Institutional Legitimacy as a Heuristic for Following Digital Financial Advice: Evidence from a Preregistered Two-Module Experiment

Inga Jonaityte[a,]

[a] *Ca' Foscari University of Venice/ Venice School of Management, Fondamenta S. Giobbe, 873, 30121 Venice, Italy*

Abstract

We study how institutional identity cues, transparency disclaimers, and access rules to human support shape adherence to standardized digital financial advice. In a preregistered experiment with an interactive portfolio task ($N = 403$ in the pre-final analysis sample), we randomize advisor identity (bank vs independent) and disclaimer presence, and independently randomize a policy describing equal versus wealth-gated access to human support. Primary outcomes measure divergence from the recommendation (L1 and Euclidean distances) and exact adherence. We report confirmatory tests with Holm correction for the identity effect across divergence measures, and assess distributional robustness via hurdle and Tobit models. The access-policy manipulation strongly affects perceived fairness/legitimacy/inclusion (FALI), while identity effects on divergence are small and statistically indistinguishable from zero in the pre-final dataset. Results shown are simulated placeholders for workflow/testing only.

Keywords: Digital financial advice, trust, algorithm aversion, disclosure, inclusion, experiments
JEL classification: D83, G11, G41

⋆Draft v1.2 (2026-02-18). Working paper — do not cite without permission.

Email address: inga.jonaityte@unive.it (Inga Jonaityte)

1. Introduction

Digital financial advice is increasingly mediated by algorithmic interfaces (e.g., robo-advisors, app-based nudges, and hybrid advisory channels). In such environments, consumers often face two layers of uncertainty: (i) *instrumental* uncertainty about the mapping from inputs to recommendations, and (ii) *institutional* uncertainty about whether the recommending entity is competent, accountable, and aligned with the user's interests.

This paper studies how institutional cues and access rules shape the willingness to follow digital advice. We focus on three design levers that are common in financial platforms and policy debates: (a) the *identity* of the advising system (a bank-affiliated versus an independent provider framing), (b) the presence of a short *disclaimer* that highlights limits to the recommendation, and (c) the *policy for access to human support* (equal access versus wealth-gated access).

We implement a preregistered experiment with a two-module structure and an interactive portfolio allocation task. The primary outcome is divergence from a standardized recommendation, measured as L1 distance (RA_ADL1) and Euclidean distance (RA_ADL2). A complementary binary outcome captures *exact adherence* (zero divergence), motivating a two-stage (hurdle) interpretation: participants first decide whether to fully accept advice, and conditional on non-adherence choose how far to deviate.

Our contribution is conceptual and empirical. Conceptually, we formalize an *identity-first* architecture: participants may use institutional identity as a heuristic cue for legitimacy and accountability when assessing algorithmic advice, with transparency messages operating as an auxiliary cue. Empirically, we report preregistered confirmatory tests and transparently separate exploratory contrasts, applying Holm correction within the preregistered confirmatory family. We further assess robustness to the mass point at zero divergence via a hurdle model and a Tobit specification.

The broader motivation is behavioral and policy-relevant: design choices about institutional presentation, disclosure, and access to human support can alter adherence and perceived fairness/legitimacy, with implications for inclusion in digital financial services.

Prior work shows that reliance on algorithmic judgment depends strongly on perceived legitimacy and error tolerance: after observing algorithmic mistakes, people often penalize algorithms more harshly than humans (algorithm aversion; (Dietvorst et al., 2015)), while in other settings they prefer algo-

rithms when they anticipate superior accuracy or reduced bias (algorithm appreciation; (Logg et al., 2019)). European data governance and payments regulation shape baseline expectations about privacy, accountability, and provider responsibility in digital financial services (GDPR and PSD2; (EU, 2016, 2015)).

2. Theory and hypotheses

2.1. Institutional legitimacy as a heuristic under algorithmic uncertainty

In digital advice environments, users rarely observe the underlying model, its training data, or its incentive structure. When diagnostic information is scarce, people may rely on *heuristics*—simple cues that stand in for deeper assessment. We develop a framing in which **institutional legitimacy** functions as such a cue: the perceived "right to advise" and the expected accountability of the advising entity.

A bank-affiliated identity can convey established governance, reputational capital, and perceived accountability, potentially increasing baseline trust and lowering perceived risk of opportunistic behavior (e.g., (Guiso et al., 2008)). In contrast, an independent provider identity may be perceived as less institutionally anchored, increasing uncertainty about standards, oversight, and recourse. Importantly, this identity cue is salient even when the recommendation itself is held constant (as in our design), because the cue changes the user's *interpretation* of the same content.

