Trust Me, I'm Human! Comparing Attitudes Towards AI versus Human Junior Colleagues in Medical Device Design

Yi Ning Chuah
Faculty of Engineering
University of Nottingham
Nottingham
UK
chuahying@hotmail.com

Glyn Lawson
School or Art and Design
Nottingham Trent University
Nottingham
UK
glyn.lawson@ntu.ac.uk

Regulatory demands for transparency and accountability conflict with opaque artificial intelligence (AI) in the medical device design industry. A mixed-methods design was used to study how trust shapes human-AI roles, where interviews with design professionals (N=7) highlighted the importance of safety-criticality in this context. This informed a 2x2 within-subjects experiment (N=24) that manipulated information source (AI, junior colleague) and task safety-criticality (high, low). Crucially, participants were presented with similar reports but were told they were created by either a junior colleague or an AI. Significantly higher trust ratings were given for reports purportedly written by colleagues than for AI. An interaction was found for perceived success, with AI rated significantly higher in non-safety-critical tasks, but not in safety-critical tasks. The results suggest that in regulated industries, where trust depends on accountability rather than just ability, efficiency is prioritised until safety is jeopardised.

Human-AI collaboration. Medical Device Design. Trust.

1. INTRODUCTION

Medical device design must consider user needs, efficacy, and safety enforced by the industry's strict regulations and complexity (Alattal et al., 2025; Money et al., 2011). Historically, a human-centered approach involving interdisciplinary teams of clinicians, engineers, and product managers has been adopted in the medical device design process (Money et al., 2011). However, this kind of safety-critical industry usually faces challenges that hinder innovation, such as heavy regulatory burden and procedural complexity (Alexander and Clarkson, 2000; Ogrodnik, 2012). This is because regulations are designed to ensure safety and effectiveness, but the resulting approval procedures often require extensive documentation, clinical testing, and adaptation to frequently changing standards. In recent years, artificial intelligence (AI), especially generative AI, is increasingly integrated into the design process, from ideation and creating 3D models to gathering information and testing new designs (Liu, 2025; Schweitzer et al., 2021; Jesso et al., 2022). It enables better design quality, supports the introduction of new ideas, and helps get products to the market faster (Onitiu et al., 2024). However, risks such as bias, opacity, and accuracy errors can be problematic in healthcare (Kostick-Quenet et al., 2024). Therefore, it is important to develop ethical and responsible AI (Matulionyte et al., 2022).

Workload and trust are also important factors when incorporating AI into design processes (Hallowell et al., 2022; Sheng, 2024; Shevtsova et al., 2024). Effective collaboration requires not only that humans believe the AI works correctly but also that they believe it supports human goals (Hatherley, 2020). In the high-stakes medical context, forming stable trust is important (Kostick-Quenet et al., 2024). Human workload can also be affected by AI: AI can reduce cognitive load through task automation, but humans will need to learn new interfaces and interpret AI outputs, thus increasing workload (Rane et al., 2023). Even though more people are using AI, there is little research about how humans and AI collaborate in designing medical devices (Bach et al., 2024). Rather than providing a comprehensive understanding of human and AI roles throughout the design process, existing studies tend to focus on specific AI tools or limited aspects of interaction (Liu, 2025). There is insufficient research to fully understand human-AI roles, workload, and trust in the medical device design process.

© Chuah et al.,
Published by BCS Learning and Development Ltd.
Proceedings of BSC HCI 2026, UK

In addition to trust and workload, collaborative design outcomes are often informally evaluated in terms of perceived success, which reflects subjective goal attainment and efficiency and may diverge from accountability-based trust judgements, particularly in safety-critical contexts (Hart and Staveland, 1988; Lee and See, 2004; Spatola, 2024; Kostick-Quenet et al., 2024).

