Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 318–328
https://doi.org/10.54941/ahfe1007091

The Transparency Paradox: How AI Explanations Reduce Perceived Autonomy in Organizational Decision-Making

Ancuta Margondai[1], Sara Willox[2], Anamaria Acevedo Diaz[3], Soraya Hani[3], Nikita Islam[3], and Mustapha Mouloua[3]

[1]College of Engineering, University of Central Florida, Orlando, FL 32816, USA
[2]College of Business, University of Central Florida, Orlando, FL 32816, USA
[3]College of Sciences, University of Central Florida, Orlando, FL 32816, USA

ABSTRACT

Artificial intelligence transparency is widely assumed to enhance user experience, yet this study reveals a paradox: detailed AI explanations reduce perceived autonomy. In a between-subjects experiment (N = 557), business students made organizational decisions with either transparent AI recommendations (detailed rationales) or basic recommendations (minimal explanation). Participants receiving detailed explanations reported significantly lower autonomy (M = 3.73) than those receiving basic recommendations (M = 3.84), d = –0.19. Personality substantially moderated effects: Openness to Experience reversed the paradox (interaction = –0.227), with intellectually curious individuals benefiting from transparency, while Extraversion amplified autonomy reduction (interaction = 0.173). Males showed twice the autonomy reduction of females (0.137 vs. 0.063), and effects disappeared by ages 23–25. Neither AI familiarity, attitudes, nor decision complexity moderated effects, suggesting fundamental psychological responses. Despite reduced autonomy, participants maintained positive AI attitudes, revealing dissociation between momentary decision control and general acceptance. Findings challenge universal transparency mandates and suggest personality-adaptive systems may better serve diverse users.

Keywords: Explainable AI, Transparency, Autonomy, Personality, Human-AI collaboration

INTRODUCTION

Artificial intelligence systems increasingly guide organizational decisions, from hiring to strategic planning (Felzmann et al., 2019). The prevailing solution to algorithmic opacity is transparency: if users understand AI reasoning, they will trust it appropriately and maintain effective oversight (Mueller et al., 2019). Evidence supports this logic, transparency improves trust, reliability perceptions, and understanding (Sullivan & Weger, 2025; Angerschmid et al., 2022; Yu et al., 2022).

However, emerging research reveals a "transparency paradox" where more information produces worse outcomes. Ngo et al. (2025) found excessive transparency triggers cognitive overload and reduces adoption. Buçinca et al. (2021) showed explanations increase rather than decrease overreliance on

Received September 15, 2025; Revised November 10, 2025; Accepted November 25, 2025; Available online February 1, 2026

AI. Bansal et al. (2020) demonstrated that explanations made users accept AI recommendations regardless of correctness without improving decision quality. Schmidt et al. (2020) found transparency can actually reduce trust. These paradoxical findings raise an unexplored question: *How does AI transparency affect perceived autonomy?* Autonomy, the sense of volition and ownership over decisions, is fundamental to motivation and organizational effectiveness (Westphal et al., 2023). Detailed AI rationales might enhance autonomy by enabling informed choice or paradoxically reduce it by creating anchoring effects and positioning the AI as an authoritative expert. No research has examined this critical question.

This study addresses this gap through an experiment comparing detailed AI rationales versus basic recommendations across organizational scenarios. The research tests whether the transparency paradox extends to autonomy and examines moderating roles of personality, demographics, and decision complexity. The findings inform AI design, regulation, and theory of human-AI collaboration.

BACKGROUND

The explainable AI (XAI) movement assumes transparency serves users by enabling understanding and supporting oversight (Wang et al., 2019). Sullivan and Weger (2025) found higher transparency significantly improved trust ($\beta = .667$), reliability perceptions ($\beta = .595$), and understanding ($\beta = 1.161$) among 216 participants. In organizational contexts, transparency increases both effectiveness perceptions and employee trust (Yu et al., 2022).

Yet Ngo et al. (2025) identified an inverted U-shape: excessive transparency triggers cognitive overload and skepticism. Buçinca et al. (2021) found explanations do not reduce overreliance and may increase it, as users develop general heuristics rather than engaging analytically. Most troubling, Bansal et al.'s (2020) study of 1,500+ participants showed explanations increased acceptance of AI recommendations without improving human-AI team performance.

Two mechanisms explain these paradoxes. First, cognitive load: explanations increase task complexity, overwhelming users with limited cognitive capacity (Westphal et al., 2023). Second, strategic disengagement: users rationally ignore explanations when processing costs exceed benefits of catching AI errors (Vasconcelos et al., 2022).

