Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 708–717
https://doi.org/10.54941/ahfe1008302

Human-Centered Design of AI Decision Support: Algorithmic Vigilance and the Role of AI Literacy

Julia N. Czerniak, Lukas Messmer, Leonie Däullary, Teuta Salii, Philipp Rüegg, and Frieder Loch

Institute for Interactive Informatics, Eastern Switzerland University of Applied Sciences, Rapperswil, CH 8640, Switzerland

ABSTRACT

Artificial intelligence is increasingly embedded in workplace decisions, yet practitioners oscillate between algorithm aversion and automation bias. Algorithmic vigilance, i.e. calibrated, critically reflective reliance on AI, is the normative design target, but it remains unclear what it looks like in everyday professional work and which conditions support it. Existing design principles largely stem from laboratory studies and treat AI literacy as the main individual moderator. This exploratory study combines interviews with 13 practitioners from construction, production planning and school administration with the Meta AI Literacy Scale (MAILS). Interview statements were quantitized into ordinal indicators and related to MAILS scores (Spearman's ρ). Practitioners unanimously retained decision authority and verified AI outputs. AI literacy tended to go with more specific transparency requirements ($\rho = .44$), whereas task complexity and decision criticality shaped explanation demand ($\rho = .76$ and $.81$) and automation affinity ($\rho = -.55$). AI literacy was unrelated to emotional security ($\rho = .19$) and emotion regulation ($\rho = -.03$); self-efficacy was associated with more systematic verification ($\rho = .59$). We derive five hypotheses and an extended model of algorithmic vigilance that adds task complexity as a second dimension next to AI literacy and locates emotional security in system design. The design implications are role- and task-sensitive transparency, active competence support and error-type-differentiated confidence communication, implemented through confirmation gates, calibrated confidence signals and configurable autonomy. A confirmatory study will test the hypotheses.

Keywords: AI literacy, Algorithmic vigilance, Automation bias, Algorithm aversion, Task complexity, MAILS, Mixed methods, Quantitizing

INTRODUCTION

Practitioners increasingly have to decide, often under time pressure, how much weight to give AI recommendations, for instance whether to accept a customer order based on an automated production schedule. Too little weight forgoes efficiency gains; too much produces automation bias and

Received September 25, 2026; Revised November 1, 2026; Accepted November 9, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

quality failures. Between these extremes lies *algorithmic vigilance*: calibrated, critically reflective reliance on AI that adjusts to task, system reliability and domain context.

Despite a growing body of work on algorithm aversion (Dietvorst et al., 2015; Jussupow et al., 2020) and automation bias (Jovchevski et al., 2026; Parasuraman and Manzey, 2010), little is known about how systems can be designed to support algorithmic vigilance. The present study therefore asks what algorithmic vigilance looks like in everyday professional decision-making and which task and design conditions support it.

The Trust Spectrum: Aversion, Algorithmic Vigilance, and Bias

Algorithm aversion denotes a disproportionate rejection of algorithmic advice, even when algorithms demonstrably outperform human judgment. Dietvorst et al. (2015) show that observing a single algorithm error is sufficient to trigger lasting distrust, independent of overall system accuracy.

Automation bias represents the opposing failure mode: uncritical deference to automated recommendations. Parasuraman and Manzey (2010) distinguish commission errors (executing erroneous AI outputs) from omission errors (ignoring correct counterinformation).

Algorithmic vigilance is defined here as the capacity to recognize, evaluate and, where warranted, challenge AI-generated outputs without defaulting to systematic rejection or uncritical acceptance. Systematic cross-domain comparisons of how trust calibration manifests across professional fields remain rare. A Dunning-Kruger dynamic in this domain is documented by Horowitz and Kahn (2024): individuals with minimal AI knowledge tend toward aversion, those with intermediate knowledge are most susceptible to automation bias, and only high expertise yields reliably calibrated judgment.

What Algorithmic Vigilance Looks Like: A Summary of Recent Work

Recent work has moved from treating trust as an attitude towards describing appropriate reliance as observable behaviour. Lee and See (2004) define appropriate reliance as trust that matches the actual capability of the automation and stress that trust has an affective as well as an analytic basis. Schemmer et al. (2023) operationalise appropriate reliance as two complementary behaviours: relying on AI advice when it is correct and relying on oneself when it is not. Zerilli et al. (2022) describe the target state as algorithmic vigilance and argue that transparency should not be reduced to explainability; dynamic task allocation and the communication of confidence and performance metrics may promote vigilance more effectively.

