

## References

- Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. “Concrete Problems in AI Safety.” July 25, 2016. <https://doi.org/10.48550/arXiv.1606.06565>.
- Anderson, T. 2007. “Visualization Tools and Argument Schemes: A Question of Standpoint.” *Law, Prob. & Risk* 6: 97. [https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/lawprisk6&section=9](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/lawprisk6&section=9).
- Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: a model of artificial intelligence development. *AI & society*, 31, 201-206.
- Babakov, N., Sivaprasad, A., Reiter, E., & Bugarín-Diz, A. (2025). Reusability of Bayesian Networks case studies: A survey. *Applied Intelligence*, 55(6), 417. <https://doi.org/10.1007/s10489-025-06289-5>
- Ban, T., Chen, L., Lyu, D., Wang, X., & Chen, H. (2023). Causal structure learning supervised by large language model. *arXiv preprint arXiv:2311.11689*.
- Benn N., Macintosh A. 2011. Argument Visualization for eParticipation: Towards a Research Agenda and Prototype Tool. 6847. 60-73. 10.1007/978-3-642-23333-3\_6.
- Bethard, Steven John. 2007. “Finding Event, Temporal and Causal Structure in Text: A Machine Learning Approach.” PhD thesis, University of Colorado at Boulder. <https://search.proquest.com/openview/405fe32503123d9b5f4836dc3be4c011/1?pq-origsite=gscholar&cbl=18750>.
- Box, G. E. P. 1976. “Science and Statistics.” *Journal of the American Statistical Association* 71 (356): 791–99. <https://doi.org/10.1080/01621459.1976.10480949>.
- Bucknall, B. S., & Dori-Hacohen, S. (2022). Current and Near-Term AI as a Potential Existential Risk Factor. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 119–129. <https://doi.org/10.1145/3514094.3534146>
- Carlsmith, J. 2024. “Is Power-Seeking AI an Existential Risk?” August 13, 2024. <https://doi.org/10.48550/arXiv.2206.13353>.
- Carlsmith, J. 2022. “Is Power-Seeking AI an Existential Risk?” <https://arxiv.org/abs/2206.13353>.
- Carlsmith, J. 2021. “Is Power-Seeking AI an Existential Risk?” 2021. <https://doi.org/10.48550/arXiv.2206.13353>.
- Chen, Lu, Ruqing Zhang, Wei Huang, Wei Chen, Jiafeng Guo, and Xueqi Cheng. 2023. “Inducing Causal Structure for Abstractive Text Summarization.” In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 213–23. Birmingham United Kingdom: ACM. <https://doi.org/10.1145/3583780.3614934>.
- Christiano, P. 2019. “What Failure Looks Like,” March. <https://www.alignmentforum.org/posts/HBxe6wdjxK239zajf/what-failure-looks-like>.
- Clarke, Sam & Cottier, Ben & Englander, Aryeh & Eth, Daniel & Manheim, David & Martin, Samuel & Rice, Issa. (2022). Modeling Transformative AI Risks (MTAIR) Project – Summary Report. 10.48550/arXiv.2206.09360.
- Cuomo, Francesca, Christine Mallin, and Alessandro Zattoni. 2016. “Corporate Governance Codes: A Review and Research Agenda.” *Corporate Governance: An International Review* 24 (3): 222–41. <https://ueaeprints.uea.ac.uk/id/eprint/57664/>.