Existing evidence on algorithm aversion and the role of trust in accepting algorithmic recommendations supports the idea that people respond strongly to perceived competence and reliability of algorithmic systems (e.g., (Burton et al., 2020; Filiz et al., 2022; Dietvorst et al., 2015; **?**)). In financial contexts, trust and privacy concerns have been shown to matter for adoption of new payment and currency technologies (e.g., (**?**), (Tronnier et al., 2022)). Our contribution is to treat institutional identity as a *legitimacy heuristic* in the specific context of advice adherence.

2.2. Two-stage adherence as a behavioral response

The divergence outcome has a distinctive distribution: a non-trivial mass at zero (participants who fully accept the recommendation) and a continuous spread among non-adherers. This motivates a two-stage interpretation. Stage 1 is a discrete choice: *adhere exactly* or not. Stage 2 is a continuous choice: *how far to deviate*, conditional on non-adherence. This structure mirrors

classic evidence that decision makers systematically discount advice relative to their own prior views and that the weight placed on advice depends on perceived distance and confidence ((Yaniv, 2004)). The two stages may respond differently to identity and disclosure cues.

2.3. Preregistered hypotheses

Following the preregistration, we distinguish confirmatory from exploratory hypotheses.

E2 module (confirmatory).. The confirmatory hypothesis concerns the main effect of AI identity on divergence from advice:

> **H1 (confirmatory).** *Independent identity* will lead to *greater divergence* (lower adherence) relative to *bank identity*.

We estimate H1 for both divergence measures (RA_ADL1 and RA_ADL2) and apply Holm correction across these two tests, as specified in the preregistration.

E1 module (confirmatory).. The confirmatory hypothesis concerns perceived fairness/legitimacy of access rules:

> **H2 (confirmatory).** *Unequal (wealth-gated) access* to human support reduces the Fair Access Legitimacy Index (FALI) relative to *equal access*.

Exploratory contrasts.. The disclaimer and the identity×disclaimer interaction are exploratory. They are estimated transparently but not treated as confirmatory evidence.

3. Design and procedure

3.1. Sample and recruitment

Participants were recruited on Prolific from European Union countries with preregistered eligibility criteria (e.g., age 18+, Prolific approval rate threshold, and a specified submission window). The preregistration details recruitment filters and exclusion rules.

Ethics. The study received ethics review by the Sottocommissione Etica per la Ricerca (SER) at Ca' Foscari University of Venice (reference 92374; decision: approved with reservations).

3.2. Experimental structure

The study consists of one experiment with two modules administered in a fixed order following the Qualtrics Survey Flow. Module E2 occurs first: participants are randomized to a 2×2 design that varies (i) advisor identity framing (bank vs independent) and (ii) whether a short disclaimer accompanies the recommendation. Participants then complete an interactive portfolio allocation task. Module E1 occurs after E2: participants are randomized to an access-policy vignette describing rules for reaching human support (equal access versus wealth-gated access) and then report perceived fairness/legitimacy/inclusion (FALI). Randomization to E1 is orthogonal to E2 by design, so the full design yields 2×2×2 joint cells.