Across studies, vigilance becomes visible in a small set of behavioural markers. Vigilant users verify outputs in proportion to the stakes, which matters most where verification is complex, because automation bias is most pronounced there (Lyell and Coiera, 2017); they form an independent judgement before or alongside the AI advice, which cognitive forcing functions deliberately elicit (Buçinca et al., 2021); they engage with explanations selectively, based on a cost–benefit trade-off (Vasconcelos et al., 2023); and

they override or adapt AI outputs where contextual knowledge contradicts them, a practice Snow (2021) terms artificing. Human–AI teams frequently fail to outperform the better agent alone, particularly in decision tasks (Vaccaro et al., 2024), so these behaviours do not emerge automatically. How they manifest in real occupational settings, and which task characteristics shape them, remains largely unexplored.

AI Literacy as a Moderator

AI literacy functions as the key moderating variable between AI exposure and vigilant behaviour. Horowitz and Kahn (2024) demonstrate that higher AI literacy reliably reduces automation bias in high-stakes decision contexts. AI literacy is not unidimensional: instruments such as MAILS (Carolus et al., 2023) distinguish declarative knowledge, procedural competence, critical appraisal and metacognitive regulation; calibrated trust draws most directly on the latter two.

Current AI systems rely on transparency and explainability to support competent decision making (Sullivan and Weger, 2025). However, neither heterogeneity in users' AI experience nor cross-domain design implications have been studied systematically. This paper addresses these gaps with an exploratory, hypothesis-generating design that combines qualitative interviews in three occupational contexts with AI literacy measurement and quantitizes the interview data to derive hypotheses on how AI literacy, task complexity and system design jointly shape algorithmic vigilance.

METHOD

Participants

Thirteen practitioners from Swiss companies and institutions took part, all male: construction (n = 4), production planning (n = 3) and school administration (n = 6). One third each were aged below 35, 35–44 and 45–54. Twelve completed the MAILS questionnaire. All reported being familiar with AI tools (M = 3.33, SD = 0.62) and using them regularly at work (M = 4.00, SD = 0.91).

Study Design

The study followed an exploratory, hypothesis-generating design. Semi-structured interviews covered decision processes, the use and perception of AI tools, critical incidents, costs and benefits, and task allocation and degree of automation.

AI literacy was measured with the Meta AI Literacy Scale (MAILS) (Carolus et al., 2023), which suits users without programming experience and is available in German. Besides the AI literacy facets Use & Apply, Know & Understand, Detect AI and AI Ethics, it covers AI self-efficacy and AI self-competency, which capture the dispositional side of vigilance. The Create AI facet was excluded, as algorithmic vigilance concerns the evaluation of AI outputs rather than the creation of AI.

Procedure

Interviews lasted about one hour and were held online or on site in Swiss German or German. Participants gave informed consent and afterwards completed demographic items and the MAILS questionnaire. Interviews were transcribed and processed anonymously in accordance with the GDPR.

Data Analysis: From Qualitative to Quantitative

Transcripts were analysed with structuring qualitative content analysis (Kuckartz, 2014). The category system was derived deductively from the interview guide (decision processes, use and perception, critical incidents, costs and benefits, task allocation) and extended inductively. Two categories emerged from the material that were not part of the guide: task complexity and emotional security.

To relate interview findings to AI literacy, the coded material was quantitized (Sandelowski et al., 2009): for each participant, seven ordinal indicators were rated on five-level anchored scales (Table 1), blind to the MAILS scores. Task complexity was anchored in Wood's (1986) distinction between component, coordinative and dynamic complexity.

Because the indicators are ordinal and the sample is small, associations were computed as Spearman rank correlations and interpreted as effect sizes rather than significance tests. Qualitative and quantitized results were integrated in a joint display; a hypothesis was formulated only where the qualitative pattern and the direction of the quantitized association converged.

Table 1: Codebook for quantitizing interview data (ordinal indicators per participant).

