow-up procedures under delegation and automation strategies. Note that $t_2 < 1 \iff \lambda < c_f/c_l$; we refer to this condition as low-prevalence case. Similarly, $t_2 \geq 1 \iff c_f/c_l \leq \lambda$ which we refer to as high-prevalence case. Moreover, we let $H(\pm\infty)$ denote the limit of a given function $H(t)$ as t approaches $\pm\infty$. When presenting our results, we use “increasing” (or “decreasing”) in the strict sense, and “nondecreasing” (or “nonincreasing”) to indicate weakly increasing or decreasing.

Theorem 1. *Given Assumptions 1 and 2, the optimal thresholds given by*

$$t_D^* = \frac{\mu_s + \mu_h}{2} - \frac{\ln(t_1)\sigma^2}{\mu_s - \mu_h} \text{ and } t_A^* = \frac{\mu_s + \mu_h}{2} - \frac{\ln(t_2)\sigma^2}{\mu_s - \mu_h}$$

minimize the healthcare organization’s expected costs under the delegation and automation strategies, respectively.

Observe that the optimal thresholds for delegation or follow-up are decreasing functions of t_1 and t_2 , respectively. In other words, as the relative adverse cost consequences of a false negative relative to that of a false positive increases, the thresholds decrease. The observation is consistent with intuition. In the following corollary, we conduct a comparative statics analysis on the information content thresholds for the algorithm’s performance with respect to cost parameters.

Corollary 2. *Given Assumptions 1 and 2, suppose for a set of initial parameters $I_{ED} < I_{DA}$ holds, hence, delegation is a possible alternative strategy within the parameter space.*

1. I_{ED} is non-decreasing in c_a .
2. I_{DA} is independent of c_a .
3. I_{ED} is non-decreasing in c_l .
4. I_{DA} is increasing in c_l when $\Phi\left(I_{DA} + \frac{\ln(t_2)}{I_{DA}}\right) < P_E(TP)\Phi\left(I_{DA} + \frac{\ln(t_1)}{I_{DA}}\right)$. It is decreasing in c_l when $\Phi\left(I_{DA} + \frac{\ln(t_2)}{I_{DA}}\right) > P_E(TP)\Phi\left(I_{DA} + \frac{\ln(t_1)}{I_{DA}}\right)$.

Corollary 2 provides further insights into the optimal strategy. Recall I_{ED} and I_{DA} are the thresholds for the AI algorithm's information content; the optimal strategy respectively switches from expert-only to delegation and from delegation to automation when the information content of the AI algorithm exceeds these thresholds. In Part (1) of the Corollary, we observe that, ceteris paribus, reductions in algorithmic costs can allow a poorer-performing algorithm to become a part of the workflow. This effect occurs through involving AI as part of a strategy combining human and machine for different tasks, i.e., the information content threshold for switching from expert-only to delegation is non-decreasing in c_a . In Part (2), however, as expected and ceteris paribus, we find that changing the algorithmic cost has no impact on the choice between the delegation and automation strategy. This result follows from the fact that both the delegation and automation strategies utilize the AI algorithm for all patients. The finding in Part (3) on the impact of litigation cost for switching from expert-only to delegation, I_{ED} , has a similar intuition as in Part (1). Yet, Part (4) presents a curious case; the results suggests that human-machine combined strategy (i.e., delegation) may become more or less desirable against the machine-only strategy (i.e., automation) with increasing litigation costs. The direction of which strategy is more preferable depends on the relative sensitivity of the AI algorithm to that of the expert.

The comparative statics analysis of optimal costs under different strategies suggests that, as would be expected, the healthcare organization can realize reductions in total costs due to changes in the workflow when either the expert's performance or the algorithm's discriminative ability increases.

Corollary 3. *Given Assumptions 1 and 2, let $I' = I_{ED}|_{c_a=0}$ and $I'' = I_{DA}$ be specific information content levels corresponding to algorithm performance and assume that $I' < I''$. Let an algorithm have information content I .*

1. *When $I < I'$, expert-only is optimal regardless of c_a .*
2. *When $I' < I < I''$, there exists c'_a such that D is optimal if $c_a < c'_a$ and E is optimal otherwise.*
3. *If $I > I''$, there exists c''_a such that A is optimal if $c_a < c''_a$ and E is optimal otherwise.*

Corollary 3 characterizes the optimal strategy from the perspective of the algorithm's performance (as indicated by the information content) against its cost. Note that $I' < I''$ ensures that both the delegation and the automation strategies can be optimal in the parameter space. When the algorithm has a low performance, the optimal design does not use the algorithm regardless of its costs (Part 1). However, when the algorithm has a moderate performance, either the delegation or the expert-only strategies can be the optimal design depending on the cost; delegation is optimal when the algorithm's cost is low and expert-only is optimal otherwise (Part 2). Delegation is no longer an optimal strategy when the algorithm has high performance. Automation is optimal when the algorithm's cost is low and expert-only is optimal otherwise (Part 3). These results suggest that the algorithm's cost and performance determine whether AI fully replaces humans or a human-machine hybrid strategy is required. Also, a high algorithm-performance does not necessarily mean full automation or even a partial automation through delegation. Instead, for a high-performing algorithm, the healthcare organization could rely solely on human experts when algorithm costs

are high. This result underscores the importance of considering economic constraints alongside technological advancements when making decisions about AI integration in healthcare.

Next, we present a result for the asymmetric liability case. We modify our model to study the case in which human vs. machine liability are different. Specifically, suppose that $c_{lh} \leq c_{la}$ holds, where c_{lh} is the litigation cost for the human expert and c_{la} is the litigation cost for the algorithm when a sick patient is deemed as healthy. Next proposition characterizes optimal strategy under this asymmetric litigation case.

Proposition 3. *Given Assumptions 1 and 2, let $c_{lh} = c_{la}$ be the benchmark scenario and problem parameters are such that, for some values of algorithm performance, delegation is the optimal strategy under this benchmark. Assume c_{lh} is fixed while c_{la} increases.*

1. *If for a given I the optimal strategy is delegation under the benchmark, then there exists c'_{la} such that for $c_{la} > c'_{la}$ the optimal strategy transitions to an expert-only strategy.*

2. *If for a given I the optimal strategy is automation under the benchmark, then;*

(a) *If*

$$P_E(TP)\lambda[c_{lh} - c_f] - P_E(FP)(1 - \lambda)c_f \leq (c_e - c_a) + \lambda c_{lh} - c_f,$$

then the optimal strategy is automation $\forall c_{la} > c_{lh}$.

(b) *Otherwise, if*

$$(c_e - c_a) + \lambda c_{lh} - c_f < P_E(TP)\lambda[c_{lh} - c_f] - P_E(FP)(1 - \lambda)c_f$$

then there exists c''_{la} such that for $c_{la} > c''_{la}$ the optimal strategy switches to the expert only strategy.

Supplementary Note 2: Proofs

We first present a preparatory Lemma that is needed to prove Theorem 1. Since, we use the finding from this Lemma in subsequent results, we present it using a generic notation. Under our general setup of conditional independence based on patients' true health status, the risk assessment r by the algorithm follows a bivariate normal distribution. Specifically, let $f_+(r) \sim \mathcal{N}(\mu_s, \sigma)$ and $f_-(r) \sim \mathcal{N}(\mu_h, \sigma)$ denote the conditional probability distribution function of r , respectively for the $+$ (analogous to patients being sick) and the $-$ (analogous to patient being healthy) class. Without loss of generality assume that $\mu_h < \mu_s$. Next, define

$$C(t) = C_0 + c_s \int_t^\infty f_+(r)dr + c_h \int_t^\infty f_-(r)dr, \quad (2)$$

where we assume that $c_h > 0$.

