Supplementary Material: For Those About to Rely—A Taxonomy of Experimental Studies on AI Reliance

Myriam Schaschek[1], Niko Spatscheck[1], and Axel Winkelmann[1]

Julius-Maximilians-Universität Würzburg, Chair of Management and Information Systems, Würzburg, Germany
{myriam.schaschek,niko.spatscheck,axel.winkelmann}@uni-wuerzburg.de

A Taxonomy Development Process

A.1 Data Collection

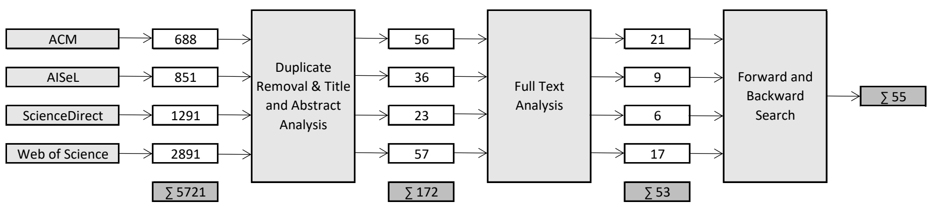

Figure A1. Literature search process overview

Survey scope and paper inclusion criteria: The focus of this survey is on empirical human-subject studies of humanAI decision making, where the goal is to evaluate, understand and/ or improve human performance and experience for a decision making task, rather than to improve the model. As such, we specify the following inclusion and exclusion criteria:

- The paper must include evaluative human-subject studies. We thus exclude purely formative studies that focus on exploring user needs to inform design of AI systems, often qualitatively.
- The paper must target a decision making task, thus we exclude tasks of other purposes (e.g., debugging and other forms of improving the model, co-creation, gaming).
- The task must involve and focus on studying human decision makers, thus we exclude papers on AI automation or other AI stakeholders (model developers, ML practitioners). However, we do not limit our studies to those that implement complete decision making processes, but also include studies that claim to evaluate some aspects of decision makers' perceptions, such as their understanding, satisfaction, and perceived fairness of the AI.

A.2 Iterative Process of Taxonomy Development

The first iteration involved a conceptual-to-empirical approach, where literature reviews provided a starting point for the initial characteristics and dimensions. In the second iteration, we chose an empirical-to-conceptual approach to capture the nuanced differences

19th International Conference on Wirtschaftsinformatik
September 2024, Würzburg, Germany

between each experiment. To do so, we randomly selected 15 out of the 55 resulting experiments (cf. data collection) to refine the taxonomy further. After the second iteration, we noticed that the modified taxonomy had become in-depth but descriptive. To counteract the issue, the third iteration again included a conceptual-to-empirical approach. We aggregated the present characteristics and dimensions by consulting existing frameworks and literature reviews from related fields. In the final iteration, we adopted the procedure of the second iteration but with a random selection of 20 experimental studies. This allowed us to refine and aggregate our dimensions and characteristics even further. Within this iteration, we met our defined ending conditions. Table A1 presents an overview of each iteration conducted during the development of our taxonomy.

Table A1. Overview of the four iterations

Iteration	Approach	Summary	#M	#D	#C	Article
1	Conceptual-to-Empirical	Literature reviews and articles with a holistic perspective provided foundational characteristics and dimensions, allowing us to establish a preliminary taxonomy for the research area.	4	35	122	(32; 24; 33; 23; 42; 10; 53)
2	Empirical-to-Conceptual	Classification of experiments highlighted subtle differences, facilitating the evaluation of quality and updating the characteristics and dimensions.	4	74	216	(2; 3; 6; 7; 11; 13; 16; 18; 21; 25; 43; 45; 57; 60; 66)
3	Conceptual-to-Empirical	Consultation of existing frameworks and literature reviews from related fields enhanced and consolidated the current dimensions and characteristics.	4	28	87	(26; 40; 36; 8; 5; 20; 63; 39)
4	Empirical-to-Conceptual	Classification of additional experiments elucidated the dimensions and characteristics, finalizing the taxonomy and fulfilling the established ending conditions.	4	18	56	(1; 4; 12; 14; 19; 22; 27; 35; 37; 38; 44; 48; 52; 54; 55; 56; 58; 59; 61; 68)
#M=Number of Metacharacteristics; #D=Number of Dimensions; #C=Number of Characteristics						

