2026 Volume: 02 Issue: 08 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-07

HUMAN TRUST IN ARTIFICIAL INTELLIGENCE: THE ROLES OF ORGANIZATIONAL TRANSPARENCY, REGULATORY COMPLIANCE, AND ETHICAL GOVERNANCE

[*1]Jalia Nabadda, [2]Oluwatosin Esther Ajewumi, [3]Susanna Ayensu Blankson

[1]Saint Louis University, USA

[2]Washington University in Saint Louis, USA

[3]Southern Illinois University Edwardsville, USA

Article Received: 11 July 2026

Article Revised: 31 July 2026

Published on: 20 August 2026

***Corresponding Author: Jalia Nabadda**

Saint Louis University, USA.

DOI: https://doi-doi.org/101555/ijrpa.2681

ABSTARCT

The idea of public trust in AI is a core element of regulatory policy in all jurisdictions. However, trust and warranted trust are not the same. The recent scholarship suggests that governance action around AI has separated the two, focusing on "trust production" rather than trustworthiness. This paper takes the critique a step further, examining the level of regulatory architecture from a doctrinal perspective in the United Kingdom and the United States, the most similar jurisdictions to consider given their regulatory histories and designs. Analysis is carried out with a focus on: locus of authority, enforceability and transparency obligations. None of the architectures provides the verification that will allow trust to be warranted, but it's a different kind of failure. The United Kingdom has a central articulation of principles with no obligations that could be shown to have been breached. Even though these commitments are not necessarily so widely recognized as legislation, they still impose enforceable obligations, and as the Colorado, Texas, and Illinois documents illustrate, the degree of enforceability is independent of the statutes' visibility. In both jurisdictions, the protections that a user can infer from the visibility of the regulations. The trust that is created should not be used to assess the regulatory design.

KEYWORDS: Trust Calibration; AI Governance; Regulatory Architecture; Comparative Law; Algorithmic

INTRODUCTION

The use of AI is becoming more common in areas such as credit, health and employment decisions, and the public's trust in AI systems is an explicit policy goal in the UK and the USA. The findings of the survey in seventeen countries suggest that the perceived effectiveness of safeguards, regulations and laws have the biggest impact on trust in AI of the modelled determinants (Gillespie et al., 2023). However, trust and warranted trust are different, the literature in the human-automation interface field has documented the systematic miscalibration of trust: when users undertrust systems that are reliable, and when users overtrust those that are unreliable (Lee and See, 2004; Glikson & Woolley, 2020).

The two jurisdictions have worked to a common goal in very different ways. In the United Kingdom, a principles-based, sector-led approach has been taken in which there are no new statutory duties, but rather, the five cross-cutting principles are embedded within the roles of existing regulators (Department for Science, Innovation and Technology, 2023, 2024). There is no specific federal legislation on AI in the United States, and enforceable commitments are instead established through state laws and through generally applicable enforcement powers.

This paper questions how these conflicting architectures relate to the calibration of trust, suggesting that both are unsuccessful in bridging trust and its warrant.

1. THEORETICAL FOUNDATION AND RELATED LITERATURE

Trust, traditionally defined as a disposition to be vulnerable on the basis of expectations regarding the trustee's behavior, refers to the attributes of the trustee that warrant this trust (Mayer et al, 1995). The two are analytically different, and their fit is the normatively desired goal, as reliance on an untrustworthy agent will be a burden, not a benefit (O'Neill, 2018).

The role of trustees, and whether it can ever be played by AI, is disputed. However, Ryan (2020) claims that the AI does not meet affective and normative criteria for trust, with no emotions, no moral responsibility, and only reliance. The objection is not on the issue of trust itself, but on the locus of trust and on the organizations and institutions that are providing such systems, which aligns with the perspectives that consider trust to be institutionally mediated (Bodó, 2020).

Calibration provides the evaluative criterion. Trust is calibrated when it matches the actual system capabilities (Glikson & Woolley, 2020), and overreliance or disuse arise when the trust is miscalibrated (Parasuraman and Riley, 1997; Lee & See, 2004). Recent scholarship

suggests that the debate around AI governance has lost the connection between trust and trustworthiness, shifted the focus to maximizing trust, and that regulatory regimes can make a contribution to this process by mimicking the acceptability of risk as a synonym for trustworthiness (Laux et al., 2024; Busuioc & Maggetti, 2026).