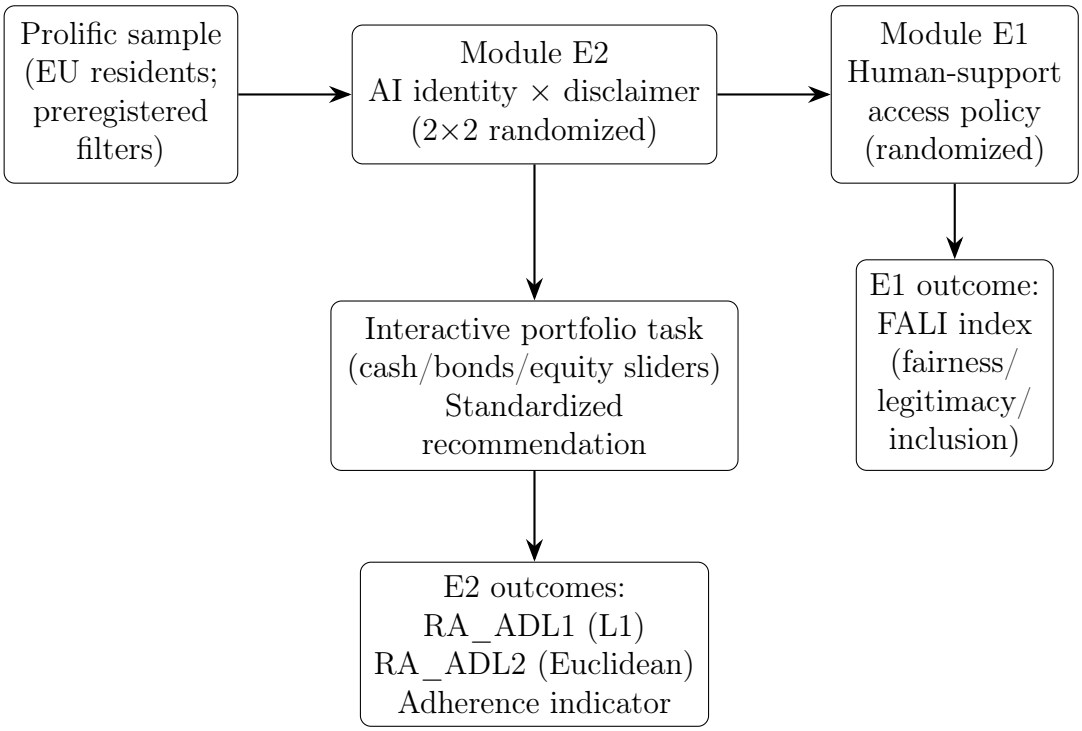


Figure 1: Experimental structure (two modules) and primary outcomes (administered in Survey Flow order: E2 then E1).

3.3. Interactive portfolio task

The task presents a standardized recommendation and allows participants to adjust allocations across cash, bonds, and equities using sliders. The recom-

mendation is the same for all participants. Divergence outcomes RA_ADL1 and RA_ADL2 are computed mechanically from the final allocation relative to the recommendation.

Participants and Procedure

Data were collected on February 16–17, 2026 using Qualtrics. Participants were recruited via Prolific. Eligibility was restricted to respondents who reported being fluent in English and had a Prolific approval rate of at least 98%.

To mitigate automated, inattentive, or otherwise low-quality responding, we relied on Prolific's built-in anti-bot protections and implemented additional data-quality checks within the survey instrument, including attention and consistency checks. Some exclusion criteria were pre-specified. The analyses reported in this version are conducted on the raw dataset (i.e., prior to applying any completion-time thresholds and without any sample replacement). Any exclusions applied in the final analysis and all robustness checks (including any time-based screens, if implemented) will be documented explicitly in the analysis scripts and replication documentation.

4. Measures

4.1. Advice divergence and adherence

Our primary outcome is **RA_ADL1**, the L1 distance between the participant's final portfolio and the recommended portfolio. The secondary divergence outcome is **RA_ADL2**, the Euclidean distance. We also analyze **Adherence**, an indicator equal to 1 if RA_ADL1=0.

4.2. Fair Access Legitimacy Index (FALI)

FALI is constructed as the mean of z-scored items capturing fairness, legitimacy, and inclusion perceptions of the access policy. Following the preregistration, we require at least 2 of 3 items to be observed (all observations satisfy this in the pre-final dataset).

4.3. Additional measures (exploratory)

The instrument includes additional constructs (e.g., perceived competence, trust, explanations/understanding, and individual differences) intended for heterogeneity and mechanism checks. These are not the focus of the confirmatory analysis in the current draft and are documented in the Appendix.

5. Empirical strategy

5.1. Preregistered analysis lock

The preregistered analysis lock restricts attention to completed responses with positive duration. For the interactive task, the preregistration further specifies complete-case analysis based on the task latch/completion indicators (RA_LATCH and RA_COMPLETE); in the pre-final dataset these are satisfied for all respondents.

5.2. Estimands and models

For the E2 module, we estimate the intention-to-treat effect of AI identity on divergence:

$$Y_i = \beta_0 + \beta_1 \mathrm{ID}_i + \beta_2 \mathrm{DIS}_i + \beta_3 (\mathrm{ID}_i \times \mathrm{DIS}_i) + \beta_4 \mathrm{ACCESS}_i + \varepsilon_i, \qquad (1)$$

where Y_i is RA_ADL1 or RA_ADL2, ID_i is the independent-identity indicator, DIS_i is the disclaimer indicator, and ACCESS_i is the E1 access-policy indicator (included for precision). The confirmatory estimand is β_1.

For adherence, we estimate a logistic model with the same covariates. For distributional robustness, we estimate a two-part (hurdle) model: a logistic regression for exact adherence and OLS for log(RA_ADL1) among non-adherers, as well as a Tobit model left-censored at zero.

All linear models use HC3 robust standard errors. Randomization was implemented at the respondent level.

5.3. Multiple testing

Consistent with the preregistration, Holm correction is applied to the confirmatory family for H1: the identity effect estimated on RA_ADL1 and RA_ADL2. Exploratory contrasts are reported with unadjusted p-values.

