A Appendix

A.1 Stage 3 Unconstrained Firm Content Threshold Decisions

Below we show the first order derivative for Equation 9.

By the fact that $M_j \geq 0$, $q < \frac{1}{2}$, $L_j < \frac{1}{2}$, we can sign the derivative of consumption N_j with respect to threshold c_j below as

$$\frac{\partial N_j(c_j)}{\partial c_j} = \begin{cases} \frac{M_j q}{1-L_j} \geq 0 & \text{if } 0 \leq c_j \leq L_j, \\ \frac{M_j(2q-1)}{1-L_j} \leq 0 & \text{if } L_j < c_j \leq 1-L_j, \\ \frac{M_j(q-1)}{1-L_j} \leq 0 & \text{if } 1-L_j < c_j \leq 1. \end{cases} \tag{24}$$

Equation 24 shows that total consumption monotonically increases in c_j in the first region and subsequently decreases in the second and third regions. As the functions are continuous, the threshold that maximizes the disclosed group's consumption is $c_j^* = L_j$. This results from the assumption that $q < \frac{1}{2}$, which means there are more consumers who prefer B than those who prefer content A [1].

Below we show the first order derivative of Equation 11.

By the fact that $M_p \leq \beta$,$M_g \leq 1$, $q < \frac{1}{2}$, $l < \frac{1}{2}$, and $\gamma < 1$, we can sign the derivative of total hidden consumer consumption N_h with respect to threshold c_h below as

$$\frac{\partial N_h(c_h)}{\partial c_h} = \begin{cases} q\left(\frac{(1-M_g)}{1-l} + \frac{(\beta-M_p)}{1-\gamma l}\right) \geq 0 & \text{if } 0 \leq c_h \leq \gamma l, \\ \frac{(\beta-M_p)(2q-1)}{1-\gamma l} + \frac{(1-M_g)q}{1-l} & \text{if } \gamma l < c_h \leq l, \\ (2q-1)\left(\frac{(1-M_g)}{1-l} + \frac{(\beta-M_p)}{1-\gamma l}\right) \leq 0 & \text{if } l < c_h \leq 1-l, \\ \frac{(\beta-M_p)(2q-1)}{1-\gamma l} - \frac{(1-M_g)(1-q)}{1-l} \leq 0 & \text{if } 1-l < c_h \leq 1-\gamma l, \\ (q-1)\left(\frac{(1-M_g)}{1-l} + \frac{(\beta-M_p)}{1-\gamma l}\right) \leq 0 & \text{if } 1-\gamma l < c_h \leq 1. \end{cases} \tag{25}$$

[1] If instead $q > \frac{1}{2}$, the threshold would be flipped, but the qualitative results remain.

Equation 25 shows that the function is continuous at the boundaries. The firm's payoff is decreasing in c_h for any $c_h > l$ and increasing in c_h for any $c_h < \gamma l$. Therefore, the optimal c_h will be in $[\gamma l, l]$. the effect of c_h on content consumption when $\gamma l \leq c_h \leq l$ depends on q, the proportion of consumers who prefer content A. The derivative of the firm's payoff with respect to the threshold c_h is negative if $q < \hat{q}$, positive if $q > \hat{q}$, and equal to zero at $q = \hat{q}$, where

$$\hat{q} = \frac{(\beta - M_p)(1-l)}{2(\beta - M_p)(1-l) + (1-M_g)(1-\gamma l)}. \tag{26}$$

A.2 Stage 2 Consumer Utility in the Unconstrained Scenario

Below we show the proof for Section 4.1.3. A fraction $1-\alpha$ of consumers have zero cost of lost privacy and make their disclosure decisions based on the utility. Recall that the threshold $\hat{q}$ depends on which consumer groups choose to hide their identities. For example, $\hat{q}_{_G} = \frac{\alpha\beta(1-l)}{2\alpha\beta(1-l)+1-\gamma l}$ is derived by setting $M_p = (1-\alpha)\beta$ and $M_g = 0$ in Equation 26. We denote this using a subscript $\{P, G, B, N\}$ to indicate whether the hiding behavior applies to only the **P**rotected group, only the **G**eneral group, **B**oth groups, or **N**either group. The thresholds are ordered as follows: $\hat{q}_{_G} < \hat{q}_{_N} = \hat{q}_{_B} < \hat{q}_{_P}$. Accordingly, we consider four distinct cases: (i) $q \leq \hat{q}_{_G}$, (ii) $\hat{q}_{_G} < q \leq \hat{q}_{_N}$, (iii) $\hat{q}_{_N} < q \leq \hat{q}_{_P}$, and (iv) $\hat{q}_{_P} < q$. Consumer utility is calculated from plugging the accuracy from Equations 10, 12, and 13 into the consumer utility function from Equation 5.

Below we present the payoffs for each consumer group in the four different cases.

Table 5. Payoffs for Consumer Groups in the Unconstrained Scenario if $q \leq \hat{q}_{_G}$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$

Table 6. Payoffs for Consumer Groups in the Unconstrained Scenario if $\hat{q}_{_G} < q \leq \hat{q}_{_N}$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$

Table 7. Payoffs for Consumer Groups in the Unconstrained Scenario if $\hat{q}_N < q \leq \hat{q}_P$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$

Table 8. Payoffs for Consumer Groups in the Unconstrained Scenario if $\hat{q}_P < q$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$

From Table 5, we see that the protected group's utility does not depend on the identity disclosure decisions. This is because in the preference heterogeneity range $q \leq \hat{q}_G$, the threshold that maximizes accuracy for the hidden group is established at the point that is optimal for hidden protected consumers. Therefore, protected consumers are indifferent between hiding and disclosing their identity and thus disclose. Comparing the general group's payoffs between disclosing and hiding, we see that

$$\frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2 - \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2 = \frac{(1-\gamma)lqv(l((\gamma+3)q-2)-2q+2)}{2(l-1)^2} > 0. \tag{27}$$

Therefore, the general group will also disclose if the preference for content A is low in the range $q \leq \hat{q}_G$. When $q > \hat{q}_G$, the general group is indifferent between hiding and disclosure, thus they disclose.

By comparing payoff levels across these cases in a similar manner, and following the same steps outlined in Section 4.1.3, we demonstrate that both groups choose to disclose their identities for all values of q. ■

A.3 Proof of Lemma 1

In this proof, we analyze the firm's Stage 1 AI learning decisions for Section 4.1.4.

To prove the equilibrium firm choice of l^* in Stage 1, note that the choice of l can determine whether the threshold for the hidden consumers will be set in Stage 3 at $c_h^* = \gamma l$ or $c_h^* = l$. In the former case, the firm's payoff function is given by Equation 16. In the latter case, the firm's payoff function is given by Equation

18. Note that we have assumed the cost of learning effort is sufficiently convex (i.e., k is sufficiently high) such that these expressions are strictly concave in l and thus each have a unique local maximum.

To be an equilibrium, the firm's Stage 1 decision, l^*, must maximize the payoff given the anticipation of the Stage 3 cutoff decision and this anticipation must be confirmed to be accurate (i.e., consistency). It must also either be that a) if the firm's Stage 1 decision is set to maximize the payoff given the anticipation of the alternative Stage 3 cutoff decision, the anticipation will not be confirmed to be accurate (i.e., inconsistency) or b) if both anticipations can be confirmed to be accurate, the firm earns a higher payoff from l^* (i.e., global optimality).

There are two claims in Lemma 1. We first prove that the firm chooses AI learning level l^*_{UNlq} when $q \leq \hat{q}_{UNhq}$. Suppose the firm instead chose l^*_{UNhq} in anticipation of setting $c^*_h = l$ in Stage 3. The fact that $q \leq \hat{q}_{UNhq}$ implies the firm would instead actually choose $c^*_h = \gamma l$. Therefore, the anticipation would not be confirmed to be rational. By definition, the optimal learning effort when anticipating $c^*_h = \gamma l$ is l^*_{UNlq}. This proves that an alternative candidate equilibrium learning effort is not consistent with the manner in which it is derived. We must next show that the proposed candidate equilibrium is consistent with the manner in which it was derived. We prove this in two parts: first showing what happens in the subregion in which $q > \frac{\beta(l-1)^2}{2\beta(l-1)^2+(\gamma l-1)^2}$ and then what happens in the subregion in which $q \leq \frac{\beta(l-1)^2}{2\beta(l-1)^2+(\gamma l-1)^2}$.

