FinTAF

FinTech and Artificial Intelligence in Finance (FINTAF)
Issue 3

**Oyenike Akinlabi**

**Sheffield Business School, Sheffield Hallam University**

**Submitted 17th November 2025**

**Accepted 9th February 2026**

**Published 26th February 2026**

## Human Intellectual Humility: Addressing the fallibility of Artificial Enabled Systems

The use of artificial intelligence cuts across diverse fields, allowing the generation of output at the speed of light. Despite its efficiency, AI is blamed for its hallucinations and biased output; it sometimes generates seemingly convincing outputs that are inaccurate. Hence, its outputs are not robust and unreliable. While this is true, society is hypercritical of this anomaly, and they tend to forget the fallibility of humans. AI systems are trained with relevant data suitable for the applicable context. But it may be reflective of ingrained human conscious or unconscious biases.

For instance, information required for customers' application forms in financial services, housing, and health is flawed with intentional use of protected characteristics (direct discrimination) or implied inference of disproportionate treatment (indirect discrimination) (Du Preez et al., 2024; Xin & Huang, 2024). The information comes with implicit biases that potentially discriminate against people with conspicuous characteristics like gender, colour, religion, age, race, and ethnicity if used as training data.

Where a protected characteristics variable is unavoidably impossible, Bartlett, Morse, Stanton and Wallace (2022) noted that the problem of omitted unobserved customers' characteristics is resolved through proxy selection. However, proxies that inaccurately represent the stakeholders can potentially result in inadvertently discriminatory outcomes. For example, the use of postcode as a proxy results in unfair treatment of people of a particular income level, religion, or race. While AI financial models are perceived to avoid human biases (Barocas &Selbst, 2016; Bartlett et al., 2022), financial models trained with data prone to implicit or explicit discrimination are biased in their judgment. Empirical study of Lindholm, Richman, Tsanakas and Wuthric (2022) shows a significant indirect discrimination against people of a particular ethnic group when a real-world insurance dataset was used. Inaccuracy in the generated results is not attributable to AI hallucination but to human fallibility.

Surprisingly, AI systems sometimes operate outside the confines of the trained data, generating false results. Like humans, AI systems are fallible and have been reported to impact decisions, resulting in financial losses. Recently, an audit firm incurred a loss of $290,000 due to the failure of human oversight in an AI-generated report (Dhanji, 2025). Until humans acknowledge their fallibility, AI hallucinations may remain an indefinite problem with no end

date. Arguably, from a psychological lens, this paper proposes intellectual humility to disrupt AI hallucinations.

Intellectual humility is the recognition of one's limit of intellectual ability when making evidence-based decisions (Leary et al., 2017; Whitcomb et al., 2017; Zmigrod et al., 2019), readiness to assess alternative views, and acceptance that the evidentiary basis of one's perceptions may lack some details (Alfano et al., 2017; Whitcomb et al., 2017). Intellectual humble individuals give up the notion that their first-order belief is the only correct and reasonable one (Alfano et al., 2017). Some may perceive intellectual humility as diffidence or Honesty-Humility. Diffidence is the quality or state of being unassertive and bashful; those who are unassertive reject their views and accept others' views due to low self-confidence. People characterised with this trait are likely to accept facts on face value and therefore not fit either as an AI developer or user.

On the other hand, Honesty-Humility is a personality trait that projects moral behaviour (Otto et al., 2021). Due to the inclination of the trait to ethical behaviour, it is an essential underpinning for ethics in AI governance (Zabel et al., 2025). This view was echoed by Rockwell, Zhu and Majumdar (2025), who stated that character traits and human behaviour are fundamental to fixing flaws in AI systems. Studies found that those high in Honesty-Humility are trustworthy (Thielmann & Hilbig, 2015) and, as a result, may regard AI systems as completely truthful (Zabel et al., 2025). However, another study found that Honesty-Humility is negatively related to general and specific acceptance and adoption of AI-enabled systems (Weger et al., 2022). While the views of authors (see Weger et al., 2022 and Zabel et al., 2025) appears to be contradictory, the differing views may be true in that the trust of Honesty-Humility individuals in AI systems (Zabel et al., 2025) emerge from the belief that AI developers are experts and have taken due diligence in the development of AI systems but yet they are disinclined to its adoption. Theoretical reasonings from literature suggest that due to the moral virtue (Angelis & Pensini, 2023) and risk-averse (Vries et al., 2009) nature of Honesty-Humility individuals, their reluctance stems from ethical concerns regarding the usage of AI (Zabel et al., 2025). They also believe AI adoption diminishes human authenticity (Zabel et al., 2025). In addition to these reasons, this present study believes they perceive machines as fallible without accepting the notion of human fallibility. Expectedly, unlike diffidence-individuals, honesty-humility individuals are not likely to accept the first-order belief of an AI system without questioning its output if they embrace its adoption.