The criticism has been formulated in a mainly normative public-administration fashion. The significance for specific regulatory architectures, however, has not been studied comparatively, and this is what this paper attempts to fill.

2. APPROACH

The study is doctrinal and conceptual in nature, rather than empirical. The UK and US are considered the most similar countries, as both are common law countries, have more established technology industries and have stated their intent to maintain public trust in AI, but have taken different approaches in their regulatory designs. By keeping these background conditions approximately the same, regulatory architecture is made the variable of interest.

There are two stages to the analysis. The regulatory tools of each jurisdiction are first built up from primary sources, which include law, official policy documents and guidance from the regulators, and augment these with peer-reviewed legal and social scientific works. Then, these are evaluated in light of the calibration criterion elaborated in section 2, on whether they produce trust, or whether they do so by verifying that would make it justifiable to count on. The comparison is arranged in an analytical way, not in terms of jurisdiction, as a simple description of the country's treatment would result.

Two limitations follow. The nature of the regimes required is determined by doctrinal analysis in this study, not by the users' reactions, and thus findings of trust effect are borrowed from the literature. Both jurisdictions also have some of the most changeable laws, with the position being as of August 2026 instead of the current month.

3. COMPARATIVE ANALYSIS: REGULATORY ARCHITECTURE IN THE UNITED KINGDOM AND THE UNITED STATES

Locus of authority. In the United Kingdom, authority is sited at the pre-established sector regulators. There are five cross-cutting principles that bodies such as the Information Commissioner's Office, the Financial Conduct Authority and the Competition and Markets Authority are expected to embed within their remit, with no new statutory powers or duties in place to regulate AI (Department for Science, Innovation and Technology, 2023, 2024). A functional distribution of authority, a central specification of normative content. In the United

States, power is delegated by region. Duties themselves are not clearly defined even if the legislation is state, because duties come from the state legislatures, and not just the application of the duties.

Enforceability. There are no independent legal powers for the UK principles and they rely on enforcement by other regulatory frameworks; where there is no other regulatory framework, there is no expectation of enforcement of the UK principles with remedies. Comparing across three, binding duties are present in the US state statutes but are not correlated with the salience of the legislation in any way. The state of Colorado is one of the first states to have passed a comprehensive AI statute in 2024, which mandates a duty of care, risk management programmes and impact assessments, with this legislation being repealed and reenacted prior to its implementation under SB 26-189, replacing it with notice, adverse-outcome explanation, and human review (Colorado General Assembly, 2026). Texas has a law that purports to be a comprehensive law for AI governance but the provisions of the law regarding the limitation of civil liberties are clearly limited to government actors and not commercially deployed systems (Texas Legislature, 2025). By comparison, Illinois has no specific statute for AI but has made it more stringent in the Human Rights Act, stating that a discriminatory effect, not intent, is a violation of civil rights (Illinois General Assembly, 2024).

Transparency obligations. Appropriate transparency is not a standalone regulatory obligation, but a principle for regulatory interpretation across the context, and no threshold for adequacy is set. While employers are required to disclose the use of AI in the hiring process in the US statutes, the rules on when, how, and what information to provide are still in the proposed stage in Illinois, which was enacted in 2024, and were proposed in mid-2026 (Illinois General Assembly, 2024).

4. DISCUSSION: ARCHITECTURE AND TRUST CALIBRATION

The comparison shows that both architectures produce the trust-signals that exceed their verification capacity but in different ways.

The United Kingdom has a system of principles that is clear to the public and coherent. Coherence is a signal in that a framework of fairness, accountability and contestability says that AI is governed. The framework does not provide any way for a user to know if a particular organisation has adhered to those principles, since they don't require any activities the user should find a lack of. The signal is articulated in the middle while verification is spread across sectoral regimes which are not designed to verify.