Indicator	Anchors (1-5)
Task complexity (TC)	1 routine, rule-based · 2 mostly routine · 3 mixed routine and novel · 4 predominantly novel, interdependent · 5 novel, interdependent, long-running
Decision criticality (DC)	1 reversible, internal · 2 operational impact · 3 substantial financial or customer impact · 4 legal or liability consequences · 5 hardly reversible consequences for individuals
Automation affinity (AA)	1 advisory only · 2 preparation with mandatory confirmation · 3 pre-decisions with human override · 4 conditional automation of sub-tasks · 5 rule-based full automation of routine tasks
Transparency specificity (TS)	1 none · 2 abstract ('no black box') · 3 direction stated, no mechanism · 4 partly specified mechanism (e.g. confidence levels) · 5 concrete implementation (e.g. source links, raw-data traceability)
Explanation demand (ED)	1 not wanted · 2 on demand only · 3 stake-dependent · 4 by default for critical outputs · 5 by default for all outputs, incl. inputs and weighting
Emotional security (ES)	1 predominant unease · 2 unease outweighs · 3 ambivalent · 4 confidence outweighs, with reservations · 5 predominant confidence
Verification intensity (VI)	1 rarely or never · 2 occasional plausibility check · 3 systematic for critical outputs · 4 as 3, plus second source or expert · 5 multi-channel verification of all outputs

RESULTS

Cross-Sector Qualitative Findings

Scope of AI use. All participants used or envisaged AI for routine and decision-preparation tasks such as drafting, summarising, research and data preparation. Where AI literacy is referred to below, participants are grouped at the MAILS median ($n = 6$ each).

AI Literacy and Transparency. Transparency with access to sources was a mandatory requirement in all sectors. Its specificity differed with AI literacy. High-literacy participants named implementation details such as clickable source links with timestamps or traceability down to raw data. Low-literacy participants expressed the same need in abstract terms, demanding "no black box" or a system that would "explain itself like a colleague".

Task complexity and automation. Construction participants, facing long-running projects and liability-relevant decisions, wanted analytical AI support under interactive human control; production participants accepted rule-based full automation of repetitive tasks ("if this occurs, do this; if neither applies, refer back to us"); school administrators requested an advisory AI within ethical boundaries. Automation affinity thus followed task complexity rather than AI literacy. All participants rejected complete decision automation for non-routine tasks and excluded tasks with legal consequences or requiring human competencies, such as customer relations or personal matters concerning pupils.

Task complexity and explanation. Explanation demand followed the stakes of the task. Production participants wanted results first and reasoning only for unclear cases ("I would first want to see the decision, and only if I don't understand it, have it explained"). In construction, explanation depth was to be qualified by confidence levels; in school administration, where decisions affect pupils, participants expected inputs and their weighting to be disclosed by default.

AI literacy and emotional security. Unease about relying on AI, such as loss of control, liability, data protection or undetected errors, occurred at all literacy levels, as did reassurance. High-literacy participants criticised inflated outputs and missing governance; low-literacy participants feared deskilling ("no longer using one's own brain") or doubted that anonymisation still suffices. Security was instead tied to system features such as controllability, data handling and liability for erroneous outputs: "if a wrong statement is made, it does not count that it was generated automatically".

Self-efficacy and vigilance. All participants intended to verify AI-prepared recommendations for non-routine tasks, but verification intensity differed. Experienced, self-confident users actively tested the limits of AI: they supplied their own facts, cross-checked outputs across tools, consulted norms and experts or made source references mandatory. Less experienced participants relied on plausibility checks based on domain experience or reported no verification strategy.

Quantitized Interview Data

Automation affinity was highest in production ($M = 4.00$) and lowest in school administration ($M = 1.83$); task complexity was highest in construction ($M = 4.25$) and lowest in production ($M = 2.00$). Explanation demand increased with task complexity ($\rho = .76$, $p = .003$) and decision criticality ($\rho = .81$, $p = .001$), whereas automation affinity decreased with task complexity ($\rho = -.55$, $p = .054$). Verification intensity was associated with transparency specificy ($\rho = .6$, $p = .02$), and task complexity ($\rho = .64$), decision criticality ($\rho = .58$), explanation demand ($\rho = .56$) and emotional security ($\rho = .56$; all $p < .05$). AI literacy correlated with self-efficacy ($\rho = .79$, $p = .002$) and task complexity ($\rho = .66$, $p = .019$); at sub-scale level, Use & Apply correlated with transparency specificity ($\rho = .58$) and Know & Understand with task complexity ($\rho = .71$, $p = .01$) . Emotion regulation was unrelated to all indicators ($|\rho| \leq .41$).

