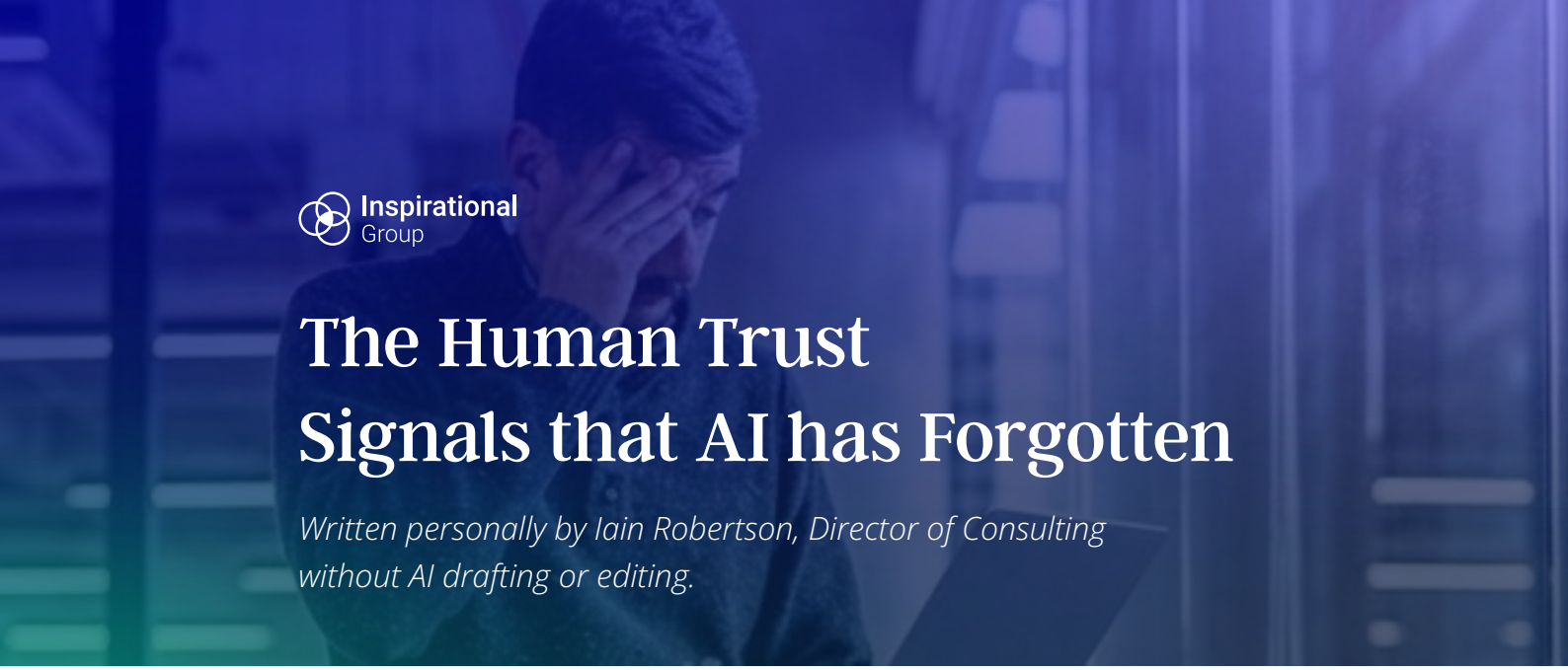


Inspirational Group

# The Human Trust Signals that AI has Forgotten

*Written personally by Iain Robertson, Director of Consulting without AI drafting or editing.*

## Executive Summary


It feels like AI is already everywhere, and yet we are consistently being told that we need more of it, in more places, doing more things. If the increase in AI is to be successful, we need to understand how much leaders trust AI and whether that trust can keep pace with the adoption and deployment we are seeing.

The clear message from our research is that trust in AI is low – or at best, conditional. In particular, we distrust it where it removes visible, human accountability and involvement. In our research, we found that three behaviours consistently restore – or at least increase – trust in AI content:

### Oversight.

We respond much better to information when it is owned by a named individual – not a policy, a process or a department. Putting my name on something also puts my reputation on it.

