

Supplementary Materials (AI Risk: Structuring the Debate)

Table – Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
1. Language: Full text is in English 2. "AI" stands for Artificial Intelligence 3. Topic: Discussion of AI as coming with risk 4. Relevance: A minimal discussion on the concept of risk in relation to AI	1. Language: all other languages 2. "AI" is not an abbreviation of Artificial Intelligence, e.g., avian influenza 3. Not on topic: Any other or the absence of a link between AI and risk (e.g., AI as solving a risk) 4. Not relevant: On the risk(s) of AI but without any reference to the meaning of AI risk. For example, texts that outline 10 prominent risks of AI, without minimally discussing what it means to be a risk or what an AI risk is. Such texts are excluded because they cannot answer the research question of how AI risk is understood and studied without making far-reaching assumptions as to what might have been implicitly meant.

Breakdown included literature per publication type

Journal Article	Conference proceeding	Book chapter	Preprint
60	18	2	1

To mitigate publication bias, conference proceedings and book chapters were included in the search criteria. Indeed, two book chapters were included from Springer and De Gruyter Mouton handbooks. Although they did not necessarily undergo double-blind peer review, they can be expected to be thoroughly edited by the books' editors. For all the texts published in conference proceedings, the conference procedures are known to have peer-review processes (CEUR Workshop Proceedings; FAccT; IEEE computer society; IEEE IJCNN; IEEE ISTAS; Proceedings of an AAAI and/or ACM Conference; Proceedings of the annual meeting of the association for computer linguistics; Proceedings Transdisciplinary AI). The included preprint is the highly cited text by Shevlane et al. (64 on list below) which was manually added.

Manually added literature

To mitigate indexing bias, the initial database was supplemented by fourteen texts at the identification phase of the review. These were identified by backward and forward snowballing (e.g., 15, 45, 66 on list below) or were otherwise come across as highly cited in the larger debate (e.g., 50, 64). All manually added literature were screened against the same inclusion/exclusion criteria as the sources from the databases. This led to the ultimate inclusion of five supplementary sources.

Alphabetical list of included studies

1. Aliman N-M, Elands P, Hürst W, Kester L, Thórisson KR, Werkhoven P, Yampolskiy R, Ziesche S (2020) Error-correction for ai safety. In: Lecture Notes in Computer Science. pp 12–22
2. Aliman N-M, Kester L, Yampolskiy R (2021) Transdisciplinary AI Observatory- Retrospective Analyses and Future-Oriented Contradistinctions. *Philosophies* 6:6. doi: [10.3390/philosophies6010006](https://doi.org/10.3390/philosophies6010006)
3. Alshehhi K, Cheaitou A, Rashid H (2024) Investigating risk elements and critical success factors for AI procurement projects in the public sector: a qualitative approach based on UAE public organisations. *Intl J Syst Assur Eng Manage*. doi: [10.1007/s13198-024-02653-9](https://doi.org/10.1007/s13198-024-02653-9)
4. Bao L, Krause NM, Calice MN, Scheufele DA, Wirz CD, Brossard D, Newman TP, Xenos MA (2022) Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior* 130. doi: [10.1016/j.chb.2022.107182](https://doi.org/10.1016/j.chb.2022.107182)
5. Bayer J (2024) The place of content ranking algorithms on the AI risk spectrum. *Telecommunications Policy* 48:102741. doi: [10.1016/j.telpol.2024.102741](https://doi.org/10.1016/j.telpol.2024.102741)
6. Bird C, Ungless EL, Kasirzadeh A (2023) Typology of Risks of Generative Text-to-Image Models. In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, AES 2023*. Montreal, Canada, pp 396–410
7. Botero Arcila B (2024) AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight? *Computer Law and Security Review* 54. doi: [10.1016/j.clsr.2024.106012](https://doi.org/10.1016/j.clsr.2024.106012)
8. Bruinsma B, Fredén A, Hansson K, Johansson M, Kisić-Merino P, Saynova D (2024) Setting the AI Agenda – Evidence from Sweden in the ChatGPT Era. In: Calegari R., Dignum V., O’Sullivan B. (eds) *CEUR Workshop Proc. CEUR-WS*
9. Bucknall BS, Dori-Hacohen S (2022) Current and Near-Term AI as a Potential Existential Risk Factor. In: *Proceedings of the 2022 AAI/ACM Conference on AI, Ethics, and Society, AIES 2022*. Oxford, England, pp 119–129
10. Burden J, Hernández-Orallo J (2020) Exploring AI safety in degrees: Generality, capability and control. In: *CEUR Workshop Proceedings*. pp 36–40
11. Chamberlain J (2023) The Risk-Based Approach of the European Union’s Proposed Artificial Intelligence Regulation: Some Comments from a Tort Law Perspective. *Eur j risk regul* 14:1–13. doi: [10.1017/err.2022.38](https://doi.org/10.1017/err.2022.38)
12. Cunneen M, Mullins M, Murphy F (2019) Autonomous Vehicles and Embedded Artificial Intelligence: The Challenges of Framing Machine Driving Decisions. *Applied Artificial Intelligence* 33:706–731. doi: [10.1080/08839514.2019.1600301](https://doi.org/10.1080/08839514.2019.1600301)

