

• Bridging Perceptions and Reality with Action-Oriented Insights

*Insights from Future of Innovation Lab x
Lee Kuan Yew Centre for Innovative Cities*

NEED

With the rapid proliferation of generative AI tools that are now accessible to the general public, trust in AI research is more critical than ever to ensure that these technologies are used responsibly, ethically, and effectively, minimising risks while maximising their potential to enhance human decision-making and productivity.

OPPORTUNITY

This article is grounded in research aimed at addressing two fundamental questions: "Will we trust AI?" and "What must be done for us to trust AI?". Our findings have broken new ground in the following areas:

1. **Measuring actions over perceptions:** We utilised economic experiments to measure behaviour, as opposed to relying solely on survey-based perceptions.
2. **Exploring AI personalities:** We examined how assigning personalities to AI impacts the level of trust people place in these systems.
3. **Fairness in AI trust:** We uncovered that fairness in trust means people are comfortable with AI benefiting from their interactions, provided the benefit is not disproportionate.

In this article, we provide nine strategies to enhance trust in AI:

- 01 **Foster Societal and Institutional Trust Alongside AI Adoption.**
- 02 **Ensure AI Benefits Are Equitably Distributed.**
- 03 **Design AI Systems to Be Socially Oriented and Consistent.**
- 04 **Implement Safeguards to Prevent AI Exploitation.**
- 05 **Enable Customisation for Increased User Control and Trust.**
- 06 **Incorporate Emotional Intelligence in AI Error Management.**
- 07 **Establish a Clear Vision for AI's Societal Benefits.**
- 08 **Adopt and Align with Institutional and International Standards.**
- 09 **Prioritise Explainability Over Performance in AI Communication.**

BIG TOPIC:

"Will we trust AI?" and "What must be done for us to trust AI?"

Trust in artificial intelligence (AI) has become a critical topic as AI technologies integrate into various aspects of daily life. Since the release of ChatGPT in December 2022, AI has sparked widespread societal discussions.

Despite its growing presence, public understanding of AI technologies often lags behind their implementation, creating gaps in trust. This disconnect raises concerns, as distrust and skepticism toward AI could hinder its future adoption and effective deployment.

To address this, our research adopts novel trust experiments rooted in experimental economics and organisational studies to measure public trust in AI technologies and organisations producing them. By focussing on trust-related actions at both technological and organisational levels, this study provides actionable insights to enhance trust in AI systems and ensure their responsible integration into society.

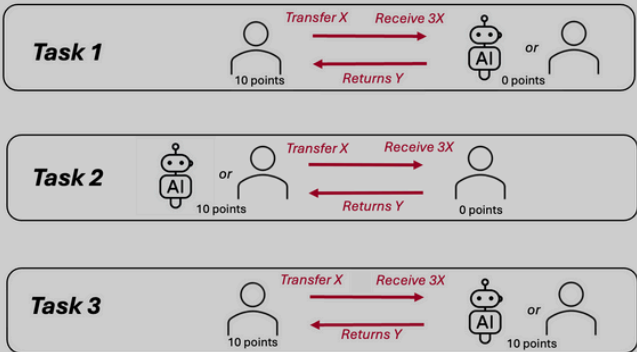
The research addresses several critical gaps in AI trust studies (see table below).

Conventional practice	What we did differently
Research on trust in AI is often fragmented.	We employ economic experiments that integrates technological, psychological, and organisational dimensions.
Rely on surveys to measure perceptions on trust and hence, susceptible to social desirability biases.	Our experiments observed actions to review behaviors, preferences, and propensities grounded in actual human behavior.
Organisational trust was a theoretical concept in the literature.	Our experiments provide empirical evidence on organisational trust.
AI is considered generic and inanimate with a focus on its technical functionality.	Our experiments gave AI different personalities to test how people trust them.

As SUTD embraces AI-driven creativity, technological innovation, and human-AI collaboration, building trust in AI is essential to unlocking its full potential and driving widespread adoption.

Here, we present key findings from our study alongside targeted policy recommendations for various stakeholders to ensure responsible and impactful AI integration.