### Transparency.

We need to be clear and honest about the scale of AI usage in our information – we want to know *where* the data has come from. Genuine disclosure of the extent of AI involvement allows me to make an informed judgement about how much I trust the output.

### Effort.

The amount of care we put into our AI content correlates with the amount of value we place on the relationship. Low effort translates into low value – higher effort to higher value.

Underneath all of this, however, is a deeper and more fundamental risk. AI undoubtedly improves information flow. What it removes at the same time are the imperfect human interactions that create relationships and trust in the first place. The friction of mistakes that make us unique and memorable. The deeper connection created by *how* something is said, not just *that* it is said.

The elimination of those building blocks of relationships will make it impossible to build trust in the first place – with or without AI.

**Five key recommendations arise from this investigation:**

- Name an owner for every output
- Raise the bar on disclosure
- Value the “effort signal”
- Develop a shared language around AI use
- Protect the relational investment

Inspirational Group

# The Trust Gamble

## Leaders are gambling, sometimes decades of hard-won public reputation and trust, on unproven and potentially unreliable AI technology – and they often don't even realise they are doing so.

Whether it's Air Canada's website chatbot making up policies[1] , or CNET publishing inaccurate stories written by its own internal AI[2] , these decisions are harming companies' reputations and lowering public trust in AI.

These aren't isolated incidents or edge cases. The last three years have seen a massive uplift in organizations deploying AI at scale. Whether that is machine learning, a website chatbot or even autonomous agents making investment or trading decisions, the involvement of AI has increased exponentially.

The fascinating thing about this from our point of view is that AI is not simply being deployed in low-visibility, low-stakes contexts. As the examples above prove, AI is having real, public impact at an organizational level for many leaders, and yet they freely admit that AI adoption is outstripping their ability to govern it (KPMG, 2024).

In a world where more adults find AI to be anxiety-inducing than exciting (Pew Research Centre, 2025) leaders have an obligation to help their people understand and trust AI and AI-generated information. For that to happen, they themselves need to understand and trust it better.

We therefore set out to discover how much leaders trust AI – and what (if anything) can be done to help them trust it more.



*[1] In February 2024 Air Canada tried to argue that a chatbot hosted on its website was, in fact, a separate legal entity and therefore that they should not be accountable for its actions. This was important because the chatbot had completely invented a policy concerning bereavement fares, resulting in a customer making a booking and incurring additional costs. Air Canada then tried to deny a refund request, leading to a tribunal where the airline was held responsible for the actions of the chatbot, setting a significant precedent that organizations are responsible for what their AI tells customers.*

*[2] In late November 2022, CNET started publishing fully AI-written financial explainer articles under the byline "By CNET Money Staff". In three months, 77 articles were published this way, 41 of which later required correction for either accuracy or plagiarism. CNET's reputation took a hit – not from the errors themselves (since the articles were under 1% of the total CNET output in that time) but from the lack of transparency and disclosure.*

Inspirational Group

# Using AI to Investigate AI – A Deliberate Choice

In May 2026, we at Inspirational Group deployed our AI Interview Avatar – Maia – to conduct primary research into AI trust among senior leaders. Those leaders are drawn from our global alumni network. We retained 19 in-depth answers from just over 40 initial sessions.

