HUMANITY LABS AI MAGAZINES

The Trust Times

ISSUE NO. 4



Owning It

IN THIS ISSUE

1. The Warmth Premium
2. The Disclosure Penalty
3. How a Machine Says Sorry

Current, peer-reviewed studies on client trust in the age of AI.

Issue #3 was about saying the hard thing. This issue takes on the question sitting on every advisory desk right now: what happens to a client's trust when AI shows up in your practice?

Full disclosure: this magazine is published by a company that builds AI for wealth management firms. So we followed the best current evidence wherever it pointed, and two of these three studies show ways AI can cost you trust. All three land on the same truth: clients do not price your tools, they price you.

The first, from *Scientific Reports*, had 400 Americans play a money game with human and AI partners. What moved trust was not competence but warmth: whether the partner seemed to be on their side. A cold AI was punished hardest of all; a warm AI was trusted like a warm human.

The second is a caution for the marketing department. In the *Journal of Medical Internet Research*, telling 1,762 Americans that a doctor uses AI heavily cut trust in that doctor roughly in half. People heard 'uses AI' as 'the machine is doing the thinking.' If your firm's homepage leads with the technology, this study is for you.

The third, from *Behavioral Sciences*, will feel familiar to Issue #2 readers. Experienced investors trusted a robo-advisor slightly more than a human expert, until it made one bad call. When the error came with a clear account of cause, limits, and path forward, trust finished above where it started. The reform signal wins again, even when the one reforming is software.

Together: bring the machine, but lead with whose side you are on, keep the tool inside the frame of your judgment, and when something slips, own it with the fix. The technology is new; the trust math underneath it is not.

Three more great studies on trust next week.

Warmly,



Marshall Kirkpatrick
Editor, *The Trust Times*

SOURCES · ISSUE 4

1. The Warmth Premium. Based on Samson & Zaleskiewicz (2026), *Scientific Reports*. N=400.

2. The Disclosure Penalty. Based on Chen & Cui (2025), *Journal of Medical Internet Research*. N=1,762.

3. How a Machine Says Sorry. Based on Han & Ko (2025), *Behavioral Sciences*. N=483.

ABOUT THIS ISSUE

Three peer-reviewed studies on how trust is built, retold in plain language. The "at your desk" sections are our editorial application, not claims made by the researchers.

What AI does to a client's trust: the warmth that opens the door, the disclosure that can close it, and the explained error that wins it back.

In Issue #3 we covered three studies on telling clients the truth.

The Warmth Premium

In a new money-game experiment, what made people trust a partner, human or machine, was not how capable it looked. It was whether it seemed to want good things for them.

Samson, K., & Zaleskiewicz, T. (2026). Low perceived warmth of AI agents reduces trust towards them. *Scientific Reports*, 16, 11775. Open access: doi:10.1038/s41598-026-42252-1

THE CONFERENCE ROOM EIGHTEEN FLOORS UP sells competence before you say a word. The diplomas are real, the view is better, and the pitch book on the table is eighty pages of proof that your firm knows what it is doing.

The couple across the table will forget most of it. According to a new experiment in *Scientific Reports*, they are running a different calculation while you talk, and it has little to do with your Sharpe ratios. They are deciding whether you are on their side.



Psychologists have long held that we size each other up on two master dimensions: warmth, meaning whether your intentions toward me are good, and competence, meaning whether you can act on them. The new study put real money against the question and added the wrinkle of the decade: sometimes the party asking for trust is not a person at all.

THE EXPERIMENT: A MONEY GAME WITH FOUR KINDS OF PARTNERS

Katarzyna Samson and Tomasz Zaleskiewicz recruited 400 US adults, matched to the country on age, sex, ethnicity, and political affiliation, and put each one into an incentivized trust game. You hold ten tokens worth real money. Whatever you send to your partner triples in value, and the partner then decides how much, if anything, comes back. The number of tokens you hand over is your trust, measured in the currency people actually guard.