Claim 1a: There exists $\tilde{q}_{UN} = \frac{\beta(l-1)^2}{2\beta(l-1)^2+(\gamma l-1)^2}$ such that if $q > \tilde{q}_{UN}$, then $\hat{q}_{UNlq} > \hat{q}_{UNhq}$. This will imply that if $q \leq \hat{q}_{UNhq}$, then setting l^*_{UNlq} in anticipation of the Stage 3 cutoff being $c^*_h = \gamma l$ will lead this anticipation to be confirmed to be accurate.

We prove $\hat{q}_{UNlq} > \hat{q}_{UNhq}$ by showing that

$$\frac{\partial \hat{q}_N}{\partial l} = \frac{\beta(\gamma - 1)}{(2\beta(l-1) + \gamma l - 1)^2} < 0, \tag{28}$$

by $\gamma < 1$, and by showing $l^*_{UNhq} > l^*_{UNlq}$ when $q > \tilde{q}_{UN}$.

To order the two learning levels, note that both Equations 16 and 18 are unimodal differentiable and continuous functions on a closed interval. We compare the derivatives of each function with respect to l.

$$\begin{aligned} \Delta_{UNhvl} &= \frac{\partial \pi_{UNhq}(l)}{\partial l} - \frac{\partial \pi_{UNlq}(l)}{\partial l} \\ &= \frac{\alpha(1-\gamma)\left(q\left(2\beta(l-1)^2 + (\gamma l - 1)^2\right) - \beta(l-1)^2\right)}{(l-1)^2(\gamma l - 1)^2}. \end{aligned} \tag{29}$$

If $\Delta_{UNhvl} > 0$, then it means $l^*_{UNhq} > l^*_{UNlq}$. This is because at l^*_{UNlq}, which is defined as the l such that $\frac{\partial \pi_{UNlq}(l)}{\partial l} = 0$, we know that the derivative $\frac{\partial \pi_{UNhq}(l)}{\partial l} > 0$. Thus, the payoff function defined Equation 18 is maximized at a greater value of l than the payoff function defined in Equation 16, which proves our claim

that $l^*_{UNhq} > l^*_{UNlq}$. Similarly, if $\Delta_{UNhvl} \leq 0$, then it means $l^*_{UNhq} \leq l^*_{UNlq}$.

Taking derivative of Δ_{UNhvl} with respect to q, by the fact that $\gamma < 1$, we can conclude that Δ_{UNhvl} increases in q as $\frac{\alpha(1-\gamma)\left(2\beta(l-1)^2+(\gamma l-1)^2\right)}{(l-1)^2(\gamma l-1)^2} > 0$. We derive $\tilde{q}_{UN}$ by finding the value of q where $l^*_{UNhq} = l^*_{UNlq}$. Thus, for any $q \leq \tilde{q}_{UN}$, we can conclude that $l^*_{UNhq} \leq l^*_{UNlq}$. And for any $q > \tilde{q}_{UN}$, we can conclude that $l^*_{UNhq} > l^*_{UNlq}$.

This concludes the proof of Claim 1a.

Claim 1b: If $q \leq \tilde{q}_{UN}$, defined above in Claim 1a, then it is also true that $q \leq \hat{q}_{UNhq}$and $q \leq \hat{q}_{UNlq}$.

We prove this claim by showing that $\tilde{q}_{UN} < \hat{q}_N$ for any l. By the fact that $\gamma < 1$ and $l < \frac{1}{2}$, we can sign the comparison below as

$$\hat{q}_N - \tilde{q}_{UN} = \frac{\beta(\gamma-1)(l-1)l(\gamma l-1)}{(2\beta(l-1)+\gamma l-1)\left(2\beta(l-1)^2+(\gamma l-1)^2\right)} > 0. \tag{30}$$

This concludes the proof of Claim 1b.

We now prove the claim from Lemma 1 that if $q > \hat{q}_{UNlq}$, then the firm will choose l^*_{UNhq}, which results in $c^*_h = l$, the hidden group threshold coincides at the point optimal for the general consumers.

From the proofs of Claims 1a and 1b above, we know $\hat{q}_{UNlq} > \hat{q}_{UNhq}$ for any $q > \hat{q}_{UNlq}$, which directly implies consistency at l^*_{UNhq} and inconsistency at l^*_{UNlq}. In other words, setting l^*_{UNhq} in anticipation of the Stage 3 cutoff being $c^*_h = l$ will lead this anticipation to be confirmed to be accurate and setting l^*_{UNlq} in anticipation of $c^*_h = \gamma l$ will not actually result in $c^*_h = \gamma l$. ■

A.4 Stage 3 Firm Content Threshold Decisions with Accuracy Parity Policy

Using the same method as in the unconstrained case, the recommendation accuracy for the general group as a function of the threshold is as follows:

$$r_{g1} = \begin{cases} \left(\frac{c_g}{1-l}\right)q + (1-q) & \text{if } c_g < l, \\ \left(\frac{c_g}{1-l}\right)q + \left(\frac{1-c_g}{1-l}\right)(1-q) & \text{if } l < c_g \leq 1-l, \\ q + \left(\frac{1-c_g}{1-l}\right)(1-q) & \text{if } 1-l < c_g \leq 1. \end{cases} \tag{31}$$

We solve for the value of c_g that will make the accuracy from Equation 31 equal to the maximized accuracy of the protected group ($\hat{r}_g(\hat{c}_g) = \bar{r}_{p1}(c^*_p)$). If both the protected and general group disclose their identities, the thresholds chosen for each will be as below.

$$c_p^* = \gamma l$$

$$\hat{c}_g = \begin{cases} \dfrac{\gamma(1-l)l}{q(1-\gamma l)} & \text{if } \frac{\gamma(1-l)}{1-\gamma l} < q, \\ \dfrac{l(-\gamma + q(\gamma l - 1) + 1)}{(2q-1)(\gamma l - 1)} & \text{if } \frac{l(\gamma l-2)+1}{(3l-2)(\gamma l-1)} < q \leq \frac{\gamma(1-l)}{1-\gamma l}, \\ \dfrac{2\gamma l^2 q - l(\gamma + (\gamma+2)q - 1) + q}{(q-1)(\gamma l - 1)} & \text{if } q \leq \frac{l(\gamma l-2)+1}{(3l-2)(\gamma l-1)}. \end{cases} \tag{32}$$

A.5 Stage 2 Consumer Utility with Accuracy Parity

The proof here follows a similar process as in Appendix A.2. Once again, we have four distinct cases to consider: (i) $q \leq \hat{q}_G$, (ii) $\hat{q}_G < q \leq \hat{q}_N$, (iii) $\hat{q}_N < q \leq \hat{q}_P$, and (iv) $\hat{q}_P < q$. The first case is analyzed in detail in the main body of the paper (see Section 4.2.2). Below we present the payoffs for each consumer group in the remaining three cases.

Table 9. Payoffs for Consumer Groups with Accuracy Parity if $q \leq \hat{q}_G$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$

Table 10. Payoffs for Consumer Groups with Accuracy Parity if $\hat{q}_G < q \leq \hat{q}_N$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2$

Table 11. Payoffs for Consumer Groups with Accuracy Parity if $\hat{q}_N < q \leq \hat{q}_P$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$

Table 12. Payoffs for Consumer Groups with Accuracy Parity if $\hat{q}_P < q$

	Protected group discloses	Protected group hides
General group discloses	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$
General group hides	$U_{p1} = \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$	$U_{p0} = \frac{v}{2}\left(\frac{ql}{1-\gamma l} + \frac{(1-q)(1-l)}{1-\gamma l}\right)^2$ $U_{g0} = \frac{v}{2}\left(\frac{ql}{1-l} + (1-q)\right)^2$

As an example, consider the case of $q \leq \hat{q}_G$, i.e., consumer preference is rather homogeneous. The utilities for this case are presented in Table 9. In this case, the protected group's utility is indifferent between disclosing and hiding as the hidden group threshold coincides with the point optimizing hidden protected consumers, so they disclose. Below, we compare the general group's payoffs between hiding and disclosing using the utilities from Table 9.

$$\frac{v}{2}\left(\frac{q\gamma l}{1-l} + (1-q)\right)^2 - \frac{v}{2}\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right)^2 = \frac{v\left(\left((1-\gamma)\gamma l^2 q\right)\left(q\left(\frac{\gamma l}{1-l} + \frac{1}{1-\gamma l} - 3\right) + 2\right)\right)}{2((1-l)(1-\gamma l))} > 0. \tag{33}$$

Since $\gamma < 1$ and $l < \frac{1}{2}$, we can conclude from Equation 33 above that the general group is better off hiding in the range $q \leq \hat{q}_G$.