A.3 Literature Classification

Table A2. Literature Collection for the Metacharacteristic Collaboration

Dimension	Characteristic	Article
Mode	Human-AI	(1; 3; 4; 7; 9; 10; 11; 12; 13; 14; 16; 17; 19; 21; 22; 25; 27; 28; 30; 29; 31; 34; 35; 37; 38; 41; 43; 44; 45; 46; 47; 48; 49; 50; 51; 53; 52; 54; 55; 56; 57; 58; 59; 60; 61; 62; 64; 65; 68; 67; 66)
	Human-Multi-AI	(2)
	Multi-Human-AI	(6; 15; 18)
Interface	Wizard-of-Oz	(1; 7; 9; 15; 18; 38; 43; 54; 56)
	Fully Functional AI	(2; 3; 4; 6; 10; 11; 12; 13; 14; 16; 17; 19; 21; 22; 25; 27; 28; 30; 29; 31; 34; 35; 37; 41; 44; 45; 46; 47; 48; 49; 50; 51; 53; 52; 55; 57; 58; 59; 60; 61; 62; 64; 65; 68; 67; 66)
Advice Count	Single	(1; 3; 4; 7; 9; 10; 11; 12; 13; 14; 16; 17; 19; 21; 22; 25; 27; 28; 30; 29; 31; 34; 35; 37; 38; 41; 43; 44; 46; 47; 48; 49; 50; 51; 53; 54; 56; 57; 58; 59; 60; 61; 64; 65; 68; 67; 66)
	Multiple	(2; 6; 15; 18; 45; 52; 55; 62)
Advice Timing	Before Manifestation	(1; 6; 7; 9; 15; 16; 17; 18; 19; 21; 25; 28; 30; 31; 35; 43; 44; 45; 46; 49; 52; 55; 56; 58; 61; 62)
	After Manifestation	(2; 3; 4; 7; 11; 12; 13; 14; 22; 27; 29; 34; 37; 38; 41; 47; 48; 50; 51; 53; 54; 57; 59; 60; 64; 65; 68; 67; 66)
	During Manifestation	(10)
Ecological Context	Business	(4; 6; 9; 11; 12; 14; 16; 18; 30; 34; 46; 48; 51; 60)
	Public	(1; 15; 17; 27; 28; 47; 49; 52; 55; 59; 62; 65)
	Private	(7; 21; 31; 35; 38; 41; 45; 50; 57; 64; 68)
	Healthcare	(19; 25; 37; 43; 44; 61)
	Gaming	(2; 3; 13; 66)
	Hypothetical	(10; 22; 29; 53; 54; 56; 58; 67)

Table A3. Literature Collection for the Metacharacteristic Agent

Dimension	Characteristic	Article
Human Characteristics	AI Literacy	(11; 12; 35; 43)
	Domain Knowledge	(2; 3; 13; 25; 27; 28; 31; 35; 37; 43; 49; 55; 59; 60; 61; 62; 68; 67; 66)
	None	(1; 4; 6; 7; 9; 10; 11; 14; 15; 16; 17; 18; 19; 21; 22; 30; 29; 34; 38; 41; 44; 45; 46; 47; 48; 50; 51; 53; 52; 54; 56; 57; 58; 64; 65)
Human Choice of Advice	Adjustable	(6; 7; 16; 18; 19; 28; 45; 47; 52)
	Predefined	(1; 2; 3; 4; 7; 9; 10; 11; 12; 13; 14; 15; 17; 21; 22; 25; 27; 30; 29; 31; 34; 35; 37; 38; 41; 43; 44; 46; 48; 49; 50; 51; 53; 54; 55; 56; 57; 58; 59; 60; 61; 62; 64; 65; 68; 67; 66)
AI Embodiment	Physical Robot	(19)
	Virtual Agent or Bot	(3; 15; 18; 21; 46; 53; 52; 54)
	Embedded inside Tool	(1; 2; 4; 6; 7; 9; 10; 11; 12; 13; 14; 16; 17; 21; 22; 25; 27; 28; 30; 29; 31; 34; 35; 37; 38; 41; 43; 44; 45; 47; 48; 49; 50; 51; 55; 56; 57; 58; 59; 60; 61; 62; 64; 65; 68; 67; 66)
AI Agency	Performative	(11; 47; 60; 68)
	Advisory	(1; 2; 3; 4; 6; 7; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 21; 22; 25; 27; 28; 30; 29; 31; 34; 35; 37; 38; 41; 43; 44; 45; 46; 48; 49; 50; 51; 53; 52; 54; 55; 56; 57; 58; 59; 60; 61; 62; 64; 65; 68; 67; 66)
AI Response Mode	Judgment	(4; 11; 12; 14; 15; 19; 27; 34; 60; 61; 65)
	Choice	(1; 2; 3; 6; 7; 9; 10; 13; 16; 17; 18; 21; 22; 25; 28; 30; 29; 31; 35; 37; 38; 41; 43; 44; 45; 46; 47; 48; 49; 50; 51; 53; 52; 54; 55; 56; 57; 58; 59; 62; 64; 68; 67; 66)