13. Cunneen M, Mullins M, Murphy F (2020) Artificial intelligence assistants and risk: framing a connectivity risk narrative. *AI & Society* 35:625–634. doi: [10.1007/s00146-019-00916-9](https://doi.org/10.1007/s00146-019-00916-9)
14. De Cooman J (2022) Humpty Dumpty and High-Risk AI Systems: The Ratione Materiae Dimension of the Proposal for an EU Artificial Intelligence Act. *Market and Competition Law Review* 6:49–88. doi: [10.34632/mclawreview.2022.11304](https://doi.org/10.34632/mclawreview.2022.11304)
15. Ebers M (2024) Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act. *Eur j risk regul* 1–20. doi: [10.1017/err.2024.78](https://doi.org/10.1017/err.2024.78)
16. Faroldi FLG (2024) Risk and artificial general intelligence. *AI & Society* 40:2541–2549. doi: [10.1007/s00146-024-02004-z](https://doi.org/10.1007/s00146-024-02004-z)
17. Fraser H, Bello Y, Villarino J-M (2024) Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough. *Eur j risk regul* 15:431–446. doi: [10.1017/err.2023.57](https://doi.org/10.1017/err.2023.57)
18. Giudici P, Centurelli M, Turchetta S (2024) Artificial Intelligence risk measurement. *Expert Systems with Applications* 235:121220. doi: [10.1016/j.eswa.2023.121220](https://doi.org/10.1016/j.eswa.2023.121220)
19. Golpayegani D, Hovsha J, Rossmair LWS, Saniei R, Misic J (2022) Towards a Taxonomy of AI Risks in the Health Domain. In: *Proceedings - 2022 4th International Conference on Transdisciplinary AI, TransAI 2022*. pp 1–8
20. Golpayegani D, Pandit HJ, Lewis D (2022) AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards. *Towards a Knowledge-Aware AI* 55:51–65. doi: [10.3233/SSW220008](https://doi.org/10.3233/SSW220008)
21. Golpayegani D, Pandit HJ, Lewis D (2023) To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards. In: *Proceedings on the 6th ACM Conference on Fairness, Accountability, and Transparency, FAccT 2023*. Chicago, pp 905–915
22. González-Arias C, López-García X (2024) Rethinking the Relation between Media and Their Audience: The Discursive Construction of the Risk of Artificial Intelligence in the Press of Belgium, France, Portugal, and Spain. *Journal Media* 5:1023–1037. doi: [10.3390/journalmedia5030065](https://doi.org/10.3390/journalmedia5030065)
23. Gupta S, Kamboj S, Bag S (2023) Role of Risks in the Development of Responsible Artificial Intelligence in the Digital Healthcare Domain. *Information Systems Frontiers* 25:2257–2274. doi: [10.1007/s10796-021-10174-0](https://doi.org/10.1007/s10796-021-10174-0)
24. Hacker P (2021) A legal framework for AI training data—from first principles to the Artificial Intelligence Act. *Law, Innovation and Technology* 13:257–301. doi: [10.1080/17579961.2021.1977219](https://doi.org/10.1080/17579961.2021.1977219)
25. Haduong N, Gao A, Smith NA (2023) Risks and NLP Design: A Case Study on Procedural Document QA. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. pp 1248–1269
26. Hassel A, Oezkiziltan D (2023) Governing the work-related risks of AI: implications for the German government and trade unions. *Transfer-European Review of Labour and Research* 29:71–86. doi: [10.1177/10242589221147228](https://doi.org/10.1177/10242589221147228)
27. Hohma E, Boch A, Trauth R, Luetge C (2023) Investigating accountability for Artificial Intelligence through risk governance: A workshop-based exploratory study. *Frontiers in Psychology* 14:1073686. doi: [10.3389/fpsyg.2023.1073686](https://doi.org/10.3389/fpsyg.2023.1073686)
28. Hutiri W, Papakyriakopoulos O, Xiang A (2024) Not My Voice! A Taxonomy of Ethical and Safety Harms of Speech Generators. In: *2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2024*. pp 359–376