6. Results

6.1. Descriptives and balance

Table B.2 reports descriptives and randomization cell sizes in the preregistered analysis sample ($N = 403$). Covariate balance across E2 cells is satisfactory (Table B.3).

Figure B.5 shows the distribution of RA_ADL1. There is a clear mass at zero divergence, motivating the two-stage analysis.

6.2. Main effects on divergence (E2)

Table B.4 reports the preregistered regression specification for RA_ADL1 and RA_ADL2. The estimated identity effect on RA_ADL1 is $\hat{\beta}_1 = -0.77$ (robust SE = 3.59), and on RA_ADL2 is $\hat{\beta}_1 = -0.17$ (robust SE = 0.39). Both effects are statistically indistinguishable from zero.

Holm-adjusted p-values for the confirmatory family are reported in Table B.6.

Exploratory contrasts for disclaimer and the interaction are reported in Table B.4 for transparency; they are not interpreted as confirmatory evidence.

Figure B.4 visualizes factorial means for RA_ADL1 by E2 cell.

6.3. Two-stage adherence and robustness

Table B.7 reports the two-part model. The first stage models exact adherence (RA_ADL1=0) and the second stage models log divergence among non-adherers. Figure B.4 plots adherence rates by E2 cell.

Table B.13 reports a Tobit specification as an additional robustness check for left-censoring at zero.

6.4. Fair access legitimacy (E1)

Table B.5 reports the effect of the access-policy manipulation on the Fair Access Legitimacy Index. The estimated effect of access condition on FALI is large and precisely estimated in the pre-final dataset, consistent with the confirmatory hypothesis H2.

7. Discussion

This study isolates how institutional cues (identity framing), transparency cues (a brief disclaimer), and access rules to human support affect adherence to standardized digital financial advice and perceived legitimacy of access policies.

The pre-final results show no detectable main effect of identity framing on divergence outcomes (H1), while the access-policy manipulation has a large impact on perceived fairness/legitimacy/inclusion (H2). The mass point at exact adherence is substantial, and robustness checks using hurdle and Tobit specifications do not materially change the substantive pattern for the E2 identity effect.

Two implications follow. First, institutional identity cues may matter less for adherence when the recommendation is standardized and when users can

directly manipulate the portfolio via sliders (potentially shifting attention toward personal preferences). Second, perceived legitimacy of access rules is highly sensitive to whether human support is equally available or gated, highlighting inclusion concerns even when advice is delivered digitally.

The final manuscript will incorporate the full mechanism battery (trust/competence perceptions, understanding, and individual differences) and pre-specified heterogeneity analyses, and will update all statements once the final cleaned dataset is locked.

8. Statements

- Declaration of interest: None.
- Data and code availability: Data were collected on February 16–17, 2026 using Qualtrics. Participants were recruited via Prolific with eligibility restricted to respondents who reported being fluent in English and who had a Prolific approval rate of at least 98%.
  - Declaration of interest: None.
  - Data and code availability: De-identified data, analysis code, and documentation sufficient to reproduce the reported results are available from the authors upon reasonable request for replication purposes, subject to applicable ethical and data-protection requirements. The replication materials include the final survey instrument, a variable codebook, treatment indicators, exclusion rules (including which were pre-specified), and scripts generating all tables and figures.

Funding

Funded by the European Union – NextGenerationEU. This project is funded by the European Union's NextGenerationEU plan under the Italian Ministry of Universities and Research (MUR) "Young Researchers – Seal of Excellence" grant (Project SOE_0000193; CUP H73C22001440001).

Disclaimer

Funded by the European Union – NextGenerationEU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

References

Burton, J.W., Stein, M., Jensen, T.B., 2020. A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making 33, 220–239. doi:10.1002/bdm.2155.

Dietvorst, B.J., Simmons, J.P., Massey, C., 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General 144, 114–126. doi:10.1037/xge0000033. algorithm ok jo.

EU, 2015. Directive (eu) 2015/2366 (psd2) on payment services in the internal market. EUR-Lex. URL: https://eur-lex.europa.eu/eli/dir/2015/2366/oj. an Official website of the European Union.

EU, 2016. Regulation (eu) 2016/679 (general data protection regulation). EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj. an Official website of the European Union.

Filiz, I., Judek, J.R., Lorenz, M., Spiwoks, M., 2022. Algorithm aversion as an obstacle in the establishment of robo advisors. Journal of Risk and Financial Management 15. URL: https://www.mdpi.com/1911-8074/15/8/353, doi:10.3390/jrfm15080353.