This study aimed to answer the research question: How should the medical device design process allocate the roles of humans and AI, and how are trust and workload related to this collaboration? By doing so, it makes two contributions to HCI research on human–AI collaboration in safety-critical design contexts. First, it demonstrates a source attribution effect - identical design summaries are trusted differently when attributed to a junior colleague versus an AI. Second, by examining how task safety-criticality shapes perceived success, mental demand, and trust, the study shows that trust calibration in regulated design work is not driven solely by perceived competence, but by socio-technical factors such as verifiability, responsibility, and defensibility.

2. RELATED WORK

Medical device design is structured and regulated, requiring every step to be transparent, traceable, and reviewed carefully (Matulionyte et al., 2022; Medina et al., 2013; Money et al., 2011). To make compliance and traceability easier, comprehensive documentation needs to be maintained. These requirements directly conflict with the opacity of many AI models (Uribe Ocampo and Kaminski, 2019). Thus, designers are legally responsible for AI outputs they may not fully understand, placing them in the position of having to justify decisions to regulators despite not fully understanding how these outputs were generated.

Trust is defined as a user's belief that technology will assist them in reaching their goals under uncertainty (Kostick-Quenet et al., 2024). When this belief is used to assess the system's actual reliability, trust calibration occurs (Lee and See, 2024). While technology acceptance models emphasize dependability, functionality, and non-opportunistic behavior (Gefen et al., 2003), trust between humans is often related to perceived capability, kindness, and honesty (Mayer and Davis, 1995). These are usually not associated with AI due to its black box nature (Zerilli et al., 2022). Moreover, research on trust in human–AI interaction has consistently highlighted that trust is not a binary state, but a dynamic judgement shaped by context, perceived consequences, and uncertainty (Lee and See, 2004; Hoff and Bashir, 2015). In safety-critical and regulated domains, this judgement is further constrained by accountability requirements, legal responsibility, and the need to justify decisions retrospectively (Durán and Jongsma, 2021; Kostick-Quenet et al., 2024). Systematic reviews of human–AI interaction in safety-critical industries indicate that opacity and limited traceability are persistent barriers to trust, even when AI performance is high (Bach et al., 2024). These findings suggest that in regulated contexts, trust calibration depends less on perceived capability alone and more on whether human operators can defend, verify, and take responsibility for AI-supported decisions. These distinctions become particularly salient in safety-critical and regulated domains, where trust judgements are constrained by legal responsibility and ethical obligation, not merely task success.

Recent work argues that opacity and limited explainability undermine trust because AI-supported decisions are difficult to justify and defend to others (Durán and Jongsma, 2021; Kostick-Quenet et al., 2024). In such settings, designers and clinicians remain responsible for outcomes even when decisions are influenced by AI, creating an accountability gap between system contribution and human liability. Systematic reviews of human–AI interaction in safety-critical industries confirm that this accountability gap persists even when AI performance is demonstrably high (Bach et al., 2024).

Another important HCI consideration is Workload, the subjective demands necessary to complete a task (Hart and Staveland, 1988), which AI has the potential to either increase or decrease. While humans can save time by using AI for repetitive tasks (Lucas et al., 2024; Tiwari and Patel, 2024; Schweitzer et al., 2021), workload often increases because humans must learn how to use new interfaces and interpret unclear outputs (Mosier et al., 1996; Luo, 2025; Trikoili et al., 2025). This leads to the burden of verification, where manual checking is necessary for outputs that are opaque or untrustworthy, adding tasks rather than removing them (Matulionyte et al., 2022). Furthermore, challenges like automation bias and the tendency to over-rely on automation outputs without critical assessment, can cause serious decision-making errors in human-AI collaboration (Holbrook et al., 2024).

Despite growing recognition of these issues, empirical work on human–AI collaboration in regulated design contexts remains limited. Existing studies tend to focus on evaluating specific AI tools or interfaces, rather than examining how designers reason about trust, responsibility, and workload when interpreting AI-generated contributions. Moreover, few studies explicitly compare AI outputs with human collaborators under controlled conditions that hold informational content constant. This gap motivates the present study, which

examines how source attribution and task safety-criticality jointly shape trust judgements in medical device design - a domain where accountability is not peripheral, but foundational.