Each participant faced one of two partners: a human, or an AI agent making its own decisions. And each partner came with a short description that moved one social dial at a time. Some partners were painted warm: cooperative, fair, interested in other people's goals. Some were painted cold. Separately, some were described as highly capable and some as error-prone. Checks confirmed the descriptions landed as intended.

Then the researchers watched the tokens move. Every token sent was a real-money bet on the partner's good intentions, and the running count showed who had earned that bet.

WHAT THEY FOUND: GOODWILL BEATS SKILL

Warmth moved the money. Partners described as warm received about 6.3 of 10 tokens on average; cold ones received about 4.2. Competence helped too, but the gap it produced was less than a third that size: skilled partners collected 5.6 tokens, unskilled ones 4.9.

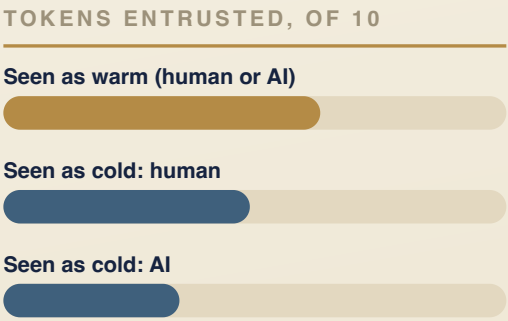
Read those numbers again from a marketing distance. A reputation for skill bought seven tenths of a token. A reputation for goodwill bought more than two full tokens, roughly three times the return on the same introduction.

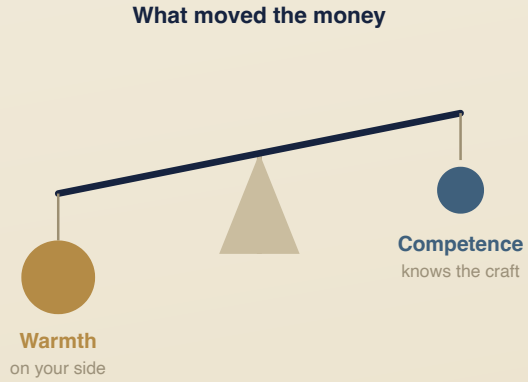
The human-versus-machine results matter most for anyone installing software between themselves and their clients. Averaged across everything, people trusted humans more than AI, 5.6 tokens to 4.9. The entire deficit, though, lived on the cold side of the ledger. A cold human still collected 4.9 tokens, while a cold AI collected 3.5, the lowest trust anywhere in the study.

A warm AI was a different creature entirely. It received the same 6.3 tokens a warm human did. Once the machine seemed to be working in the participant's interest, the machine penalty disappeared.

Participants put real stakes behind the warm machine. Faced with an algorithm described as cooperative and interested in their goals, people committed the same share of their money they would have handed a warm human stranger. And the same pool of participants gave a cold AI less than anyone, so this was not a crowd that trusted machines on principle.

The study found an asymmetry worth memorizing: coldness costs a machine more than it costs a person, and warmth buys a machine exactly what it buys a person.





N = 400

Why would coldness hurt an AI more? The researchers point to what each party is suspected of by default. A cold person is unpleasant, but still recognizably one of us. A cold machine invites the darker read: a system pursuing goals that are not yours, with nobody home to appeal to. Warmth answers the question people ask of any new counterparty: whose interests does this thing serve?

Underneath this sits an old habit. Before anyone asks whether a stranger is capable, they ask what the stranger wants from them. A machine that gives no answer gets the worst assumption, and in this game the worst assumption cost it nearly three tokens.

WHAT THIS COULD MEAN AT YOUR DESK

Open with whose side you are on. Before the deck and the track record, the first minutes of a meeting could do the work this study says matters most: restate the family's goals in their own words, and name what you would do first for them and why. Competence claims land only after the warmth question is settled.

Do not ask credentials to carry a cold room. In this game, a description of fairness and cooperativeness outbought a description of skill by a wide margin. If clients praise your patience before your performance, this study says their instincts have the order right.