By comparing payoff levels across these cases, and following the same steps outlined in Section 4.2.2, we demonstrate that the protected group chooses to disclose, and the general group chooses to hide for all values of q. ■

A.6 Proof of Lemma 3

In this section, we analyze the firm learning effort when the accuracy parity policy is imposed for Section 4.2.3. The proof is similar to the analysis when the firm is unconstrained in Appendix A.3.

Similarly, the choice of l can determine whether the threshold for the hidden consumers will be set in Stage 3 at $c_h^* = \gamma l$ or $c_h^* = l$. In the former case, the firm's payoff function is given by Equation 20. In the latter case, the firm's payoff function is given by Equation 22. We have assumed the cost of learning effort

is sufficiently convex such that these expressions are strictly concave in l and thus each have a unique local maximum.

There are two claims in Lemma 3. We first prove that the firm chooses AI learning level l^*_{PPlq} when $q \leq \hat{q}_{PPhq}$. Suppose the firm instead chose l^*_{PPhq} in anticipation of setting $c^*_h = l$ in Stage 3. The fact that $q \leq \hat{q}_{PPhq}$ implies the firm would instead actually choose $c^*_h = \gamma l$. Therefore, the anticipation would not be confirmed to be rational. By definition, the optimal learning effort when anticipating $c^*_h = \gamma l$ is l^*_{PPlq}. This proves that an alternative candidate equilibrium learning effort is not consistent with the manner in which it is derived. We must next show that the proposed candidate equilibrium is consistent with the manner in which it was derived. We prove this in two parts: first showing what happens in the subregion in which $q > \frac{\alpha\beta(l-1)^2}{2\alpha\beta(l-1)^2+(\gamma l-1)^2}$ and then what happens in the subregion in which $q \leq \frac{\alpha\beta(l-1)^2}{2\alpha\beta(l-1)^2+(\gamma l-1)^2}$.

Claim 3a: Similar to Claim 1a in Appendix A.3, we prove below that if $q > \tilde{q}_{PP} = \frac{\alpha\beta(l-1)^2}{2\alpha\beta(l-1)^2+(\gamma l-1)^2}$, then $\hat{q}_{PPlq} > \hat{q}_{PPhq}$, which will imply that if $q < \hat{q}_{PPhq}$, then setting l^*_{PPlq} in anticipation of the Stage 3 cutoff being $c^*_h = \gamma l$ will lead this anticipation to be confirmed to be accurate.

We prove $\hat{q}_{PPlq} > \hat{q}_{PPhq}$ by showing that

$$\frac{\partial \hat{q}_G}{\partial l} = \frac{\alpha\beta(\gamma-1)}{(2\alpha\beta(l-1)+\gamma l-1)^2} < 0 \tag{34}$$

and $l^*_{PPhq} > l^*_{PPlq}$ when $q > \tilde{q}_{PP}$.

To order the two learning levels, note that both Equations 20 and 22 are unimodal differentiable and continuous functions on a closed interval. We compare the derivatives of each function with respect to l.

$$\begin{aligned} \Delta_{PPhvl} &= \frac{\partial \pi_{PPhq}(l)}{\partial l} - \frac{\partial \pi_{PPlq}(l)}{\partial l} \\ &= \frac{(1-\gamma)\left(q\left(2\alpha\beta(l-1)^2+(\gamma l-1)^2\right)-\alpha\beta(l-1)^2\right)}{(l-1)^2(\gamma l-1)^2}. \end{aligned} \tag{35}$$

Similar to Claim 1a in Appendix A.3, if $\Delta_{PPhvl} > 0$, then it means $l^*_{PPhq} > l^*_{PPlq}$. If $\Delta_{PPhvl} < 0$, then it means $l^*_{PPhq} < l^*_{PPlq}$. Taking derivative of Δ_{PPhvl} with respect to q, by the fact that $\gamma < 1$, we can conclude that Δ_{PPhvl} increases in q as $\frac{(1-\gamma)\left(2\alpha\beta(l-1)^2+(\gamma l-1)^2\right)}{(l-1)^2(\gamma l-1)^2} > 0$. We derive $\tilde{q}_{PP}$ by finding the point of q where $l^*_{PPhq} = l^*_{PPlq}$. Thus, for any $q < \tilde{q}_{PP}$, we can conclude that $l^*_{PPhq} < l^*_{PPlq}$. And for any $q > \tilde{q}_{PP}$, we can conclude that $l^*_{PPhq} > l^*_{PPlq}$.

This concludes the proof of Claim 3a.

Claim 3b: Similar to Claim 1b in Appendix A.3, we prove consistency below by showing that if $q < \tilde{q}_{PP}$, then it is also true that $q < \hat{q}_{PPhq}$ and $q < \hat{q}_{PPlq}$. We prove this by showing that $\tilde{q}_{PP} < \hat{q}_G$ for any l. By the fact that $\gamma < 1$ and $l < \frac{1}{2}$, we can sign the comparison below as

$$\hat{q}_G - \tilde{q}_{PP} = \frac{\alpha\beta(\gamma-1)(l-1)l(\gamma l-1)}{(2\alpha\beta(l-1)+\gamma l-1)\left(2\alpha\beta(l-1)^2+(\gamma l-1)^2\right)} > 0. \tag{36}$$

This concludes the proof that if consumer preference for content A is sufficiently low (i.e., $q < \hat{q}_{PPhq}$) then the firm will choose l^*_{PPlq}, which results in $c^*_h = \gamma l$.

We now prove the claim from Lemma 3 that if $q > \hat{q}_{PPlq}$, then the firm will choose l^*_{PPhq}, which results in $c^*_h = l$, the hidden group threshold coincides at the point optimal for the general consumers.

From proof of Claim 3b above, we know $\hat{q}_{PPlq} > \hat{q}_{PPhq}$ for any $q > \hat{q}_{PPlq}$, which directly implies setting l^*_{PPhq} in anticipation of the Stage 3 cutoff being $c^*_h = l$ will lead this anticipation to be confirmed to be accurate. It also implies that setting l^*_{PPlq} in anticipation of $c^*_h = \gamma l$ will not actually result in $c^*_h = \gamma l$. ■

A.7 Proof of Proposition 1

In this proof, we compare the learning efforts under the accuracy parity policy vs. under the unconstrained scenario. We show that in the range of $\underline{q} < q < \hat{q}_{UNhq}$, accuracy parity leads to a higher learning level, that is $l^*_{PPhq} > l^*_{UNlq}$.

According to Lemmas 1 and 3, when $\underline{q} < q < \hat{q}_{UNhq}$, the firm chooses l^*_{UNlq} in the unconstrained condition and l^*_{PPhq} with Accuracy Parity policy according to the proofs.

Because the firm's payoff functions $\pi(l)$ are unimodal functions, if the marginal benefit of learning levels are such that $\frac{\partial \pi_{PPhq}(l)}{\partial l} > \frac{\partial \pi_{UNlq}(l)}{\partial l}$ $\forall l$, then we can conclude that the learning efforts $l^*_{PPhq} > l^*_{UNlq}$. Therefore, in the following section we compare the learning efforts by comparing the marginal benefit of learning across policies.

As defined in Table 2, $\underline{q} = \max\{\check{q}(l^*_{UNhq}), \hat{q}_{PPlq}\}$. Below we first start with the range $\hat{q}_{PPlq} < q < \hat{q}_{UNhq}$, then introduce how we arrive at $\check{q}(l^*_{UNhq})$, and next prove the existence of the range $\underline{q} < q < \hat{q}_{UNhq}$.