Another human trait character that may depict human infallibility is intellectual arrogance. Samuelson and Church (2015) noted that holding to one's belief without considering other views depicts humans as infallible. Those with this improper higher-order epistemic attitude are intellectually arrogant (Zmigrod et al., 2019). A developer of an AI system characterised with this trait may be heedless to potential problems in the development process. Intellectual humility is the virtue that helps to discern between intellectual arrogance and diffidence (Johnson, 2017). Likewise, Honesty-Humility individuals may need to be intellectually humble to embrace AI systems.

Intellectually humble individuals recognise and accept that humans are fallible and therefore engage the two processes of human thinking and reasoning systems. The first process generates fast, automatic and intuitive information, and the second is a conscious and deliberate analysis of the output from the former (Alfano et al., 2017). When humans rely on heuristics and disengage the latter system, they are perceived as intellectually arrogant, discounting the possibility of misinformation in their output. Similarly, AI systems engage the first process of human cognition that retrieves fast, automated information and its users should regard this as

a first-order belief. Because AI systems cannot critically analyse their output, an intellectually humble user will recognise their fallibility and assess their output for potential hallucination.

Rockwell et al. (2025) call for more qualitative research to understand the development and application of AI in lived environments as opposed to experimental methods found in (Sommaggio & Marchiori, 2020; Weger et al., 2022). Hence, this paper proposes to recruit AI developers and users to answer questions in the Comprehensive Intellectual Humility Scale (CIHS) to assess participants' self-perception of their intellectual humility. Subsequently, behavioural interview questions will be asked about their successes and mistakes to check to validate the self-reported evaluation. Finally, they will be interviewed to confirm if they engage the two processes of human thinking and reasoning to assess their belief about human fallibility.

The core argument of this paper is that Intellectual Humility should be a psychological attribute essential for AI developers to minimise flaws in development processes that cause AI hallucinations. A developer who believes humans are infallible accepts the model developed as perfect without admitting flaws in training data or biases in design choices. These oversights cause output of false information from AI systems. On the other hand, users of AI systems that lack Intellectual Humility will accept the output without assessing its accuracy. Therefore, the findings of this research will highlight how intellectual humility shapes developers' accountability of AI's development and users' responsibility for communicating true output from the AI system, challenging humans for AI fallibility.

**References**

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). Development and validation of a multi-dimensional measure of intellectual humility. *PLOS ONE*, *12*(8), e0182950. https://doi.org/10.1371/journal.pone.0182950

Angelis, S., & Pensini, P. (2023). HONESTY-HUMILITY predicts humanitarian prosocial behavior via social connectedness: A parallel mediation examining connectedness to community, nation, humanity, and nature. *Scandinavian Journal of Psychology*, *64*(6), 810–818. https://doi.org/10.1111/sjop.12932

Barocas, Solon; Selbst, Andrew D. (2016). *Big Data□s Disparate Impact*. https://doi.org/10.1577/Z38BG31

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech Era. *Journal of Financial Economics*, *143*(1), 30–56. https://doi.org/10.1016/j.jfineco.2021.05.047

Dhanji, K. (2025, October 6). Deloitte AI Hallucinations Report [The guardian.com/]. *Deloitte to Pay Money Back to Albanese Government after Using AI in $440,000 Report*. https://www.theguardian.com/australia-news/2025/oct/06/deloitte-to-pay -money -back-to-albanese- government-after-using-ai-in-440000-report