29. Inverardi N, Bertaina S, Biganzoli I, Cosentini A, Desiante R, Fontanella D, Penco IG (2024) Fundamental Rights and AI Impact Assessment: A proposal for a new quantitative approach. In: Proc Int Jt Conf Neural Networks
30. Jia K, Zhang N (2022) Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines. *Electronic Markets* 32:59–71. doi: [10.1007/s12525-021-00480-5](https://doi.org/10.1007/s12525-021-00480-5)
31. Jordan SR (2019) Designing Artificial Intelligence Review Boards Creating Risk Metrics for Review of AI. In: Cunningham M, Cunningham P (eds) 2019 IEEE International Symposium on Technology and Society (ISTAS)
32. Jordan SR (2020) A Costly Emergencies Approach to Estimating Costs for Artificial Intelligence Review. In: Proceedings of the 2020 IEEE International Symposium on Technology and Society (ISTAS). *ELECTR NETWORK*, pp 180–191
33. Kaminski ME (2023) Regulating the Risks of AI. *Boston University Law Review* 103:1347–1411
34. Kilian KA, Ventura CJ, Bailey MM (2023) Examining the differential risk from high-level artificial intelligence and the question of control. *Futures* 151:103182. doi: [10.1016/j.futures.2023.103182](https://doi.org/10.1016/j.futures.2023.103182)
35. Kim J (2024) Defining risk in AI legislation: perspectives and implications in the Korean context. *Communication Research and Practice* 10:316–327. doi: [10.1080/22041451.2024.2346417](https://doi.org/10.1080/22041451.2024.2346417)
36. Krieger JB, Boudier F, Wibral M, Almeida RJ (2024) A systematic literature review on risk perception of Artificial Narrow Intelligence. *Journal of Risk Research*. doi: [10.1080/13669877.2024.2350725](https://doi.org/10.1080/13669877.2024.2350725)
37. Kusche I (2024) Possible harms of artificial intelligence and the EU AI act: fundamental rights and risk. *Journal of Risk Research*. doi: [10.1080/13669877.2024.2350720](https://doi.org/10.1080/13669877.2024.2350720)
38. Lahiri S, Saltz J (2024) AI Project and Deployment Risk: Articulation and Legitimization. In: Bui T.X. (ed) Proc. Annu. Hawaii Int. Conf. Syst. Sci. IEEE Computer Society, pp 5836–5845
39. Lee SU, Perera H, Xia B, Liu Y, Lu Q, Zhu L, Salvado O, Whittle J (2024) QB4AIRA: A Question Bank for Responsible AI Risk Assessment. *IEEE Software*. doi: [10.1109/MS.2024.3512577](https://doi.org/10.1109/MS.2024.3512577)
40. Mahler T (2022) Between Risk Management and Proportionality: The Risk-Based Approach in the EU’s Artificial Intelligence Act Proposal. *IRI* 247–270. doi: [10.53292/208f5901.38a67238](https://doi.org/10.53292/208f5901.38a67238)
41. McLean S, Read GJM, Thompson J, Baber C, Stanton NA, Salmon PM (2023) The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental and Theoretical Artificial Intelligence* 35:649–663. doi: [10.1080/0952813X.2021.1964003](https://doi.org/10.1080/0952813X.2021.1964003)
42. Müller VC, Cannon M (2021) Existential risk from AI and orthogonality: Can we have it both ways? *RATIO* 35:25–36. doi: [10.1111/rati.12320](https://doi.org/10.1111/rati.12320)
43. Neri H, Cozman F (2020) The role of experts in the public perception of risk of artificial intelligence. *AI & Society* 35:663–673. doi: [10.1007/s00146-019-00924-9](https://doi.org/10.1007/s00146-019-00924-9)
44. Neyazi TA, Ng SWT, Hobbs M, Yue A (2023) Understanding user interactions and perceptions of AI risk in Singapore. *Big Data & Society*:1-13. doi: [10.1177/20539517231213823](https://doi.org/10.1177/20539517231213823)