3. METHODOLOGY

A two-phase, mixed-method design was employed. In phase one, professionals were interviewed to obtain a contextually rich understanding of current practices, attitudes, and concerns. In phase two, a controlled experiment was used to quantitatively test the framework developed from phase one.

3.1 Phase One: Interviews

Phase one was designed as an exploratory elicitation study, using interviews to probe current AI usage, perceived risks, and responsibility for AI-supported decisions, with the specific aim of informing the design of a controlled experimental manipulation in phase two. The qualitative data were therefore used to sensitise variables and construct experimental contrasts.

The only inclusion criterion was that the participants were professionals with experience in the medical device design process. In total, seven professionals were recruited, ranging from in-house designers and consultants to academics. Microsoft Teams was used to carry out remote semi-structured interviews, each lasting approximately 30 minutes. Thematic analysis was conducted on all transcripts using the six-phase method of Braun and Clarke (2006).

3.2 Phase Two: Experiment

A 2×2 within-subjects experimental design was employed to control for individual differences. The independent variables were information source (AI vs. junior colleague) and task safety-criticality (high vs. low). Summary reports were therefore constructed to reflect realistic medical device design artefacts that varied in safety-criticality while holding informational content constant across sources. Non-safety-critical summaries described formative usability and user-needs findings for consumer health devices (a body composition scale and a home fitness monitor), whereas safety-critical summaries reported use-related risk analyses and post-market surveillance findings for regulated medical devices (a brain stimulation system and an infusion pump). For each task scenario, two versions of the same summary were created: one was labelled as produced by an "AI" and one was labelled as produced by a "junior colleague". Each participant saw only one version of each scenario, while source labels were counterbalanced across participants using a Latin square design. This ensured that content was comparable across source conditions at the group level, while avoiding repeated exposure to the same material within participants. Presenting identical content under different source labels allowed trust judgements to be examined independently of actual information quality or performance. This approach is consistent with prior trust-in-automation research, which emphasises the role of framing, authority cues, and perceived agency in shaping reliance decisions (Jian et al., 2000; Lee and See, 2004).

Task safety-criticality also had two levels: high safety-criticality and low safety-criticality. This was manipulated by the description of the task: two scenarios represented high safety-critical tasks (brain stimulation device, infusion pump) and two represented low safety-critical tasks (body composition scale, home fitness monitor). To mitigate order effects, a Latin Square design was used to counterbalance the four conditions. The study intentionally operationalised AI as an unspecified system to capture participants' baseline assumptions and default trust heuristics, rather than reactions to a particular tool with known explainability or audit features.

Participants were 24 university students with relevant academic backgrounds (for example, human factors, product design) and an understanding of English. Participants were between 22 to 42 years old (M = 27.43, SD = 6.00).

Participants attended one-on-one remote experiment sessions where they were instructed to review four short summary reports related to medical devices (*high-safety-criticality/AI; high-safety-criticality/junior colleague; low-safety-criticality/AI; low-safety-criticality/junior colleague*). Post-condition questionnaires were completed immediately after reading each summary to measure subjective trust, perceived information quality, workload, and conclusion support. Subjective trust was measured as the mean score on the adapted five-item Trust in Automated Systems (TIAS) scale (Jian et al., 2000), measured on a seven-point Likert scale, where a higher mean score indicated a higher subjective trust on the source. Perceived information quality was measured as the mean score on the four-item Information Quality (InfoQual) subscale of the Computer System Usability Questionnaire (CSUQ), measured on a seven-point Likert scale, where a higher mean score indicated a higher perceived information quality (Lewis, 1995). Workload was measured using the separate scores on the six NASA Task Load Index (NASA-TLX) subscales (Hart and Staveland, 1988). For mental demand, physical demand, temporal demand, effort, and frustration subscales, a higher score indicated a greater workload. For the success subscale, a higher score indicated a greater success feeling in accomplishing the task's goals. Conclusion support was defined as the score on a seven-point Likert scale for a single-item that asked

how likely they were to support the summary's recommendations. To help explain the quantitative findings, a brief, semi-structured interview was conducted to gather qualitative data at the end of the experiment.