First, consider $\hat{q}_{PPlq} < q < \hat{q}_{UNhq}$. In this case:

$$\frac{\partial \pi_{PPhq}}{\partial l} - \frac{\partial \pi_{UNlq}}{\partial l} = \hat{\Delta} = \frac{\alpha(1-\gamma)\left(q\left(2\beta(l-1)^2+(\gamma l-1)^2\right)-\beta(l-1)^2\right)}{(l-1)^2(\gamma l-1)^2}. \tag{37}$$

Next, taking the derivative of $\hat{\Delta}$ with respect to q, we can sign $\frac{\partial \hat{\Delta}}{\partial q} = \frac{\alpha(1-\gamma)\left(2\beta(l-1)^2+(\gamma l-1)^2\right)}{(l-1)^2(\gamma l-1)^2} > 0$. This means $\hat{\Delta}$ increases in q. Also $\hat{\Delta}$ equals to zero at a $\check{q} = \frac{\beta(1-l)^2}{2\beta(1-l)^2+(1-\gamma l)^2}$. Therefore, $l^*_{PPhq} > l^*_{UNlq}$, when $q > \check{q}$. Thus, the accuracy parity policy increases AI learning effort when $\max\{\check{q}, \hat{q}_{PPlq}\} < q < \hat{q}_{UNhq}$.

To verify that the range of $\max\{\check{q}, \hat{q}_{PPlq}\} < q < \hat{q}_{UNhq}$ exists, we first prove that the range of $\check{q} < q < \hat{q}_{UNhq}$ exists, and then prove that the range of $\hat{q}_{PPlq} < q < \hat{q}_{UNhq}$ exists.

First, we show that $\hat{q}_{UNhq} > \check{q}(l^*_{UNhq})$ in the unconstrained scenario. By the fact that $l^*_{UNhq} < \frac{1}{2}$, and $\gamma < 1$, we can sign below as

$$\begin{aligned}\hat{q}_{UNhq} - \breve{q}(l^*_{UNhq}) &= \frac{\beta(1-l^*_{UNhq})}{2\beta(1-l^*_{UNhq})+1-\gamma l^*_{UNhq}} - \frac{\beta(1-l^*_{UNhq})^2}{2\beta(1-l^*_{UNhq})^2+(1-\gamma l^*_{UNhq})^2} \\ &= \frac{\beta(\gamma-1)(l^*_{UNhq}-1)l^*_{UNhq}(\gamma l^*_{UNhq}-1)}{(2\beta(l^*_{UNhq}-1)+\gamma l^*_{UNhq}-1)\Big(2\beta(l^*_{UNhq}-1)^2+(\gamma l^*_{UNhq}-1)^2\Big)} > 0.\end{aligned} \tag{38}$$

Next, we show that the range of $\hat{q}_{PPlq} < q < \hat{q}_{UNhq}$ exists. We know that both functions $\hat{q}_N(l) = \frac{\beta(1-l)}{2\beta(1-l)+1-\gamma l}$ and $\hat{q}_G(l) = \frac{\alpha\beta(1-l)}{2\alpha\beta(1-l)+1-\gamma l}$ decrease in learning l from previous analysis in Appendix A.3 and A.6, respectively. Thus, the highest $\hat{q}$ cutoff for the accuracy parity policy is when $\hat{q}_G(l=0)$ and the lowest $\hat{q}$ cutoff for the unconstrained scenario is when $\hat{q}_N(l=\frac{1}{2})$. Thus, the distance between the two cutoffs is $\hat{q}_{UNhq} - \hat{q}_{PPlq} > \hat{q}_N(l=\frac{1}{2}) - \hat{q}_G(l=0) = \beta\left(\frac{1}{2\beta-\gamma+2} - \frac{\alpha}{2\alpha\beta+1}\right)$. The distance decreases in α, equals to zero when $\alpha = \frac{1}{2-\gamma}$, and is thus greater than zero when $\alpha < \frac{1}{2-\gamma}$. Note that this is the strictest case with the smallest possible distance between $\hat{q}_N$ and $\hat{q}_G$. When $l^* \in (0, \frac{1}{2})$, there is a broader range of α such that $\hat{q}_{PPlq} < q < \hat{q}_{UNhq}$. ■

A.8 Proof of Proposition 2

In this proof, we compare the general group's combined group-level accuracy under the accuracy parity policy to that under the unconstrained scenario. Note that the combined group-level accuracy is a weighted average across both the α proportion of consumers who have high costs of privacy and thus never disclose their type and the $(1-\alpha)$ proportion of consumers who have zero cost of privacy. As shown in the previous analysis, the latter consumers only choose to disclose in the unconstrained benchmark. This weighted average is defined mathematically below.

$$r_j = \alpha r_{j0} + (1-\alpha) r_{j1}. \tag{39}$$

We show below that compared to the unconstrained scenario, an accuracy parity policy leads to an increase in the combined group-level accuracy of the general consumers when $\underline{q} < q < \hat{q}_{UNhq}$, where the bounds on this range are defined in Table 2. Given this range of q, mathematically our claim can be written as: $\bar{r}_{g0}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$. We prove this claim in three steps: (i) First, establish that at a fixed learning level l^*_{PPhq}, the following holds: $\bar{r}_{g0}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{PPhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPhq})$. (ii) Next, we show that both accuracy levels, $\underline{r}_{g0}$ and $\bar{r}_{g1}$, increase in l, which implies our third point: (iii) Since $l^*_{PPhq} > l^*_{UNlq}$, $\alpha \underline{r}_{g0}(l^*_{PPhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$. From these three steps, By transitivity, we conclude that $\bar{r}_{g0}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$.

Step i: First, we compare the combined general group's accuracy under accuracy parity policy to the unconstrained scenario evaluated at the learning level l^*_{PPhq}. Specifically, we plug l^*_{PPhq} into $\bar{r}_{g0} - (\alpha \underline{r}_{g0} +$

$(1-\alpha)\bar{r}_{g1}$). By the fact that $l^*_{PPhq} < \frac{1}{2}$ and $\gamma < 1$, we can sign the comparison as

$$\begin{aligned} &\frac{ql^*_{PPhq}}{1-l^*_{PPhq}} + (1-q) - \\ &\left(\alpha\left(\frac{q\gamma l^*_{PPhq}}{1-l^*_{PPhq}} + (1-q)\right) + (1-\alpha)\left(\frac{ql^*_{PPhq}}{1-l^*_{PPhq}} + (1-q)\right)\right) = \frac{(1-\gamma)\alpha q l^*_{PPhq}}{1-l^*_{PPhq}} > 0. \end{aligned} \tag{40}$$

Step ii: By the fact that $q > 0$ and $\gamma > 0$, we can sign the following derivatives:

$$\frac{\partial \underline{r}_{g0}}{\partial l} = \frac{\gamma q}{(l-1)^2} > 0, \tag{41}$$

$$\frac{\partial \bar{r}_{g1}}{\partial l} = \frac{q}{(l-1)^2} > 0. \tag{42}$$

Step iii: Last, we know from the proof of A.7 Equation 37 that the AI learning level is higher under the accuracy parity policy compared to the unconstrained scenario $l^*_{PPhq} > l^*_{UNlq}$ if $\underline{q} < q < \hat{q}_{UNhq}$. In conclusion, we have proven $\bar{r}_{g0}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{PPhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPhq}) > \alpha \underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$, which proves that the combined general group's accuracy is higher under the accuracy parity policy than in the unconstrained benchmark. ■

A.9 Proof of Proposition 3

This proof is similar to the proof of Proposition 2 in Section A.8. We compare the protected group's combined group-level accuracy under the accuracy parity policy to that under the unconstrained scenario. Note that the combined group-level accuracy is a weighted average across both the α proportion of consumers who have high costs of privacy and thus never disclose their type and the $(1-\alpha)$ proportion of consumers who have zero cost of privacy. As shown previously the disclosure decisions are the same in the accuracy parity policy condition as in the unconstrained benchmark. Mathematically, the weighted average accuracy is greater in the accuracy parity policy condition than the unconstrained benchmark when $q < \hat{q}_{PPhq}$ if $(1-\alpha)\bar{r}_{p1}(l^*_{PPlq}) + \alpha\bar{r}_{p0}(l^*_{PPlq}) < (1-\alpha)\bar{r}_{p1}(l^*_{UNlq}) + \alpha\bar{r}_{p0}(l^*_{UNlq})$.