Table A4. Literature Collection for the Metacharacteristic Task

Dimension	Characteristic	Article
Input	Image	(2; 3; 7; 9; 13; 16; 17; 21; 30; 35; 37; 38; 41; 46; 47; 50; 55; 58; 64; 67; 66)
	Text	(1; 4; 6; 11; 12; 14; 15; 16; 18; 21; 22; 27; 28; 30; 29; 31; 34; 38; 43; 44; 46; 48; 49; 50; 51; 53; 52; 54; 55; 56; 57; 59; 60; 61; 62; 64; 65; 68)
	Video	(25; 45)
	Audio	(19; 54)
	3D Model	(10; 47)
Solution Nature	Subjective	(1; 6; 7; 28; 29; 38; 41; 46; 49; 50; 54; 55; 64)
	Objective	(2; 3; 4; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 21; 22; 25; 27; 30; 31; 34; 35; 37; 43; 44; 45; 47; 48; 51; 53; 52; 56; 57; 58; 59; 60; 61; 62; 65; 68; 67; 66)
Importance	Low-Stake	(2; 3; 4; 6; 7; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 21; 22; 30; 31; 34; 38; 41; 45; 46; 47; 48; 49; 50; 51; 53; 52; 54; 56; 57; 58; 59; 60; 62; 64; 65; 68; 67; 66)
	High-Stake	(1; 25; 27; 28; 29; 35; 37; 43; 44; 55; 61)

Table A5. Literature Collection for the Metacharacteristic Precondition

Dimension	Characteristic	Article
Manipulation	Performance	(9; 10; 11; 12; 13; 14; 15; 21; 25; 27; 29; 31; 38; 48; 50; 53; 52; 62; 64; 68; 66)
	Transparency	(2; 3; 7; 9; 11; 22; 25; 27; 35; 38; 41; 43; 45; 46; 48; 50; 51; 53; 52; 55; 57; 58; 59; 60; 61; 64; 65; 68; 67)
	Interaction	(1; 4; 6; 15; 16; 18; 19; 25; 28; 31; 34; 37; 56; 65; 68; 67; 66)
	Representation	(2; 4; 18; 21; 27; 30; 37; 44; 46; 54; 62; 67)
	Individuality	(1; 3; 4; 11; 12; 17; 22; 28; 29; 35; 38; 44; 45; 55; 56; 59; 60)
	Task	(7; 10; 28; 47; 49; 58; 64)
Setup	Single Trial	(1; 15; 18; 19; 29; 31; 46; 65)
	Multiple Trials	(2; 3; 4; 6; 7; 9; 10; 11; 12; 13; 14; 16; 17; 21; 22; 25; 27; 28; 30; 34; 35; 37; 38; 41; 43; 44; 45; 47; 48; 49; 50; 51; 53; 52; 54; 55; 56; 57; 58; 59; 60; 61; 62; 64; 68; 67; 66)
Feedback	Not Given	(1; 2; 6; 7; 9; 10; 11; 15; 16; 17; 18; 19; 21; 22; 25; 28; 29; 31; 35; 37; 38; 43; 44; 45; 46; 47; 48; 49; 51; 53; 52; 54; 55; 57; 58; 59; 60; 61; 62; 65)
	Given-Immediate	(12; 13; 14; 22; 25; 27; 34; 41; 47; 56; 68; 67; 66)
	Given-Delayed	(3; 4; 11; 12; 30; 38; 50; 64)
User-AI Onboarding	None	(3; 6; 10; 11; 13; 14; 15; 17; 18; 19; 21; 22; 25; 29; 31; 38; 43; 48; 53; 52; 55; 56; 57; 60; 61; 62; 65; 67)
	Instruction	(1; 2; 9; 16; 27; 28; 30; 35; 41; 44; 45; 46; 47; 49; 50; 51; 59; 64)
	Pre-Experimental Usage	(4; 7; 12; 16; 34; 37; 41; 45; 47; 49; 51; 54; 58; 68; 66)
Reward Structure	Monetary Performance	(2; 3; 4; 6; 7; 10; 11; 13; 22; 30; 37; 38; 41; 49; 50; 51; 52; 58; 59; 64; 65; 68; 67; 66)
	Monetary Non-Performance	(9; 12; 14; 21; 27; 29; 35; 46; 48; 60)
	None	(1; 16; 18; 19; 28; 31; 43; 44; 45; 47; 54; 55; 56; 57; 61; 62)
	Non-Monetary Performance	(25; 34)
	Non-Monetary Non-Performance	(15; 17; 53)