The authors used AI to assist in constructing the short summary reports used as prompts in Phase Two. AI-generated content was reviewed, verified and revised by the authors where necessary to meet the study objectives. The authors take full responsibility for the final content.

4. RESULTS

4.1. Phase one

Thematic analysis of the semi-structured interviews with seven professionals with medical device design experience revealed three major themes on the potential integration of AI in the field.

4.1.1. Theme one: junior colleague metaphor
Participants compared AI to a junior colleague who knows the basics but does not have enough experience. P1 explained, "…wouldn't necessarily trust them until they have got more experience…". In a similar theme, participants reported AI is not always dependable, and they must therefore carefully check each result. P2 noted that AI can hallucinate answers without admitting uncertainty: "It's hard to detect when AI is just filling in the gaps because it doesn't really admit when it's not sure". In a field governed by exhaustive documentation, the inability to ask AI to explain itself or be held responsible is seen as a fatal flaw. As P7 warned, providing only "the LLM said so" as evidence of a design decision is unacceptable.

4.1.2. Theme two: calibrated trust
Participants calibrate their trust in AI depending on how important the task is, how easy it is to check the results, and what could go wrong if mistakes are made. Professionals trust AI more for simple and everyday tasks like email or summarising, since there is a low chance of mistakes and the results are easy to check. However, the calibration needs to be much more accurate as the importance of the situation grows. In the high-pressure regulatory environment of medical device design, where professionals must "prove that when the... concept is being used by representative users... there's no serious harm and that all risks are mitigated" (P2), trust is granted with caution. Accountability becomes a problem. A higher risk is introduced when an AI output is difficult to validate or cannot be easily linked to a trustworthy source. As P3 put it, "imagine that I'm using an AI tool... And it makes a mistake... Whose responsibility is it? Is it both? Is it just mine that I didn't check?".

4.1.3. Theme three: possible applications
Despite the lack of trust, it was generally agreed that AI's current application is enhancing the inputs and outputs of the design process. AI is used to accelerate tasks that P7 termed "administrative grunt work". Among the specific uses mentioned were ideation and brainstorming, rapid domain learning to extract the gist of lengthy scholarly articles, and the production of preliminary documentation drafts. Participants also thought that the creation of specialised AI tools and empirical validation could foster trust.

4.2 Phase Two

A reliability analysis conducted on the five-item trust scale showed high internal consistency (Cronbach's α=0.95). As there were less than 50 participants in this study, Shapiro-Wilk test was used to ensure normality. The trust variable satisfied the normality assumption, but other DVs did not. Therefore, trust data were analysed using a 2x2 repeated measures ANOVA, while other DVs were analysed using an ART ANOVA.

There was a significant main effect for information source on trust ($F(1,23)=12.44$, $p=.002$, *partial* $\eta^2=.351$). Participants reported significantly higher trust in summaries attributed to a junior colleague than those attributed to AI. Similarly, participants were significantly more likely to support the conclusions from a junior colleague than from an AI ($F(1,21)=4.98$, $p=.029$). The main effect of task safety-criticality on trust was not significant, and there was no significant interaction effect on trust.

Participants rated perceived information quality significantly lower for high safety-critical tasks compared to low safety-critical tasks ($F(1,23)=8.20$, $p=.006$). Furthermore, high safety-critical tasks were perceived as requiring significantly more mental demand ($F(1,23)=7.64$, $p=.007$). There was no significant main effect of information sources on perceived quality.