Give your machines a warm introduction. Clients now meet your firm through portals, chatbots, and machine-drafted emails, and cold is the factory setting for all of them. Introduce any client-facing tool the way you would a junior colleague: what it does for the client, whose interests it serves, who is accountable for it. The study suggests an AI presented as being on the client's side can be trusted like a person, and one that feels indifferent will be trusted less than anyone in the building.

Audit the coldest touchpoint. The automated margin email, the do-not-reply address, the chatbot that cannot say who to call: each one is a cold machine wearing your firm's name, and cold machines took the deepest discount in the study.

THE BOTTOM LINE

The trust decision is a warmth decision first, whether the counterparty breathes or boots. People in this experiment forgave a lack of skill far more readily than a lack of goodwill, and they extended a warm machine the same trust they extended a warm human. That order held whether the counterparty was a person or a program.

Clients will meet more machines at your firm every year, in the portal, in the paperwork, in the follow-up note. The standard they will hold each one to is the standard you already know how to meet: show whose side you are on before you show what you know.

The Disclosure Penalty

Tell people a professional uses AI and they trust that professional less. A preregistered experiment with 1,762 Americans measured how much less, and the answer should slow down every firm drafting an AI announcement.

Chen, C., & Cui, Z. (2025). Impact of AI-assisted diagnosis on American patients' trust in and intention to seek help from health care professionals: Randomized, web-based survey experiment. *Journal of Medical Internet Research*, 27, e66083. Open access: doi:10.2196/66083

She is choosing a doctor the way everyone does now, from a directory on her laptop after the house has gone quiet. Two profiles from the same practice: same training, same years of experience, same reassuring paragraph about listening to patients.

One profile carries an extra line about using artificial intelligence extensively to evaluate symptoms and decide on treatments. She reads the line twice. In the morning she books the other doctor.



Researchers Chen and Cui measured that pause. Their preregistered experiment in the *Journal of Medical Internet Research* is one of the cleanest looks yet at what disclosing AI use does to trust in the professional who discloses it. They ran it in medicine, the field where expert judgment carries the highest stakes and the deepest reserves of public trust. What they found should concern anyone whose product is judgment.

THE EXPERIMENT: ONE PROFILE, FOUR SENTENCES

In June 2024, the researchers showed 1,762 US adults, quota-matched to the country on age, gender, and political affiliation, a profile of 'Doctor M,' a physician of average experience and expertise treating the reader's moderate symptoms. Everyone saw the same doctor with the same credentials. The only thing that changed, across four randomized groups, was a single line about AI.

A quarter of participants read no mention of AI at all. A quarter read that Doctor M does not use AI, such as ChatGPT, when evaluating symptoms and deciding on treatments. The remaining half read that Doctor M uses it either moderately or extensively in that same work.

Then everyone answered the questions a practice lives on: how much do you trust this doctor as a professional, how much do you trust her as a person, and how willing would you be to put your own symptoms in her hands?

WHAT THEY FOUND: THE WORD ALONE CARRIED A PRICE

Mentioning AI lowered all three measures, with effects the authors report as large. Trust in the doctor as a professional roughly halved between the profile that disavowed AI and the profile that leaned on it heavily, falling from 0.63 to 0.30 on the study's zero-to-one scale.

Moderate use bought little mercy. Most of the penalty arrived the moment AI entered the profile at all, and it deepened as the described reliance grew. There was no dosage low enough to be free.



Trust in the professional, 0 to 1 scale: 0.63 down to 0.30.
Preregistered; N = 1,762 US adults.

The profile that explicitly avoided AI outscored even the profile that said nothing, as if the absence of the technology had become a feature worth advertising. In this experiment, restraint functioned as a credential, and the heaviest reported use paid the steepest price.

And the reaction crossed every line the researchers checked. Young and old, left and right, college degree or none: the discount held at every crossing. The authors looked for a group that shrugged it off and did not find one. This will not retire with your older clients.

Willingness to act fell furthest of all. The same participants were much less inclined to seek Doctor M's help, the measure closest to what a firm would call the pipeline. The penalty reached past feelings into the decision to show up, which for a practice is the line between a hesitant client and no client.

