THE DECISION SCIENCE SERIES

The Decision Sequence

Designing Engagement Strategies Based on How Consumers Decide.

When Buyers Trust the Machine More Than You



DS 16

Trust in AI and Automation

–A Strategic Resource by Joe Worden | More @ www.jworden.com

DECISION ARCHITECTURE RESEARCH LIBRARY

The AI-Era Research · Entry 16

Trust in Automation and Human-AI Decision Making

Madhavan & Wiegmann · Lee & See · Parasuraman & Riley · Goddard et al. · Dietvorst · Logg · NIST

Questions this paper answers

- How do we respond when a prospect brings an inaccurate AI summary of our firm?
- How can we help buyers verify an AI-generated claim without turning the meeting into an argument?
- How much should our own team rely on automated recommendations?
- Why might one person trust AI too much while another dismisses it after an error?

Definition. "Trust in automation" has been around for decades, quietly tracking a simple question: when do people choose to lean on a machine, and when do they keep their distance? The headline result is less moral than practical. The aim isn't to push people into trusting machines more, or to scare them into trusting less. It's to get the match right, trust sized to what the system can truly deliver. That fit is what researchers mean by calibration. AI, as the latest and most capable kind of automation, inherits the whole bundle of issues. It reshapes the decision setting. It doesn't erase the need for trust, it raises the stakes and makes "getting it right" harder than it looks.

The Confident Wrong Answer

The meeting was going fine until the buyer looked up from his notes and said, "But my research shows you don't really work with organizations our size."

His research. He'd asked an AI assistant about the firm on the drive over, and the machine, confident as a game-show host, had served up a description that was two pivots out of date. The room got quiet in a particular way I've come to recognize, because here's the trap: the machine's version reads as neutral. Nobody paid it to say anything. The firm's correction, meanwhile, reads as exactly what a salesperson would say.

You can't out-argue a machine the room believes has no agenda. You're the one selling something. The machine, apparently, is just telling it like it is.

What struck me afterward wasn't the error, errors happen. It was the asymmetry. That buyer would have shrugged off the same mistake from a colleague: "Bob's thinking of the old days." From the machine, the wrong answer arrived wearing a lab coat, and it took twenty minutes of the meeting to take the coat off.

It turns out there's forty years of research on exactly this, on when people trust machines too much, too little, and almost never the right amount. For marketers, it's suddenly required reading.

Trusting a Machine Is Like Trusting a Person, Almost

In everyday life, people hand out trust to machines the way they hand it out to other people: expectations form, reliance follows, experience adjusts the dial. Yet Madhavan and Wiegmann, in their review of human-automation trust, point to an imbalance that keeps showing up. Machines are judged more harshly. Many of us carry a half-formed expectation that a machine should be close to flawless, so when it fails in a visible way, trust collapses faster than it would after the same mistake from a person. A person who gets it wrong is, well, a person. A machine that gets it wrong feels defective. That difference lingers and it shapes what happens next.

The Goal Is Appropriate Trust, Not Maximum Trust

Lee and See's classic piece, "Trust in Automation: Designing for Appropriate Reliance," sits at the center of this literature. Their definition is plain and useful: trust is the attitude that a system will help you reach your goal in a situation you can't fully control. Then they separate that attitude from reliance, the behavior it produces, and the space between the two is where the trouble lives. The point that matters most is also the one that gets lost in popular talk about AI. The target is not maximum trust. The target is appropriate reliance, calibration that tracks the system's actual capability.

Overrate a system and you'll lean on it when you shouldn't. Underrate it and you'll refuse help that would have improved the decision.

Parasuraman and Riley laid out the major failure patterns even earlier, and they didn't hide them behind soft phrasing. "Humans and Automation: Use, Misuse, Disuse, Abuse" names four modes. Use is the success case, reliance that fits. Misuse is over-reliance, letting the machine carry decisions you should still be checking. Disuse is under-reliance, brushing off the machine even when it's correct, often after too many false alarms. Abuse shifts the blame upstream: builders and deployers putting systems into the world without regard for how humans will really interact with them. Most automation trouble lands in one of these buckets, and only one of them is the outcome you want.

Two Ways to Get It Wrong, and a Twist

Over-reliance has a familiar label, automation bias. Goddard, Roudsari, and Wyatt, looking at medical decision systems, describe two error types that come with it. First, people miss what the system doesn't flag. Second, they accept what the system recommends even when the evidence in front of them points the other way. The system's confidence can quietly outrank the user's judgment. That is what "too much trust" costs.

My reading of that drive-over research is that it resembled both patterns: accepting what the machine said, and not noticing what it never mentioned.

Under-reliance is no less real, and the evidence around it refuses to behave neatly, which is part of the value of the research. Dietvorst, Simmons, and Massey documented algorithm aversion: show people an algorithm making a mistake and many will drop it quickly, even if it still beats the human option. We tend to grant people more grace than we grant machines.