From Equation 12 and Table 5, $\bar{r}_{p1} = \bar{r}_{p0} = \frac{q\gamma l}{1-\gamma l} + (1-q)$. By the fact that $(1-\alpha)\bar{r}_{p1}(l) + \alpha\bar{r}_{p0}(l)$ is increasing in l and the proof of A.7 Equation 43 shows that $l^*_{PPlq} < l^*_{UNlq}$, we can conclude $(1-\alpha)\bar{r}_{p1}(l^*_{PPlq}) + \alpha\bar{r}_{p0}(l^*_{PPlq}) < (1-\alpha)\bar{r}_{p1}(l^*_{UNlq}) + \alpha\bar{r}_{p0}(l^*_{UNlq})$. ■

A.10 Proof of Table 3

In this section, we prove the remaining claims from Table 3 that have yet to be established from the proofs for Propositions 1 and 2. We focus on the ranges of q in which the equilibrium learning levels are established in Lemmas 1 and 3.

A.10.1 Learning level comparisons

We first show the impact of an accuracy parity policy on learning levels. The proofs below are similar to the proof of Proposition 1 in Section A.7. We analyze the learning levels first in a low q condition, and then in a high q condition.

1. Suppose $q < \hat{q}_{PPhq}$. Below we show that if $q < \hat{q}_{PPhq}$, then $l^*_{PPlq} < l^*_{UNlq}$. The firm's payoff function π_{PPlq} is in Equation 20, and π_{UNlq} is in Equation 16. We compare the marginal benefit of learning under the accuracy parity policy vs. the unconstrained scenario. By the fact $\alpha < 1$ and $\gamma < 1$, we can sign the comparison below as:

$$\frac{\partial \pi_{PPlq}}{\partial l} - \frac{\partial \pi_{UNlq}}{\partial l} = \frac{q(\alpha - 1)(1 - \gamma)}{(l-1)^2} < 0, \tag{43}$$

thus, $l^*_{PPlq} < l^*_{UNlq}$.

This concludes the proof for the range $q < \hat{q}_{PPhq}$. ■

2. Suppose $q > \hat{q}_{UNlq}$. Below we show that if $q > \hat{q}_{UNlq}$, then accuracy parity policy does not change AI learning effort; that is $l^*_{PPhq} = l^*_{UNhq}$. The firm's payoff function π_{PPhq} is in Equation 22, and π_{UNhq} is in Equation 18. We compare the marginal benefit of learning under the accuracy parity policy vs. the unconstrained scenario below:

$$\frac{\partial \pi_{PPhq}}{\partial l} - \frac{\partial \pi_{UNhq}}{\partial l} = 0, \tag{44}$$

thus, $l^*_{PPhq} = l^*_{UNhq}$.

This concludes the proof for the range $q > \hat{q}_{UNlq}$. ■

A.10.2 Impact on general consumers

Next, we show the impact of an accuracy parity policy on general group's combined group-level accuracy, defined in Equation 39. The proofs below are similar to the proof of Proposition 2 in Section A.8. We analyze the accuracy 1, first in a low q condition, then 2, in a high q condition.

1. Suppose $q < \hat{q}_{PPhq}$. We show below that compared to the unconstrained scenario, an accuracy parity policy leads to a lower combined group-level accuracy when $q < \hat{q}_{PPhq}$, where the bound on this

range is defined in Table 2. Given this range of q, mathematically our claim can be written as: $\underline{r}_{g0}(l^*_{PPlq}) < \alpha\underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$. We prove this claim in three steps: (i) First, establish that at a fixed learning level l^*_{PPlq}, the following holds: $\underline{r}_{g0}(l^*_{PPlq}) < \alpha\bar{r}_{g0}(l^*_{PPlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPlq})$. (ii) Next, we show that both accuracy levels, $\underline{r}_{g0}$ and $\bar{r}_{g1}$, increase in l, which implies our third point: (iii) Since $l^*_{PPlq} < l^*_{UNlq}$, $\alpha\bar{r}_{g0}(l^*_{PPlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPlq}) < \alpha\bar{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$. From these three steps, By transitivity, we conclude that $\underline{r}_{g0}(l^*_{PPlq}) < \alpha\underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$.

When $q < \hat{q}_{PPhq}$, the hidden group threshold coincides at the point optimal for the protected consumers in both the unconstrained scenario and the scenario with the accuracy parity policy. Thus, the accuracy for the entire hidden general group consumers with the accuracy parity policy is $\underline{r}_{g0} = \frac{q\gamma l}{1-l} + (1-q)$ from Equation 13 and Table 9. The accuracy for the general group consumers in the unconstrained scenario, with $(1-\alpha)$ who disclose is $\bar{r}_{g1} = \frac{ql}{1-l} + (1-q)$ from Equation 10, and α who always hide, given that the hidden group threshold coincides at the point optimal for the protected consumers, has accuracy is $\underline{r}_{g0} = \frac{q\gamma l}{1-l} + (1-q)$ from Equation 13 and Table 5.

Step i: First, we compare the combined general group's accuracy under the accuracy parity policy to the unconstrained scenario at a common learning level l^*_{PPlq}. By the fact that $l^*_{PPlq} < \frac{1}{2}$ and $\gamma < 1$, we can sign the comparison as

$$\begin{aligned}\underline{r}_{g0} - (\alpha\underline{r}_{g0} + (1-\alpha)\bar{r}_{g1}) = \frac{q\gamma l^*_{PPlq}}{1-l^*_{PPlq}} + (1-q) - \\ \left(\alpha\left(\frac{q\gamma l^*_{PPlq}}{1-l^*_{PPlq}} + (1-q)\right) + (1-\alpha)\left(\frac{ql^*_{PPlq}}{1-l^*_{PPlq}} + (1-q)\right)\right) = \frac{(\alpha-1)(\gamma-1)l^*_{PPlq}q}{l^*_{PPlq}-1} < 0.\end{aligned} \tag{45}$$

Step ii: By the fact that $q > 0$ and $\gamma > 0$, we can sign the following derivatives:

$$\frac{\partial\underline{r}_{g0}}{\partial l} = \frac{\gamma q}{(l-1)^2} > 0, \tag{46}$$

$$\frac{\partial\bar{r}_{g1}}{\partial l} = \frac{q}{(l-1)^2} > 0. \tag{47}$$

Step iii: Last, we know from the proof of A.7 Equation 43 that AI learning level is lower under the accuracy parity policy compared to the unconstrained scenario $l^*_{PPlq} < l^*_{UNlq}$ if $q < \hat{q}_{PPhq}$. In conclusion, we have proven $\underline{r}_{g0}(l^*_{PPlq}) < \alpha\bar{r}_{g0}(l^*_{PPlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPlq}) < \alpha\underline{r}_{g0}(l^*_{UNlq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNlq})$, which proves that the combined general group's accuracy is lower under the accuracy parity policy than in the unconstrained benchmark. ■

2. Suppose $q > \hat{q}_{UNlq}$. We show below that compared to the unconstrained scenario, the accuracy

parity policy leads to no change to the combined group-level accuracy when $q > \hat{q}_{UNlq}$, where the bound on this range is defined in Table 2. Given this range of q, mathematically our claim can be written as: $\bar{r}_{g0}(l^*_{PPhq}) = \alpha\bar{r}_{g0}(l^*_{UNhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNhq})$. We prove this claim in two steps: (i) First, establish that at a fixed learning level l^*_{PPhq}, the following holds: $\bar{r}_{g0}(l^*_{PPhq}) = \alpha\bar{r}_{g0}(l^*_{PPhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPhq})$. (ii) Next, since $l^*_{PPhq} = l^*_{UNhq}$, we conclude that $\bar{r}_{g0}(l^*_{PPhq}) = \alpha\bar{r}_{g0}(l^*_{UNhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNhq})$.

When $q > \hat{q}_{UNlq}$, the hidden group threshold coincides at the point optimal for the general consumers in both the unconstrained and the accuracy parity policy scenarios, thus the accuracy for the entire hidden general group consumers, under the accuracy parity policy is $\bar{r}_{g0} = \frac{ql}{1-l} + (1-q)$ is from Equation 13 and Table 12. The accuracy for the general group consumers in the unconstrained scenario, with $(1-\alpha)$ who disclose is $\bar{r}_{g1} = \frac{ql}{1-l} + (1-q)$ from Equation 10, and α who always hide is $\bar{r}_{g0} = \frac{ql}{1-l} + (1-q)$ from Equation 13 and Table 8. Note that $\bar{r}_{g1} = \bar{r}_{g0}$.