Lemma 1. *For the cost function in (2), the following holds.*

1. *If $c_s \geq 0$, then $C(t)$ is decreasing in t , with $C(\infty) = C_0$.*
2. *If $c_s < 0$, then $C(t)$ is first decreasing then increasing in t . Moreover, the closed form solution for the optimal threshold and cost that minimizes $C(t)$ is given as:*

$$\begin{aligned} t^* &= \frac{\mu_h + \mu_s}{2} - \frac{\ln(c_s/c_h)\sigma^2}{\mu_s - \mu_h}, \\ C^*(I) &= C_0 + c_s \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) \\ &\quad + c_h \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right). \end{aligned} \quad (3)$$

Proof of Lemma 1: We first take the derivative of $C(t)$ with respect to t :

$$\frac{dC(t)}{dt} = \frac{c_s}{\sigma} f_+(r) \Big|_{r=t}^{r=\infty} + \frac{c_h}{\sigma} f_-(r) \Big|_{r=t}^{r=\infty}$$

$$= -c_s \frac{e^{-\frac{(t-\mu_s)^2}{2\sigma^2}}}{\sigma} - c_h \frac{e^{-\frac{(t-\mu_h)^2}{2\sigma^2}}}{\sigma}, \quad (4)$$

where we used the fact that normal distribution approaches 0 as t approaches ∞ ($f_+(\infty) = f_-(\infty) = 0$). When $c_s \geq 0$, then the derivative of $C(t)$ is negative since $c_h > 0$. Hence, $C(t)$ is decreasing in t while it converges to C_0 at infinity, i.e., $C(\infty) = C_0$. When $c_s < 0$, for clarity of exposition, substitute c_s with $-c_s$ as such $c_s > 0$ holds. Let us define:

$$G(t) := \frac{c_s}{c_h} e^{\frac{(\mu_h^2 - \mu_s^2)}{2\sigma^2}} e^{\frac{t(\mu_s - \mu_h)}{\sigma^2}} - 1.$$

Then, we rewrite the derivative of $C(t)$ as

$$\begin{aligned} \frac{dC(t)}{dt} &= c_s \frac{e^{-\frac{(t-\mu_s)^2}{2\sigma^2}}}{\sigma} - c_h \frac{e^{-\frac{(t-\mu_h)^2}{2\sigma^2}}}{\sigma} \\ &= \underbrace{c_h e^{-\frac{(t-\mu_h)^2}{2\sigma^2}}}_{\geq 0} \left(\frac{c_s}{c_h} e^{\frac{(\mu_h^2 - \mu_s^2)}{2\sigma^2}} e^{\frac{t(\mu_s - \mu_h)}{\sigma^2}} - 1 \right) \end{aligned} \quad (5)$$

$$= c_h e^{-\frac{(t-\mu_h)^2}{2\sigma^2}} G(t). \quad (6)$$

From (6), we see that the sign of $\frac{dC(t)}{dt}$ is the same as $G(t)$. So we analyze the sign of $G(t)$ next. Due to our assumption that $\mu_h < \mu_s$, we have

$$\lim_{t \rightarrow -\infty} G(t) = \lim_{t \rightarrow -\infty} \left(\frac{c_s}{c_h} e^{\frac{(\mu_h^2 - \mu_s^2)}{2\sigma^2}} e^{\frac{t(\mu_s - \mu_h)}{\sigma^2}} - 1 \right) = -1,$$

and

$$\lim_{t \rightarrow \infty} G(t) = \lim_{t \rightarrow \infty} \left(\frac{c_s}{c_h} e^{\frac{(\mu_h^2 - \mu_s^2)}{2\sigma^2}} e^{\frac{t(\mu_s - \mu_h)}{\sigma^2}} - 1 \right) = \infty,$$

which implies that $\lim_{t \rightarrow -\infty} \frac{dC(t)}{dt} < 0$ and $\lim_{t \rightarrow \infty} \frac{dC(t)}{dt} > 0$ (since $\frac{dC(t)}{dt}$ and $G(t)$ have the same sign). Moreover, observe that for any t

$$\frac{dG(t)}{dt} = \frac{c_s}{c_h} e^{\frac{(\mu_h^2 - \mu_s^2)}{2\sigma^2}} e^{\frac{t(\mu_s - \mu_h)}{\sigma^2}} \frac{\mu_s - \mu_h}{\sigma^2} > 0, \quad (7)$$

where we used our assumptions that $\mu_s > \mu_h$, $c_s > 0$ (due to reversal of sign above), and $c_h > 0$. Equation (7) implies that $G(t)$ is increasing in t . Therefore, from the Intermediate Value Theorem, there exists a unique t^* such that

$$G(t) < 0 \text{ for } t < t^* \text{ and } 0 < G(t) \text{ for } t^* < t.$$

Since $G(t)$ and $\frac{dC(t)}{dt}$ have the same sign, $\frac{dC(t)}{dt} < 0$ for $t < t^*$ and $\frac{dC(t)}{dt} > 0$ for $t > t^*$. Therefore, $C(t)$ is first decreasing and then increasing in t . Moreover, t^* is the value of t such that $\frac{dC(t)}{dt} = 0$. Hence, t^* minimizes the cost function $C(t)$. Using (6), we can calculate the closed-form solution for t^* as

$$t^* = \frac{\mu_h + \mu_s}{2} - \frac{\ln(c_s/c_h)\sigma^2}{\mu_s - \mu_h}.$$

Finally, by substituting $t = t^*$ into (2) and using the definition of $\Phi(\cdot)$, the cumulative distribution function of the standard normal distribution, we can calculate the optimal cost $C^*(I)$ as

$$C^*(I) = C_0 + c_s \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) + c_h \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right),$$

which completes the proof of the Lemma. $\square$

We are now ready to prove Theorem 1.

Proof of Theorem 1: For a given decision threshold, the cost for delegation and automation strategies are respectively given as:

$$\begin{aligned} C_D(t_D) &= c_a + \lambda c_l \\ &+ P(r \geq t_D | s) \lambda [c_e + P_E(TP)(c_f - c_l)] \\ &+ P(r \geq t_D | h) (1 - \lambda) [c_e + P_E(FP)c_f], \end{aligned} \tag{8}$$

$$\begin{aligned}
C_A(t_A) &= c_a + \lambda c_l \\
&+ P(r \geq t_A | s) \lambda (c_f - c_l) \\
&+ P(r \geq t_A | h) (1 - \lambda) c_f.
\end{aligned} \tag{9}$$

Due to Assumption 1 for $C_A(t_A)$ and Assumption 2 for $C_D(t_D)$, $C_A(t_A)$ and $C_D(t_D)$ are in the form specified in (2). Therefore, the closed form solutions for t_D^* and t_A^* follow from Lemma 1. $\square$

To prove Proposition 1, we need two preparatory Lemmas which we present next. As these results will apply to multiple subsequent results, we again present them in generic form and drop the subscripts for cost, $C(\cdot)$ and thresholds t .

Lemma 2. *The optimal cost $C^*(I)$ given in (3) is decreasing in I for $I \in [0, \infty]$.*

Proof of Lemma 2: Note that the closed-form solution for the optimal cost $C^*(I)$ given in (3) holds when $c_s < 0$. For clarity of exposition, again substitute c_s with $-c_s$ as such $c_s > 0$ holds. We can rewrite $C^*(I)$ as

$$\begin{aligned}
C^*(I) &= C_0 - c_s \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) + c_h \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) \\
&= C_0 + c_h \left(-\frac{c_s}{c_h} \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) + \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) \right).
\end{aligned}$$

Define

$$G(I) := -\frac{c_s}{c_h} \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right) \right) + 1 - \Phi \left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I} \right),$$

then we have

$$C^*(I) = C_0 + c_h G(I). \tag{10}$$

From (10), we see that the derivative of $C^*(I)$ with respect to I has the same sign as $G(I)$ since $c_h > 0$ and C_0 does not depend on I . Therefore, if $G(I)$ is decreasing

in I , then $C^*(I)$ is also decreasing in I . To examine the behavior and see whether it is a decreasing function, let us take the derivative of $G(I)$ with respect to I . The derivative of $G(I)$ is given as:

$$\begin{aligned}
\frac{dG(I)}{dI} &= \left[\frac{c_s}{c_h} e^{-\frac{\left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I}\right)^2}{2}} \left(-\frac{1}{2} + \frac{\ln(c_s/c_h)}{I^2} \right) - e^{-\frac{\left(\frac{I}{2} - \frac{\ln(c_s/c_h)}{I}\right)^2}{2}} \left(\frac{1}{2} + \frac{\ln(c_s/c_h)}{I^2} \right) \right] \\
&= e^{-\frac{\left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I}\right)^2}{2}} \left[\frac{c_s}{c_h} \left(-\frac{1}{2} + \frac{\ln(c_s/c_h)}{I^2} \right) - \frac{c_s}{c_h} \left(\frac{1}{2} + \frac{\ln(c_s/c_h)}{I^2} \right) \right] \\
&= -e^{-\frac{\left(-\frac{I}{2} - \frac{\ln(c_s/c_h)}{I}\right)^2}{2}} \frac{c_s}{c_h} < 0, \quad \forall I \in [0, \infty],
\end{aligned} \tag{11}$$

which implies that $G(I)$ is decreasing in I and as such $C^*(I)$ is also decreasing in I . $\square$

Lemma 3. *For a given I , optimal system cost under different strategies are given as follows:*

$$\begin{aligned}
C_E &= c_e + \lambda c_l \\
&\quad - \lambda(c_l - c_f)P_E(TP) + (1 - \lambda)c_f P_E(FP) \\
C_D^*(I) &= c_a + \lambda c_l \\
&\quad - \lambda [P_E(TP)(c_l - c_f) - c_e] \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(t_1)}{I} \right) \right) \\
&\quad + (1 - \lambda) [P_E(FP)c_f + c_e] \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(t_1)}{I} \right) \right) \\
C_A^*(I) &= c_a + \lambda c_l - \lambda [c_l - c_f] \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(t_2)}{I} \right) \right) \\
&\quad + (1 - \lambda)c_f \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(t_2)}{I} \right) \right).
\end{aligned} \tag{12}$$

Moreover, $C_D^*(I)$ and $C_A^*(I)$ are decreasing in I .