The Autonomy Gap

Despite extensive transparency research, no studies examine the effects on perceived autonomy, users' sense of decision ownership, and control. This omission is striking given autonomy's centrality to motivation and organizational functioning. When AI provides detailed rationales, two competing processes may occur. Transparency could enhance autonomy by empowering informed choice. Alternatively, detailed explanations might reduce autonomy through anchoring effects, increased perceived complexity, or subtle messaging that the AI has completed the analytical work, positioning humans as merely ratifying pre-made decisions.

Yu et al. (2022) found that transparency simultaneously increased effectiveness perceptions and discomfort, suggesting transparency can have conflicting psychological effects. Yet whether these extend to autonomy remains unknown. Given mixed evidence and theoretical ambiguity, we test competing hypotheses:

- **H1a:** AI transparency will increase perceived autonomy (enhancement hypothesis).
- **H1b:** AI transparency will decrease perceived autonomy (paradox hypothesis).
- **H2:** Personality traits will moderate transparency-autonomy relationships.
- **H3:** Decision complexity will moderate these effects.

METHOD

Five hundred fifty-seven undergraduate students (54.8% male, 44.2% female, Mage = 21.62, SD = 3.52) from a large public university participated for course credit. They were mainly junior (58.2%) and senior (39.1%) business students, majoring in Finance (21.7%), Accounting (14.2%), Integrated Business (13.8%), and Marketing (8.1%). Most used AI tools daily (48.7%) or weekly (43.4%), and 55.3% viewed AI positively.

The study used a between-subjects design with random assignment to two conditions: Transparent (n = 278) or Basic (n = 279). Both received identical AI recommendations in decision scenarios but differed in explanation depth.

Participants evaluated ten decision scenarios in human resources, finance, strategy, and operations. Scenarios involved routine decisions with clear criteria (e.g., hiring, budgeting, supplier selection) or complex decisions with ambiguity and multiple stakeholders (e.g., executive hiring, capital allocation).

Each scenario included a business situation, candidate options, and an AI recommendation. In the Transparent condition, recommendations had detailed rationales with six to eight bullet points explaining reasoning. In the Basic condition, recommendations consisted of a brief statement.

Measures

Perceived Autonomy

After each scenario, participants rated their agreement with seven statements assessing decision ownership and control using 5-point Likert scales (1 = Strongly Disagree, 5 = Strongly Agree): “I would feel this decision is MY decision,” “I would have control over the final outcome,” “I would be personally responsible for the consequences,” “My professional judgment would be important,” “I would feel independent in making this choice,” “This decision reflects my values,” and “I have enough authority over this decision” ($\alpha = .92$).

The Ten-Item Personality Inventory (TIPI; Gosling et al., 2003) measured Big Five dimensions: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience. Additional items from validated Big Five scales supplemented the TIPI to enhance reliability.

Demographics and AI Experience

Participants reported age, gender, academic standing, major, AI usage frequency (Never to Daily), and general attitude toward AI (Negative to Positive).

Participants completed the study online via Qualtrics. After providing informed consent, they answered demographic questions and completed the personality inventory. Participants were then randomly assigned to condition and presented with the ten decision scenarios in randomized order. For each scenario, they read the situation, reviewed the AI recommendation (with or without detailed rationale depending on condition), and rated their perceived autonomy. Following all scenarios, participants completed post-experiment measures assessing their feelings about working with AI and their attitudes toward human-AI collaboration. The session lasted approximately 20–30 minutes. All procedures were approved by the university's Institutional Review Board.

RESULTS

The sample comprised 557 participants randomly assigned to Transparent (n = 278, detailed AI rationales) or Basic (n = 279, minimal explanation) conditions. Participants were predominantly junior (58.2%) and senior (39.1%) business students (Mage = 21.62, SD = 3.52; 54.8% male, 44.2% female) majoring in Finance (21.7%), Accounting (14.2%), Integrated Business (13.8%), and Marketing (8.1%). Most reported daily (48.7%) or weekly (43.4%) AI usage, with 55.3% holding positive AI attitudes.