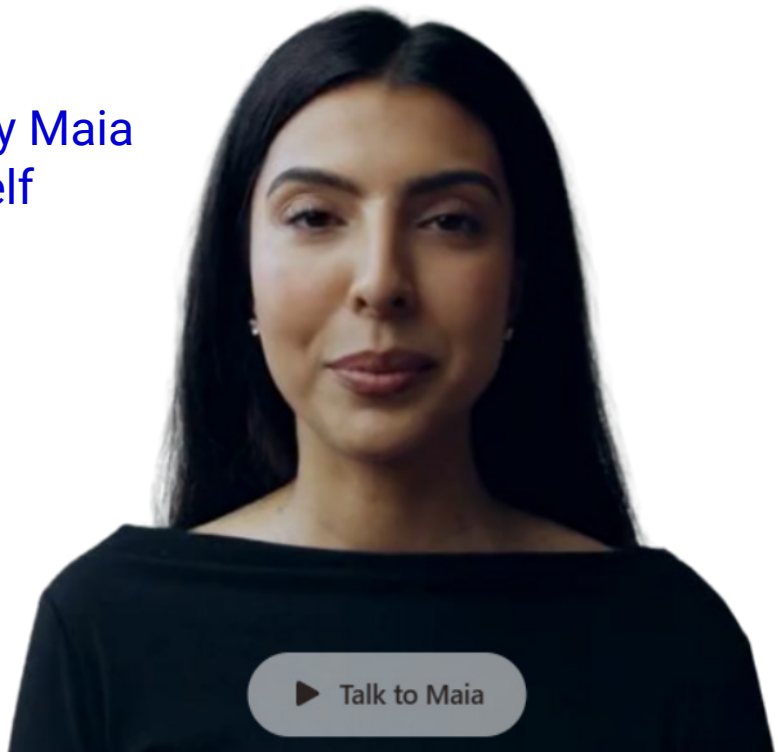
The choice of methodology was itself part of the argument. A standard survey might have generated *more* responses, but almost certainly not *better* responses. In this case, the quality of the data is significantly more important than the quantity of the data – as evidenced by the complex, nuanced responses that Maia was able to elicit from the interviews.

Consequently, rather than just a set of numbers, we have a series of insights. We know that N allocated a dual score – 3 for pure AI, but 8 for AI with visible human involvement. We know that M objected to the premise of a numerical rating in and of itself. And we found that for many, the real trust problem is not the technology itself, but what the using of that technology says about the person using it, and their opinion of us. Each of these observations came through conversation, not from a forced judgement to a scale or an open-ended text box.

It is clear that trust in AI is complex, contextual and sometimes contradictory. Open ended, flowing conversation (even with an AI avatar) allows for more measured reflection by the respondents, leading to longer, more valuable responses that often have greater emotional openness. The method was a deliberate choice for exactly those reasons.

The flexibility of the conversational format mirrors the complexity of the issue. It avoids the risk of simplifying a variable and conditional answer into a single digit. Instead, it allowed our participants (referred to by their initials only in this report) to explore and consider more deeply – leading to more valuable and more interesting conclusions.

Click to try Maia for yourself



# What We Found

The first thing to note about the responses is the conditionality within the sample. Almost all of our respondents attached riders to their scores – whether that is about the amount of human oversight, or whether (in at least one case) it is a "nothing could raise this score for me – I simply will never trust AI".

This reinforces the point above – reducing the nuance and complexity of both individual and collective responses to a mean, mode or median misses the point entirely. Nobody gave an unconditional score, the spread of responses is large and it would be far too easy to say "trust is 5.4 out of 10" without examining the underlying drivers of that number.

Even when we do look at the numbers carefully, the story remains complex and layered. No-one gave a score above an 8 – in our sample, therefore there is no unconditional trust in AI[3]. Most scores cluster around the 5-7 mark, although almost all of those come with a condition. At least one person was adamant that nothing could move their score above a 2.

Two of our participants even pushed back quite strongly against the premise itself – for them the idea of reducing a complex and varied response to a single number simply made no sense. This further underlines the value of our qualitative data gathering approach.



> *If it's fully AI generated... maybe three. But if I know it's been collaborated with a human and there's a human name attached to it, then maybe eight.*
>
> *N, MAIA Session, May 2026*

This one quote gets to the heart of our respondents' views. For N, there is a *five point* gap between pure AI and human-oversight AI. That's not a rounding error, or an incremental improvement. Visibly including humans in the creation process fundamentally changes trust from "I don't trust this" to "I pretty much do". N has put their finger on the key issue that all of our participants were grappling with.