The Trust Game



Level of trust measured by \$ transferred and returned

Nine Strategies to Enhance Trust in AI

Based on the insights from the trust game experiments with AI personalities: Social, Exploratory, and Competitive.

- 01 Foster Societal and Institutional Trust Alongside AI Adoption.**
- 02 Ensure AI Benefits Are Equitably Distributed.**
- 03 Design AI Systems to Be Socially Oriented and Consistent.**
- 04 Implement Safeguards to Prevent AI Exploitation.**

Based on the insights from learning AI.

- 05 Enable Customisation for Increased User Control and Trust.**
- 06 Incorporate Emotional Intelligence in AI Error Management.**

Based on the insights from the vignette experiment.

- 07 Establish a Clear Vision for AI's Societal Benefits.**
- 08 Adopt and Align with Institutional and International Standards.**
- 09 Prioritise Explainability Over Performance in AI Communication.**

01 - 04

Key recommendations to raise trust in AI technologies based on the insights from the trust game experiments with AI personalities: Social, Exploratory, and Competitive.

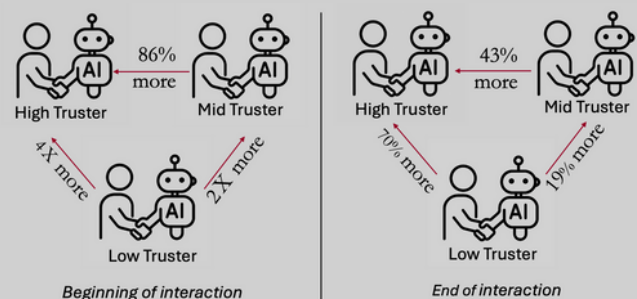
01 Foster Societal and Institutional Trust Alongside AI Adoption

Organisational efforts to increase trust in AI solutions must go in tandem with efforts to increase trust in individuals and society at large.

Individuals are prepared to trust AI and humans alike but what seems to matter is the initial trust level.

At the beginning of the interaction, high-trust individuals send 4× more to the AI than low-trusters and 86% more than mid-trusters. Mid-trusters send 2× more than low-trusters. Over time, low-trusters improve by 70%, but still lag behind mid-trusters.

This suggests that high-trust individuals are not only more trusting initially but also more likely to maintain their trust over time. Mid-trusters gradually increase their trust, while low-trust individuals improve too, but never catch up.



02 Ensure AI Benefits Are Equitably Distributed

To build trust in AI, organisations must ensure that people do not regard the technology as benefiting more than its users disproportionately.

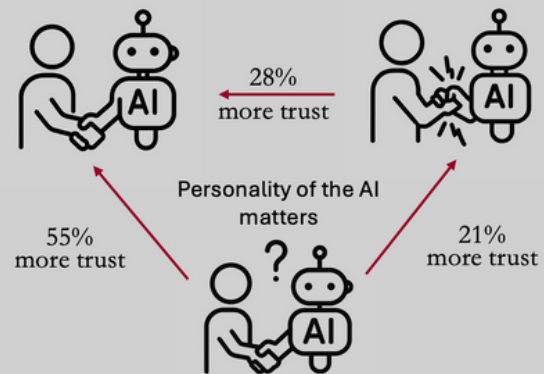
People's trust in AI is influenced by the balance of initial benefits. When AI is seen as gaining more benefits than its users, it often leads to defiance and reduced trust.

03 Design AI Systems to Be Socially Oriented and Consistent.

To increase trust, organisations should design AI systems to be socially oriented and consistent. In our experiment we gave AI three personalities: Social AI (consistent and rewarding), Competitive AI (fair but self-serving), and Exploratory AI (unpredictable). AI that demonstrates socially oriented behavior earns greater trust — about 55% more than unpredictable AI and 28% more than competitive AI.

This highlights the importance of designing AI systems that prioritise social alignment and user benefits. Consistency is crucial because trust in AI is highly sensitive to breaches, especially among users who initially have high levels of trust.