WHAT PATIENTS HEARD

The profiles never said the AI decided anything. Participants filled that in themselves. The researchers read the result as a substitution story: when a professional credits a machine, observers reassign the thinking to the machine and discount the human's own contribution to the work. 'I use AI' arrives as 'less of me shows up in your file,' and the more essential the judgment, the more there is to lose.

That reading fits a pattern researchers call algorithm aversion. People hold machines to a stricter standard than humans in consequential domains, and they discount professionals who seem to hand over the parts of the job that were supposed to be theirs. Familiarity may soften the reaction over time, but nobody should bet a client list on when.

The experiment tested only the bare disclosure. There was no arm in which Doctor M explained that the AI double-checks her reading while every decision stays hers, no framing built to answer the substitution fear head-on. One line stood alone, and alone it cost half the trust.

That untested space is where an advisor gets to operate.

WHAT THIS COULD MEAN AT YOUR DESK

Treat 'AI-powered' as a claim clients hear about you, not the software. The word did the damage in this study before any detail followed. The phrase is meant to signal modernity; these results say it can price like an admission. A homepage that leads with the technology could read, to the exact families you most want, as a professional stepping back from his own judgment.

Frame the judgment, not the tool. The editorial move this research invites: answer the substitution fear in the same breath as the disclosure. Say what the machine widens, then say what only you do. 'Our software lets me check three hundred scenarios instead of thirty. Every recommendation still carries my name.' Judgment first, machinery second; the order is the message.

Never let disclosure happen to you. Nothing here argues for hiding AI use; it argues for authoring the sentence yourself, on your own schedule and in your own frame. A client who discovers quietly automated work later meets the substitution fear and a candor problem at once, and Issue #3 covered how expensive those are.

Count the disclosure surfaces you do not control. Custodian portals, planning tools, and note-taking apps announce their AI freely, inside work that carries your name. An unmanaged banner is a disclosure you did not write. When a client meets one, your framing should already have gotten there first.

THE BOTTOM LINE

The study measured the cost of a sentence: the same professional, minus half the trust, once heavy AI use entered the profile. The sentence was bare, and readers filled the gap themselves: in every corner of the sample they concluded the machine was doing the thinking.

What clients are asking is not whether you have good tools. They are asking whether you are still the one doing the thinking. Every AI sentence your firm publishes, on the site, in the deck, in the disclosure stack, should answer that question in the same breath it raises it.

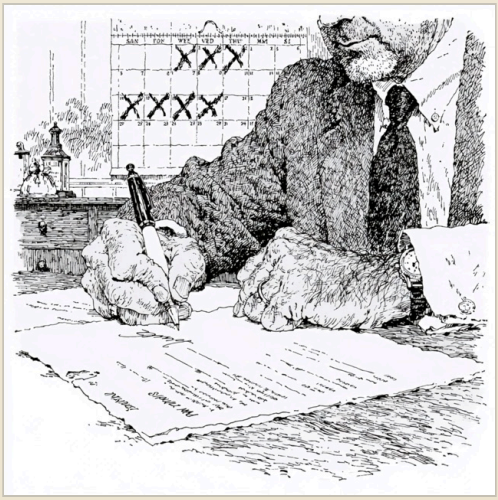
How a Machine Says Sorry

Investors handed a robo-advisor their trust faster than they handed it to a human expert. One bad call took it back. What happened next depended entirely on the quality of the explanation.

Han, J., & Ko, D. (2025). Trust formation, error impact, and repair in human-AI financial advisory: A dynamic behavioral analysis. *Behavioral Sciences*, 15(10), 1370. Open access: doi:10.3390/bs15101370

The recommendation was wrong, and not in the arguable way. The kind of wrong a client sees on one statement, circles in pen, and brings to the next meeting. Every practice eventually mails that statement. The research question is what you send with it.

Two experiments in *Behavioral Sciences* by researchers Han and Ko put that question to 483 experienced Korean investors, and along the way documented something uncomfortable: people extended trust to a machine advisor more readily than to a human one.