Step i: First, we compare the combined general group's accuracy under the accuracy parity policy to the unconstrained scenario at a common learning level l^*_{PPhq}.

$$\bar{r}_{g0} - (\alpha\bar{r}_{g0} + (1-\alpha)\bar{r}_{g1}) = \frac{ql^*_{PPhq}}{1-l^*_{PPhq}} + (1-q) - \left(\frac{ql^*_{PPhq}}{1-l^*_{PPhq}} + (1-q)\right) = 0, \tag{48}$$

Step ii: Next, we know from the proof of A.7 Equation 44 that $l^*_{PPhq} = l^*_{UNhq}$ if $q > \hat{q}_{UNlq}$. In conclusion, we have proven $\bar{r}_{g0}(l^*_{PPhq}) = \alpha\bar{r}_{g0}(l^*_{PPhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{PPhq}) = \alpha\bar{r}_{g0}(l^*_{UNhq}) + (1-\alpha)\bar{r}_{g1}(l^*_{UNhq})$, which proves that the combined general group's accuracy is the same under the accuracy parity policy as in the unconstrained benchmark. ■

A.11 Proof of Extension 6.1

In the main model, we assumed that the general and protected groups exhibited the same level of heterogeneity in content preferences. In this extension, we relax that assumption and allow for differences between the groups.

Intuitively, as q decreases, content preferences within a group become more homogeneous, making it easier for the firm to deliver content that aligns with the majority consumers' preferences within that group. Conversely, as q approaches $\frac{1}{2}$, preference entropy increases, making it more challenging for the firm to accurately match content to consumers. Here, we explore a scenario where the degree of heterogeneity (or homogeneity) in content preferences varies between groups, allowing us to examine its impact on the firm's decision-making and outcomes.

We begin with the same model primitives as those described in Section 3. The key difference in this extension is that we now allow content preferences to differ between the protected group and the general group. Specifically, we denote the fraction of consumers who prefer content A in the protected group as q_p,

and the fraction in the general group as q_g. Since q values range between 0 and $\frac{1}{2}$, smaller q values indicate greater homogeneity in preferences within a group. Therefore, different values of q_p and q_g reflect varying degrees of preference difference between the two groups.

Following the solution process outlined in Section 4, we first analyze the unconstrained scenario and then the scenario with the accuracy parity policy. The solution follows the same backward induction process, moving through the four stages in the timeline depicted in Figure 2, but with q replaced by q_p for the protected group and q_g for the general group. The first notable difference in the solution process is in defining the condition that determines when the hidden group threshold coincides with the point optimal for the general consumers over the hidden protected consumers, or vice versa, among all the hidden consumers, specifically when deciding the content threshold in Stage 3. As a reminder, the condition is expressed as a threshold that represents the values of q below which the hidden group threshold coincides with the point optimal for the the hidden protected consumers, where $c_h = \gamma l$. Conversely, when q exceeds the threshold, the hidden group threshold coincides at the point optimal for the general consumers, where $c_h = l$. Additionally, recall that the specific threshold depends on which groups hide: whether neither the general nor the protected group hides, only the general group hides, only the protected group hides, or both groups hide; these cases correspond to the thresholds $\hat{q}_N$, $\hat{q}_G$, $\hat{q}_P$, and $\hat{q}_B$, respectively.

In the main model, these thresholds were defined solely in terms of q. In this extension, however, the threshold depends on both q_p and q_g. For instance, when neither group hides, the hidden group threshold coincides at the point optimal for the protected consumers if $q_p \leq \hat{q}_N$, where $\hat{q}_N = \frac{1}{2}\left(1 - \frac{q_g(1-\gamma l)}{\beta(1-l)}\right)$. Similarly, when only the general group hides, the hidden group threshold coincides at the point optimal for the protected group if $q_p \leq \hat{q}_G$, where $\hat{q}_G = \frac{1}{2}\left(1 - \frac{q_g(1-\gamma l)}{\alpha\beta(1-l)}\right)$. These two thresholds play a crucial role in demonstrating the robustness of the main result later.

For the unconstrained scenario in Stage 2, by comparing accuracy levels across all values of q_p (since optimal consumer utility is proportional to accuracy), we find that the optimal decision for both the protected and general groups is to disclose their identity. The solution process and the resulting observation closely mirror those in the main model.

In Stage 1, the profit function that the unconstrained firm optimizes for is also analogous to that of the main model as specified by Equations 16 and 18 with q replaced with q_p and q_g, in respective places. For instance, since neither the protected group nor the general group hides, the threshold of interest is $\hat{q}_N$, and the profit expression for $q_p < \hat{q}_N$ is:

$$\begin{aligned}\pi_{_{UNlq}} =&(1-\alpha)\left(\beta\left(\frac{q_p\gamma l}{1-\gamma l}+1-q_p\right)+\left(\frac{q_g l}{1-l}+1-q_g\right)\right)+\\ &\alpha\left(\beta\left(\frac{q_p\gamma l}{1-\gamma l}+1-q_p\right)+\frac{q_g\gamma l}{1-l}+1-q_g\right)-\tau(l),\end{aligned} \tag{49}$$

and for $q_p > \hat{q}_{_N}$ is:

$$\begin{aligned}\pi_{_{UNhq}} =&(1-\alpha)\left(\beta\left(\frac{q_p\gamma l}{1-\gamma l}+1-q_p\right)+\left(\frac{q_g l}{1-l}+1-q_g\right)\right)+\\ &\alpha\left(\beta\left(\frac{q_p l}{1-\gamma l}+\frac{(1-q_p)(1-l)}{1-\gamma l}\right)+\frac{q_g l}{1-l}+1-q_g\right)-\tau(l).\end{aligned} \tag{50}$$

We then determine the optimal level of learning that maximizes the firm's profit. Given that the expressions for optimal learning are complex, we instead analyze the marginal return on learning by computing the first-order derivative of each profit function with respect to l, denoted as $\frac{\partial \pi_{_{UNlq}}}{\partial l}$ and $\frac{\partial \pi_{_{UNhq}}}{\partial l}$. For the purpose of this extension, we focus on the optimal learning level l^* derived from solving the profit maximization problem based on the expression for $\pi_{_{UNlq}}$. Specifically, when $q_p < \hat{q}_{_N}(l^*)$, the optimal learning level is determined by the respective marginal return on learning, $\frac{\partial \pi_{_{UNlq}}}{\partial l}$.

Now, moving on to the accuracy parity policy scenario, in Stage 2, the primary effect of the accuracy parity policy is that it forces the accuracy of the disclosed group with higher accuracy to match that of the group with lower accuracy. Since there are now two distinct q values, one for the general group (q_g) and one for the protected group (q_p), it is possible for the disclosed general group to have higher accuracy than the disclosed protected group, or vice versa.

When the general group is more homogeneous than the protected group (i.e., $q_p > q_g$), or the protected group is not highly homogeneous (i.e., $q_p > q_{pvg}$, where q_{pvg} is derived from comparing the accuracy levels of the disclosed protected and general groups in Stage 2), the accuracy of the disclosed general group is always higher than that of the disclosed protected group, making the accuracy parity policy a binding constraint, just as in the main model.

However, if the protected group is significantly more homogeneous than the general group (i.e., $q_p < q_g$ and $q_p < q_{pvg}$), then the disclosed protected group can exhibit higher accuracy than the disclosed general group. This outcome renders the accuracy parity policy irrelevant, as the protected group naturally attains a higher accuracy without any fairness constraint. Thus, when the protected group is highly homogeneous, accuracy parity does not serve a meaningful role. A practical interpretation of this observation is that when the protected group's preferences are highly homogeneous and significantly more so than that of the general group, content matching accuracy remains high despite the disparity in learning. In this case, the disparity

in learning is more than compensated for by the greater predictability of a highly homogeneous group's preferences. We formally summarize this observation below:

Observation 1. *When the protected group is sufficiently homogeneous, its accuracy can exceed that of the general group. In such cases, the accuracy parity policy becomes irrelevant in practice.*

Moving forward, we focus on the case where q_p is above q_{pvg} (i.e., when the protected group is not highly homogeneous) to test the robustness of our key result. As in the main model, we find that the general group can still find it beneficial to hide, whereas the protected group always prefers to disclose. Focusing on the cases where only the general group chooses to hide, the profit function then depends on the $\hat{q}_{_G}$ threshold. We analyze the profit function above this threshold, which follows a similar form to Equation 22 in the main model, with q replaced by q_p and q_g in their respective positions:

$$\begin{aligned}\pi_{_{PPhq}} =&(1-\alpha)\beta\left(\frac{q_p\gamma l}{1-\gamma l}+1-q_p\right)+\\ &\alpha\beta\left(\frac{q_p l}{1-\gamma l}+\frac{(1-q_p)(1-l)}{1-\gamma l}\right)+\left(\frac{q_g l}{1-l}+1-q_g\right)-\tau(l).\end{aligned} \tag{51}$$