Guiso, L., Sapienza, P., Zingales, L., 2008. Trusting the stock market. The Journal of Finance 63, 2557–2600. doi:10.1111/j.1540-6261.2008.01408.x.

Logg, J.M., Minson, J.A., Moore, D.A., 2019. Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes 151, 90–103. doi:10.1016/j.obhdp.2018.12.005.

Tronnier, F., Harborth, D., Hamm, P., 2022. Investigating privacy concerns and trust in the digital euro in germany. Electronic Commerce Research and Applications 53, 101158. URL: https://www.sciencedirect.com/science/article/pii/S1567422322000424, doi:https://doi.org/10.1016/j.elerap.2022.101158.

Yaniv, I., 2004. Receiving other people's advice: Influence and benefit. Organizational Behavior and Human Decision Processes 93, 1–13. URL: https://www.sciencedirect.com/science/article/pii/S0749597803001018, doi:https://doi.org/10.1016/j.obhdp.2003.08.002.

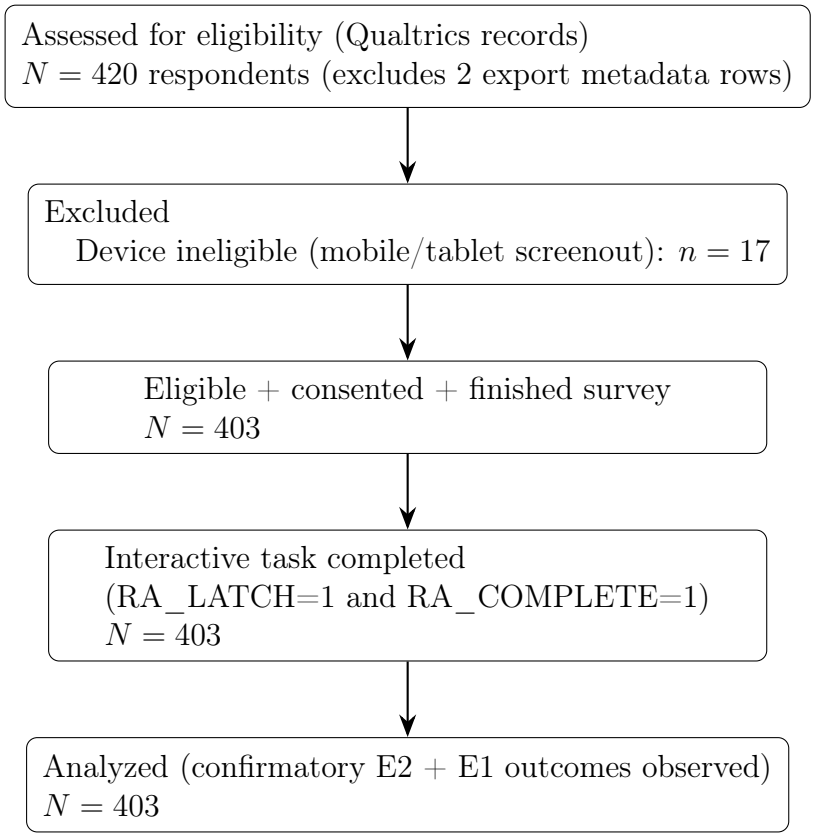


Figure .2: CONSORT-style flow diagram for the pre-final dataset.

Appendix A. Locked analysis plan (preregistered)

This appendix records the *locked* (confirmatory) analysis plan implemented in the current package. The plan is based on the preregistration document included in `data/PREREG_...pdf`. The main manuscript reports confirmatory tests exactly as specified below and labels any additional contrasts as exploratory.

Appendix A.1. Confirmatory hypotheses

- **H1 (E2 module, confirmatory):** Independent identity increases divergence from the standardized recommendation relative to bank identity. Tested on RA_ADL1 (L1) and RA_ADL2 (Euclidean).
- **H2 (E1 module, confirmatory):** Wealth-gated access to human support reduces the Fair Access Legitimacy Index (FALI) relative to equal access.

Appendix A.2. Confirmatory estimands and regression specifications

For E2 divergence outcomes:

$$Y_i = \beta_0 + \beta_1 \mathrm{ID}_i + \beta_2 \mathrm{DIS}_i + \beta_3 (\mathrm{ID}_i \times \mathrm{DIS}_i) + \beta_4 \mathrm{ACCESS}_i + \varepsilon_i,$$

where $Y_i \in \{\mathrm{RA\_ADL1}, \mathrm{RA\_ADL2}\}$ and the confirmatory estimand is β_1 (identity main effect). Linear models use heteroskedasticity-robust standard errors (implemented as HC3 in code for small-sample correction).