Proof of Lemma 3: The formula for C_E follows directly from the definition. To find the closed form solutions for the delegation and automation strategies for a given algorithm, we rely on intermediate results presented earlier. Accordingly, notice that

both $C_D(t_D)$ and $C_A(t_A)$ respectively defined in (8) and (9) are in the form given in (2) (see Lemma 1). Therefore, the optimal costs for delegation and automation strategies, respectively $C_D^*(I)$ and $C_A^*(I)$, can be calculated using (3), which gives the desired result. To complete the proof, we need to show that $C_D^*(I)$ and $C_A^*(I)$ are decreasing in I . For this, we will use Lemma 2, which claims that any cost function in the form (2) is decreasing in I . Since both $C_D^*(I)$ and $C_A^*(I)$ are in the form of (2), the desired result follows. $\square$

We are now prepared to prove the first result presented in the main manuscript, which will be shown next.

Proof of Proposition 1: To prove this result, we need to compare C_E , $C_D^*(I)$, and $C_A^*(I)$ using three claims. In the following, we sequentially state each claim followed by its proof.

Claim 1: If $P_E(TP)$ and $P_E(FP)$ are such that

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e - c_a, \quad (13)$$

then

$$C_D^*(I) < C_E, \quad \forall I \in [0, \infty]. \quad (14)$$

Otherwise, if

$$c_e - c_a \leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f, \quad (15)$$

then there exists a unique $I_{ED} \in [0, \infty]$ s.t.

$$C_E \leq C_D^*(I), \text{ for } 0 \leq I \leq I_{ED}, \text{ and } C_D^*(I) < C_E, \text{ for } I > I_{ED}. \quad (16)$$

Proof of Claim 1: The closed form formula for the optimal cost for the collaboration strategy $C_D^*(I)$ is given in (12). From Lemma 3, we know that $C_D^*(I)$ is decreasing in I for all $I \in \mathbb{R}^+$. Next, observe from (12) that

$$C_D^*(\infty) = c_a + \lambda c_e + \lambda c_l + \lambda P_E(TP)[c_f - c_l].$$

Moreover, we can verify that

$$\begin{aligned} C_D^*(\infty) - C_E &= c_a - (1 - \lambda)c_e - (1 - \lambda)P_E(FP)c_f \\ &< c_a - (1 - \lambda)c_e \\ &< 0, \end{aligned} \tag{17}$$

$$\implies C_D^*(\infty) < C_E, \tag{18}$$

where the inequality in (17) follows from Assumption 1. Recall the design parameter t_1 defined in (1) as:

$$t_1 = \frac{\lambda}{1 - \lambda} \times \frac{P_E(TP)(c_l - c_f) - c_e}{c_e + P_E(FP)c_f}. \tag{19}$$

Observe that

$$\lim_{I \rightarrow 0} C_D^*(I) = \begin{cases} c_a + \lambda c_l, & \text{if } t_1 < 1 \\ c_a + \lambda c_l + c_e - P_E(TP)\lambda[c_l - c_f] + P_E(FP)(1 - \lambda)c_f, & \text{if } 1 \leq t_1. \end{cases} \tag{20}$$

Now assume that inequality (13) in the claim holds. That is, we have

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e - c_a. \tag{21}$$

Then

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e - c_a$$

$$\implies P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e \quad (22)$$

$$\implies P_E(TP)\lambda[c_l - c_f] - \lambda c_e < P_E(FP)(1 - \lambda)c_f + (1 - \lambda)c_e$$

$$\implies P_E(TP)\lambda[c_l - c_f] - \lambda c_e < P_E(FP)(1 - \lambda)c_f + (1 - \lambda)c_e$$

$$\implies \frac{\lambda}{1 - \lambda} \times \frac{P_E(TP)(c_l - c_f) - c_e}{c_e + P_E(FP)c_f} < 1$$

$$\implies t_1 < 1, \quad (23)$$

where the last inequality follow from the definition of t_1 given in (19). Therefore, since $t_1 < 1$, from (20), we have $\lim_{I \rightarrow 0} C_D^*(I) = c_a + \lambda c_l$. Now, observe that

$$\begin{aligned} \lim_{I \rightarrow 0} C_D^*(I) - C_E &= c_a + \lambda c_l - c_e - \lambda c_l \\ &\quad + P_E(TP)\lambda[c_f - c_l] - P_E(FP)(1 - \lambda)c_f \\ &= c_a - c_e + P_E(TP)\lambda[c_f - c_l] - P_E(FP)(1 - \lambda)c_f \\ &< 0, \end{aligned} \quad (24)$$

where the inequality in (24) follows from (21). The inequality in (24) implies that $\lim_{I \rightarrow 0} C_D^*(I) < C_E$. Hence, since $C_D^*(I)$ is decreasing in I , we have

$$C_D^*(I) \leq \lim_{I \rightarrow 0} C_D^*(I) < C_E, \quad \forall I \in [0, \infty],$$

proving (14). Next, assume the inequality (13) in the claim does not hold. Then, we have

$$c_e - c_a \leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f. \quad (25)$$

First, consider the case

$$c_e - c_a \leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e, \quad (26)$$

which from (22) implies that $t_1 < 1$ and as such we again have $\lim_{I \rightarrow 0} C_D^*(I) = c_a + \lambda c_l$.

Then, we see that

$$\begin{aligned} \lim_{I \rightarrow 0} C_D^*(I) - C_E &= c_a + \lambda c_l - c_e - \lambda c_l \\ &+ P_E(TP)\lambda[c_f - c_l] - P_E(FP)(1 - \lambda)c_f \\ &= c_a - c_e + P_E(TP)\lambda[c_f - c_l] - P_E(FP)(1 - \lambda)c_f \\ &\geq 0, \end{aligned} \quad (27)$$

where (27) follows from (25). Inequality (27) implies that $C_E \leq \lim_{I \rightarrow 0} C_D^*(I)$. Hence, combining (17) and (27), we get that $C_D^*(\infty) < C_E \leq \lim_{I \rightarrow 0} C_D^*(I)$. Since we know from Lemma 3 that $C_D^*(I)$ is decreasing in I and a continuous function, by Intermediate Value Theorem $\exists I_{ED} \geq 0$ such that

$$C_E \leq C_D^*(I), \text{ for } 0 \leq I \leq I_{ED}, \quad C_D^*(I) < C_E, \text{ for } I > I_{ED},$$

proving (16). Finally, let us assume

$$\begin{aligned} c_e &\leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f, \\ \implies (1 - \lambda)(c_e + P_E(FP)c_f) &\leq \lambda(P_E(TP)(c_l - c_f) + c_e) \\ \implies 1 &\leq t_1, \end{aligned} \quad (28)$$

where the last inequality follow from the definition of t_1 given in (19). Then, from (20), we have

$$\lim_{I \rightarrow 0} C_D^*(I) = c_a + \lambda c_l + c_e - P_E(TP)\lambda[c_l - c_f] + P_E(FP)(1 - \lambda)c_f, \quad (29)$$

and

$$\begin{aligned} \lim_{I \rightarrow 0} C_D^*(I) - C_E &= c_a + \lambda c_l + c_e - P_E(TP)\lambda[c_l - c_f] + P_E(FP)(1 - \lambda)c_f \\ &\quad - c_e - \lambda c_l + P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f \\ &= c_a \\ &> 0, \end{aligned} \quad (30)$$

follow, which implies that $\lim_{I \rightarrow 0} C_D^*(I) > C_E$. Hence, combining (18) and (30), we obtain $C_D^*(\infty) < C_E < \lim_{I \rightarrow 0} C_D^*(I)$. As $C_D^*(I)$ is decreasing in I and is a continuous function, by Intermediate Value Theorem, $\exists I_{ED} \geq 0$ such that

$$C_E \leq C_D^*(I) \text{ for } 0 \leq I \leq I_{ED}, \text{ and } C_D^*(I) < C_E \text{ for } I > I_{ED},$$

which proves (16) and completes the proof of Claim 1.

