Is Your Co-Worker Human Enough to Trust? The Impact of Anthropomorphism and AI Accuracy on User Responses to AI Recommendations

Yağız Talay PhD Researcher, Leeds University Business School.

Supervisors: Prof. Andy Charlwood, Dr. Helen Hughes

The Challenge: AI in High-Stakes Decisions

AI systems are increasingly embedded in our workplaces, acting as "digital colleagues" that influence critical decisions, especially in Human Resources recruitment. They promise enhanced efficiency and objectivity. But a critical challenge remains: how do we build appropriate user trust? A global study found that a staggering 66% of users rely on AI output without even evaluating its accuracy (Global AI Trust Study, 2025). This blind reliance is particularly concerning. This research tackles this by examining two crucial elements: how an AI's human-like design (anthropomorphism) and its demonstrable performance (accuracy) shape user trust, ultimately impacting collaboration when the stakes, like selecting the right candidate, are high.

(Chart Description: A line chart titled "Trust in AI Has Dipped From 2022 to 2024" shows that the percentage of users who perceive AI systems as trustworthy has decreased from 63% in 2022 to 56% in 2024. Conversely, the percentage of users worried about AI systems has increased from 49% in 2022 to 62% in 2024. Those willing to rely on AI systems has gone from 52% in 2022, to 43% in 2024. The source is the "Global AI Trust Study, 2025, Univ. of Melbourne & KPMG".)

Theoretical Context

Ever notice how a human-like AI can feel more like a 'someone' than a 'something'? That's the **CASA Paradigm** (Reeves & Nass, 1996) in action—we often treat tech socially if it has human-like cues. This helps us explore if an ai's **anthropomorphic design** (its look and feel) might directly influence the **user trust** we place in it.

But a friendly face isn't everything. We also rely on basic trust principles (like those in **Mayer et al.'s 1995 model**): we need to see that an AI is competent. This means its **accuracy** is a vital ingredient for building that **user trust**.

Our study proposes that **user trust** acts as a crucial mediator: it's potentially shaped by both how an AI *appears* and how it *performs*. This trust, in turn, is proposed to shape **actual behaviour** and even a user's **psychophysiologic state**. We also investigate if **gender** influences how strongly these initial design and accuracy perceptions affect the trust individuals form.

Key Questions

- Does giving AI a more human-like persona genuinely deepen our trust, or is it just a superficial charm?

- When it comes to trusting an AI, what truly wins us over—its flawless performance or its relatable, human-like design?

- How does our level of trust in an AI translate into real-world decisions, our efficiency, and even the subtle signals from our bodies (like stress or confidence)?

- Do men and women perceive and respond to an AI's human-like qualities and its competence differently when deciding who—or what—to trust?

Our Proposed Model: Unpacking Trust

This model outlines how we believe trust in AI is formed and its consequences. We investigate how an AI's human-like qualities and its performance initially build user trust. This trust is then proposed to be the crucial factor driving user responses, from their actions to their physiological state, with user gender potentially influencing how strongly those initial AI characteristics shape trust.

(Flowchart description: AI Accuracy and Anthropomorphic Design influence Perceived User Trust. Perceived User Trust then leads to Compliance, Decision Latency, and Physiological Responses. User Gender is shown as a factor that can influence the relationship between the AI accuracy and anthropomorphic design.)

(Stacked barchart description: Trust and acceptance of AI systems.

Trust in AI overall – 35% unwilling to trust, 19% ambivalent, 46% willing to trust. Human Resources AI – 39% unwilling to trust, 19% ambivalent, 42% willing to trust. Healthcare AI - 30% unwilling to trust, 18% ambivalent, 52% willing to trust.

Source: Global AI Trust study, 2025. University of Melbourne and KPMG.)

An Experimental Multi-Method Approach

Our experiment immerses participants in a realistic hiring challenge where they act as HR professionals, evaluating job candidates with the help of an AI assistant. We've specifically designed the task to be challenging and ambiguous, making their trust in the AI a critical factor in their decision-making.

We systematically test two core AI aspects:

- **AI's Human-like Qualities (Anthropomorphic Cues):** Comparing a human-like avatar to a basic chatbot.

- **AI's "Performance":** Contrasting high versus lower accuracy.

To capture a complete picture of trust, we measure:

- **What People Say:** Explicit trust levels (via HCTS).

- **What People Do:** Actual behaviour (compliance with AI, decision speed).

- **What Their Bodies Reveal:** Subtle physiological stress/confidence signals (EDA & HRV using MindWare Mobile).

Why This Matters: Expected Contributions

Understanding how we trust AI isn't just academic—it's about building better, safer, and fairer "digital colleagues." This research aims to make a real-world difference by:

- **Improving AI Design:** Guiding the creation of AI that fosters well-calibrated trust, ensuring users don't over-rely on flawed systems or underutilize truly capable ones.

- **Understanding User Reactions:** Revealing how AI's human-like design, performance, and user gender shape trust, reliance, and even subconscious physiological responses.

- **Advancing Trust Measurement:** Showing how combining behaviour with physiological data (EDA/HRV) offers deeper insights into trust.

- **Enhancing Real-World AI Use:** Providing practical evidence-based advice for using AI effectively, ultimately aiming to build more productive and reliable human-AI collaboration.

Acknowledgments

References

- Lee, J. D. and See, K. A. (2004) Trust in automation: designing for appropriate reliance. *Human Factors*, 46(1).

- Mayer, R. C., Davis, J. H. and Schoorman, F. D. (1995) An integrative model of organizational trust. *Academy of Management Review*, 20(2), pp. 709-734.

- Reeves, B. and Nass, C. (1996) *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

- Global AI Trust Study, 2025. Univ. of Melbourne & KPMG.