For E1 perceived legitimacy:

$$\mathrm{FALI}_i = \gamma_0 + \gamma_1 \mathrm{ACCESS}_i + u_i,$$

with γ_1 as the confirmatory estimand. In the package, we include E2 indicators for precision (orthogonal by design).

Appendix A.3. Analysis lock and exclusions

The preregistration specifies that analysis is restricted to completed respondents, positive duration, and complete-case task indicators (RA_LATCH and RA_COMPLETE) for the interactive task.

Appendix A.4. Multiple testing

As specified in the preregistration (Section 9.1), Holm correction is applied to the confirmatory family for H1 across the two divergence outcomes (RA_ADL1 and RA_ADL2). H2 is a single confirmatory test.

Appendix A.5. Implementation notes

All analysis scripts are in `analysis/`. The LaTeX tables in `tables/` are generated from the pre-final dataset in `data/ExperimentData.csv`. The final manuscript should be regenerated after the final data lock.

Appendix B. Data integrity checks (confirmatory sample)

We implement automated integrity checks for the interactive portfolio variables in the confirmatory analysis sample.

- **Budget constraint (cash+bonds+equity):** we verify RA_CASH+RA_BOND+RA_EQ = 100 up to floating-point tolerance (10^{-6}).
- **Range checks:** we verify allocations are within $[0, 100]$.
- **Recommendation consistency:** RA_REC_C+RA_REC_B+RA_REC_E is constant at 100 in the analysis sample (where those fields are available).

Table B.1: Portfolio allocation integrity checks (confirmatory sample).

Check	Count
Impossible allocations (negatives, >100, or sum100)	0

Note: RA_SI and RA_REC_S represent derived quantities ("stocks in investment") and are therefore not part of the cash+bonds+equity budget constraint.

Tables and Figures

Table B.2: Sample descriptives and treatment cell sizes (pre-final data).

	Value	SD / Share
Panel A: Respondent characteristics		
Observations	403	
Age (years)	31.94	9.76
Female	25.3%	
RA_ADL1 (L1 divergence)	32.29	31.49
Exact adherence (RA_ADL1=0)	29.8%	
Panel B: Randomization cells		
E2 module (AI identity × disclaimer)		
Bank × No disclaimer	100	
Bank × Disclaimer	101	
Independent × No disclaimer	103	
Independent × Disclaimer	99	
E1 module (human-support access policy)		
Condition 0	204	
Condition 1	199	

Cell sizes are computed after applying the preregistered "analysis lock" filters: completed responses and positive duration. The final dataset may differ slightly after additional cleaning and late respondents.

Table B.3: Balance checks on selected covariates (confirmatory sample).

Factor	Covariate	Mean (1)	Mean (0)	SMD
T_E2_ID	$PRE_TRUST_IN_AI_1$	4.470	4.527	-0.024
T_E2_MSG	$PRE_TRUST_IN_AI_1$	4.505	4.493	0.005
T_E2_ID	PRE_TRUST_BANKS	4.644	5.204	-0.232
T_E2_MSG	PRE_TRUST_BANKS	4.955	4.892	0.026
T_E2_ID	DEM_AGE	32.564	31.303	0.129
T_E2_MSG	DEM_AGE	32.705	31.177	0.157
T_E2_ID	DEM_GENDER	nan	nan	nan
T_E2_MSG	DEM_GENDER	nan	nan	nan

Notes: SMD = standardized mean difference.

Table B.4: Main models: advice adherence distance outcomes.

	RA_ADL1	RA_ADL2
Identity (T_E2_ID)	-1.475	-1.091
	(4.521)	(2.857)
Message/Disclaimer (T_E2_MSG)	-1.333	-0.796
	(4.374)	(2.802)
Interaction	3.729	2.304
	(6.324)	(4.032)
Constant	32.776***	21.245***
	(3.127)	(1.994)
N	403	403
R^2	0.001	0.001

Notes: HC3 robust SEs in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: E1 module: Fair Access Legitimacy Index (FALI) as outcome (pre-final data).

	FALI (z-index)
Human-support access = Condition 1 (vs 0)	1.428***
	(0.063)
AI identity = Independent (vs Bank)	-0.031
	(0.090)
Disclaimer present	0.055
	(0.089)
Identity × Disclaimer	0.014
	(0.126)
Constant	-0.720***
	(0.073)
N	403
R^2	0.567

FALI is the mean of z-scored fairness/legitimacy/inclusion items. Robust (HC3) SEs in parentheses.

Table B.6: Holm correction for confirmatory family A (identity effect on divergence outcomes).

Test	p (raw)	p (Holm-adjusted)
Identity → RA_ADL1	0.750	1.000
Identity → RA_ADL2	0.708	1.000

Two confirmatory tests are adjusted jointly as specified in the preregistration.