Claim 2: If $P_E(TP)$ and $P_E(FP)$ are such that

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e - c_a + \max\{0, \lambda c_l - c_f\}, \quad (31)$$

then

$$C_A^*(I) < C_E, \quad \forall I \in [0, \infty]. \quad (32)$$

Otherwise, if

$$c_e - c_a + \max\{0, \lambda c_l - c_f\} \leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f, \quad (33)$$

then there exists a unique $I_{EA} \in [0, \infty]$ s.t.

$$C_E \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{EA}, \text{ and } C_A^*(I) < C_E, \text{ for } I > I_{EA}. \quad (34)$$

Proof of Claim 2: Based on Lemma 3, we restate the optimal cost for the automation strategy as:

$$\begin{aligned} C_A^*(I) = & c_a + \lambda c_l + \lambda \left(c_f - c_l \right) \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(t_2)}{I} \right) \right) \\ & + (1 - \lambda) \left(c_f \right) \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(t_2)}{I} \right) \right). \end{aligned} \quad (35)$$

Again from Lemma 3, $C_A^*(I)$ is decreasing in I and from the definition of $C_A^*(I)$ given in (35), we can calculate

$$C_A^*(\infty) = c_a + \lambda c_f < C_A^*(I), \quad \forall I \in \mathbb{R}^+. \quad (36)$$

Moreover, observe that

$$\begin{aligned} C_A^*(\infty) - C_E &= c_a - c_e - \lambda(1 - P_E(TP))[c_l - c_f] \\ &\quad - (1 - \lambda)P_E(FP)c_f \\ &\leq c_a - c_e < 0, \end{aligned} \quad (37)$$

$$\implies C_A^*(\infty) < C_E, \quad (38)$$

where inequality (37) follows from Assumption 1. Next, we will take the limit of $C_A^*(I)$ as I approaches 0. First, assume that $c_f - \lambda c_l \leq 0$, then

$$\frac{c_f}{c_l} \leq \lambda \iff t_2 = \frac{\lambda(c_l - c_f)}{(1 - \lambda)c_f} \geq 1,$$

which implies that

$$\lim_{I \rightarrow 0} C_A^*(I) = c_a + c_f. \quad (39)$$

Next, assume that $c_f - \lambda c_l > 0$, then

$$\frac{c_f}{c_l} > \lambda \iff t_2 = \frac{\lambda(c_l - c_f)}{(1 - \lambda)c_f} < 1,$$

which implies that

$$\lim_{I \rightarrow 0} C_A^*(I) = c_a + \lambda c_l. \quad (40)$$

Combining (39) and (40), we get

$$\lim_{I \rightarrow 0} C_A^*(I) = c_a + \lambda c_l + \min\{0, c_f - \lambda c_l\}. \quad (41)$$

First, assume that

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < c_e - c_a + \max\{0, \lambda c_l - c_f\}. \quad (42)$$

In this case, observe that

$$\begin{aligned} \lim_{I \rightarrow 0} C_A^*(I) - C_E &= c_a + \lambda c_l + \min\{0, c_f - \lambda c_l\} \\ &\quad - c_e - \lambda c_l + \lambda(c_l - c_f)P_E(TP) - (1 - \lambda)c_f P_E(FP) \\ &< \min\{0, c_f - \lambda c_l\} + \max\{0, \lambda c_l - c_f\} \end{aligned} \quad (43)$$

$$= \min\{0, c_f - \lambda c_l\} - \min\{0, c_f - \lambda c_l\} = 0, \quad (44)$$

where inequality in (43) follows from the assumption in (42) and the equality in (44) follows from the fact that for any two numbers a and b , $\max\{a, b\} = -\min\{-a, -b\}$. Equation (44) implies that

$$\lim_{I \rightarrow 0} C_A^*(I) < C_E, \quad (45)$$

which combined with Lemma 3 ($C_A^*(I)$ is decreasing in I) implies that

$$C_A^*(I) \leq \lim_{I \rightarrow 0} C_A^*(I) < C_E, \quad \forall I \in \mathbb{R}^+$$

proving (32). Next, assume that

$$c_e - c_a + \max\{0, \lambda c_l - c_f\} \leq P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f. \quad (46)$$

Then,

$$\begin{aligned} \lim_{I \rightarrow 0} C_A^*(I) - C_E &= c_a + \lambda c_l + \min\{0, c_f - \lambda c_l\} \\ &\quad - c_e - \lambda c_l + \lambda(c_l - c_f)P_E(TP) - (1 - \lambda)c_f P_E(FP) \\ &\geq \min\{0, c_f - \lambda c_l\} + \max\{0, \lambda c_l - c_f\} \end{aligned} \quad (47)$$

$$= \min\{0, c_f - \lambda c_l\} - \min\{0, c_f - \lambda c_l\} = 0, \quad (48)$$

where inequality in (47) follows from the assumption in (46) and equality in (48) follows from the fact that for any two numbers a and b , $\max\{a, b\} = -\min\{-a, -b\}$. Equation (48) implies that

$$\lim_{I \rightarrow 0} C_A^*(I) \geq C_E, \quad (49)$$

which combined with (38) results in

$$C_A^*(\infty) < C_E \leq \lim_{I \rightarrow 0} C_A^*(I). \quad (50)$$

Inequality (50) combined with the fact that $C_A^*(I)$ is decreasing in I (from Lemma 3) implies, by the Intermediate Value Theorem, that $\exists$ a unique I_{EA} such that

$$C_E \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{EA}, \quad C_A^*(I) < C_E, \text{ for } I > I_{EA},$$

which completes the proof of Claim 2.

Claim 3: If prevalence is low, i.e. $\lambda < c_f/c_l \iff t_2 < 1$, consider two cases. When $P_E(TP)$ and $P_E(FP)$ are such that

$$\frac{c_e c_l}{c_f(c_l - c_f)} \leq P_E(TP) - P_E(FP),$$

then there exists a unique I_{DA} such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \text{ and } C_A^*(I) < C_D^*(I), \text{ for } I > I_{DA}. \quad (51)$$

Otherwise, when $P_E(TP)$ and $P_E(FP)$ are such that

$$P_E(TP) - P_E(FP) < \frac{c_e c_l}{c_f(c_l - c_f)},$$

then

$$C_A^*(I) < C_D^*(I), \quad \forall I > 0. \quad (52)$$

If prevalence is high $c_f/c_l \leq \lambda \iff t_2 \geq 1$, consider two cases. When $P_E(TP)$ and $P_E(FP)$ are such that

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f \geq \lambda[c_l - c_f] - (1 - \lambda)c_f + c_e,$$

then, there exists a unique I_{DA} such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \text{ and } C_A^*(I) < C_D^*(I), \text{ for } I > I_{DA}. \quad (53)$$

Otherwise, when $P_E(TP)$ and $P_E(FP)$ are such that

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f < \lambda[c_l - c_f] - (1 - \lambda)c_f + c_e,$$

then

$$C_A^*(I) < C_D^*(I), \quad \forall I \in \mathbb{R}^+. \quad (54)$$

Proof of Claim 3:

Using the closed form solutions for $C_D^*(I)$ and $C_A^*(I)$ from Lemma 3, we calculate

$$\begin{aligned} C_D^*(\infty) &= c_a + \lambda c_l + \lambda \left(c_e + P_E(TP)[c_f - c_l] \right), \\ C_A^*(\infty) &= c_a + \lambda c_l + \lambda \left(c_f - c_l \right). \end{aligned}$$