Then the story turns. Logg, Minson, and Moore found something close to the reverse in other contexts, calling it algorithm appreciation: people sometimes give algorithmic advice more weight than identical advice attributed to a human. Which way it goes depends on the task and on the decision maker's expertise. The take-away is not "people trust machines too much" or "people don't trust them enough." It's that trust is jumpy, situational, and sensitive to framing, which is exactly why calibration is so difficult.

Why This Belongs in Decision Architecture

This body of work backs a claim Decision Architecture states plainly: AI can change how uncertainty gets examined, but it doesn't remove the need for judgment. A machine's output doesn't validate itself. Someone still has to decide what it's worth believing, and how far to act on it.

Newer frameworks try to build conditions for that judgment instead of leaving it to chance. The NIST AI Risk Management Framework, for instance, organizes "trustworthy AI" around governing, mapping, measuring, and managing risk. It also points to qualities, validity, transparency, accountability, that are meant to make reliance easier to size correctly. In other words, its aim aligns with the field's aim: appropriate reliance.

This helps explain a tension that keeps surfacing in practice. AI can compress parts of research, gathering, summarizing and comparing at speed. That doesn't establish how much confidence a particular buyer should place in the result. It can't decide, for the buyer, how believable the

summary is. Field behavior lines up with lab findings: buyers often start with AI, then check the output against trusted humans, because they sense the need for calibration.

Put in the framework's language, the Trust Gate hasn't disappeared. It's multiplied. Instead of calibrating trust in one thing, the buyer now calibrates two at once: the firm itself, and the machine's story about the firm.

And the calibration challenge is getting harder, not easier, for a basic reason. Older automation looked like automation. Autopilots, warning lights, flagged transactions, they announced themselves as machines, and that kept a small guardrail in place. Generative AI does the opposite. It speaks in fluent, confident, human-sounding prose, and in doing so it strips away the cues people once relied on to remember, "this is a machine." The result is an invitation to the overtrust that automation-bias work warns about, paired with a quieter problem: the automation can be easy to miss.

For decades, the research message has been consistent, calibrate your trust in machines. The newest machine is also the most skilled, so far, at making us forget to.

The Marketing Read

The machine will be in your next sales meeting whether you invite it or not. The only question left is whether it walks in as your witness or as opposing counsel.

Start by accepting the asymmetry, because it's the strategic fact of the decade: the machine's account of your firm reads as neutral, and yours reads as interested. That means the worst place to correct the machine is in the room, where every rebuttal sounds like spin. Argue with the machine before the meeting, through the public record it reads, the audit ritual from the last entry, the coherence work, the third-party sources. In the room, the play is different: don't fight the summary, calibrate it. "That's a fair description of who we were three years ago, here's what changed and here's where you can verify it" treats the machine as a witness to be cross-examined, and hands the buyer the thing the research says they're missing: help sizing their trust.

Then make yourself easy to verify, because the field data shows buyers instinctively run machine output past trusted humans and checkable sources. Every claim that can be confirmed in thirty seconds converts the buyer's calibration step into a point for you. Every claim that can't gets filed with the machine's version, and the machine speaks with more confidence than you do.

Know which trust error you're facing, because the research describes both possibilities. One participant may accept the AI research too readily; another may dismiss useful assistance after seeing it fail. Job title doesn't tell you which response you'll meet. Same meeting, opposite

calibrations, and your material has to work for both: verifiable enough for the skeptic, human enough to add what the machine can't.

And hold your own tools to the standard the research sets, because Dietvorst's finding cuts against you too. One confident wrong answer from your firm's client-facing chatbot, your AI-generated report, your automated summary, can damage the client's willingness to rely on it, even where the system remains useful. The size and direction of that response depend on the task and the person. If you deploy machine judgment under your brand, ship it calibrated: sourced, hedged where hedging is honest, and reviewed by someone with a name.

What backfires: winning the argument with the machine in front of the buyer. Even when you're right, the room remembers that you fought the neutral party. Fix the record, not the referee.

The hard call: your firm is currently being described, confidently, to buyers you haven't met, by a system that can be wrong without making the uncertainty clear. One practical move is clear: become easier to verify than the machine is to believe.

Primary Sources

- Poornima Madhavan & David A. Wiegmann, "Similarities and Differences Between Human-Human and Human-Automation Trust." Theoretical Issues in Ergonomics Science, vol. 8, 2007, pp. 277–301.
- John D. Lee & Katrina A. See, "Trust in Automation: Designing for Appropriate Reliance." Human Factors, vol. 46, 2004, pp. 50–80.
- Raja Parasuraman & Victor Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse." Human Factors, vol. 39, 1997, pp. 230–253.
- Kate Goddard, Abdul Roudsari & Jeremy C. Wyatt, "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators." Journal of the American Medical Informatics Association, vol. 19, 2012, pp. 121–127.
- Berkeley J. Dietvorst, Joseph P. Simmons & Cade Massey, "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." Journal of Experimental Psychology: General, vol. 144, 2015, pp. 114–126.
- Jennifer M. Logg, Julia A. Minson & Don A. Moore, "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment." Organizational Behavior and Human Decision Processes, vol. 151, 2019, pp. 90–103.
- National Institute of Standards and Technology, AI Risk Management Framework (AI RMF 1.0). NIST, 2023.