Table 2: Joint display of qualitative patterns, quantitized associations, and derived hypotheses.

Dimension	Qualitative Pattern	Association	Hyp.
AI literacy × transparency	High literacy: concrete implementation requirements; low literacy: abstract ("no black box")	ρ(MAILS, TS) = .44 ρ(UA, TS) = .58*	H1
Task complexity × automation	Routine tasks: rule-based automation accepted, also at low literacy; complex tasks: human control	ρ(TC, AA) = .465 ρ(MAILS, AA) = –.48	H2
Task complexity × explanation	Reasoning on demand for routine tasks; disclosure of inputs and weighting as baseline for high-stakes decisions	ρ(TC, ED) = .693 ρ(DC, ED) = .757**	H3
AI literacy × security	Unease at all literacy levels; security tied to system features and liability	ρ(MAILS, ER) = –.03 ρ(ES, VI) = .461 ρ(ES, VI) = .56*	H4
Self-efficacy × vigilance	Experienced users test AI boundaries and keep verification routines	ρ(MAILS, SE) = .79** ρ(SE, VI) = .59*	H5

Note. Spearman's ρ; †$p < .10$, *$p < .05$, **$p < .01$. UA = MAILS sub-scale Use & Apply; SE = self-efficacy, ER = emotion regulation.

Integration and Hypothesis Generation

Table 2 juxtaposes the qualitative patterns with the quantitized associations; given the sample size, these indicate plausibility rather than confirmed effects. Five hypotheses follow. *H1: Higher AI literacy is associated with more specific transparency requirements*; as literacy also correlated with task complexity, this may partly reflect role demands. *H2: Automation affinity decreases with task complexity, independently of AI literacy*; the negative association with literacy likely reflects sector. *H3: Explanation demand increases with task complexity and decision criticality. H4: Emotional security in AI-supported decision-making depends on system properties rather than on AI literacy*; security went with more verification, suggesting that confidence arose from control rather than replacing it. *H5: Higher AI self-efficacy is associated with more systematic verification of AI outputs.*

DISCUSSION

Sources of Influence on Algorithmic Vigilance

The findings distinguish three sources of influence on algorithmic vigilance. User competence, captured by AI literacy and self-efficacy, primarily affects how users articulate requirements and verify outputs (H1, H5). Task characteristics, namely complexity and decision criticality, determine the accepted level of automation and the demanded depth of explanation (H2, H3). Emotional security depends on system properties rather than on the user (H4).

Neither failure mode dominated. No participant showed signs of automation bias (Parasuraman and Manzey, 2010): higher competence went with more rather than less verification. Participants at the lower end of the literacy spectrum reported the least verification and the lowest emotional security, a profile closer to algorithm aversion (Dietvorst et al., 2015). Algorithmic vigilance thus appears less as a trait that grows with knowledge than as a practice that combines confidence with verification.

Fig. 1 summarises these relations as a corridor whose position shifts with task complexity and decision criticality: the more complex and consequential the task, the lower the reliance that remains appropriate. Construction lies within the corridor. Production planning shows the highest reliance, appropriate for routine tasks but close to the automation-bias boundary given its low verification intensity. School administration lies at the aversion boundary; its advisory-only stance follows from decisions affecting pupils and is paired with the highest verification intensity, suggesting deliberate caution. As AI literacy and task complexity were correlated ($\rho = .66$), their effects cannot yet be separated.