Observe that,

$$\begin{aligned} C_D^*(\infty) - C_A^*(\infty) &= \lambda(c_e + (P_E(TP) - 1)(c_f - c_l)) \\ &= \lambda(c_e + (1 - P_E(TP))(c_l - c_f)) \\ &> 0, \end{aligned} \quad (55)$$

where (55) follows from $c_f < c_l$ based on Assumption 1. The inequality in (55) implies that

$$C_A^*(\infty) < C_D^*(\infty). \quad (56)$$

Hence, the expected cost of automation strategy is lower than that of the delegation strategy if the performance of the AI algorithm is high enough. Next, let the difference between the delegation and automation strategy defined as

$$D(I) := C_D^*(I) - C_A^*(I);$$

from (56), we have $D(\infty) > 0$. Let K and $G(I)$ be defined as

$$\begin{aligned} K &:= \frac{e^{-\frac{\ln^2(t_2)}{2I^2} - \frac{I^2}{8}}}{\sqrt{2\pi}} \times \sqrt{\lambda(1-\lambda)c_f(c_l - c_f)}, \\ G(I) &:= \left(1 - \underbrace{\frac{\sqrt{P_E(TP)[c_l - c_f] - c_e}}{\sqrt{c_l - c_f}}}_{< 1 \text{ since } P_E(TP) < 1} \times \underbrace{\frac{\sqrt{c_e + P_E(FP)c_f}}{\sqrt{c_f}}}_{< 1 \text{ due to Assumption 2}} \times e^{\frac{\ln^2(t_2) - \ln^2(t_1)}{2I^2}} \right) \end{aligned} \quad (57)$$

The derivative of $D(I)$ with respect to I is

$$\begin{aligned} D'(I) &= \underbrace{K}_{> 0} \times \left(1 - \underbrace{\frac{\sqrt{P_E(TP)[c_l - c_f] - c_e}}{\sqrt{c_l - c_f}}}_{< 1 \text{ since } P_E(TP) < 1} \times \underbrace{\frac{\sqrt{c_e + P_E(FP)c_f}}{\sqrt{c_f}}}_{< 1 \text{ due to Assumption 2}} \times e^{\frac{\ln^2(t_2) - \ln^2(t_1)}{2I^2}} \right), \\ &= KG(I). \end{aligned} \quad (58)$$

Observe that

$$0 < G(\infty) = \left(1 - \frac{\sqrt{P_E(TP)[c_l - c_f] - c_e}}{\sqrt{c_l - c_f}} \frac{\sqrt{c_e + P_E(FP)c_f}}{\sqrt{c_f}} \right) < 1$$

and note that since K is always positive $D'(I)$ has the same sign as $G(I)$ as such $D'(\infty) > 0$. Next, we can find the derivative of $G(I)$ with respect to I as

$$G'(I) = \left(\frac{\sqrt{P_E(TP)[c_l - c_f]} - c_e}{\sqrt{c_l - c_f}} \frac{\sqrt{c_e + P_E(FP)c_f}}{\sqrt{c_f}} \frac{\ln^2(t_2) - \ln^2(t_1)}{I^3} e^{\frac{(\ln^2(t_2) - \ln^2(t_1))}{2I^2}} \right). \quad (59)$$

If $\ln^2(t_2) < \ln^2(t_1)$, then $0 < 1 = \lim_{I \rightarrow 0} G(I)$ and equation (59) implies that $G(I)$ is decreasing ($G'(I) < 0, \forall I \in \mathbb{R}^+$) in I . This implication, combined with the fact that $G(\infty) > 0$, suggests $G(I) > 0, \forall I \in \mathbb{R}^+$. Similarly, if $\ln^2(t_2) = \ln^2(t_1)$, observe from equation (57), $G(I)$ does not depend on I and $G(I) > 0, \forall I \in \mathbb{R}^+$. Since $G(I)$ and $D'(I)$ have the same sign, we have $D'(I) > 0 \forall I \in \mathbb{R}^+$ when $\ln^2(t_2) \leq \ln^2(t_1)$. Therefore, $D(I)$ is increasing in I when $\ln^2(t_2) \leq \ln^2(t_1)$. Otherwise, if $\ln^2(t_2) > \ln^2(t_1)$, then $\lim_{I \rightarrow 0} G(I) = -\infty$, $G(\infty) > 0$ and from equation (59), $G(I)$ is increasing in I ($G'(I) > 0 \forall I \in \mathbb{R}^+$), which implies that there exists $I_G > 0$ such that $G(I) \leq 0$ for $0 \leq I \leq I_G$ and $G(I) > 0$ for $I > I_G$. Since $D'(I)$ and $G(I)$ have the same sign, we have $0 \leq D'(I) \leq 0$ for $I \leq I_G$ and $D'(I) > 0$ for $I > I_G$. Therefore, $D(I)$ is first decreasing then increasing in I . To continue the proof, we will analyze the cases of low prevalence ($t_2 < 1 \iff \lambda < c_f/c_l$) and high prevalence ($t_2 \geq 1 \iff \lambda \geq c_f/c_l$) separately.

Low-prevalence case, $t_2 < 1$:

If $t_1 \leq t_2 < 1$, then from (20) and (41), we have $D(0) = 0$ and since $\ln^2(t_2) - \ln^2(t_1) \leq 0$, we have $D(I)$ increasing in I . Hence,

$$D(I) > 0 \implies C_D^*(I) > C_A^*(I) \quad \forall I > 0.$$

Otherwise, if $t_2 < t_1 < 1$, then again from (20) and (41), we have $D(0) = 0$ and $D(\infty) > 0$ but since $\ln^2(t_2) - \ln^2(t_1) > 0$, $D(I)$ is first decreasing and then increasing in I . Therefore, there exist a unique $0 < I_{DA}$ such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \quad C_A^*(I) < C_D^*(I), \text{ for } I > I_{DA}.$$

Finally, if $t_2 < 1 \leq t_1$, then $D(0) < 0$ and depending on whether $1 \leq t_1 < 1/t_2$ or $1/t_2 < t_1$, we have $D(I)$ first decreasing then increasing in I or $D(I)$ is always increasing in I , respectively. Note that $D(0) < 0$ implies that $\lim_{I \rightarrow 0} C_D^*(I) < \lim_{I \rightarrow 0} C_A^*(I)$. Moreover, $D(\infty) > 0$ implies that $C_A^*(\infty) < C_D^*(\infty)$. Hence, in both cases, we have a unique $I_{DA} > 0$ such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \quad C_A^*(I) < C_D^*(I), \text{ for } I > I_{DA}.$$

We have shown that for the low prevalence case, the existence of I_{DA} (the level of information content below which the delegation strategy is optimal) requires $t_2 < t_1$. By definition of t_1 and t_2 , this condition can be re-written as

$$\begin{aligned} \frac{\lambda(c_l - c_f)}{(1 - \lambda)(c_f)} &< \frac{\lambda(P_E(TP)(c_l - c_f) - c_e)}{(1 - \lambda)(P_E(FP)c_f + c_e)} \\ \implies \frac{c_e c_l}{c_f(c_l - c_f)} &< P_E(TP) - P_E(FP). \end{aligned}$$

High-prevalence case, $t_2 \geq 1$:

Let us define

$$\begin{aligned} D_1 &:= \lambda[P_E(TP)(c_l - c_f) - c_e], \quad D_2 = (1 - \lambda)[P_E(FP)c_f + c_e] \\ A_1 &:= \lambda(c_l - c_f), \quad A_2 := (1 - \lambda)c_f \\ \tau(I, t) &:= -t \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(t)}{I} \right) \right) + 1 - \Phi \left(\frac{I}{2} - \frac{\ln(t)}{I} \right). \end{aligned} \tag{60}$$

Note that $t_2 = \frac{A_1}{A_2}$, $t_1 = \frac{D_1}{D_2}$,

$$D(I) = C_D^*(I) - C_A^*(I) = D_2 \tau(I, t_1) - A_2 \tau(I, t_2), \tag{61}$$

and when $t_2 > 1$, we have

$$D(0) = \begin{cases} A_2(1 - t_2) = \lambda c_l - c_f & t_1 < 1 \\ D_2(1 - t_1) - A_2(1 - t_2) = \lambda c_l - c_f + c_e - \lambda P_E(TP)(c_l - c_f) + (1 - \lambda)P_E(FP)c_f & 1 \leq t_1. \end{cases} \quad (62)$$