B Descriptive Pattern Analysis

Table B1. Descriptive frequency analysis of dimensions (1).

MC	Dimension	Characteristic	Absolute	Relative
Collaboration	Mode	Human-AI	51	0,93
		Human-Multi-AI	1	0,02
		Multi-Human-AI	3	0,05
	Interface	Wizard-of-Oz	9	0,16
		Fully Functional AI	46	0,84
	Advice Count	Single	47	0,85
		Multiple	8	0,15
	Advice Timing	Before Manifestation	26	0,47
		After Manifestation	29	0,53
		During Manifestation	1	0,02
	Ecological Context	Business	14	0,25
		Public	12	0,22
		Private	11	0,2
		Healthcare	6	0,11
		Gaming	4	0,07
		Hypothetical	8	0,15
Agent	Human Characteristics	AI Literacy	4	0,07
		Domain Knowledge	19	0,35
		None	35	0,64
	Human Choice of Advice	Adjustable	9	0,16
		Predefined	47	0,85
	AI Embodiment	Physical Robot	1	0,02
		Virtual Agent or Bot	8	0,15
		Embedded inside Tool	47	0,85
	AI Agency	Performative	4	0,07
		Advisory	54	0,98
	AI Response Mode	Judgment	11	0,2
		Choice	44	0,8

Table B2. Descriptive frequency analysis of dimensions (2).

MC	Dimension	Characteristic	Absolute	Relative
Task	Input	Image	21	0,38
		Text	38	0,69
		Video	2	0,04
		Audio	2	0,04
		3D Model	2	0,04
	Solution Nature	Subjective	13	0,24
		Objective	42	0,76
	Importance	Low-Stake	44	0,8
		High-Stake	11	0,2
Precondition	Manipulation	Performance	21	0,38
		Transparency	29	0,53
		Interaction	17	0,31
		Representation	12	0,22
		Individuality	17	0,31
		Task	7	0,13
	Setup	Single Trial	8	0,15
		Multiple Trials	47	0,85
	Feedback	Not Given	40	0,73
		Given-Immediate	13	0,24
		Given-Delayed	8	0,15
	User-AI Onboarding	None	28	0,51
		Instruction	18	0,33
		Pre-Experimental Usage	15	0,27
	Reward Structure	Monetary Performance	24	0,44
		Monetary Non-Performance	10	0,18
		None	16	0,29
		Non-Monetary Performance	2	0,04
		Non-Monetary Non-Performance	3	0,05