[3] *This is in stark contrast to a recent experience of a colleague of mine with a group of young professionals, one of whom confidently stated "AI doesn't make mistakes. It's just wrong prompts".*

# This Isn't New – But It Is Different

N is not alone in this. Both "Ga" and "Gr" in our survey latched on to something that we have long known to be true – our trust in anything is dependent on our trust in the person or people it came from. My friends and I have a long-standing reference point of "man down pub said, gotta be right" as a shortcut for information or opinion that is under-researched or simply made up.

Aristotle spoke of ethos as a direct parallel here. We've always needed a shortcut to establish quickly and easily whether an argument or opinion is worthy of our trust – or at least, our further investigation. ethos recognises that our perception of the author – and in particular their character – is one of those shortcuts. If I trust you, if I believe you are the sort of person who actually bothers to do their own thinking and their own research, then I will trust your output.

As information has become more distributed, and we can't personally know everyone who has written on a subject, the discipline of academic referencing acts as that trust. If you can point to the original research, I have the ability to check your work. Most of the time I won't do so, but knowing that I can gives me the confidence that the work has been done.

Even Wikipedia recognises this – the famous "citation needed" review point on countless pages stands testament to the fact that we want to know how someone knows what they know, rather than just having them state it to us as a fait accompli.

AI therefore is not actually creating a new form of mistrust. We don't mistrust AI because it's AI – we mistrust AI because it removes two core elements that we have relied on for hundreds of years. In one fell swoop, it removes the named, attributable person (whose ethos we may have been relying on) and the security of a verifiable, traceable source. No wonder trust scores are low.

In addition, AI content *sounds* convincing. LLMs are designed to be persuasive. They are built to articulate arguments clearly and concisely. They take advantage of our desire for things to be simple, and easy, and clear. They gloss over nuance and complexity with confident platitudes *designed* to make people believe them.

This is where the new challenge lies. A poorly-researched, badly-drafted report written by a human looks and sounds like a poorly-researched, badly-drafted report. We can all spot that more work is needed, and the conclusions are at best, unreliable. A completely hallucinated report from an LLM looks and sounds exactly like a wholly reliable report. There are no tells. Consequently, our response is to mistrust *all* AI content as a defence mechanism.

We need to give people back the support structure that they are used to and that has worked for years – namely ethos and a clear path to verifiable supporting data.



# The Three Conditions for Trust

Our research shows clearly that there are three key things that we can do that will go a long way towards remedying the trust deficit – oversight, transparency and effort.




## Oversight

A named person who is taking responsibility for the outcome – gives people back ethos. It returns to the oldest trust mechanism humans have – the knowledge that *this* individual is the person owning the information. It's why leaders have always used face-to-face communication when it's important. Seeing your boss *in person* adds their reputation and their credibility to the message. Where in person is not possible, knowing that they've put their name to it – and therefore own the risk of it being wrong – goes a long way.

> *If I know it's been collaborated with a human and there's a human name attached to it. Score: 8.*
>
> *N, MAIA Session, May 2026*

## Transparency

Being clear and honest about where (and how much) AI has been used – gives people back the citation. Reading something that I *know* has been created or supported by AI tells me what lens to use when reviewing it, and how careful I need to be when trusting it. An email that says "here's what me and my invisible friend came up with" (I, MAIA Session, May 2026) tells me something both about the content I'm about to review, and the person who sent it to me.

> *The conversation we're not having is about what it signals when you use AI without saying so. That's the trust problem. Not the technology itself.*
>
> *T, MAIA Session, May 2026*

Inspirational Group

This also points to an additional element that emerged clearly from the research and is what we're calling the...

## Effort Signal

It links to both oversight and transparency. Unlike the first two, this is not about restoring something that AI has removed. Rather, it's about recognising what the fact that you're using AI in the first place communicates to the other party.

When someone sends you content that is clearly AI generated (or even AI edited) without care, attention and disclosure they are not just failing to provide ethos or a reference. They are indirectly communicating how valuable you are to them by telling you how much they value (or don't value) your time and attention. One respondent illustrated this with an anecdote of a client who sent an email with the prompt still visible:

> *If you haven't even got the effort to copy and paste properly, I'm not impressed with you...*
>
> *I, MAIA Session, May 2026*

This is where intention and outcome can be furthest apart. People assume that the speed or polish that an AI-drafted response allows is the most valuable thing to the recipient. Instead, that same recipient is (consciously or subconsciously) recognising the telltale AI signals and immediately downgrading their view of both the message and the messenger.