Du Preez, V., Bennet, S., Byrne, M., Couloumy, A., Das, A., Dessain, J., Galbraith, R., King, P., Mutanga, V., Schiller, F., Zaaiman, S., Moehrke, P., & Van Heerden, L. (2024). From bias to black boxes: Understanding and managing the risks of AI – an actuarial perspective. *British Actuarial Journal*, *29*, e6. https://doi.org/10.1017/S1357321724000060

Johnson, A. R. (2017). Intellectual Humility: An Introduction to the Philosophy and Science, Ian M.Church and Peter L.Samuelson, Bloomsbury, 2017 (ISBN 978-1-4742-3674-4), xii + 356 pp., pb £21.99. *Reviews in Religion & Theology*, *24*(4), 676–679. https://doi.org/10.1111/rirt.13050

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and Interpersonal Features of Intellectual

Humility. *Personality and Social Psychology Bulletin*, *43*(6), 793–813. https://doi.org/10.1177/0146167217697695

Lindholm, M., Richman, R., Tsanakas, A., & Wuthrich, M. V. (2022). A Multi-Task Network Approach for Calculating Discrimination-Free Insurance Prices. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4155585

Merriam-Webster. (n.d.). Diffidence. In *Merriam-Webster*.com Dictionary. Retrieved October 10, 2025 from https://www.merriam-webster.com/dictionary /diffidence

Otto, S., Pensini, P., Zabel, S., Diaz-Siefer, P., Burnham, E., Navarro-Villarroel, C., & Neaman, A. (2021).

The prosocial origin of sustainable behavior: A case study in the ecological domain. *Global Environmental Change*, *69*, 102312. https://doi.org/10.1016/j.gloenvcha.2021.102312

Rockwell, F. C., Zhu, Q., & Majumdar, S. (2025). Exploring AI ethics in global contexts: A culturally responsive, psychologically realist approach. *AI and Ethics*. https://doi.org/10.1007/s43681-025-00821-6

Samuelson, P. L., & Church, I. M. (2015). When cognition turns vicious: Heuristics and biases in light of virtue epistemology. *Philosophical Psychology*, *28*(8), 1095–1113. https://doi.org/10.1080/09515089.2014.904197

Sommaggio, P., & Marchiori, S. (2020). Moral Dilemmas in the A.I. Era: A New Approach. *Journal of Ethics and Legal Technologies*, *2*(04/2020), 89–102. https://doi.org/10.14658/pupj-JELT-2020-1-5

Thielmann, I., & Hilbig, B. E. (2015). The Traits One Can Trust: Dissecting Reciprocity and Kindness as Determinants of Trustworthy Behavior. *Personality and Social Psychology Bulletin*, *41*(11), 1523–1536. https://doi.org/10.1177/0146167215600530

Vries, R. E. D., Vries, A. D., & Feij, J. A. (2009). Sensation seeking, risk-taking, and the HEXACO model of personality. Personality and Individual Differences, 47(6), 536–540. https://doi.org/10.1016/j .paid.2009.05.029

Weger, K., Easley, T., Branham, N., Tenhundfeld, N., & Mesmer, B. (2022). Individual Differences in the Acceptance and Adoption of AI-enabled Autonomous Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 241–245. https://doi.org/10.1177/1071181322661154

Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual Humility: Owning Our Limitations. *Philosophy and Phenomenological Research*, *94*(3), 509–539. https://doi.org/10.1111/phpr.12228

Xin, X., & Huang, F. (2024). Antidiscrimination Insurance Pricing: Regulations, Fairness Criteria, and Models. *North American Actuarial Journal*, *28*(2), 285–319. https://doi.org/10.1080/10920277.2023.2190528

Zabel, S., Pensini, P., & Otto, S. (2025). Unveiling the role of honesty-humility in shaping attitudes towards artificial intelligence. *Personality and Individual Differences*, *238*, 113072. https://doi.org/10.1016/j.paid.2025.113072

Zmigrod, L., Zmigrod, S., Rentfrow, P. J., & Robbins, T. W. (2019). The psychological roots of intellectual humility: The role of intelligence and cognitive flexibility. *Personality and*

*Individual Differences*, *141*, 200–208. https://doi.org/10.1016/j.paid.2019.01.016