Moreover, from Assumption 2, we have

$$P_E(FP) < \frac{c_f - c_e}{c_f} \iff P_E(FP)c_f + c_e < c_f,$$

which implies that

$$\frac{A_2}{D_2} = \frac{c_f}{P_E(FP)c_f + c_e} > 1. \quad (63)$$

We can calculate the partial derivative of $\tau(I, t)$ with respect to t as

$$\frac{\partial \tau(I, t)}{\partial t} = 0.5 \left(-1 - \text{Erf} \left(\frac{I^2 + 2 \log(t)}{2\sqrt{2}I} \right) \right),$$

where Erf is the error function which takes values between -1 and 1 (note the error function takes the value ± 1 at $\pm \infty$) as such for any $t \in \mathbb{R}^+$

$$-1 < \frac{\partial \tau(I, t)}{\partial t} < 0 \implies \left| \frac{\partial \tau(I, t)}{\partial t} \right| < 1,$$

which implies that $\tau(I, t)$ is decreasing in t for any I . Furthermore, note that for any $I \in \mathbb{R}^+$

$$\tau(I, 1) = \Phi \left(-\frac{I}{2} \right) - \Phi \left(\frac{I}{2} \right) \leq 0, \quad (64)$$

where in (64) we use the fact that $\Phi(\cdot)$ is an increasing function in its argument. Hence, (64) combined with the fact that $\tau(I, t)$ is decreasing in t implies that $\tau(I, t) < 0$ for $1 \leq t$. Assume that $t_1 < t_2$ and there exists I_{DA} such that $D(I_{DA}) = D_2\tau(I_{DA}, t_1) -$

$A_2\tau(I_{DA}, t_2) = 0$ which implies that

$$\frac{\tau(I_{DA}, t_1)}{\tau(I_{DA}, t_2)} = \frac{A_2}{D_2} > 1 \quad (65)$$

$$\implies \tau(I_{DA}, t_1) < \tau(I_{DA}, t_2) < 0, \quad (66)$$

where (65) follows from (63) and (66) follows from the fact that $t_2 > 1$ as such $\tau(I, t_2) < 0$. But, we showed that $\tau(I, t)$ is decreasing in t for all I and we assumed that $t_1 < t_2$. Hence, (66) is a contradiction. Therefore, whenever $t_1 < t_2$ we have $D(I) > 0 \implies C_A^*(I) < C_D^*(I)$ for all $I \in \mathbb{R}^+$. Now, assume that $1 \leq t_2 \leq t_1$, then $\ln^2(t_2) \leq \ln^2(t_1)$ and as such, $D(I)$ is increasing in I . Therefore, if $D(0) > 0$ then $C_A^*(I) < C_D^*(I)$ for all I . Finally, if $1 \leq t_2 \leq t_1$ and $D(0) < 0$, then since $D(\infty) > 0$ and $D(I)$ is increasing in I , there exists a unique I_{DA} such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \quad C_A^*(I) < C_D^*(I), \text{ for } I_{DA} < I. \quad (67)$$

Hence, for the high-prevalence case, there exists I_{DA} such that (70) holds if $1 \leq t_2 < t_1$ and $D(0) < 0$. From Equation (62), we see that when $1 \leq t_2$ the condition $D(0) < 0$ holds if and only if

$$P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f > c_e + \lambda c_l - c_f, \quad (68)$$

which is equivalent to the following condition

$$D_2(1 - t_1) < A_2(1 - t_2) \iff A_2(t_2 - 1) < D_2(t_1 - 1). \quad (69)$$

Observe that

$$\begin{aligned} A_2(t_2 - 1) < D_2(t_1 - 1) &\iff 1 < \frac{A_2}{D_2} < \frac{t_1 - 1}{t_2 - 1} \\ &\implies t_2 < t_1, \end{aligned}$$

which implies that when $1 < t_2$ and $D(0) < 0$, we have $t_2 < t_1$. Therefore, if equation (68) holds, then there exists a unique I_{DA} such that

$$C_D^*(I) \leq C_A^*(I), \text{ for } 0 \leq I \leq I_{DA}, \quad C_A^*(I) < C_D^*(I), \text{ for } I_{DA} < I. \quad (70)$$

Finally, note that the result shows that when I is high enough, then $C_A^*(I) < C_D^*(I)$ is true. Therefore, the delegation region disappears when $I_{DA} < I_{ED}$. Note that for both low- and high-prevalence cases, a necessary condition for the existence of delegation to be the optimal strategy is $t_2 < t_1$, a fact which we use in subsequent proofs.

Using Claims 1,2, and 3, we will now construct Supplementary Table 1 of the main manuscript given in Proposition 1 of the main manuscript. Recall our definitions of BA (Balanced Accuracy) and NFB (Net Financial Benefit) as

$$BA = \frac{P_E(TP) + P_E(TN)}{2}, \quad (71)$$

$$NFB = P_E(TP)\lambda[c_l - c_f] - P_E(FP)(1 - \lambda)c_f.$$

Note that BA can also be written as

$$BA = 0.5(P_E(TP) + P_E(TN)) = 0.5(P_E(TP) + 1 - P_E(FP)),$$

using which we can write the condition in (51) and (52) respectively as

$$BA \geq 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}, \quad BA < 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}.$$

Note that *Table 1(a)* represents the low liability case and we will prove it in four steps.

Step 1: When $NFB \leq c_e - c_a$ and $BA < 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}$, from (14) and (52), we have

$$C_A^*(I) < C_D^*(I) < C_E, \quad \forall I \in [0, \infty], \text{ proving the left top part of Table 1 (a).}$$

Step 2: When $NFB < c_e - c_a$ and $BA \geq 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}$, from (14) and (51), we have

$C_D^*(I) < C_E$, $\forall I \in [0, \infty]$, and $C_D^*(I) \leq C_A^*(I)$, for $0 \leq I \leq I_{DA}$, and $C_A^*(I) < C_D^*(I)$, for $I_{DA} < I$, proving bottom left of the Table 1 (a).

Step 3: When $c_e - c_a \leq NFB$ and $BA < 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}$, from (34), we have $C_E \leq C_A^*(I)$,

for $0 \leq I \leq I_{EA}$, and $C_A^*(I) < C_E$, for $I > I_{EA}$, and from (52), we have $C_A^*(I) < C_D^*(I)$, $\forall I \in \mathbb{R}^+$, proving top right of Table 1 (a).

Step 4: Suppose $c_e - c_a \leq NFB$, $BA \geq 0.5 + \frac{c_e c_l}{2c_f(c_l - c_f)}$, and $I_{ED} < I_{DA}$ holds. Then,

for $I < I_{ED}$ from (16), we have $C_E \leq C_D^*(I)$. Moreover, $C_D^*(I) < C_A^*(I)$, due to our assumption that $I_{ED} < I_{DA}$. Hence, for $I < I_{ED}$, we have $C_E < \min\{C_D^*(I), C_A^*(I)\}$. Next, for $I_{ED} < I < I_{DA}$ from (16) and our assumption $I_{ED} < I_{DA}$, $C_D^*(I) < C_E$, $C_D^*(I) < C_A^*(I)$. For $I_{DA} < I$, from (16) and (51), $C_D^*(I) < C_E$, $C_A^*(I) < C_D^*(I)$, which proves the bottom right of Table 1 (a).

We next prove the results in Table 1 (b) in two steps. Note that when liability is high, we have $0 \leq \lambda c_l - c_f$ and $c_e - c_a < c_e - c_a + \lambda c_l - c_f$.

Step 1: When $NFB < c_e - c_a + \lambda c_l - c_f$, from (32), we have $C_A^*(I) < C_E, \forall I \in [0, \infty]$,

and from (54), we have $C_A^*(I) < C_D^*(I)$, $\forall I \in \mathbb{R}^+$. Hence, for all $I \in \mathbb{R}^+$, we have $C_A^*(I) < \min\{C_D^*(I), C_E\}$, proving left side of Table 1 (b).