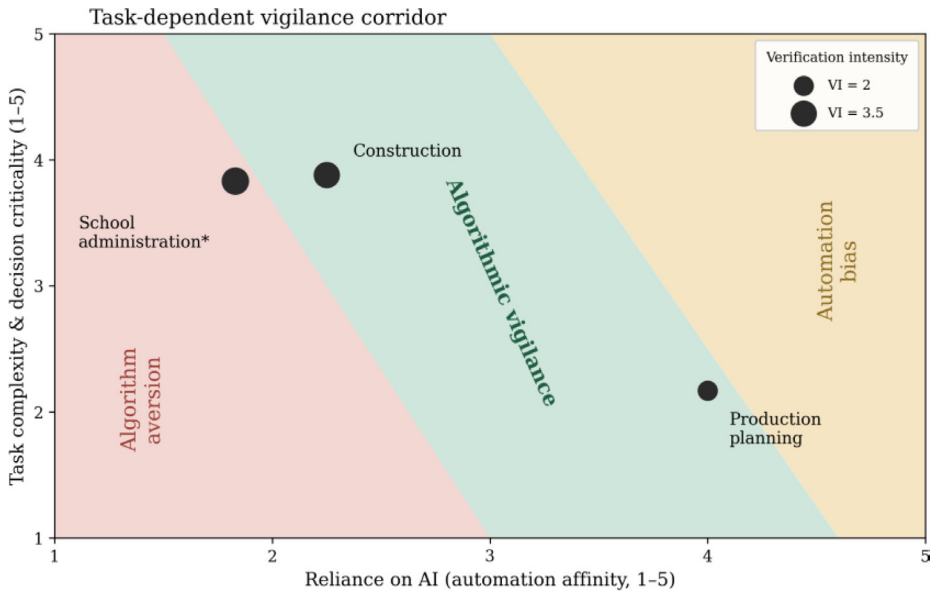


Figure 1: Task-dependent vigilance corridor. Use cases are positioned by mean automation affinity (x) and mean task complexity and decision criticality (y); marker size indicates verification intensity. Corridor boundaries are conceptual; *see Section 4.1.

Implications for Design

Established guidelines for human–AI interaction (Amershi et al., 2019) call for transparency, explainability and human oversight. Our findings confirm these principles as necessary but not sufficient for algorithmic vigilance; three extensions follow.

Role- and task-sensitive transparency. Explanation depth and automation level should adapt to task complexity and decision criticality (H2, H3): results first, reasoning on demand and configurable rule-based automation for routine tasks; default disclosure of inputs and weighting and an advisory role for high-stakes or person-related decisions.

Active competence support. Because AI literacy and self-efficacy shape how users articulate needs and verify outputs (H1, H5), verification support such as linked sources, cross-checks against original documents and prompts for critical outputs should be built in for all users rather than presupposed.

Error-type-differentiated confidence communication. As emotional security depended on system properties and error consequences (H4), systems should communicate confidence per output together with the type of potential error (financial, legal, person-related) and make explicit where responsibility lies.

Limitations and Future Work

The analysis is exploratory: with $n = 12–13$, effect sizes and directions are more informative than p-values. A sensitivity analysis with 3-level coding yielded the same directions for all associations except the near-zero one between AI literacy and emotional security. Sector and AI literacy are partly confounded; two interviews were coded from handwritten notes, and all codes were assigned by a single coder.

CONCLUSION

This paper examined how decision-support AI must be designed to achieve algorithmic vigilance. Across three occupational contexts, appropriate reliance depended on task complexity and decision criticality rather than on AI literacy, self-efficacy went with more rather than less verification, and emotional security was tied to system properties.

Taken together, the findings support a reframing of the design objective. Principle-based guidance, from Amershi et al. (2019) to the EU AI Act (2024), implicitly treats appropriate trust as a downstream outcome of transparent, controllable and explainable systems. Our data suggest that appropriate trust also requires users who are positioned, by competence, role and institutional context, to exercise judgment; system design can support this positioning but cannot substitute for it.

Hypotheses H1–H5 will be tested in a confirmatory, preregistered study with a larger and more diverse sample, varying task complexity in scenario-based decision tasks, manipulating design features (confirmation gates, confidence signals, configurable autonomy) and measuring algorithmic vigilance behaviourally as appropriate reliance (Schemmer et al., 2023), alongside MAILS and emotional security.

ACKNOWLEDGMENT

This research was conducted as part of the Innosuisse innovation projects Athena: AI Assistant for Efficient Construction Project Management (130.110 IP-ICT), KIdsPUPILpro (132.281 IP-SBM), and KODI NX-AI: Adaptive Production Optimization for Industry 5.0 SMEs (132.282 IP-ICT). The authors gratefully acknowledge the support of Innosuisse, the Swiss Innovation Agency.