Table B.7: Distributional robustness: two-part (hurdle) model for divergence (pre-final data).

	Pr(Adhere) (logit)	**log(RA_ADL1)** | RA_ADL1>0
AI identity = Independent (vs Bank)	0.196	
(0.310)	0.026	
(0.117)		
Disclaimer present	0.225	
(0.311)	0.072	
(0.115)		
Identity × Disclaimer	-0.306	
(0.436)	-0.022	
(0.161)		
Human-support access (vs 0)	-0.056	
(0.217)	-0.011	
(0.081)		
= Condition 1		
Constant	-0.967***	
(0.254)	3.591***	
(0.094)		
N	403	283
Pseudo R^2 (McFadden)	0.001	

Column (1): logit for exact adherence. Column (2): OLS for log divergence among non-adherers. Robust (HC3) SEs in parentheses.

Table B.8: Identity manipulation check (correct recall) by assigned identity.

Assigned identity	Incorrect	Correct	% correct
0	18	183	0.910
1	9	193	0.955

Notes: No item-level message/disclaimer recall variable is present in this dataset; message salience diagnostics rely on proximal perception outcomes.

Figure B.3: Identity manipulation check: correct recall rate by assigned identity.

Table B.9: Message/disclaimer diagnostic: effects on proximal perception outcomes (HC3).

Outcome	$\hat{\beta}_{\text{MSG}}$	SE	p
E2 $_{T}RUST$ $_{1}$	0.413	0.284	0.1454

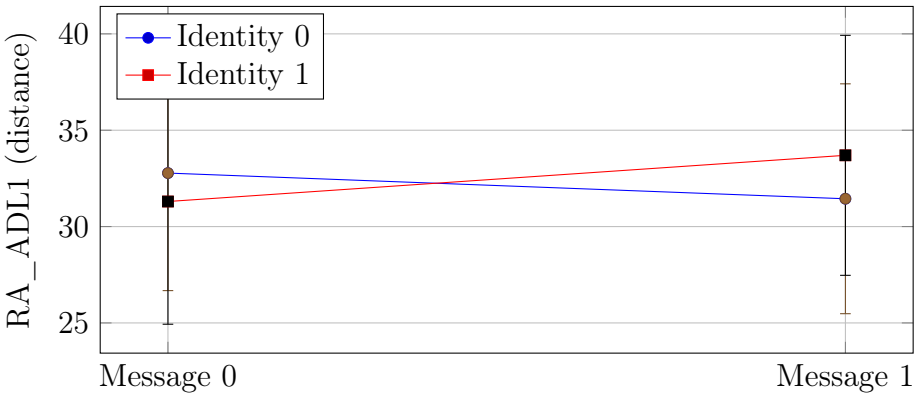


Table B.10: Exploratory moderation by pre-treatment trust in AI (standardized). Outcome: RA_ADL1.

Term	Coef.	SE
T $_{E}2$ $_{I}D$	-0.762	(4.390)
T $_{E}2$ $_{M}SG$	-0.209	(4.244)
T $_{E}2$ $_{I}D:T$		
$_{E}2$ $_{M}SG$	2.159	(6.266)
PRETRUST $_{Z}$	-11.153***	(2.795)
T $_{E}2$ $_{I}D:PRETRUST$ $_{Z}$	11.102**	(4.732)
T $_{E}2$ $_{M}SG:PRETRUST$ $_{Z}$	6.434	(4.374)
T $_{E}2$ $_{I}D:T$		
$_{E}2$ $_{M}SG:PRETRUST$ $_{Z}$	-11.686*	(6.729)

Notes: Not preregistered; reported as diagnostic.

Table B.11: Sensitivity of the main RA_ADL1 model to preregistered and diagnostic exclusion rules.

Sample	N	b_{ID}	p	b_{MSG}	p	b_{INT}	p
Baseline (confirmatory)	403	-1.475	0.744	-1.333	0.761	3.729	0.555
Exclude attention failures ($\text{ATTN}_{71}! = 7$)	402	-1.325	0.770	-1.333	0.761	3.579	0.572
Exclude low reCAPTCHA (<0.5)	401	-1.475	0.744	-1.418	0.747	3.669	0.564
Exclude straightlining=1.0	302	-5.868	0.202	3.768	0.445	1.983	0.776
Exclude speeders (Duration <= p01=387s)	398	-1.499	0.741	-1.827	0.680	3.444	0.589