Main Effect: The Transparency Paradox

Contrary to predictions, participants receiving detailed AI rationales reported lower perceived autonomy than those receiving basic recommendations. This pattern emerged across all six tested scenarios, reaching significance in two. Scenario 8 (Risk Assessment): Transparent M = 3.70 (SD = 0.73) vs. Basic M = 3.84 (SD = 0.66), t(555) = –2.43, p = .016, d = –0.21. Scenario 9 (Product Development): Transparent M = 3.72 (SD = 0.75) vs. Basic M = 3.85 (SD = 0.65), t(555) = –2.18, p = .030, d = –0.19. Scenario 1 (Hiring) showed a marginal trend: Transparent M = 3.73 (SD = 0.63) vs. Basic M = 3.82 (SD = 0.60), t(555) = –1.74, p = .082, d = –0.15.

Overall autonomy aggregated across scenarios showed Transparent M = 3.73 (SD = 0.59) vs. Basic M = 3.84 (SD = 0.55), d = –0.19, confirming a robust transparency paradox: detailed AI explanations reduced decision ownership, see Figure 1.

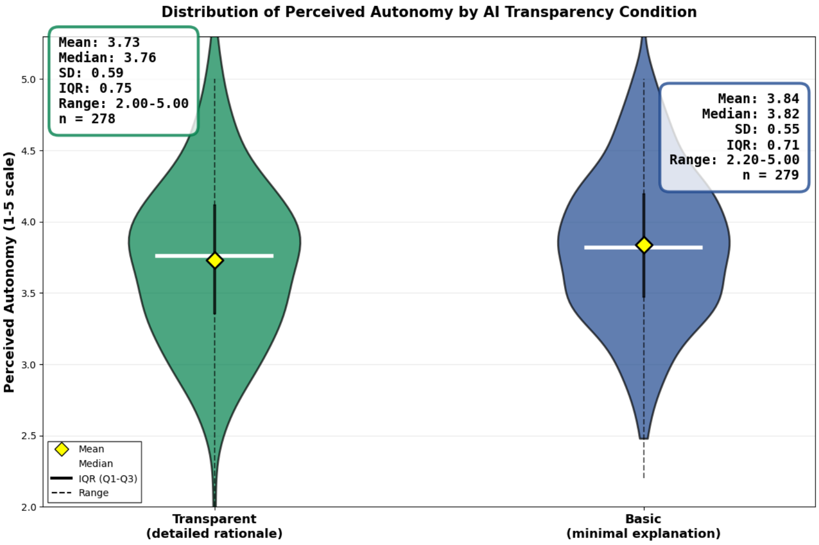

Figure 1: Distribution of perceived autonomy by AI transparency condition.

Personality Moderators

Extraversion (interaction = 0.173): High extraverts showed strong effects (Transparent M = 3.70, SD = 0.58, n = 105; Basic M = 3.91, SD = 0.55, n = 129; difference = 0.202), while low extraverts showed minimal response (Transparent M = 3.75, SD = 0.60, n = 173; Basic M = 3.77, SD = 0.55, n = 150; difference = 0.029), see Figure 2.

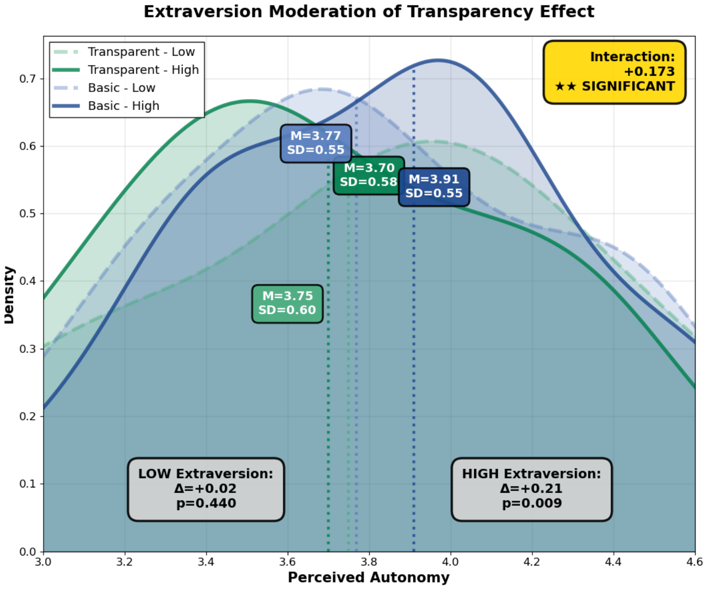

Figure 2: Personality moderation effects on the transparency paradox.