The United States creates the opposite problem. Duties do exist and are in locations that do require them, as the Illinois effects-based standard demonstrates. However, the link between visibility and enforceability is fragile: the most visible state AI statute retracted before taking effect, a comprehensive statute of AI governance focuses its core prohibitions on government bodies, and the most onerous requirements are found in an amendment to an existing civil rights statute. A user who finds his or herself in an organization that is implementing AI in the United States cannot assume that any of the laws relevant to AI are in force.

Ethical governance comes in as the space between the two regimes. The problem is it's self-proclaimed. In the absence of regulatory architecture providing external verification, 'observationally equivalent' internal governance and its performance result in the decoupling of trust and trustworthiness (Busuioc & Maggetti, 2026).

Both architectures, then, are not able to fill the calibration gap. The UK signals but doesn't bind. The US binds but not necessarily and not consistently. In both, the things users see are a bad indicator of what limits behaviour.

5. CONCLUSION

Research comparing two jurisdictions that have tried to promote public trust in AI via different approaches to designing regulation suggests that neither reliably yields warranted public trust. The United Kingdom lays down principles and doesn't impose obligations; the United States imposes obligations, and a principle cannot be inferred from its legislative salience. The paper takes the trust-trustworthiness critique of normative scholarship one step further and asks the question of how much trust a regime generates or, if it does supply it, does it supply the verification that would justify it?

There were three restrictions that surrounded such claims. This analysis is doctrinal and sets the requirements for the regimes, not the reaction of users of the regime to the architecture, and so the relationship between architecture and calibration is theoretical. The findings are only relevant to two jurisdictions, both common law and do not apply to the more detailed statutory approach taken elsewhere. Both are heavily subject to change, as Colorado illustrates with its law, and the account links to the position as of August 2026.

The following two research priorities are listed. Empirical designs are required that differentiate trust and calibrated trust to assess if a user's confidence is aligned with the protections at play, rather than with the ones visible to them in a regime. A comparative approach using the EU would allow demonstrating either that a comprehensive and statutory verification gap is closed, or that it is merely moved elsewhere.

REFERENCES

1. Bodó, B. (2020). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media & Society*, *23*(9), 146144482093992. https://doi.org/10.1177/1461444820939922
2. Busuioc, M., & Martino, M. (2026, July 26). *Worthy of trust? AI governance and the role of (DIS)trust*. Perspectives on Public Management and Governance. https://doi.org/10.1093/ppmgov/gvag007
3. Colorado General Assembly. (2026). SB26-189 automated decision-making technology | Colorado General Assembly. In *Colorado.gov*. https://leg.colorado.gov/bills/sb26-189
4. Department for Science, Innovation and Technology. (2023). A pro-innovation approach to AI regulation. In *GOV.UK*. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
5. Department for Science, Innovation and Technology. (2024). *A pro-innovation approach to AI regulation government response to consultation*. https://assets.publishing.service.gov.uk/media/65c1e399c43191000d1a45f4/a-pro-innovation-approach-to-ai-regulation-amended-governement-response-web-ready.pdf
6. Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). Trust in artificial intelligence: A global study. *Trust in Artificial Intelligence*. https://doi.org/10.14264/00d3c94
7. Glikson, E., & Woolley, A. W. (2020). Human trust in Artificial intelligence: Review of Empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057
8. Illinois General Assembly. (2024). Illinois General Assembly - bill status for hB 3773. In *Ilga.gov*. https://www.ilga.gov/ftp/legislation/103/BillStatus/HTML/10300HB3773.html
9. Laux, J., Wachter, S., & Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European Union <scp>AI</scp> act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, *18*(1). https://doi.org/10.1111/rego.12512
10. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
11. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of Organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/https://doi.org/10.2307/258792

12. O'Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies*, *26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637
13. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886
14. Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26*(5). https://doi.org/10.1007/s11948-020-00228-y
15. Texas Legislature (2025) *House Bill 149: Texas Responsible Artificial Intelligence Governance Act*, 89th Legislature, Regular Session. Austin: Texas Legislature. https://capitol.texas.gov/tlodocs/89R/billtext/pdf/HB00149F.pdf