REFERENCES

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E. (2019) 'Guidelines for Human-AI Interaction', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, ACM, pp. 1–13 [Online]. DOI: 10.1145/3290605.3300233.

Buçinca, Z., Malaya, M. B. and 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*, vol. 5, CSCW1, article 188 [Online]. DOI: 10.1145/3449287.

Carolus, A., Koch, M. J., Straka, S., Latoschik, M. E. and Wienrich, C. (2023) 'MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies', *Computers in Human Behavior: Artificial Humans*, vol. 1, no. 2, p. 100014 [Online]. DOI: 10.1016/j.chbah.2023.100014.

Dietvorst, B. J., Simmons, J. P. and Massey, C. (2015) 'Algorithm aversion: People erroneously avoid algorithms after seeing them err', *Journal of Experimental Psychology: General*, vol. 144, no. 1, pp. 114–126.

European Union (2024) 'Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)', *Official Journal of the European Union*, L series, 12 July 2024 [Online]. Available at http://data.europa.eu/eli/reg/2024/1689/oj.

Horowitz, M. C. and Kahn, L. (2024) 'Bending the Automation Bias Curve: A Study of Human and AI-Based Decision Making in National Security Contexts', *International Studies Quarterly*, vol. 68, no. 2.

Jovchevski, P., Buijsman, S. and Neerincx, M. (2026) 'What is Wrong With Automation Bias?', *Philosophy & Technology*, vol. 39, no. 2, p. 84 [Online]. DOI: 10.1007/s13347-026-01090-9.

Jussupow, E., Benbasat, I. and Heinzl, A. (2020) 'Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion', in Rowe, F. (ed) *Proceedings of the 28th European Conference on Information Systems (ECIS): An Online AIS Conference, June 15-17, 2020* [Online], Atlanta, GA, AISeL, p. 168. Available at https://aisel.aisnet.org/ecis2020_rp/168 (Accessed 6 July 2026).

Kuckartz, U. (2014) *Qualitative Text Analysis: A Guide to Methods, Practice and Using Software*, London, Sage.

Lee, J. D. and See, K. A. (2004) 'Trust in automation: designing for appropriate reliance', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, no. 1, pp. 50–80.

Lyell, D. and Coiera, E. (2017) 'Automation bias and verification complexity: Asystematic review', *Journal of the American Medical Informatics Association*, vol. 24, no. 2, pp. 423–431.

Parasuraman, R. and Manzey, D. H. (2010) 'Complacency and bias in human use of automation: An attentional integration', *Human factors*, vol. 52, no. 3, pp. 381–410 [Online]. DOI: 10.1177/0018720810376055.

Sandelowski, M., Voils, C. I. and Knafl, G. (2009) 'On quantitizing', *Journal of Mixed Methods Research*, vol. 3, no. 3, pp. 208–222 [Online]. DOI: 10.1177/1558689809334210.

Schemmer, M., Kuehl, N., Benz, C., Bartos, A. and Satzger, G. (2023) 'Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations', pp. 410–422.

Snow, T. (2021) 'From satisficing to artificing: The evolution of administrative decision-making in the age of the algorithm', *Data & Policy*, vol. 3 [Online]. DOI: 10.1017/dap.2020.25.

Sullivan, V. and 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*, vol. 69, no. 1, pp. 1827–1833.

Vaccaro, M., Almaatouq, A. and Malone, T. (2024) 'When combinations of humans and AI are useful: A systematic review and meta-analysis', *Nature Human Behaviour*, vol. 8, no. 12, pp. 2293–2303 [Online]. DOI: 10.1038/s41562-024-02024-1.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. S. and Krishna, R. (2023) 'Explanations Can Reduce Overreliance on AI Systems During Decision-Making', *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, CSCW1, article 129 [Online]. DOI: 10.1145/3579605.

Wood, R. E. (1986) 'Task complexity: Definition of the construct', Organizational Behavior and Human Decision Processes, vol. 37, no. 1, pp. 60–82.

Zerilli, J., Bhatt, U. and Weller, A. (2022) 'How transparency modulates trust in artificial intelligence', *Patterns*, vol. 3, no. 4, p. 100455.