Openness (interaction = –0.227): High Openness reversed the paradox (Transparent M = 3.95, SD = 0.59, n = 81; Basic M = 3.89, SD = 0.58, n = 101; effect = –0.062), while low Openness showed the strongest negative effect (Transparent M = 3.64, SD = 0.57, n = 197; Basic M = 3.80, SD = 0.54, n = 178; effect = 0.166). Creative, intellectually curious individuals benefited from transparency; conventional thinkers found it constraining.

Emotional Stability (interaction = 0.120): Emotionally stable participants showed stronger effects (0.180) than those lower in stability (0.060). Agreeableness (0.087) and Conscientiousness (–0.062) showed minimal moderation.

Demographic Moderators

Gender: Males showed significant effects (Transparent M = 3.68, SD = 0.58, n = 165; Basic M = 3.82, SD = 0.54, n = 140; t(303) = –2.12, p = .035, effect = 0.137), while females showed smaller, non-significant differences (Transparent M = 3.79, SD = 0.60, n = 111; Basic M = 3.86, SD = 0.56, n = 135; t(244) = –0.84, p = .40, effect = 0.063). Males experienced twice the autonomy reduction.

Age: Effects were present in younger students but vanished by mid-twenties. Ages 18–20: effect = 0.124 (p = .087); ages 21–22: effect = 0.110 (p = .168); ages 23–25: effect = –0.009 (p = .952); ages 26+: effect = 0.142 (ns, n = 35). Decision-making experience appears to buffer against autonomy threats. Academic standing patterns mirrored age findings (juniors: effect = 0.133; seniors: effect = 0.083).

Non-Moderators

AI Experience: Daily users (effect = 0.110, n = 271), weekly users (effect = 0.117, n = 242), and infrequent users (effect = 0.086, n = 42) showed identical patterns. AI familiarity did not protect against the paradox. General AI attitudes also showed no moderation.

Decision Complexity: Routine scenarios (Transparent M = 3.745, SD = 0.599; Basic M = 3.850, SD = 0.556; t(555) = –2.16, p = .031, d = 0.183) and complex scenarios (Transparent M = 3.714, SD = 0.610; Basic M = 3.804, SD = 0.613; t(555) = –1.74, p = .082, d = 0.148) showed nearly identical effects (interaction = –0.015). The paradox operated consistently regardless of task difficulty, suggesting a fundamental psychological response rather than context-dependent reaction.

AI Acceptance Dissociation

Post-experiment measures revealed no condition differences in feelings about working with AI. Capability: Transparent M = 3.67 (SD = 0.91) vs. Basic M = 3.77 (SD = 0.92), t(335) = –1.01, p = .31. Confidence: Transparent M = 3.86 (SD = 0.87) vs. Basic M = 3.80 (SD = 1.00), t(340) = 0.66, p = .51. Autonomy with AI: Transparent M = 3.38 (SD = 0.99) vs. Basic M = 3.37 (SD = 1.16), t (365) = 0.14, p = .89. Responsibility: Transparent M = 3.45 (SD = 0.87) vs. Basic M = 3.54 (SD = 1.01), t(359) = –0.92, p = .36.

This dissociation indicates the paradox operates as a localized, in-the-moment phenomenon affecting decision ownership without altering broader AI attitudes or self-assessed competence. Users experienced autonomy loss from detailed explanations while still valuing AI as a confidence-enhancing tool, ruling out simple "backlash" explanations.

DISCUSSION

This study examined whether AI transparency's benefits extend to perceived autonomy or whether detailed explanations paradoxically undermine decision ownership. The findings reveal a robust transparency paradox: contrary to the assumption that transparency empowers users, detailed AI rationales consistently reduced perceived autonomy relative to basic recommendations (d = –0.19). This effect was substantially moderated by personality traits and demographics, suggesting the paradox operates differentially across user populations. Critically, reduced autonomy did not translate to decreased AI acceptance, indicating a dissociation between momentary decision control and general attitudes toward AI collaboration.

The Transparency Paradox: Mechanisms and Implications

The autonomy-reducing effect of detailed AI explanations challenges the prevailing "more transparency is better" paradigm in XAI research and regulation. While Sullivan and Weger (2025) demonstrated that transparency improves trust and understanding, and Yu et al. (2022) found that it increases effectiveness perceptions, our findings reveal a psychological cost: users feel less ownership over decisions when AI reasoning is comprehensively explained. This aligns with Ngo et al.'s (2025) identification of an inverted U-shape relationship and extends it to a new domain, perceived autonomy.