There was a significant interaction effect between information source and task safety-criticality on perceived success ($F(1,23)=5.31$, $p=.024$). Specifically, in low safety-critical tasks, participants felt significantly more successful using AI than a junior colleague ($p = .017$). No significant effect was found for task safety-criticality regarding conclusion support.

Participants were significantly more likely to support conclusions from a junior colleague than from an AI ($F (1, 21)=4.98$, $p=0.029$).

Thematic analysis of post-experiment interviews provided context for these results. Significant lower trust for AI can be explained by the black box dilemma and trust as a multi-faceted calibration. P11 stated, "AI tends to lie quite badly", fearing AI hallucinations. AI does not have social context for

trust, unlike a human colleague, where most participants assumed that the colleague's competence is "vouched for" by the company. The interview also suggested that when the tasks are high in safety-criticality, participants reported "thinking harder" due to the ethical and regulatory demands, explaining the increased mental demand in high safety-critical tasks.

5. DISCUSSION

The experiment revealed that participants rated summaries attributed to a junior colleague significantly higher in subjective trust than identical summaries attributed to AI, supporting phase one's junior colleague metaphor. While task safety-criticality did not significantly affect trust, it significantly increased mental demand and decreased perceived information quality. Furthermore, a significant interaction effect was found regarding perceived success, where in low safety-critical tasks, participants felt significantly more successful with AI.

5.1 Findings interpretation

Post-experiment interviews suggest participants relied on social context and organisational affiliation, not just information content. Specifically, P3 assumed the company "vouches for" the junior colleague's competence, creating cues of accountability and goodwill that the AI did not convey. This suggests that the significantly lower trust in AI is not necessarily a perception of inability, but a reflection of its inability to be an accountable partner in a regulated, safety-critical domain (Gefen et al., 2003). This aligns with prior work on algorithm aversion (Dietvorst et al., 2015), where humans prefer human advice over algorithmic advice even when both are equally accurate. Although participants characterised junior colleagues as inexperienced, they nonetheless attributed them with organizational embedding and accountability, making them more trustworthy than an unaccountable AI source.

The significantly higher mental demand in safety-critical tasks suggests participants switched from automatic thinking to more active and careful thinking (Kahneman, 2012). This suggests a recalibration of internal standards. Specifically, participants explained during post-experiment interviews that they perceived the short summaries were high quality for comprehension in low safety-critical tasks, but it was perceived as lower quality for high safety-critical tasks as it lacked depth for a defensible decision. This shift is consistent with evaluation criteria described by Trikoili et al. (2025).

The significant interaction effect on perceived success might be explained by the trade-off between efficiency and accountability. Participants explained in the post-experiment interviews that they perceived success in low safety-critical tasks as related to efficiency, speed, and clear presentation (AI's strengths), whereas success in high safety-critical tasks was perceived as related to accountability, verifiability, and integrity.