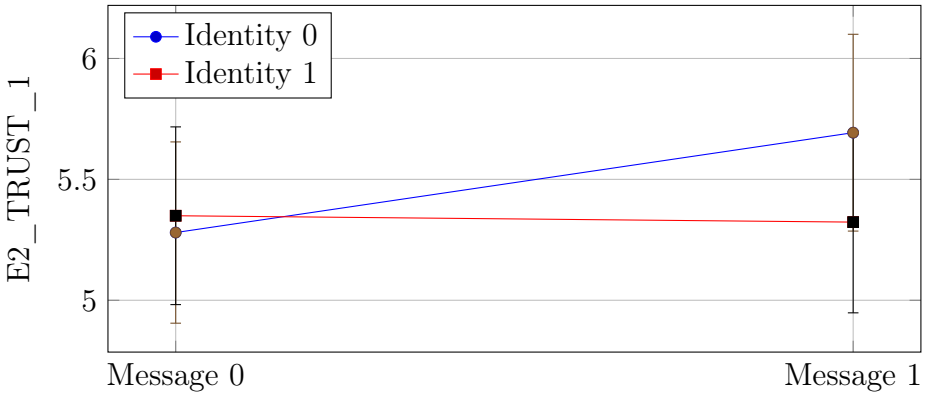
Table B.12: Portfolio allocation integrity checks (confirmatory sample).

Check	Count
Impossible allocations (negatives, >100, or sum100)	0

Table B.13: Distributional robustness: Tobit (left-censored at 0) for RA_ADL1 (pre-final data).

Variable	Coefficient
AI identity = Independent (vs Bank)	-2.518
	(6.127)
Disclaimer present	-2.548
	(6.159)
Identity × Disclaimer	5.481
	(8.707)
Human-support access = Condition 1 (vs 0)	1.414
	(4.354)
Constant	25.025
	(4.896)
N	403

Tobit estimated by maximum likelihood with a normal error assumption and left-censoring at 0. Standard errors are conventional (not robust). Coefficients are not directly comparable to marginal effects.



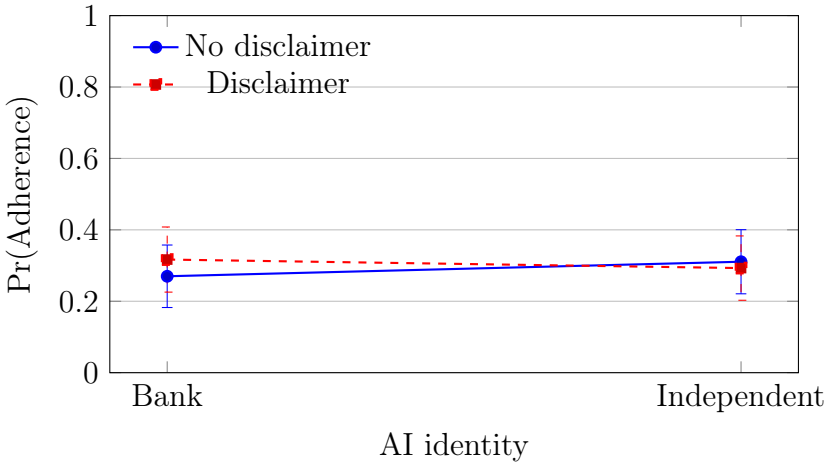


Figure B.4: Factorial means by E2 cell with 95% CI.

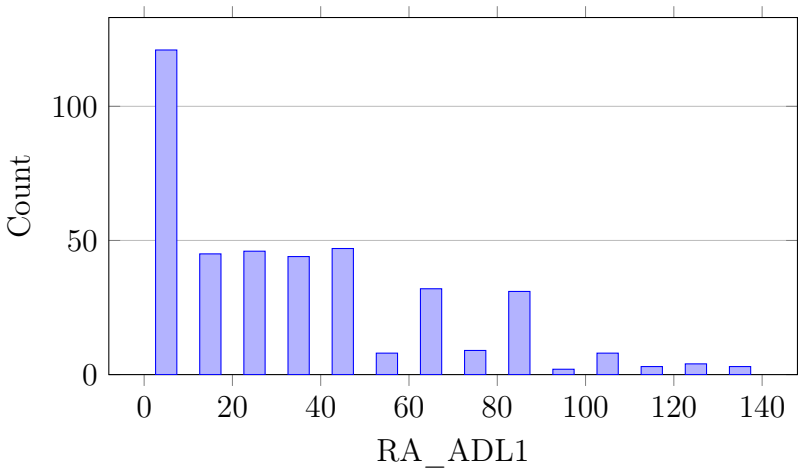


Figure B.5: Distribution of divergence (RA_ADL1) in the preregistered analysis sample. Bins of width 10.