To demonstrate the robustness of the results, the ordering between $\hat{q}_{_N}(l^*)$ in the unconstrained scenario and $\hat{q}_{_G}(l^*)$ in the accuracy parity scenario is crucial. Since expressing the optimal learning level (l^*) explicitly is difficult, instead, we evaluate the values of $\hat{q}_{_N}$ and $\hat{q}_{_G}$ at extreme values of l. Both thresholds decrease with l (as indicated by the negative first-order derivatives, i.e., $\frac{\partial \hat{q}_{_N}}{\partial l}=-\frac{q_g(1-\gamma)}{2\beta(1-l)^2}<0$ and $\frac{\partial \hat{q}_{_G}}{\partial l}=-\frac{q_g(1-\gamma)}{2\alpha\beta(1-l)^2}<0$), meaning the upper limit of $\hat{q}_{_G}(l)$ occurs at $l=0$, and the lower limit of $\hat{q}_{_N}(l)$ occurs at $l=\frac{1}{2}$.

We can show that, even at these extreme values, $\hat{q}_{_G}(0)<\hat{q}_{_N}(\frac{1}{2})$ holds under certain conditions, in which case there exists a range of q_p values where the optimal learning l^* is defined by setting $\frac{\partial \pi_{_{PPhq}}}{\partial l}=0$ in the accuracy parity scenario and by $\frac{\partial \pi_{_{UNlq}}}{\partial l}=0$ in the unconstrained scenario. Furthermore, we also find that $\frac{\partial \pi_{_{PPhq}}}{\partial l}>\frac{\partial \pi_{_{UNlq}}}{\partial l}$ when q_p is sufficiently large ($q_p>\frac{1}{2}-\frac{q_g(\gamma l-1)^2}{2\beta(l-1)^2}$), which lies between $\hat{q}_{_G}(0)$ and $\hat{q}_{_N}(\frac{1}{2})$ when $\hat{q}_{_G}(0)<\hat{q}_{_N}(\frac{1}{2})$ holds. This echoes our key result described in Proposition 1: there exists conditions that the accuracy parity policy can incentivize the firm to increase its investment in learning to enhance the accurate classification of consumer preferences. ■

A.12 Proof of Extension 6.2

In the main model, privacy sensitivity, and consequently the fraction of consumers who choose to hide no matter what, was independent of group membership. A fraction α of consumers from both the protected and general groups were inherently privacy-sensitive and always chose to hide their identity prior to the first stage. The remaining consumers, being less concerned about privacy, based their subsequent decision to

disclose or hide on the utility they derived from consuming content that matched their preferences.

Now, we relax this assumption, allowing for the possibility that privacy sensitivity varies across groups, leading to different proportions of individuals choosing to hide their identities. While we initially treated privacy sensitivity as a fixed parameter for the overall consumer population, it is plausible that certain groups exhibit greater privacy sensitivity than others. For instance, if the protected group represents a minority population that faces social stigma in certain cultures (e.g., age), its members would have a stronger preference for privacy. Additionally, differences in privacy sensitivity can arise from disparities in awareness—variations in understanding the importance of privacy and how organizations use personal data can contribute to group-level differences in privacy concerns. In this extension, we assess the robustness of our main result in Proposition 1 when privacy sensitivity differs between the protected and general groups.

We maintain the same model primitives as in the main model for ease of comparison. However, a key difference is that the fraction of highly privacy-sensitive consumers now varies between groups: in the protected group, this fraction remains α, while in the general group, it is $\theta\alpha$, where $0 < \theta < \frac{1}{\alpha}$. When $\theta = 1$, the model aligns with the main model, ensuring consistency while allowing for group differences in privacy sensitivity.

Following the same solution approach as in the previous extension and as outlined in Section 4, we first analyze the unconstrained scenario and then the scenario with the accuracy parity policy. The solution process follows the same backward induction method, moving through the four stages in the timeline depicted in Figure 2, but with α replaced by $\theta\alpha$ for the general group. The first notable difference in the solution process arises in defining the condition that determines whether the hidden group threshold coincides at the point optimal for the general consumers over the hidden protected consumers, or vice versa, among all the hidden consumers, specifically when setting the content threshold in Stage 3. The thresholds, $\hat{q}_N$, $\hat{q}_G$, $\hat{q}_P$, and $\hat{q}_B$, that determine which hidden consumers receive an optimal threshold now incorporate θ. For instance, when neither group hides, the hidden group threshold coincides at the point optimal for the protected consumers if $q \leq \hat{q}_N$, where $\hat{q}_N = \frac{\beta(1-l)}{2\beta(1-l)+\theta(1-l\gamma)}$. Similarly, when only the general group hides, the hidden group threshold coincides at the point optimal for the protected consumers if $q \leq \hat{q}_G$, where $\hat{q}_G = \frac{\alpha\beta(1-l)}{2\alpha\beta(1-l)+1-l\gamma}$. As in the main model and the previous extension, these two thresholds play a crucial role in demonstrating the robustness of our main result.

For the unconstrained scenario in Stage 2, comparing accuracy levels across all values of q discloses that the optimal decision for both the protected and general groups is to disclose their identity. The solution process and the resulting observation closely align with those in the main model and the previous extension.

In Stage 1, the firm's profit function in the unconstrained scenario remains structurally similar to that in the main model, as specified by Equations 16 and 18, with the general group's privacy sensitivity parameter now adjusted to $\theta\alpha$. Specifically, when neither group chooses to hide, the relevant threshold is $\hat{q}_N$, and the

profit function for $q < \hat{q}_N$ is:

$$\begin{aligned}\pi_{UNlq} = &\left((1-\alpha)\beta\left(\frac{q\gamma l}{1-\gamma l}+1-q\right)+(1-\alpha\theta)\left(\frac{ql}{1-l}+1-q\right)\right)+\\ &\alpha\left(\beta\left(\frac{q\gamma l}{1-\gamma l}+1-q\right)+\theta\left(\frac{q\gamma l}{1-l}+1-q\right)\right)-\tau(l),\end{aligned} \tag{52}$$

and for $q > \hat{q}_N$ is:

$$\begin{aligned}\pi_{UNhq} = &\left((1-\alpha)\beta\left(\frac{q\gamma l}{1-\gamma l}+1-q\right)+(1-\alpha\theta)\left(\frac{ql}{1-l}+1-q\right)\right)+\\ &\alpha\left(\beta\left(\frac{ql}{1-\gamma l}+\frac{(1-q)(1-l)}{1-\gamma l}\right)+\theta\left(\frac{ql}{1-l}+1-q\right)\right)-\tau(l).\end{aligned} \tag{53}$$

Next, we determine the optimal learning level that maximizes the firm's profit. Similar to the main model, we instead analyze the marginal return on learning by computing the first-order derivative of each profit function with respect to l, denoted as $\frac{\partial \pi_{UNlq}}{\partial l}$ and $\frac{\partial \pi_{UNhq}}{\partial l}$. For the purpose of this extension, we focus on the optimal learning level l^* obtained by solving the profit maximization problem using the expression for π_{UNlq}. Specifically, when $q_p < \hat{q}_N(l^*)$, the optimal learning level is determined by its corresponding marginal return on learning, $\frac{\partial \pi_{UNlq}}{\partial l}$.

In the accuracy parity policy scenario, Stage 2 follows a similar pattern. Since accuracy does not depend on the fraction of privacy-sensitive consumers, the disclosed general group consistently achieves a higher accuracy than the disclosed protected group, as in the main model. Consequently, as previously observed, the general group prefers to hide while the protected group chooses to disclose. Thus, the firm's profit function depends on the threshold $\hat{q}_G$. We analyze the profit function above this threshold, which corresponds exactly to Equation 22 in the main model. Note that the parameter θ does not appear in this expression, as all general group consumers, both those who are privacy-sensitive ($\alpha\theta$ fraction) and those who are not ($1-\alpha\theta$ fraction), choose to hide.

