3rd Edition

Data Research meetup by MagIC

AI Self-Efficacy, Trust, and Human Reliance on AI-Generated Advice: A Systematic Literature Review

André Ricou Tavares Carreiro – 20231149@novaims.unl.pt

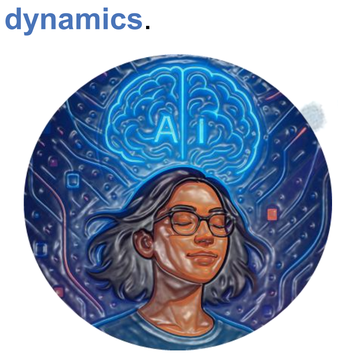
NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

INTRODUCTION

Artificial Intelligence (AI) systems increasingly influence human judgment and decision-making across professional and everyday contexts. Recommendation engines, decision-support systems, and especially large language models have intensified human–AI interaction, making it essential to understand **how individuals perceive their own ability to work with AI** and how this perception shapes downstream behaviors such as trust, advice-taking, and **reliance**.

Research in technology acceptance and digital literacy suggests that **self-efficacy** influences technology use, confidence, and adaptation; however, its specific role in **AI-supported decision-making** remains conceptually fragmented.

This review synthesizes evidence aiming to clarify how these constructs are connected, identifying emerging directions for future research on **human–AI decision dynamics**.

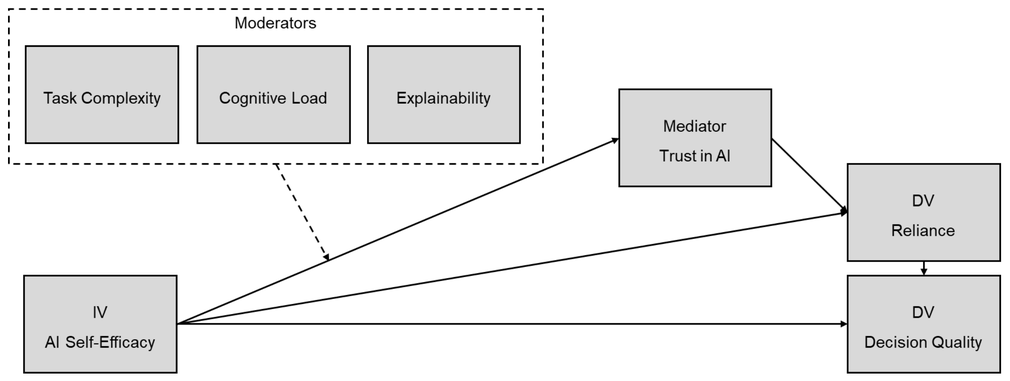


RESULTS & DISCUSSION

Qualitative Findings: Trust emerged as the most frequently studied construct. Transparency, Explainability, Complexity, and Competence were often present. AI self-efficacy was rarely measured empirically, only referenced indirectly. Overreliance occurred in conditions of ambiguity.

Quantitative Patterns: Models found trust mediates system characteristics (e.g., explainability) and reliance. Self-efficacy showed positive correlations with confidence, perceived usefulness, and willingness to engage with AI, though findings were dispersed.

Interpretation: Taken together, the literature suggests a fragmented but converging model: self-efficacy shapes how users interpret AI outputs; trust determines whether users rely on them; and reliance influences decision quality.

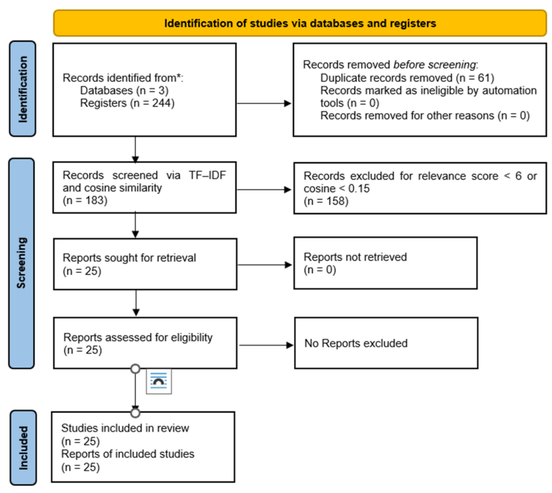


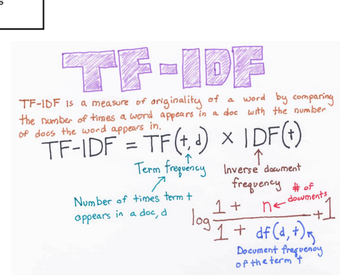
METHODS AND MATERIALS

PRISMA 2020 was followed. Searches were conducted in **Scopus**, **Web of Science** and **Science Direct**, using controlled keywords related to AI self-efficacy, trust in AI, advice-taking, human–AI interaction, and decision-making.

Only journal articles were included. Records were merged, deduplicated, and screened through title and abstract review. Twenty-five studies were subject to full-text review. This screening was done using the TF-IDF algorithm with a threshold of 0.15 cosine similarity and a score.

Data extraction included **theoretical frameworks**, **constructs** examined, **operational definitions**, **research designs**, **measurement instruments**, **sample characteristics**, and **key findings**.





CONCLUSION

This review synthesizes current knowledge on the **interplay between AI self-efficacy, trust, and reliance in human decision-making**. Results indicate that while trust-related constructs dominate the literature, AI self-efficacy remains conceptually acknowledged but insufficiently operationalized, leaving its role in AI-supported decisions unclear. Existing studies consistently suggest that individuals who feel more competent with AI perceive systems as more useful, engage more confidently, and potentially calibrate trust more appropriately—but rigorous empirical validation is limited.

The findings reinforce the need for integrated models that include psychological and system-level variables, and experimental studies testing moderating and mediating effects across different task complexities. This review highlights opportunities for future research to examine how explainability, and cognitive load interact with user-level factors to shape advice-taking and reliance.

Understanding these mechanisms is increasingly important as AI systems become embedded in organizational and societal decision processes. Clearer models of self-efficacy and trust will support better AI design, improved user training, and more responsible adoption. Ultimately, this review lays the groundwork for a research agenda aimed at. **enhancing human–AI collaboration and ensuring effective, trustworthy, and equitable use of AI-generated advice.**

REFERENCES

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.

Blundell, C., Thompson, P., & Zhao, Y. (2023). The illusion of understanding: Overreliance on large language models in human–AI collaboration. AI & Society. Advance online publication. https://doi.org/10.1007/s00146-023-01786-9

Chee, B., He, L., & Park, J. (2023). Transparency and fairness in explainable AI: Implications for user trust. Computers in Human Behavior, 147, 107729. https://doi.org/10.1016/j.chb.2023.107729

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated decision-making: Effects on trust and compliance. Computers in Human Behavior, 99, 277–289. https://doi.org/10.1016/j.chb.2019.05.017

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

ACKNOWLEDGEMENTS

This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS - https://doi.org/10.54499/UID/04152/2025 (2025-01-01/2028-12-31) and UID/PRR/04152/2025 https://doi.org/10.54499/UID/PRR/04152/2025 (2025-01-01/ 2026-06-30).

Additional thanks to NOVA IMS faculty and colleagues for their guidance and support throughout this research, especially Professor Tiago Oliveira, my supervisor, Professor Marco Painho and Professor Diego Costa.

Funded by:

Fundação para a Ciência e a Tecnologia · PRR Plano de Recuperação e Resiliência · República Portuguesa · Financiado pela União Europeia NextGenerationEU

NOVA Information Management School, 18 December 2025

ONLINE VERSION