Step 2: Suppose $c_e - c_a \leq c_e - c_a + \lambda c_l - c_f < NFB$ and $I_{ED} < I_{DA}$ is true. When $I < I_{ED}$,

from (16), we have $C_E \leq C_D^*(I)$, and $C_D^*(I) < C_A^*(I)$ from our assumption that $I_{ED} < I_{DA}$. Hence, for $I < I_{ED}$, we have $C_E < C_D^*(I) < C_A^*(I)$. Next, when $I_{ED} < I < I_{DA}$, from (16) and our assumption $I_{ED} < I_{DA}$, $C_D^*(I) < C_E$ and $C_D^*(I) < C_A^*(I)$. When $I_{DA} < I$, from (16) and (53) $C_D^*(I) < C_E$, $C_A^*(I) < C_D^*(I)$, finishing the proof of bottom right of Table 1 (b)

□

Proof of Corollary 1: From the proof of Proposition 1, in Equation (38), we show that $C_A^*(\infty) < C_E$, and in (56), we show that $C_A^*(\infty) < C_D^*(\infty)$, which implies that

there exists $I_A \geq 0$ such that for $I > I_A$, we have

$$C_A^*(\infty) < \min\{C_D^*(\infty), C_E\}.$$

□

Proof of Corollary 2: The following shows the proof of each part.

Part 1: Note that C_E is independent of c_a . From (12), we see that $C_D^*(I)$ is increasing in c_a and in Lemma 3, we have shown it is decreasing in I . Therefore, if c_a increases I_{ED} should increase.

Part 2: The results follows from the fact that $C_D^*(I) - C_A^*(I)$ is independent of c_a .

Part 3: Define the difference between the expected costs for a given algorithm with information content I as $\Delta_{ED}(I, c_l) = C_E - C_D^*(I)$. Taking the partial derivative of $\Delta_{ED}(I, c_l)$ with respect c_l , we obtain

$$\begin{aligned} \frac{\partial \Delta_{ED}(I, c_l)}{\partial c_l} &= -\frac{1}{2} \lambda P_E(TP) \\ &\quad \times \operatorname{erfc} \left(\frac{2 \log \left(\frac{\lambda(c_e + c_f P_E(TP) - c_l P_E(TP))}{(l-1)(c_e + c_f P_E(TP))} \right) + I^2}{2\sqrt{2}I} \right) \\ &< 0, \end{aligned} \tag{72}$$

where (72) follows from erfc , the complementary error function, being always positive. Inequality (72) implies that for each I , $\Delta_{ED}(I, c_l)$ is decreasing in c_l . From Lemma 3, we know that $C_D^*(I)$ is decreasing in I for any c_l . Therefore, I_{ED} is non-decreasing in c_l .

Part 4: Let for a given set of parameters $I_{ED} < I_{DA}$ holds and define the difference between the expected optimal costs of delegation and automation strategies for a given algorithm with information content I as $\Delta_{DA}(I, c_l) = C_D^*(I) - C_A^*(I)$. The derivative

of $\Delta_{DA}(I, c_l)$ with respect to c_l is given as

$$\frac{\partial \Delta_{DA}(I, c_l)}{\partial c_l} = 0.5\lambda \left(\Phi \left(0.5I + \frac{\ln(t_2)}{I} \right) - P_E(TP) \Phi \left(0.5I + \frac{\ln(t_1)}{I} \right) \right).$$

Assume for a set of base parameters, we have

$$0.5\Phi \left(0.5I + \frac{\ln(t_2)}{I} \right) < P_E(TP) \Phi \left(0.5I + \frac{\ln(t_1)}{I} \right),$$

as such $\frac{\partial \Delta_{DA}(I_{DA}, c_l)}{\partial c_l} < 0$. By definition of the derivative, there exists $\epsilon > 0$ such that whenever $c'_l \in (c_l, c_l + \epsilon]$, we have $\Delta_{DA}(I_{DA}, c'_l) < 0$. This implies that whenever $c'_l \in (c_l, c_l + \epsilon]$, we have $C_D^*(I_{DA}) < C_A^*(I_{DA})$. In the proof of Claim 3 of Proposition 1, we demonstrated that for a given c_l , $\Delta_{DA}(I, c_l)$ is either always increasing in I or first decreasing and then increasing in I and $\Delta_{DA}(\infty, c_l) > 0$. Hence, since $\Delta_{DA}(I_{DA}, c'_l) < 0$ when $c'_l \in (c_l, c_l + \epsilon]$ and $\Delta_{DA}(\infty, c'_l) > 0$, there exists a value $I'_{DA} > I_{DA}$ such that $\Delta_{DA}(I'_{DA}, c'_l) = 0$. A similar argument holds for the case when $\frac{\partial \Delta_{DA}(I, c_l)}{\partial c_l} > 0$. $\square$

Proof of Corollary 3: The corollary uses $I' = I_{ED}|_{c_a=0}$ and $I'' = I_{DA}$ as specific information content levels corresponding to algorithm performance along with the assumption $I' < I''$. The following shows the proof of each part.

Part 1: First, note that from (12) both $C_D^*(I)$ and $C_A^*(I)$ are increasing in c_a , $C_D^*(I) - C_A^*(I)$ and C_E are independent of c_a . Moreover, when $c_a = 0$, by definition of $I' = I_{ED}$ being the information content threshold above which the optimal strategy switches from expert-only to delegation and our assumption $I' = I_{ED} < I'' = I_{DA}$, we have

$$C_E < C_D^*(I) < C_A^*(I), \quad \forall I < I'. \quad (73)$$

Hence, as we increase c_a both $C_D^*(I)$ and $C_A^*(I)$ will increase, their difference will remain the same and C_E will remain constant. Therefore, (73) will continue to hold.

Part 2: Remember from Corollary 2 that I'' and $C_D^*(I) - C_A^*(I)$ is independent of c_a .

Assume that $I' < I < I''$; then, when $c_a = 0$ we have $C_D^*(I) < C_E$. However, as we increase c_a , $C_D^*(I)$ will increase but C_E will stay constant. Moreover, as c_a approaches infinity so will $C_D^*(I)$. Therefore, there exists c'_a such that D is optimal if $c_a < c'_a$ and E is optimal otherwise.

Part 3: Assume that $I'' < I$. We know that I'' is independent of c_a , so we have

$$C_A^*(I) < C_D^*(I), \quad \forall c_a > 0.$$

However, $C_A^*(I)$ is increasing in c_a and $C_A^*(I)$ approaches infinity as c_a goes to infinity, whereas C_E remains unchanged. Therefore, there exists a c''_a such that if $c_a < c''_a$, we have

$$C_A^*(I) < C_E,$$

and if $c_a > c''_a$, we have

$$C_E < C_A^*(I),$$

which completes the proof of the corollary. $\square$

Proof of Proposition 3:

In the case of asymmetric liability, the optimal costs for each strategy are given as follows:

$$C_E = c_e + \lambda c_{lh} - \lambda(c_{lh} - c_f)P_E(TP) + (1 - \lambda)c_f P_E(FP) \quad (74)$$

$$C_D^*(I) = c_a + \lambda c_{la} - \lambda[P_E(TP)(c_{lh} - c_f) - c_e + c_{la} - c_{lh}] \left(1 - \Phi\left(-\frac{I}{2} - \frac{\ln(t_{dl})}{I}\right)\right)$$

$$+ (1 - \lambda) [P_E(FP)c_f + c_e] \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(t_{dl})}{I} \right) \right) \quad (75)$$

$$\begin{aligned} C_A^*(I) &= c_a + \lambda c_{la} \\ &- \lambda [c_{la} - c_f] \left(1 - \Phi \left(-\frac{I}{2} - \frac{\ln(t_{al})}{I} \right) \right) \\ &+ (1 - \lambda)c_f \left(1 - \Phi \left(\frac{I}{2} - \frac{\ln(t_{al})}{I} \right) \right), \end{aligned} \quad (76)$$

where we define the terms

$$t_{dl} := \frac{\lambda}{1 - \lambda} \times \frac{P_E(TP)(c_{lh} - c_f) - c_e + c_{la} - c_{lh}}{c_e + P_E(FP)c_f} \text{ and } t_{al} := \frac{\lambda}{1 - \lambda} \times \frac{c_{la} - c_f}{c_f}. \quad (77)$$

Note that C_E does not depend on c_{la} and the parameters t_{dl} and t_{al} have the same interpretation as t_1 and t_2 , as defined in (1), except the incorporation of asymmetric liability cost. In the proof of Claim 3 of Proposition 1, we have demonstrated that a necessary condition for the existence of delegation strategy is $t_2 < t_1$. Hence, since delegation is a possible strategy for the benchmark case (the case litigation costs for humans and algorithms are symmetric), we have $t_{al} < t_{dl}$ when $c_{la} = c_{lh}$. Next, define

$$\Delta_{ED}(I, c_{la}) := C_E - C_D^*(I), \quad \Delta_{DA}(I, c_{la}) := C_D^*(I) - C_A^*(I).$$