References

1. Alon-Barkat, S. and Busuioc, M. (2023), 'Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice', *Journal of Public Administration Research and Theory* **33**(1), 153–169.
2. Banovic, N., Yang, Z., Ramesh, A. and Liu, A. (2022), 'Being Trustworthy is Not Enough: How Untrustworthy Artificial Intelligence (AI) Can Deceive the End-Users and Gain Their Trust', *Proceedings of the ACM on Human-Computer Interaction* **0**(CSCW0), 1–16.
3. Bayer, S., Gimpel, H. and Markgraf, M. (2021), 'The role of domain expertise in trusting and following explainable AI decision support systems', *Journal of Decision Systems* **00**(00), 1–29.
4. Berger, B., Adam, M., Rühr, A. and Benlian, A. (2021), 'Watch me improve—algorithm aversion and demonstrating the ability to learn', *Business & Information Systems Engineering* **63**(1), 55–68.
5. Bonaccio, S. and Dalal, R. S. (2006), 'Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences', *Organizational behavior and human decision processes* **101**(2), 127–151.
6. Brachman, M., Ashktorab, Z., Desmond, M., Duesterwald, E., Dugan, C., Joshi, N. N., Pan, Q. and Sharma, A. (2022), 'Reliance and automation for human-ai collaborative data labeling conflict resolution', *Proceedings of the ACM on Human-Computer Interaction* **6**(CSCW2), 1–27.
7. 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* **5**(CSCW1), 1–21.
8. Burton, J. W., Stein, M.-K. and Jensen, T. B. (2020), 'A systematic review of algorithm aversion in augmented decision making', *Journal of behavioral decision making* **33**(2), 220–239.
9. Cabrera, Á. A., Perer, A. and Hong, J. I. (2023), 'Improving human-ai collaboration with descriptions of ai behavior', *Proceedings of the ACM on Human-Computer Interaction* **7**(CSCW1), 1–21.
10. Cao, S. and Huang, C.-M. (2022), 'Understanding user reliance on ai in assisted decision-making', *Proceedings of the ACM on Human-Computer Interaction* **6**(CSCW2), 1–23.
11. Chiang, C.-W. and Yin, M. (2021), You'd better stop! understanding human reliance on machine learning models under covariate shift, *in* 'Proceedings of the 13th ACM Web Science Conference 2021', pp. 120–129.
12. Chiang, C.-W. and Yin, M. (2022), Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning models, *in* '27th International Conference on Intelligent User Interfaces', pp. 148–161.
13. Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K. and Cagan, J. (2022), 'Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of ai advice', *Computers in Human Behavior* **127**, 107018.

14. Daschner, S. and Obermaier, R. (2022), ‘Algorithm aversion? On the influence of advice accuracy on trust in algorithmic advice’, *Journal of Decision Systems* **31**(S1), 77–97.
15. Dennis, A. R., Lakhiwal, A. and Sachdeva, A. (2023), ‘Ai agents as team members: Effects on satisfaction, conflict, trustworthiness, and willingness to work with’, *Journal of Management Information Systems* **40**(2), 307–337.
16. Dikmen, M. and Burns, C. (2022), ‘The effects of domain knowledge on trust in explainable ai and task performance: A case of peer-to-peer lending’, *International Journal of Human-Computer Studies* **162**, 102792.
17. Elson, J. S., Derrick, D. and Ligon, G. (2018), ‘Examining trust and reliance in collaborations between humans and automated agents’, *Proceedings of the Annual Hawaii International Conference on System Sciences* .
18. Flathmann, C., McNeese, N. J., Schelble, B., Knijnenburg, B. and Freeman, G. (2023), ‘Understanding the impact and design of ai teammate etiquette’, *Human–Computer Interaction* pp. 1–28.
19. Giorgi, I., Minutolo, A., Tirotto, F., Hagen, O., Esposito, M., Gianni, M., Palomino, M. and Masala, G. L. (2023), ‘I am robot, your health adviser for older adults: Do you trust my advice?’, *International Journal of Social Robotics* pp. 1–20.
20. Glikson, E. and Woolley, A. W. (2020), ‘Human trust in artificial intelligence: Review of empirical research’, *Academy of Management Annals* **14**(2), 627–660.
21. Gupta, A., Basu, D., Ghantasala, R., Qiu, S. and Gadiraju, U. (2022), To trust or not to trust: How a conversational interface affects trust in a decision support system, *in* ‘Proceedings of the ACM Web Conference 2022’, pp. 3531–3540.
22. He, G., Kuiper, L. and Gadiraju, U. (2023), Knowing about knowing: An illusion of human competence can hinder appropriate reliance on ai systems, *in* ‘Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems’, pp. 1–18.
23. Hemmer, P., Schemmer, M., Vössing, M. and Kühl, N. (2021), ‘Human-ai complementarity in hybrid intelligence systems: A structured literature review.’, *PACIS* p. 78.
24. Hoff, K. A. and Bashir, M. (2015), ‘Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust’, *Human Factors* **57**(3), 407–434.
25. Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., VepŘek, L. H. and Quinn, G. (2023), ‘Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making’, *ACM Transactions on Computer-Human Interaction* **30**(5), 1–29.
26. Jussupow, E., Benbasat, I. and Heinzl, A. (2020), ‘Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion’, *ECIS 2020 Research Papers* p. 168.
27. Kahr, P. K., Rooks, G., Willemsen, M. C. and Snijders, C. C. (2023), It seems smart, but it acts stupid: Development of trust in ai advice in a repeated legal decision-making task, *in* ‘Proceedings of the 28th International Conference on Intelligent User Interfaces’, pp. 528–539.
28. Kaufmann, E. (2021), ‘Algorithm appreciation or aversion? comparing in-service and pre-service teachers’ acceptance of computerized expert models’, *Computers and Education: Artificial Intelligence* **2**, 100028.