When users repeatedly experience breaches of trust, regaining their confidence becomes significantly more difficult. Additionally, the more people interact with AI systems, the more likely they are to trust the technology over time.

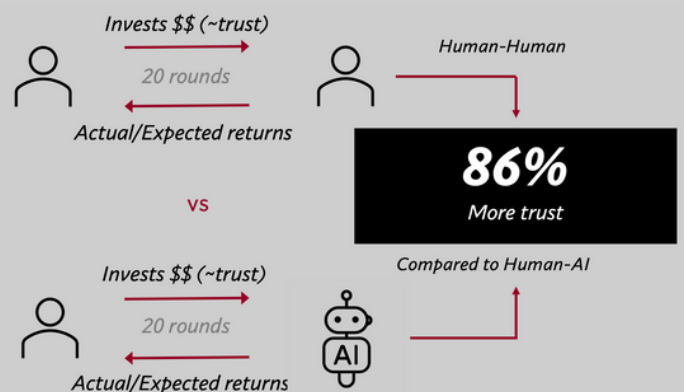


04 Implement Safeguards to Prevent AI Exploitation

At the same time, to prevent AI from being exploited by humans, it is essential to implement safeguards within the AI system and include humans in the loop.

Humans are more likely to take advantage of AI when they perceive there are no consequences for doing so. This behavior was observed in our study, by measuring how much humans reciprocated to different AI personalities. On average, approximately 86% of the participants chose to reciprocate more to humans than to AI, likely due to social rules of reciprocity.

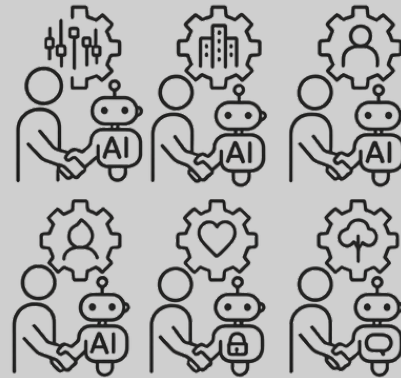
These results suggest that incorporating social norms into the design of AI systems is critical and/or we can design humans to be part of the loop to act as safeguards.



05 Enable Customisation for Increased User Control and Trust

Organisations can build trust in AI by allowing users to customise it. People are sensitive to whether the AI is benefiting them and being able to customise the AI increases the benefits, and vice-versa.

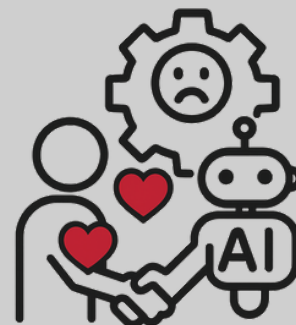
We see increasing use of this for example on our smart phones, web platforms, and Gen AI tools, where you can customise interfaces, settings, and preferences.



06 Incorporate Emotional Intelligence in AI Error Management

To rebuild trust after an AI has made a mistake, organisations must also include strategies that account for people's emotional reactions to the mistakes.

This is because participants reacted emotionally to the mistakes of the AI by reducing their trust in the subsequent interactions, which hinders further optimal interactions with the AI.



07 Establish a Clear Vision for AI's Societal Benefits

To increase trust in organisations, companies should present a clear vision to how the technology benefits the society and actively ask customers for their feedback. We found that 60% of the participants trust companies that care about customer feedback and with a clear vision of how their technology benefits society.

The fact that none of the leading AI companies has laid out compelling vision of how the AI could benefit users/workers instead of displacing them causes much anxiety.

60%

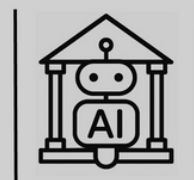
participants trust companies that care about customer feedback and with clear visions of the benefits to society.

08 Adopt and Align with Institutional and International Standards

Organisations are more trusted when they have institutional standards and guidelines in place, especially international ones. We found that 75% of the participants express higher levels of trust when clear institutional standards and guidelines are established, particularly favoring frameworks that are international rather than national.