THE EXPERIMENTS: THREE ROUNDS AND ONE PLANTED ERROR

In the first study, 189 investors evaluated an investment recommendation from either a human financial expert or an AI robo-advisor. The advice was identical to the word, yet the robo-advisor drew more initial trust, 4.79 to 4.44 on a seven-point scale. The edge was small but real, and the machine had done nothing yet to earn it.

The second study, with 294 investors, tested the relationship. Advice arrived across three rounds, and for some participants, round two contained one deliberately inaccurate recommendation. After the error, half of those participants received an explanation covering what caused the mistake, what the system can and cannot do, and how to proceed. The other half got the error with no account of it.

Trust, satisfaction, and willingness to rely were measured after every round, giving the researchers a motion picture of confidence breaking and mending rather than the usual single snapshot taken after the fact.

WHAT THEY FOUND: THE EXPLANATION OUTWEIGHED THE ERROR

The error cost was immediate and steep. One bad recommendation in round two sent trust sharply down, the largest single movement anywhere in the study, and satisfaction and willingness to rely fell with it.

Then the paths split. Where the error came with an explanation, trust climbed back in round three and did not stop at recovery: it finished above where it had started in round one. Where the error came bare, trust stayed down.

The sharpest swings belonged to the most financially literate participants. The people best equipped to catch the error punished it hardest and then gave back the most trust when the account of it was honest.

An explained failure, in these experiments, was worth more trust than an unblemished record. That echoes what Issue #2 found about corporate apologies: clients rebuild confidence around the visible fix, and here the thing doing the learning was software.

WHAT THIS COULD MEAN AT YOUR DESK

Write the error protocol before the error. The explanation that worked had three parts in a fixed order: what caused it, what the system's limits are, what happens next. That is a template you can draft this week and keep in a drawer. The day something in your stack misfires, the difference between a trust dip and a trust gain could be whether that note goes out in hours or in weeks.

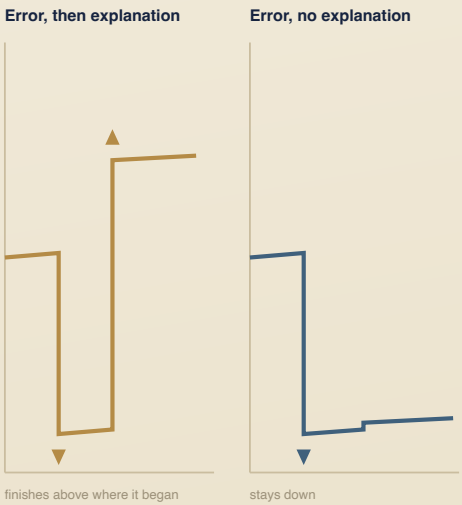
Your most sophisticated clients are your best repair audience. The financially literate dropped hardest and rebounded hardest. A knowledgeable family that catches a slip is not lost; by these results, it is the client most responsive to a rigorous, specific accounting.

When the tool errs, you author the explanation. Clients arrived trusting the robo-advisor, the error tested that trust, and the explanation decided it. If a platform in your stack misfires, its generic error banner should never be the only account your client reads. The cause, the limits, and the path forward should arrive under your signature.

THE TAKEAWAY

In these studies, trust that survived an explained error ended higher than trust that was never tested. Some tool in your practice will eventually be wrong; what you can write today is the note that goes out when it is.

ONE ERROR, TWO ENDINGS



The AI Workforce
for the Human Firm.

Humanity
Labs

LIVE ZOOM WEBINAR · JULY 31, 2026

Ask Me Anything



Andrei Pop
Founder & CEO, Humanity Labs

11:00 AM PT · 2:00 PM ET
60 MINUTES

Register Here

ABOUT HUMANITY LABS

Humanity Labs gives wealth management firms an AI Workforce that works inside existing systems to complete work across the front, middle, and back office, helping firms serve more clients without adding cost at the same rate. Every job strengthens a shared memory that builds the firm's Organizational General Intelligence (OGI).

Humanity Labs works with leading firms overseeing more than $750 billion in client assets and is backed by Google Ventures and Vestigo Ventures.

Learn more at **HumanityLabs.ai**