To assess the robustness of our results, the relationship between $\hat{q}_N(l^*)$ in the unconstrained scenario and $\hat{q}_G(l^*)$ in the accuracy parity policy scenario is critical. Following the same approach as in the previous extension, we evaluate these thresholds at extreme values of l, given the complexity of expressing l^* explicitly. Since both thresholds decrease with l (as indicated by their negative first-order derivatives, $\frac{\partial \hat{q}_N}{\partial l} = \frac{\beta\theta(\gamma-1)}{(2\beta(1-l)+\theta(1-l\gamma))^2} < 0$ and $\frac{\partial \hat{q}_G}{\partial l} = \frac{\alpha\beta(\gamma-1)}{(2\alpha\beta(1-l)+1-l\gamma)^2} < 0$ by $\gamma < 1$), the maximum value of $\hat{q}_G(l)$ occurs at $l = 0$, while the minimum value of $\hat{q}_N(l)$ occurs at $l = \frac{1}{2}$. We can show that under certain conditions (i.e., $\alpha < \frac{1}{\theta(2-\gamma)}$, as opposed to the main model where $\alpha < \frac{1}{2-\gamma}$), even at these extreme values, $\hat{q}_G(0) < \hat{q}_N(\frac{1}{2})$ holds. This suggests that there exists a range of q values where the optimal learning level l^* is determined by

$\frac{\partial \pi_{PPhq}}{\partial l} = 0$ in the accuracy parity policy scenario and by $\frac{\partial \pi_{UNlq}}{\partial l} = 0$ in the unconstrained scenario. Furthermore, we find that $\frac{\partial \pi_{PPhq}}{\partial l} > \frac{\partial \pi_{UNlq}}{\partial l}$ when q exceeds a certain threshold ($q > \frac{\beta(1-l)^2}{2\beta(1-l)^2+\theta(1-\gamma l)^2}$, as opposed to the main model where $q > \frac{\beta(1-l)^2}{2\beta(1-l)^2+(1-\gamma l)^2}$), which lies between $\hat{q}_G(0)$ and $\hat{q}_N(\frac{1}{2})$ when $\hat{q}_G(0) < \hat{q}_N(\frac{1}{2})$ holds. This reinforces our key result that the accuracy parity policy can incentivize the firm to invest more in learning, thereby improving the accurate classification of consumer preferences. ■

A.13 Proof of Extension 6.3

In the main model, privacy sensitivity, and consequently the fraction of consumers who choose to hide no matter what, was independent of group membership. Now, we relax this assumption, allowing for the possibility that privacy sensitivity varies across groups, leading to different proportions of individuals choosing to hide their identities: in the protected group, this fraction remains α, while in the general group, it is $\theta\alpha$, where $0 < \theta < \frac{1}{\alpha}$. The analysis is in Appendix A.12.

We find that AI learning levels $\frac{\partial \pi_{PPhq}}{\partial l} > \frac{\partial \pi_{UNlq}}{\partial l}$ when q exceeds a certain threshold ($q > \frac{\beta(1-l)^2}{2\beta(1-l)^2+\theta(1-\gamma l)^2}$, as opposed to the main model where $q > \frac{\beta(1-l)^2}{2\beta(1-l)^2+(1-\gamma l)^2}$). This reinforces our key result that the accuracy parity policy can incentivize the firm to invest more in learning, thereby improving the accurate classification of consumer preferences.

Besides the key difference of μ, we maintain the same model primitives as in the main model for ease of comparison. Following the same analysis approach as outlined in Section 4, we first analyze the unconstrained scenario and then the scenario with the accuracy parity policy. The solution process follows the same backward induction, moving through the four stages in the timeline depicted in Figure 2. Now, in addition to the revealed group(s) and the hidden group, we also have a guessed general group and a guessed protected group.

Stage 2, 3, and 4 analysis are similar to the main model.

In the unconstrained condition, the firm's profit function in Stage 1 for $q \leq \hat{q}_N$ is:

$$\begin{aligned}\pi_{UNlq} =& \alpha(1-\mu)\left(\frac{q\gamma l}{1-l} + (1-q)\right) + \beta\left(\frac{q\gamma l}{1-\gamma l} + (1-q)\right) \\ & (1-\alpha(1-\mu))\left(\frac{lq}{1-l} + 1(1-q)\right) - \tau(l),\end{aligned} \tag{54}$$

and for $q > \hat{q}_N$ is:

$$\begin{aligned}\pi_{_{UNhq}} =&(\beta-\alpha\beta(1-\mu))\left(\frac{q(\gamma l)}{1-\gamma l}+(1-q)\right)+\left(\frac{lq}{1-l}+1(1-q)\right)\\&+\alpha\beta(1-\mu)\left(\frac{(1-l)(1-q)}{1-\gamma l}+\frac{lq}{1-\gamma l}\right)-\tau(l).\end{aligned}\tag{55}$$

In the constrained condition, the guessed general group has a higher recommendation accuracy than the protected groups, thus the parity constraint would require the guessed general group to match the revealed protected group's accuracy.

The profit function for $q \leq \hat{q}_{_G}$ is:

$$\begin{aligned}\pi_{_{PPlq}} =&\beta\left(\frac{q(\gamma l)}{1-\gamma l}+(1-q)\right)+\mu\left(\frac{q(\gamma l)}{1-\gamma l}+(1-q)\right)\\&+(1-\mu)\left(\frac{q(\gamma l)}{1-l}+(1-q)\right)-\tau(l),\end{aligned}\tag{56}$$

and for $q > \hat{q}_{_G}$ is:

$$\begin{aligned}\pi_{_{PPhq}} =&\alpha\beta(1-\mu)\left(\frac{(1-l)(1-q)}{1-\gamma l}+\frac{lq}{1-\gamma l}\right)+(\alpha\beta\mu+(1-\alpha)\beta)\left(\frac{q(\gamma l)}{1-\gamma l}+(1-q)\right)\\&(+\mu)\left(\frac{q(\gamma l)}{1-\gamma l}+(1-q)\right)+(1-\mu)\left(\frac{lq}{1-l}+(1-q)\right)-\tau(l).\end{aligned}\tag{57}$$

Next, we determine the optimal learning level that maximizes the firm's profit. Similar to the main model, we analyze the marginal return on learning by computing the first-order derivative of each profit function with respect to l, denoted as $\frac{\partial \pi_{_{UNlq}}}{\partial l}$, $\frac{\partial \pi_{_{UNhq}}}{\partial l}$, $\frac{\partial \pi_{_{PPlq}}}{\partial l}$, and $\frac{\partial \pi_{_{PPlq}}}{\partial l}$.

Similar to the main model, we focus on the comparison of $\frac{\partial \pi_{_{UNlq}}}{\partial l}$ and $\frac{\partial \pi_{_{PPlq}}}{\partial l}$ to assess the robustness of our results. And similar to the main model, their comparison depends on q, specifically, our results hold ($\frac{\partial \pi_{_{UNlq}}}{\partial l} < \frac{\partial \pi_{_{PPlq}}}{\partial l}$) when $\max\{\frac{\alpha\beta(1-l)^2(1-\mu)}{\gamma\mu l^2+\alpha(1-\mu)(2\beta(l-1)^2+(\gamma l-1)^2)-\mu}, \hat{q}_{_G}\} \leq q < \hat{q}_{_N}$ (or q is in the range of $\{\underline{q}', \hat{q}_{_{UNhq}}\}$), with an additional condition where $\mu \leq \bar{\mu} = \frac{\alpha(1-\gamma)l(1-\gamma l)}{l(\alpha(1-\gamma)(1-\gamma l)-\gamma l)+1}$. ■

This new boundary condition indicates that when the proportion of estimated hidden consumers is sufficiently small, our key results remain where the accuracy parity policy can incentivize the firm to increase investment in AI learning, thus improving consumer preference classification. If the firm can estimate the identities of all or a sufficiently large proportion of hidden consumers, endogenous consumer identity disclosure, which drives our key mechanism, would become irrelevant.

In addition, given that $\frac{\partial \bar{\mu}}{\partial l} > 0$, as AI learning increases, the constraint on estimation proportion μ relaxes. Given that $\frac{\partial \bar{\mu}}{\partial \alpha} > 0$, as the proportion of consumers who never reveal identity (i.e., consumer privacy concern) increases, the constraint on μ also relaxes.