Three mechanisms likely explain this paradox. First, cognitive anchoring: detailed rationales create strong mental anchors that make deviating from AI recommendations psychologically difficult, even when users disagree (Bansal et al., 2020). When AI articulates six to eight compelling reasons supporting its recommendation, users may feel their own intuitions are inadequate by comparison. Second, authority positioning: comprehensive explanations implicitly communicate that the AI has completed the analytical work, positioning humans as ratifiers rather than decision-makers. The AI becomes the expert, and the human becomes the executor. Third, perceived task complexity: detailed explanations may increase cognitive load, making users feel the decision exceeds their independent judgment capacity (Westphal et al., 2023). Rather than empowering users with information, transparency can overwhelm them, reducing confidence in their autonomous decision-making ability.

Importantly, the paradox operated consistently across both routine and complex decisions (interaction = –0.015), undermining predictions that transparency would particularly benefit ambiguous scenarios. Whether selecting office suppliers or navigating strategic planning, detailed AI explanations similarly diminished autonomy. This suggests the phenomenon reflects fundamental psychological responses to authoritative information presentation rather than task-specific reactions, making it more pervasive and concerning than anticipated.

Individual Differences

The most theoretically significant finding is that personality fundamentally alters transparency's effects. Openness to Experience produced a complete reversal: intellectually curious, creative individuals were the only group to experience increased autonomy with detailed rationales (effect = −0.062 for high Openness vs. +0.166 for low Openness, interaction = −0.227). This suggests open-minded users integrate AI reasoning as cognitive scaffolding that enhances rather than supplants their judgment. They appear to engage with explanations analytically, extracting useful information while maintaining decision ownership. Conversely, conventional thinkers experience detailed rationales as overwhelming or constraining, perhaps lacking the cognitive flexibility to critically evaluate complex AI reasoning without feeling dominated by it.

Extraversion showed the opposite pattern: action-oriented, socially confident individuals experienced the strongest autonomy reduction (effect = 0.202 vs. 0.029 for introverts, interaction = 0.173). Extraverts may prefer rapid, intuitive decision-making and find lengthy explanations incompatible with their decision-making style. Detailed rationales force deliberative processing that conflicts with their natural approach, creating frustration and diminished ownership. This aligns with Buçinca et al.'s (2021) finding that cognitive forcing interventions benefit high-Need-for-Cognition individuals more, extending it to trait-level personality differences.

Gender and age effects reveal important developmental and cultural dimensions. Males experienced twice the autonomy reduction of females (0.137 vs. 0.063), possibly reflecting documented gender differences in decision-making styles or social conditioning around collaborative versus independent judgment. The complete disappearance of effects by ages 23–25 suggests that decision-making experience or cognitive maturity buffers against autonomy threats. Older students may possess sufficient expertise to integrate AI rationales without feeling supplanted, whereas younger students lack the decisional confidence to maintain ownership when confronted with comprehensive AI reasoning.

Critically, neither AI familiarity nor general attitudes moderated effects. Even daily AI users experienced autonomy reduction, indicating mere exposure does not inoculate against the paradox. This contradicts assumptions that transparency concerns primarily affect AI novices and suggests the phenomenon operates at a deeper psychological level than simple unfamiliarity.

The Autonomy-Acceptance Dissociation

Perhaps most practically significant is the finding that reduced autonomy did not translate into decreased AI acceptance, confidence, or perceptions of capability. Participants who felt less autonomous during specific decisions nevertheless maintained positive views about working with AI generally (all ps > .31). This dissociation has important implications: it suggests transparency reduces *momentary* decision ownership without damaging *long-term* AI relationship quality. Organizations concerned about employee autonomy can potentially address this through design choices, such as allowing users to hide/show explanations (Vasconcelos et al., 2022) or providing briefer summaries, without fundamentally undermining AI adoption or trust.

However, this dissociation should not be interpreted as evidence that autonomy concerns are trivial. Accumulated experiences of reduced autonomy across many decisions may eventually erode job satisfaction, intrinsic motivation, and professional identity even if general AI attitudes remain positive. The lack of immediate backlash does not preclude longer-term psychological costs. Moreover, perceived autonomy is intrinsically valuable and organizationally consequential regardless of whether it correlates with AI acceptance. Employees deserve to feel genuine ownership over their decisions as a matter of dignity and professional respect, independent of productivity considerations.