29. Kim, T. and Song, H. (2023*a*), ‘Communicating the limitations of ai: The effect of message framing and ownership on trust in artificial intelligence’, *International Journal of Human-Computer Interaction* **39**(4), 790–800.
30. Kim, T. and Song, H. (2023*b*), ‘“i believe ai can learn from the error. or can it not?”: The effects of implicit theories on trust repair of the intelligent agent’, *International Journal of Social Robotics* **15**(1), 115–128.
31. Lacroux, A. and Martin-Lacroux, C. (2022), ‘Should i trust the artificial intelligence to recruit? recruiters’ perceptions and behavior when faced with algorithm-based recommendation systems during resume screening’, *Frontiers in Psychology* **13**, 895997.
32. Lai, V., Chen, C., Smith-Renner, A., Liao, Q. V. and Tan, C. (2023), Towards a science of human-ai decision making: An overview of design space in empirical human-subject studies, *in* ‘Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency’, pp. 1369–1385.
33. Lee, J. D. and See, K. A. (2004), ‘Trust in automation: Designing for appropriate reliance’, *Human factors* **46**(1), 50–80.
34. Leffrang, D. (2023), ‘The broken leg of algorithm appreciation: An experimental study on the effect of unobserved variables on advice utilization’, *Wirtschaftsinformatik 2023 Proceedings* .
35. Leichtmann, B., Humer, C., Hinterreiter, A., Streit, M. and Mara, M. (2023), ‘Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task’, *Computers in Human Behavior* **139**, 107539.
36. Li, Y. and Hahn, J. (2022), ‘Review of research on human trust in artificial intelligence’.
37. Lohoff, L. and Rühr, A. (2021), ‘Introducing (Machine) Learning Ability as Antecedent of Trust in Intelligent Systems’, *ECIS 2021 Research Papers* (June), 1–16.
38. Lu, Z. and Yin, M. (2021), ‘Human reliance on machine learning models when performance feedback is limited: Heuristics and risks’, *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* pp. 1–16.
39. Lukyanenko, R., Maass, W. and Storey, V. C. (2022), ‘Trust in artificial intelligence: From a foundational trust framework to emerging research opportunities’, *Electronic Markets* **32**(4), 1993–2020.
40. Mahmud, H., Islam, A. N., Ahmed, S. I. and Smolander, K. (2022), ‘What influences algorithmic decision-making? a systematic literature review on algorithm aversion’, *Technological Forecasting and Social Change* **175**, 121390.
41. Mehrotra, S., Jorge, C. C., Jonker, C. M. and Tielman, M. L. (2024), ‘Integrity-based explanations for fostering appropriate trust in ai agents’, *ACM Transactions on Interactive Intelligent Systems* **14**(1), 1–36.
42. Memmert, L. and Bittner, E. (2022), ‘Complex problem solving through human-ai collaboration: Literature review on research contexts’.
43. Naiseh, M., Al-Thani, D., Jiang, N. and Ali, R. (2023), ‘How the different explanation classes impact trust calibration: The case of clinical decision support systems’, *International Journal of Human-Computer Studies* **169**, 102941.
44. Narayanan, S., Yu, G., Ho, C.-J. and Yin, M. (2023), How does value similarity affect human reliance in ai-assisted ethical decision making?, *in* ‘Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society’, pp. 49–57.