45. Nguyen D (2023) How news media frame data risks in their coverage of big data and AI. *Internet Policy Review* 12. doi: [10.14763/2023.2.1708](https://doi.org/10.14763/2023.2.1708)
46. Niklas J, Dencik L (2024) Data justice in the “twin objective” of market and risk: How discrimination is formulated in EU’s AI policy. *Policy & Internet*. doi: [10.1002/poi3.392](https://doi.org/10.1002/poi3.392)
47. Novelli C, Casolari F, Rotolo A, Taddeo M, Floridi L (2023) Taking AI risks seriously: a new assessment model for the AI Act. *AI & Society*. doi: [10.1007/s00146-023-01723-Z](https://doi.org/10.1007/s00146-023-01723-Z)
48. Novelli C, Casolari F, Rotolo A, Taddeo M, Floridi L (2024) AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act. *DISO* 3:13. doi: [10.1007/s44206-024-00095-1](https://doi.org/10.1007/s44206-024-00095-1)
49. Onitui D, Wachter S, Mittelstadt B (2024) How AI challenges the medical device regulation: patient safety, benefits, and intended uses. *Journal of Law and the Biosciences*. doi: [10.1093/jlb/lxae007](https://doi.org/10.1093/jlb/lxae007)
50. Paul R (2023) European artificial intelligence “trusted throughout the world”: Risk-based regulation and the fashioning of a competitive common AI market. *Regulation & Governance*. doi: [10.1111/rego.12563](https://doi.org/10.1111/rego.12563)
51. Ploug T, Holm S (2023) The right to a second opinion on Artificial Intelligence diagnosis-Remedying the inadequacy of a risk-based regulation. *Bioethics* 37:303–311. doi: [10.1111/bioe.13124](https://doi.org/10.1111/bioe.13124)
52. Png M-T (2022) At the Tensions of South and North: Critical Roles of Global South Stakeholders in AI Governance. In: *ACM International Conference Proceeding Series*. pp 1434–1445
53. Recki L, Lawo D, Krauß V, Pins D, Boden A (2024) “You Can either Blame Technology or Blame a Person…” - A Conceptual Model of Users’ AI-Risk Perception as a Tool for HCI. *Proc ACM Hum Comput Interact* 8. doi: [10.1145/3686996](https://doi.org/10.1145/3686996)
54. Ricciardi Celsi L (2023) The Dilemma of Rapid AI Advancements: Striking a Balance between Innovation and Regulation by Pursuing Risk-Aware Value Creation. *Information* 14:645. doi: [10.3390/info14120645](https://doi.org/10.3390/info14120645)
55. Sattlegger A, Bharosa N (2024) Embedding ethical AI risks in public sector risk management practice An explorative analysis using the Three Lines of Defense risk management model. In: *ACM International Conference Proceeding Series*. pp 70–80
56. Schiff DS, Kelley S, Camacho Ibáñez J (2024) The emergence of artificial intelligence ethics auditing. *Big Data Soc* 11. doi: [10.1177/20539517241299732](https://doi.org/10.1177/20539517241299732)
57. Schnitzer R, Hapfelmeier A, Gaube S, Zillner S (2024) AI Hazard Management: A Framework for the Systematic Management of Root Causes for AI Risks. In: *Frontiers of Artificial Intelligence, Ethics, and Multidisciplinary Applications*. pp 359–375
58. Schuett J (2023) Risk Management in the Artificial Intelligence Act. *European Journal of Risk Regulation* 15:367–385. doi: [10.1017/err.2023.1](https://doi.org/10.1017/err.2023.1)
59. Schuett J (2023) Three lines of defense against risks from AI. *AI & Society*. doi: [10.1007/s00146-023-01811-0](https://doi.org/10.1007/s00146-023-01811-0)
60. Schuett J (2024) Frontier AI developers need an internal audit function. *Risk Anal*. doi: [10.1111/risa.17665](https://doi.org/10.1111/risa.17665)
61. Schwarz A (2024) The mediated amplification of societal risk and risk governance of artificial intelligence: technological risk frames on YouTube and their impact before and after ChatGPT. *J Risk Res*. doi: [10.1080/13669877.2024.2437629](https://doi.org/10.1080/13669877.2024.2437629)