5.2 Interpretive framework of trust reasoning in regulated human–AI collaboration

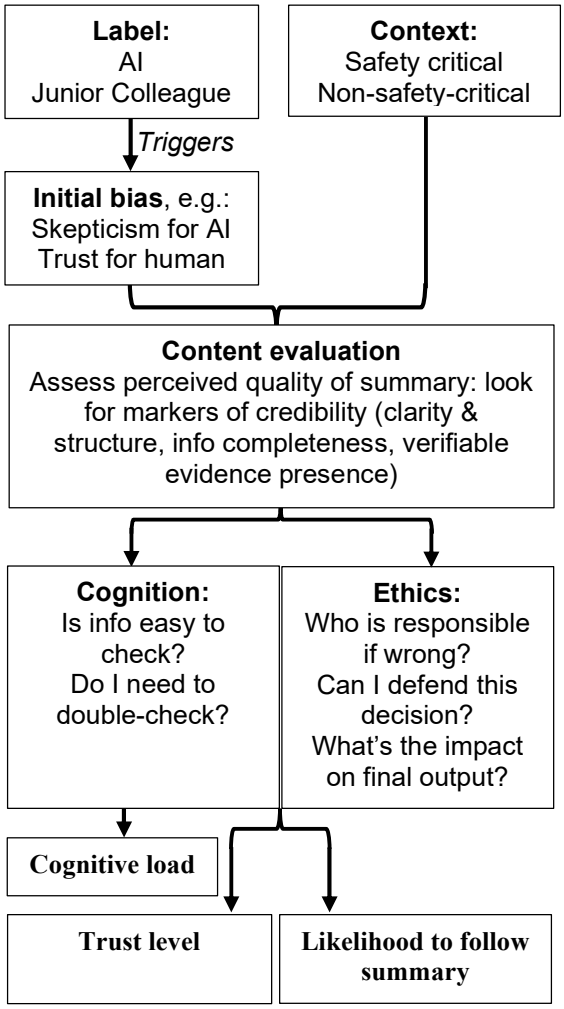

Figure 1: *Interpretive framework for trust reasoning in medical device design.*

Figure 1 represents an interpretive synthesis of how participants articulated trust judgements when evaluating design summaries attributed to either a junior colleague or an AI under varying levels of safety-criticality. For clarity, this should not be read as a cognitive process model or a causal account of internal decision-making, but as an illustration of recurring dimensions of reasoning - such as source-based heuristics, perceived accountability, verifiability, and anticipated consequences - that participants drew upon when judging whether reliance on a given source could be justified if challenged. The framework distinguishes efficiency-oriented reasoning from

accountability-oriented reasoning to explain why AI contributions were perceived as more successful in low-risk contexts yet consistently less trustworthy in high safety-critical ones. The framework draws on interview findings, post-experiment explanations, and existing trust in automation literature. Its purpose is to make explicit the interpretive logics that participants drew upon - often implicitly - when deciding whether an output could be relied upon, defended, or justified to others in a regulated design context.

By framing trust as a form of justificatory reasoning rather than as confidence in technical capability, the framework highlights how source attribution acts as a cue for accountability, responsibility, and institutional embedding. In doing so, it offers a conceptual lens for understanding why AI contributions may be perceived as efficient yet difficult to trust in high safety-critical contexts, even when informational content is held constant. These findings contribute to HCI by highlighting how perceptions of accountability shape human responses to AI collaborators, even when informational content is held constant.

5.3 Limitations and Future Research

This study has several limitations that should be considered when interpreting the findings. First, while practising medical device professionals informed the exploratory qualitative phase, the experimental phase relied on graduate-level participants with design-relevant training rather than professionals operating under formal regulatory responsibility. This sampling decision reflects the practical difficulty of recruiting regulated-industry professionals for controlled experimental studies, particularly where experimental abstraction is required. As a result, the findings should be understood as capturing interpretive trust heuristics rather than reflecting workplace decision-making under direct regulatory accountability (Jian et al., 2000). Future work should seek to validate and extend these findings through field studies involving practising medical device designers embedded within organisational and regulatory workflows.

Second, the experimental task abstracted the medical device design process into short summary evaluations, rather than embedding AI within a full design workflow. While this limits ecological fidelity, it enabled tight experimental control by holding informational content constant and isolating the effect of source attribution. Such abstraction is necessary to disentangle trust in outputs from trust in sources, but future studies could examine how these dynamics evolve over longer-term interaction and iterative use.

Finally, AI was deliberately defined in abstract terms to capture participants' baseline perceptions of AI as a general category of collaborator, rather than responses to a specific system. This choice precludes conclusions about particular AI tools or explainability features. Future work should investigate whether explainable AI, provenance tracking, or accountability scaffolding reduces the trust gap observed here, particularly in high safety-critical tasks where current AI systems are viewed as difficult to defend. This work could also address the lack of transparency in AI, which makes it harder to trust (Durán and Jongsma, 2021).