This represents a missed opportunity, as no such comprehensive framework currently exists, highlighting a critical gap in the global governance of AI. The absence of a unified, globally recognised AI governance framework leads to fragmented approaches in addressing ethical concerns related to AI development.

This lack of cohesion creates significant obstacles to establishing universally accepted ethical norms and standards.



75%

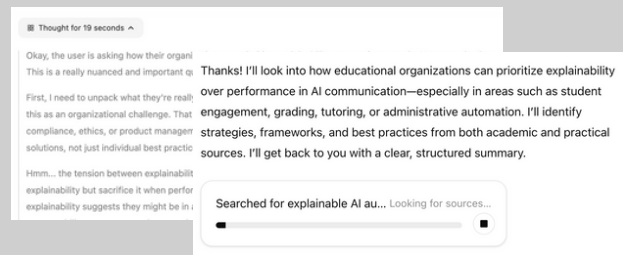
of participants express higher levels of trust when clear institutional standards and guidelines are established

09 Prioritise Explainability Over Performance in AI Communication

To build trust, it is essential to focus on explaining how the technology works rather than emphasising its performance. Offering selected explanations about the technology significantly increases the trust of participants in the company, with full explanations having the greatest impact on trust.

Notably, it is the act of providing explanations, rather than participants' comprehension of them, that plays a crucial role in building trust. Participants do not indicate that performance factors, such as data quality or the number of environments in which the technology operates, are important to them.

It suggests they may perceive their use of the technology as requiring minimal functionality and may not fully understand the factors that influence performance. For example, 72% preferred explainability while only 17% said data quality was important for trust over other dimensions of organisational reputation.



Screenshots of AI platforms providing explanations of how they are reasoning and analysing input prompts

References

Reference to study:

1. Jayasekara, D., Prissé, B., Deng, R., & Ho, J. Q. (2025, April 25). *Exploring trust in artificial intelligence (AI) systems: Insights from a repeated trust game*. SSRN. <https://doi.org/10.2139/ssrn.5229860>
2. Prissé, B., Mussagulova, A., & Ho, J. Q. (2025, May 22). *Transparency and public communication foster trust in AI companies*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5268718
3. Prissé, B., Jayasekara, D., Deng, R., & Ho, J. Q. (2025, May 1). *Building trust with a teachable artificial intelligence: The case of repeated trust games*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5268700

Reference to other related studies:

4. Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, 35(2), 530–549. <https://doi.org/10.1108/JEIM-08-2021-0364>
5. Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal*, 40(5), 685–706. <https://doi.org/10.1016/j.emj.2022.03.006>
6. 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(1), 114–126. <https://doi.org/10.1037/xge0000033>
7. Farjam, M. (2019). On whom would I want to depend; humans or computers? *Journal of Economic Psychology*, 72, 219–228. <https://doi.org/10.1016/j.joep.2019.04.001>
8. Kreps, S., George, J., Lushenko, P., & Rao, A. (2023). Exploring the artificial intelligence “trust paradox”: Evidence from a survey experiment in the United States. *PLOS ONE*, 18(7), e0288109. <https://doi.org/10.1371/journal.pone.0288109>
9. 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. <https://doi.org/10.1016/j.obhdp.2018.12.005>
10. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.5465/amr.1995.9508080335>
11. Oksanen, A., Savela, N., Latikka, R., & Koivula, A. (2020). Trust toward robots and artificial intelligence: An experimental approach to human-technology interactions online. *Frontiers in Psychology*, 11, 568256. <https://doi.org/10.3389/fpsyg.2020.568256>
12. Upadhyaya, N., & Galizzi, M. M. (2023). In bots we trust? Personality traits and reciprocity in human-bot trust games. *Frontiers in Behavioral Economics*, 2, 1164259. <https://doi.org/10.3389/fbeco.2023.1164259>
13. von Schenk, A., Klockmann, V., & Köbis, N. (2023). Social preferences toward humans and machines: A systematic experiment on the role of machine payoffs. *Perspectives on Psychological Science*. Advance online publication. <https://doi.org/10.1177/17456916231194949>