Then,

$$\frac{\partial \Delta_{ED}(I, c_{la})}{\partial c_{la}} = -\lambda \left(1 - \Phi \left(\frac{I}{2} + \frac{\ln(t_{dl})}{I} \right) \right), \quad (78)$$

$$\frac{\partial \Delta_{DA}(I, c_{la})}{\partial c_{la}} = \lambda \left(\Phi \left(\frac{I}{2} + \frac{\ln(t_{al})}{I} \right) - \Phi \left(\frac{I}{2} + \frac{\ln(t_{dl})}{I} \right) \right). \quad (79)$$

From (78), since $\Phi(x) < 1$ for any x , we see that for any value of t_{dl} , we have $\frac{\partial \Delta_{ED}(I, c_{la})}{\partial c_{la}} < 0$. Since $\Phi(\cdot)$ is an increasing function, from (79), we can show that $\frac{\partial \Delta_{DA}(I, c_{la})}{\partial c_{la}} < 0$ holds whenever $t_{al} < t_{dl}$ which is true when $c_{la} = c_{lh}$. For $c_{lh} < c_{la}$,

observe that

$$\begin{aligned}\frac{t_{dl}}{t_{al}} &= \frac{c_f}{c_e + P_E(FP)c_f} \times \frac{P_E(TP)(c_{lh} - c_f) - c_e + c_{la} - c_{lh}}{c_{la} - c_f} \\ &= \frac{c_e + P_E(FP)c_f}{c_f} \times \left(1 + \frac{(P_E(TP) - 1)(c_{lh} - c_f) - c_e}{c_{la} - c_f}\right).\end{aligned}\quad (80)$$

Note that, from Assumption 1, we have $c_{lh} > c_f$ as such

$$(P_E(TP) - 1)(c_{lh} - c_f) - c_e = -(1 - P_E(TP))(c_{lh} - c_f) - c_e < 0.$$

Hence, Equation (80) implies that $\frac{t_{dl}}{t_{al}}$ is increasing in c_{la} when $c_{la} > c_{lh} > c_f$. At the baseline where $c_{lh} = c_{la}$, we have $t_{dl}/t_{al} > 1$; we just demonstrated that t_{dl}/t_{al} is increasing in c_{la} when $c_{la} > c_{lh}$. Therefore, $t_{dl}/t_{al} > 1$ holds whenever $c_{la} > c_{lh}$ as such $\frac{\partial \Delta_{DA}(I, c_{la})}{\partial c_{la}} < 0$ and $\frac{\partial \Delta_{ED}(I, c_{la})}{\partial c_{la}} < 0$ whenever $c_{la} > c_{lh}$. Hence, we can conclude that as c_{la} increases, both $\Delta_{ED}(I, c_{la})$ and $\Delta_{DA}(I, c_{la})$ are decreasing. Next, note that when c_{la} approaches ∞ , under the delegation strategy, all cases will be delegated to the human expert as the liability of algorithm missing a cancer will approach to ∞ .

Therefore, we have

$$\lim_{c_{la} \rightarrow +\infty} C_D^*(I) = c_a + C_E > C_E. \quad (81)$$

To prove *Part 1* of the proposition, assume that for a given I the optimal strategy under the benchmark case, when $c_{la} = c_{lh}$, is delegation; this implies that we have $C_D^*(I) < C_A^*(I)$. For $c_{la} > c_{lh}$, $\Delta_{DA}(I, c_{la})$ is decreasing in c_{la} as such $C_D^*(I) < C_A^*(I)$ will continue to hold when $c_{la} > c_{lh}$. Hence, automation can not become the optimal strategy for a given I as we increase c_{la} . Next, we have shown that $\Delta_{ED}(I, c_{la})$ is decreasing in c_{la} and we know that C_E is independent of c_{la} . Therefore, $C_D^*(I)$ is increasing in c_{la} , which combined with (81) implies that there exist a threshold c'_{la}

such that delegation is optimal when $c_{la} < c'_{la}$ and human expert is optimal when $c'_{la} < c_{la}$.

For *Part 2a* of the proposition, note that

$$\frac{\partial C_A^*(I)}{\partial c_{la}} = \lambda \left(1 - \Phi \left(\frac{I}{2} + \frac{\ln(t_{al})}{I} \right) \right) > 0, \quad \lim_{c_{la} \rightarrow +\infty} C_A^*(I) = c_a + c_f, \quad (82)$$

where we used the fact that $\Phi(x) < 1$ for any $x \in \mathbb{R}$ and when the algorithms liability of missing a cancer (c_{la}) approaches to ∞ all cases will be sent for follow-up. Equation (82) implies that $C_A^*(I)$ is increasing in c_{la} and we also know from *Part 1* that $\Delta_{DA}(I, c_{la})$ and $\Delta_{ED}(I, c_{la})$ are decreasing in c_{la} when $c_{la} > c_{lh}$. Next, assume that

$$\begin{aligned} & P_E(TP)\lambda[c_{lh} - c_f] - P_E(FP)(1 - \lambda)c_f \leq (c_e - c_a) + \lambda c_{lh} - c_f \\ \iff & c_a + c_f \leq c_e + \lambda c_{lh} - P_E(TP)\lambda[c_{lh} - c_f] + P_E(FP)(1 - \lambda)c_f \\ \implies & \lim_{c_{la} \rightarrow +\infty} C_A^*(I) \leq C_E, \end{aligned} \quad (83)$$

where (83) follows from (74) and (82). Hence, combining (81) and (83), we get

$$\lim_{c_{la} \rightarrow +\infty} C_A^*(I) \leq C_E < \lim_{c_{la} \rightarrow +\infty} C_D^*(I).$$

But, since $\Delta_{DA}(I, c_{la})$ is decreasing in c_{la} and $C_A^*(I)$ is increasing in c_{la} , we have

$$C_A^*(I) < C_E, \quad C_A^*(I) < C_D^*(I), \forall c_{la} > c_{lh}.$$

Hence, automation strategy is optimal for all $c_{la} > c_{lh}$.

For *Part 2b* of the Proposition, assume that

$$\begin{aligned}
(c_e - c_a) + \lambda c_{lh} - c_f &< P_E(TP)\lambda[c_{lh} - c_f] - P_E(FP)(1 - \lambda)c_f. \\
\iff c_e + \lambda c_{lh} - P_E(TP)\lambda[c_{lh} - c_f] + P_E(FP)(1 - \lambda)c_f &< c_a + c_f \\
\implies C_E &< \lim_{c_{la} \rightarrow +\infty} C_A^*(I),
\end{aligned} \tag{84}$$

where (84) follows from (74) and (82). Then, from (81) and (84), we have

$$C_E < \lim_{c_{la} \rightarrow +\infty} C_D^*(I), \quad C_E < \lim_{c_{la} \rightarrow +\infty} C_A^*(I).$$

Therefore, there exists c_{la}'' such that for $c_{la}'' < c_{la}$ the optimal strategy is expert only. $\square$

Proof of Proposition 2: This proposition is a corollary of Proposition 3. *Part 1* of Proposition 3 shows that if, for a given I , the optimal strategy is delegation under the benchmark, then there exists c_{la}' above which the optimal strategy is expert-only. Similarly, *Part 2b* of Proposition 3 shows that if, for a given I , the optimal strategy is automation under the benchmark, then there exists c_{la}'' above which the optimal strategy is expert-only. This concludes the proof of the proposition. $\square$

Supplementary Note 3: Dream Digital

Mammography Challenge

The digital mammography challenge aimed to develop AI systems for mammography-based breast cancer detection. Specifically, it asked participating teams to develop machine learning algorithms to match or beat the radiologist's performance in assessing digital mammograms. The ultimate objective was to determine whether state-of-the-art AI systems could tip the balance in transforming the clinical practice

in radiological imaging [1]. For this purpose, the DREAM consortium provided mammography data from the Kaiser Permanente Washington and Karolinska Institute in Sweden. The dataset contained 640,000 mammograms. The holdout sample had 25,657 mammograms used for testing the performance of algorithms against the radiologists.

References

- [1] Trister, A.D., Buist, D.S.M., Lee, C.I.: Will machine learning tip the balance in breast cancer screening? *JAMA Oncology* **3**(11), 1463–1464 (2017)