Theoretical Contributions

This research makes three key contributions. First, it extends the transparency paradox beyond trust and accuracy to include autonomy, showing that the psychological costs are broader than previously known. Second, it identifies personality traits, especially Openness and Extraversion, as key moderators, revealing that "one-size-fits-all" transparency favors some users but harms others. Third, it shows that task complexity doesn't affect transparency effects, indicating the paradox stems from general psychological mechanisms rather than specific contexts. These findings challenge current regulations mandating transparency. If detailed explanations reduce autonomy for certain groups (extraverts, traditional thinkers, males, younger workers), blanket transparency may harm their well-being while benefiting a minority (open, curious individuals). Regulations should promote adaptive transparency, tailoring explanations to user traits and preferences instead of uniform rules disclosure.

Practical Implications

For AI designers, findings suggest offering user control over explanation depth, with summaries as defaults (Vasconcelos et al., 2022). Personality-adaptive interfaces could provide comprehensive rationales to open users while offering action-focused summaries to extraverts. Framing AI as collaborative rather than authoritative, using language positioning explanations as "considerations" not "conclusions", preserves agency. Organizations should train employees to critically evaluate AI rationales as hypotheses rather than expert opinions and explicitly affirm human decision authority. For policymakers, findings challenge blanket transparency mandates. While transparency serves accountability, detailed explanations may psychologically undermine the human oversight they aim to support. Regulation should permit adaptive transparency, balancing accountability with the preservation of autonomy.

Limitations and Future Research

Several limitations warrant consideration. The student sample may not generalize to experienced professionals, though age findings suggest experience buffers autonomy threats. The study examined perceived autonomy rather than objective decision quality; whether transparency reduces both or creates a tradeoff remains unknown. The scenario methodology cannot capture the long-term consequences of repeated autonomy reduction. Future research should test personality-adaptive transparency, training interventions for critical AI engagement, and longitudinal effects on job satisfaction. Research should also explore optimal explanation lengths and examine actual organizational implementations where stakes and expertise differ substantially.

CONCLUSION

This research demonstrates that AI transparency's benefits come with psychological costs: detailed explanations, while potentially improving understanding, systematically reduce perceived autonomy. This transparency paradox is robust across decision types but moderated by personality and demographics, suggesting transparency's effects are more complex and individualized than current theory and regulation acknowledge. The finding that intellectually curious individuals benefit from transparency while conventional thinkers suffer challenges, universal transparency mandates, and suggests personalized approaches may better serve diverse user populations. As AI systems assume greater roles in organizational decision-making, understanding and addressing transparency's psychological consequences becomes essential for designing systems that genuinely empower rather than inadvertently undermine human judgment.

REFERENCES

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in AI-assisted decision-making. *Machine Learning and Knowledge Extraction, 4*(2), 556–579. https://doi.org/10.3390/make4020026

Bansal, G., Wu, T. S., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2020). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3313831.3376283

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamó-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society, 6*(1), 1–14. https://doi.org/10.1177/2053951719860542

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *arXiv preprint arXiv:1902.01876*. https://arxiv.org/abs/1902.01876

Ngo, V. M., Nguyen, H. V., Tran, N. P., Nguyen, H. H., & Nguyen, P. V. (2025). The AI transparency dilemma: When more is less for trust and adoption. *Technology in Society, 80*, 102345. https://doi.org/10.1016/j.techsoc.2024.102345

Schmidt, P., Biessmann, F., & Teubner, T. (2020). Transparency and trust in artificial intelligence systems. *Journal of Decision Systems, 29*(4), 260–278. https://doi.org/10.1080/12460125.2020.1819094

Shin, D., Zhong, B., & Biocca, F. A. (2021). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management, 62*, 102429. https://doi.org/10.1016/j.ijinfomgt.2021.102429

Sullivan, V., & Weger, K. (2025). Transparency and explainability in AI-assisted decision making: Effects on trust, perceived reliability, confidence, and ease of understanding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 69*(1), 1–5. https://doi.org/10.1177/10711813251234567

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2022). Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction, 7* (CSCW1), 1-38. https://doi.org/10.1145/3579605

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300831

Wang, X., Yin, M., & Chen, H. (2021). Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression. *Nature Communications, 12*, 3219. https://doi.org/10.1038/s41467-021-23479-0

Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior, 144*, 107714. https://doi.org/10.1016/j.chb.2023.107714

Yu, L., Nickerson, J. V., & Sakamoto, Y. (2022). The bright and dark sides of AI transparency on employee trust and discomfort: A parallel mediation model. *Proceedings of the 55th Hawaii International Conference on System Sciences*, 1473–1482. https://doi.org/10.24251/HICSS.2022.182

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. https://doi.org/10.1145/3351095.3372852