6. CONCLUSION

This mixed-methods study examined how trust in human versus AI collaborators is shaped by accountability, verifiability, and task safety-criticality in the context of medical device design. Results suggest that successful human-AI collaboration depends on socio-technical factors such as accountability, verifiability, and perceived responsibility, rather than on assumptions of raw competence alone. In high safety-critical contexts (errors cause harm or regulatory consequences), perceived accountability and verifiability might explain the significantly higher trust in junior colleague than AI. Furthermore, task safety-criticality shifts cognition. High safety-critical tasks were found to increase effortful, critical evaluation and mental demand. A recalibration of standards is suggested in the increase in mental demand and decrease in perceived information quality, which shift from rapid evaluation in low safety-critical tasks to thorough evaluation in high safety-critical tasks. When combined, these results suggest a context-dependent allocation model, which calls for human oversight for accountability in high-criticality, convergent phases and AI for efficiency in low-criticality, divergent phases, as long as AI are built with transparency, verifiability, and support for responsible human oversight in mind. These interpretations do not claim to measure cognitive processes directly but offer a theoretically grounded account of how participants articulated and rationalised trust decisions in regulated contexts. For HCI researchers and practitioners, the findings highlight the importance of designing human-AI systems that support accountability and trust calibration rather than focusing solely on performance improvements.

DATA AVAILABILITY STATEMENT

Due to ethical restrictions regarding participant confidentiality, the data underlying this publication are not openly available. A metadata record of the dataset has been created with NTU Open Research Archive (NORA), detailing data collected.

REFERENCES

Alattal, D., Khoshravan Azar, A., Myles, P., Branson, R., Abdulhussein, H. and Tucker, A. (2025) Integrating Explainable AI in medical devices: technical, clinical and regulatory insights and recommendations. arXiv preprint arXiv:2505.06620.

Alexander, K. and Clarkson, P.J. (2000) Good design practice for medical devices and equipment, Part II: design for validation. Journal of Medical Engineering & Technology, 24(2), 53–62.

Bach, T.A., Kristiansen, J.K., Babic, A. and Jacovi, A. (2024) Unpacking human–AI interaction in safety-critical industries: a systematic literature review. IEEE Access, 12, 106385–106414.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: Gen*eral*, 144(1), 114–126

Durán, J.M. and Jongsma, K.R. (2021) Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. Journal of Medical Ethics, 47(5), 329-335.

Gefen, D., Karahanna, E. and Straub, D.W. (2003) Trust and TAM in online shopping: an integrated model. MIS Quarterly, 27(1), 51–90.

Hallowell, N., Badger, S., Sauerbrei, A., Nellåker, C. and Kerasidou, A. (2022) "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms in the diagnosis of rare disease. BMC Medical Ethics, 23(1), 112.

Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology. Elsevier, Amsterdam, pp. 139–183.

Hatherley, J.J. (2020) Limits of trust in medical AI. Journal of Medical Ethics, 46(7), 478–481.

Holbrook, C., Holman, D., Clingo, J. and Wagner, A.R. (2024) Overtrust in AI recommendations about whether or not to kill: evidence from two human–robot interaction studies. Scientific Reports, 14(1), 19751.

Hoff, K.A. and Bashir, M. (2015) Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407–434.

Jesso, S.T., Kelliher, A., Sanghavi, H., Martin, T. and Parker, S.H. (2022) Inclusion of clinicians in the development and evaluation of clinical artificial intelligence tools: a systematic literature review. Frontiers in Psychology, 13, 830345

Jian, J.Y., Bisantz, A.M. and Drury, C.G. (2000) Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71.

Kahneman, D. (2012) Thinking, fast and slow. Penguin Books, London.

Kostick-Quenet, K., Lang, B.H., Smith, J., Hurley, M. and Blumenthal-Barby, J. (2024) Trust criteria for artificial intelligence in health: normative and epistemic considerations. Journal of Medical Ethics, 50(8), 544–551.