45. Nourani, M., Roy, C., Block, J. E., Honeycutt, D. R., Rahman, T., Ragan, E. and Gogate, V. (2021), Anchoring bias affects mental model formation and user reliance in explainable ai systems, *in* '26th International Conference on Intelligent User Interfaces', pp. 340–350.
46. Ochmann, J., Michels, L., Zilker, S., Tiefenbeck, V. and Laumer, S. (2020), 'The influence of algorithm aversion and anthropomorphic agent design on the acceptance of ai-based job recommendations.', *ICIS 2020 Research Papers* p. 4.
47. Okamura, K. and Yamada, S. (2020), 'Adaptive trust calibration for human-ai collaboration', *Plos one* **15**(2), e0229132.
48. Papenmeier, A., Kern, D., Englebienne, G. and Seifert, C. (2022), 'It's Complicated: The Relationship between User Trust, Model Accuracy and Explanations in AI', *ACM Transactions on Computer-Human Interaction* **29**(4).
49. Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A. and Tomsett, R. (2022), 'Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making', *Proceedings of the ACM on Human-Computer Interaction* **6**(CSCW1), 1–22.
50. Rechkemmer, A. and Yin, M. (2022), When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models, *in* 'Proceedings of the 2022 chi conference on human factors in computing systems', pp. 1–14.
51. Schemmer, M., Kuehl, N., Benz, C., Bartos, A. and Satzger, G. (2023), Appropriate reliance on ai advice: Conceptualization and the effect of explanations, *in* 'Proceedings of the 28th International Conference on Intelligent User Interfaces', pp. 410–422.
52. Schmitt, A., Wambsganss, T. and Janson, A. (2022), 'Designing for conversational system trustworthiness: the impact of model transparency on trust and task performance', *ECIS 2022 Research Papers* p. 172.
53. Schmitt, A., Wambsganss, T., Söllner, M. and Janson, A. (2021), 'Towards a trust reliance paradox? exploring the gap between perceived trust in and reliance on algorithmic advice', *ICIS 2021 Research Papers* pp. 1–17.
54. Schreuter, D., van der Putten, P. and Lamers, M. H. (2021), 'Trust me on this one: Conforming to conversational assistants', *Minds and Machines* **31**, 535–562.
55. Selten, F., Robeer, M. and Grimmelikhuijsen, S. (2023), ''just like i thought': Street-level bureaucrats trust ai recommendations if they confirm their professional judgment', *Public Administration Review* **83**(2), 263–278.
56. Sharan, N. N. and Romano, D. M. (2020), 'The effects of personality and locus of control on trust in humans versus artificial intelligence', *Heliyon* **6**(8).
57. Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H. and Seaborn, K. (2022), 'Trust in human-ai interaction: Scoping out models, measures, and methods', *CHI Conference on Human Factors in Computing Systems Extended Abstracts* pp. 1–7.
58. 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* **7**(CSCW1), 1–38.
59. Wang, X. and Yin, M. (2021), 'Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making', *26th international conference on intelligent user interfaces* pp. 318–328.

60. Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B. and Rafaeli, A. (2023), ‘Decision control and explanations in human-ai collaboration: Improving user perceptions and compliance’, *Computers in Human Behavior* **144**, 107714.
61. Wysocki, O., Davies, J. K., Vigo, M., Armstrong, A. C., Landers, D., Lee, R. and Freitas, A. (2023), ‘Assessing the communication gap between AI models and healthcare professionals: Explainability, utility and trust in AI-driven clinical decision-making’, *Artificial Intelligence* **316**, 103839.
62. Xu, L., Pardos, Z. A. and Pai, A. (2023), Convincing the expert: Reducing algorithm aversion in administrative higher education decision-making, *in* ‘Proceedings of the Tenth ACM Conference on Learning@ Scale’, pp. 215–225.
63. Yang, R. and Wibowo, S. (2022), ‘User trust in artificial intelligence: A comprehensive conceptual framework’, *Electronic Markets* **32**(4), 2053–2077.
64. Yin, M., Vaughan, J. W. and Wallach, H. (2019), ‘Understanding the effect of accuracy on trust in machine learning models’, *Conference on Human Factors in Computing Systems* pp. 1–12.
65. You, S., Yang, C. L. and Li, X. (2022), ‘Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation?’, *Journal of Management Information Systems* **39**(2), 336–365.
66. Zhang, G., Chong, L., Kotovsky, K. and Cagan, J. (2023), ‘Trust in an ai versus a human teammate: The effects of teammate identity and performance on human-ai cooperation’, *Computers in Human Behavior* **139**, 107536.
67. Zhang, Q., Lee, M. L. and Carter, S. (2022), You complete me: Human-ai teams and complementary expertise, *in* ‘Proceedings of the 2022 CHI conference on human factors in computing systems’, pp. 1–28.
68. Zhang, Y., Liao, Q. V. and Bellamy, R. K. (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* pp. 295–305.