# What We Didn't Expect

All of the findings above, while interesting, are not necessarily surprising. What else emerged from the data, however, was a genuinely interesting insight that we're calling the relational displacement concern.

> *…[AI is] dangerous in the potential detriment of people investing in human-to-human relationships, which offer more challenge and the opportunity for break and repair and development of relationships that you cannot get with AI-generated content.*
>
> *I, MAIA Session, May 2026*

Put simply, this concern is that even if we solve the trust problem (clear disclosure and oversight, attribution and observable effort) we are then left with something more insidious and concerning. Yes, you may gain speed and even accuracy, but you lose something more fundamental and human – the friction of imperfect person to person communication.

This insight frames the fact that we *do* communicate imperfectly as a feature, not a bug. That message that has a mistake in it, causing you to pick up the phone; a phone call that meanders and goes off topic; even the frustrating "meeting that could have been an email" – all of these are not just information delivery systems to be optimised, they are the building blocks of human relationships. It's an insight echoed by another respondent, N, who noted that one way to increase their trust in AI would be for it to have the pauses and silences that are part of everyday conversation.

Every time we need to seek additional clarity because someone has struggled to articulate their point, every time we have to wade through unstructured preamble to get to the point, every time we roll our eyes on a Teams call as a colleague rambles on *again* - these are the moments that represent our investment in relationships. We take the time *because* there is a human on the other side of the exchange, and that human has value to us beyond their ability to provide us with information.

So even if AI can smooth over the cracks of our incomplete communication and speed up the flow of information around our organizations, will it at the same time damage our relationships, making them shallower and more transactional?

And as those relationships become shallower, and our trust in our colleagues diminishes, we come back to the problem of ethos. If one of the ways I can bring myself to trust AI is because I trust the person who has oversight of that AI, anything that systematically erodes that trust will also reduce my faith in AI content.



Inspirational Group

# The Three Tiers of Trust

We bring all of this together in a simple but effective three-tier taxonomy of AI usage. This is something we've been explicitly using internally as we increase our adoption of AI, and the data from our interviews has validated it. Many of our interviewees were talking about this, even if they didn't articulate it as clearly.

| Category | Description | Trust implication |
|---|---|---|
| **1. Pure AI** | AI generated, unreviewed, undisclosed | **Low** – reputational risk to the sender |
| **2. AI +Human oversight** | AI generated, reviewed by a named individual, disclosed | **High** – the threshold most participants accept |
| **3. Human + AI assist** | Human-led, AI-supported, disclosed | **Highest** – most participants are most comfortable here |

This links clearly to our earlier findings. Pure AI content fails to deliver on any of oversight, transparency and the effort signal – it is therefore no wonder that trust is low. Human generated with AI assistance, by contrast, delivers on all three – provided the AI assistance is disclosed and actually assistance, not generation. Trust here is likely to be highest.

There are explicit echoes of the framework in some of our respondents' words – the idea that people should "not just prompt something and deliver it without double-checking" (Ba, MAIA Session, May 2026) clearly illustrates the difference between tier one and tier two – and the impact that difference has.

Another interviewee makes the distinction between their own use of AI and others'. Because they know the checks and balances that they put into play, "BI" trusts their own AI-supported output much more than that created by others. Without understanding or sight of process, "not all citations are equal" (BI, MAIA Session, May 2026). This reinforces the point above – transparency without clear oversight and an effort signal is not enough on its own.

The challenges leaders and organizations have faced (including Air Canada and CNET) have come from deploying at tier one. What we've shown so far is that their customers and their teams are only comfortable at tiers two and three. This is a leadership and culture issue, not simply a technology one.