Lee, J.D. and See, K.A. (2004) Trust in automation: designing for appropriate reliance. Human Factors, 46(1), 50–80.

Lewis, J.R. (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human–Computer Interaction, 7(1), 57–78.

Liu, Z., 2025. Human-AI co-creation: a framework for collaborative design in intelligent systems. arXiv preprint arXiv:2507.17774.

Lucas, G.M., Becerik-Gerber, B. and Roll, S.C. (2024) Calibrating workers' trust in intelligent automated systems. Patterns, 5(9), 101045.

Luo, Y. (2025) Designing with AI: a systematic literature review on the use, development, and perception of AI-enabled UX design tools. Advances in Human-Computer Interaction, 2025, Article 3869207.

Matulionyte, R., Nolan, P., Beheshti, A. and Magrabi, F. (2022) Should AI-enabled medical devices be explainable? International Journal of Law and Information Technology, 30, 151–180.

Mayer, R.C. and Davis, J.H. (1995) An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.

Medina, L.A., Okudan Kremer, G.E. and Wysk, R.A. (2013) Supporting medical device development: a standard product design process model. Journal of Engineering Design, 24(2), 83–119.

Money, A.G., Barnett, J., Kuljis, J., Craven, M.P., Martin, J.L. and Young, T. (2011) The role of the user within the medical device design and development process: medical device manufacturers' perspectives. BMC Medical Informatics and Decision Making, 11(1), 15.

Money, A.G., Barnett, J., Kuljis, J., Craven, M.P., Martin, J.L. and Young, T. (2011) The role of the user within the medical device design and development process: medical device

manufacturers' perspectives. BMC Medical Informatics and Decision Making, 11(1), 15.

Mosier, K.L., Skitka, L.J. and Heers, S.T. (1996) Automation bias: decision making and performance in high-risk environments. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 40, 204–208.

Ogrodnik, P. (2012) Medical device design: innovation from concept to market. Elsevier, San Diego.

Onitiu, D., Wachter, S. and Mittelstadt, B. (2024) How AI challenges the medical device regulation: patient safety, benefits, and intended uses. Journal of Law and the Biosciences, 11(1), lsae007.

Rane, N., Choudhary, S. and Rane, J. (2023) Enhanced product design and development using artificial intelligence, virtual reality, augmented reality, additive manufacturing, IoT and blockchain: a review. SSRN Working Paper.

Schweitzer, G.M., Bitzer, M. and Vielhaber, M. (2021) Artificial intelligence in engineering: evolution of virtual product development in the context of medical device industry. Procedia CIRP, 100, 349–354.

Sheng, Y. (2024) c. Advances in Creative Engineering, 34(1), 23–30.

Shevtsova, D., Ahmed, A., Boot, I.W.A., Sanges, C., Hudecek, M., Jacobs, J.J.L., Hort, S. and Vrijhoef, H.J.M. (2024) Trust in and acceptance of artificial intelligence applications in medicine: a mixed methods study. JMIR Human Factors, 11, e47031

Tiwari, A. and Patel, M.H. (2024) Impact of AI in product lifecycle. In: Proceedings of the 2024 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, August 2024. IEEE, pp. 1–10.

Trikoili, A., Georgiou, D., Pappa, C.I. and Pittich, D. (2025) Critical thinking assessment in higher education: a mixed-methods comparative analysis of AI and human evaluators. International Journal of Human–Computer Interaction, pp. 1–14.

Uribe Ocampo, J. and Kaminski, P.C. (2019) Medical device development, from technical design to integrated product development. Journal of Medical Engineering & Technology, 43(5), 287–304.

Zerilli, J., Bhatt, U. and Weller, A. (2022) How transparency modulates trust in artificial intelligence, Patterns, 3(4), p. 100455.