62. Schwarz A, Faj T (2023) Communicating and perceiving risks of artificial intelligence as an emerging technology. In: *Communicating Risk and Safety*. pp 503–526
63. Schwesig R, Brich I, Buder J, Huff M, Said N (2023) Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people’s willingness to use AI. *Journal of Risk Research* 26:1053–1084. doi: [10.1080/13669877.2023.2249927](https://doi.org/10.1080/13669877.2023.2249927)
64. Shevlane T, Farquhar S, Garfinkel B, Phuong M, Whittlestone J, Leung J, Kokotajlo D, Marchal N, Anderljung M, Kolt N, Ho L, Siddarth D, Avin S, Hawkins W, Kim B, Gabriel I, Bolina V, Clark J, Bengio Y, Christiano P, Dafoe A (2023) Model evaluation for extreme risks. doi:10.48550/arXiv.2305.15324
65. Siapka A (2024) A virtue ethics approach to AI-induced risk. *Teoria* 44:95–116. doi: [10.4454/wj35mz94](https://doi.org/10.4454/wj35mz94)
66. Slattery P, Saeri AK, Grundy EAC, Graham J, Noetel M, Uuk R, Dao J, Pour S, Casper S, Thompson N (2024) The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. *AGI* 1. doi: [10.70777/agi.v1i1.10881](https://doi.org/10.70777/agi.v1i1.10881)
67. Sotala K, Yampolskiy R (2015) Responses to catastrophic AGI risk: A Survey. *Physica Scripta* 90. doi: [10.1088/0031-8949/90/6/069501](https://doi.org/10.1088/0031-8949/90/6/069501)
68. Steimers A, Schneider M (2022) Sources of Risk of AI Systems. *International Journal of Environmental Research and Public Health* 19:3641. doi: [10.3390/ijerph19063641](https://doi.org/10.3390/ijerph19063641)
69. Stein AL (2022) Assuming the Risks of Artificial Intelligence. *Boston University Law Review* 102:979–1035
70. Thekdi S, Tatar U, Santos J, Chatterjee S (2023) Disaster risk and artificial intelligence: A framework to characterize conceptual synergies and future opportunities. *Risk Analysis* 43:1641–1656. doi: [10.1111/risa.14038](https://doi.org/10.1111/risa.14038)
71. Tjoa S, Kieseberg P, Temper M, Zanol J, Wagner M, Holzinger A (2022) AIRMan: An Artificial Intelligence (AI) Risk Management System. In: *2022 International Conference on Advanced Enterprise Information System*. London, ENGLAND, pp 72–81
72. Tong ST, Sopory P (2019) Does integral affect influence intentions to use artificial intelligence for skin cancer screening? A test of the affect heuristic. *Psychology & Health* 34:828–849. doi: [10.1080/08870446.2019.1579330](https://doi.org/10.1080/08870446.2019.1579330)
73. Turchin A (2019) Assessing the future plausibility of catastrophically dangerous AI. *Futures* 107:45–58. doi: [10.1016/j.futures.2018.11.007](https://doi.org/10.1016/j.futures.2018.11.007)
74. Turchin A, Denkenberger D (2020) Classification of global catastrophic risks connected with artificial intelligence. *AI & Society* 35:147–163. doi: [10.1007/s00146-018-0845-5](https://doi.org/10.1007/s00146-018-0845-5)
75. Vähäkainu P, Lehto M (2022) Use of Artificial Intelligence in a Cybersecurity Environment. In: *Artificial Intelligence and Cybersecurity: Theory and Applications*. pp 3–27
76. Vulpe S-N, Rughiniş R, Țurcanu D, Rosner D (2024) AI and cybersecurity: a risk society perspective. *Frontier Comput Sci* 6. doi: [10.3389/fcomp.2024.1462250](https://doi.org/10.3389/fcomp.2024.1462250)
77. Wilczek B, Thäsler-Kordonouri S, Eder M (2024) Government regulation or industry self-regulation of AI? Investigating the relationships between uncertainty avoidance, people’s AI risk perceptions, and their regulatory preferences in Europe. *AI Soc*. doi: [10.1007/s00146-024-02138-0](https://doi.org/10.1007/s00146-024-02138-0)

78. Wirtz BW, Weyerer JC, Kehl I (2022) Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly* 39:101685. doi: [10.1016/j.giq.2022.101685](https://doi.org/10.1016/j.giq.2022.101685)
79. Wissing BG, Reinhard M-A (2018) Individual differences in risk perception of artificial intelligence. *Swiss Journal of Psychology* 77:149–157. doi: [10.1024/1421-0185/a000214](https://doi.org/10.1024/1421-0185/a000214)
80. Xia B, Lu Q, Perera H, Zhu L, Xing Z, Liu Y, Whittle J (2023) Towards Concrete and Connected AI Risk Assessment (C2AIRA): A Systematic Mapping Study. In: 2023 IEEE/ACM 2nd International Conference on AI Engineering - Software Engineering for AI. pp 104–116
81. Zanotti G, Chiffi D, Schiaffonati V (2024) AI-Related Risk: An Epistemological Approach. *Philosophy and Technology* 37. doi: [10.1007/s13347-024-00755-7](https://doi.org/10.1007/s13347-024-00755-7)