# What Leaders Should Do

1 **Name an owner.** Ethos has worked for centuries because we trust the individual and we believe in their character. Most clearly evidenced in N's dual score (3 vs 8), this requires that a real person puts their real name (and hence their real reputation) next to any AI generated content. Early in my career I worked in professional services, and my work was reviewed and signed off by a named manager and partner before any decisions were made based on it – that discipline should persist today.



2 **Disclose fully, disclose clearly.** The small cost of early disclosure is always lower than the cost of later discovery – as CNET found to their cost. In our sample, T told us that the absence of transparency was a bigger hurdle to trust than any inherent limitation in the technology. Honest disclosure gives us back the citation.

3 **Role model the effort signal.** How leaders individually and collectively use AI (and how they treat its output) tells their teams and organizations what they consider to be acceptable. If the cultural norm that they set is that tier one usage is fine, then tier one usage will be ubiquitous before they know it. By contrast, careful consideration and visible tier three usage will create much more trusted output and trusting team members.

4 **Use a clear taxonomy.** Our three-tier approach is one – but not the only one. Having a shared vocabulary to describe how we are using AI internally helps to reinforce the value of the effort signal noted above. It also provides an easily understood way to audit your current AI use and provide your team with a way of highlighting where they may be uncomfortable with it.

5 **Protect your relational investment.** At Anthropy26 I heard Nationwide describe how they have AI in their contact centre processes, saving on average 20 seconds per call. In contact centre terms, where improvement is measured in fractions of a second, that's huge. Rather than simply bank the saving, they chose to deliberately reinvest it in allowing their agents to have longer, human to human conversations with their customers. They prioritised the relationship, not the process.

Leaders need to make similar, *conscious* choices about which human interactions should be preserved both within their organizations and between their organizations and the outside world. Investing time in building those relationships is the force multiplier that's needed to support elements one, two and three in this list.

In essence, if AI has removed our trust signals – it's the job of leaders to put them back.

# What We Don't Know (Yet)...

Our research is definitely a starting point rather than a definitive answer. Whilst 19 rich, varied conversations with measured, considered participants have given us valuable insight and interesting data, they have of course also raised a number of new, equally interesting questions.

## The impact of social media

Badly written, easily identifiable AI slop dominates many social media platforms – and affects our trust of AI more widely. As J put it "... the main experience I've had is through social media, which is a very untrustworthy source." This exposure may well further contaminate our view of *all* AI content. The risk is that organizations that are producing AI content that is well-researched, human-influenced and valid are losing trust to nonsense created by LinkedIn chatbots.

## Individual AI literacy

"BI"'s reflection that their trust in AI output is very different when they have created it in person versus when others have done so. "BI" trusts their AI-supported content "almost wholeheartedly" – thanks to the checks and rigour they know have gone into it. This may create a further trust distinction that is not easily addressed by our suggestions above – a fourth variable in addition to oversight, transparency and effort.

## Sector and role variation

Our data does not distinguish between sectors or job roles – even only asking for respondents' names resulted in some refusals! Collecting more comprehensive source data would allow for more detailed analysis (although given the levels of trust on display, this may be a challenge!)

With that additional data, we predict that leaders in highly regulated industries will already be setting a materially higher bar for their trust in AI data. Similarly, those in operational roles who are being constantly challenged to do more with less are more likely to overlook the odd hallucination

## The asymmetry of relational displacement

There are two issues here: the variety of relationships I hold, and the incentive to develop those relationships on either side. A friend I've known for 30 years will laugh off a clumsy AI copy and paste without any damage to our relationship. A potential client may well see it as evidence that they don't want to be a client after all.

The second asymmetry is in unbalanced incentives. I may not care what someone in a contact centre who I'll never speak to again thinks of me (and therefore feel that slapdash AI use doesn't matter). By contrast, the Nationwide example mentioned earlier illustrates that some organizations *do* care about those one-off interactions. Drilling down into these uneven elements will help establish the impact on our reservoirs of existing trust.

It's interesting that in looking for answers about what makes us trust AI, we've uncovered more about what makes us trust – or not trust – *anything*, coupled with a list of new questions to explore.