- Dafoe A. 2021. "AI Governance: A Research Agenda." 2021. <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>.
- Dafoe, A. 2018. "AI Governance: A Research Agenda." Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK 1442: 1443. <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf>.
- Demirag, Istemi, Sudi Sudarsanam, and MIKE WRIGHT. 2000. "Corporate Governance: Overview and Research Agenda." *The British Accounting Review* 32 (4): 341–54. <https://www.academia.edu/download/49469624/bare.2000.014620161009-3955-1dt4aq5.pdf>.
- De Villiers, Charl, and Ruth Dimes. 2021. "Determinants, Mechanisms and Consequences of Corporate Governance Reporting: A Research Framework." *Journal of Management and Governance* 25 (1): 7–26. <https://doi.org/10.1007/s10997-020-09530-0>.
- Di Vito, Jackie, and Kim Trottier. 2022. "A Literature Review on Corporate Governance Mechanisms: Past, Present, and Future\*." *Accounting Perspectives* 21 (2): 207–35. <https://doi.org/10.1111/1911-3838.12279>.
- Duhem, Pierre Maurice Marie. 1954. *The Aim and Structure of Physical Theory*. 1. Princeton University Press.
- Gruetzmacher, R. 2022. "Bayesian Networks Vs. Conditional Trees for Creating Questions for Forecasting Tournaments."
- Hallegatte, S., Shah A., Lempert, R. Brown C., Gill S. 2012. "Investment Decision-Making Under Deep Uncertainty-Application to Climate Change." Policy Research Working Paper 6193. <https://enpc.hal.science/hal-00802049/document>.
- Heinze-Deml, Christina, Marloes H. Maathuis, and Nicolai Meinshausen. 2018. "Causal Structure Learning." *Annual Review of Statistics and Its Application* 5 (1): 371–91. <https://doi.org/10.1146/annurev-statistics-031017-100630>.
- Hunt, T. 2025. "The Insane 'Logic' of the AI Arms Race." Medium. March 3, 2025. <https://tamhunt.medium.com/the-insane-logic-of-the-ai-arms-race-45a5f79f4c0e>.
- Kaur, Kawaljit. 2024. "Corporate Governance and Legal Accountability: A Critical Review of Global Practices." *Journal of Law* 2 (6): 1–7. <https://joi.shodhsagar.org/index.php/SSJOI/article/view/16>.
- Khartabil, D., C. Collins, S. Wells, B. Bach, and J. Kennedy. 2021. "Design and Evaluation of Visualization Techniques to Facilitate Argument Exploration." *Computer Graphics Forum* 40 (6): 447–65. <https://doi.org/10.1111/cgf.14389>.
- Khartabil, Dana. 2020. "Visualisation Techniques to Facilitate Argument Exploration." PhD thesis. <https://napier-repository.worktribe.com/output/2694675>.
- Koller, Daphne, and Nir Friedman. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT press. [https://books.google.ca/books?hl=en&lr=&id=7dzpHCHzNQ4C&oi=fnd&pg=PR9&dq=Koller,+D.,+%26+Friedman,+N.+\(2009\).+Probabilistic+Graphical+Models&ots=py2HAh0VAL&sig=gpaID3x6-TY8x5SOopuXpZDXfzs](https://books.google.ca/books?hl=en&lr=&id=7dzpHCHzNQ4C&oi=fnd&pg=PR9&dq=Koller,+D.,+%26+Friedman,+N.+(2009).+Probabilistic+Graphical+Models&ots=py2HAh0VAL&sig=gpaID3x6-TY8x5SOopuXpZDXfzs).
- Lindley, Dennis V. 2013. *Understanding Uncertainty*. John Wiley & Sons. [https://books.google.ca/books?hl=en&lr=&id=Tfk8AgAAQBAJ&oi=fnd&pg=PR11&dq=Lindley,+D.+\(2006\).+Understanding+Uncertainty&ots=55HS6lTOVP&sig=0gKCDvRu5rUKhuPyJqhOzW23upU](https://books.google.ca/books?hl=en&lr=&id=Tfk8AgAAQBAJ&oi=fnd&pg=PR11&dq=Lindley,+D.+(2006).+Understanding+Uncertainty&ots=55HS6lTOVP&sig=0gKCDvRu5rUKhuPyJqhOzW23upU).
- List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press.
- Manheim, David. 2021. "Modeling Transformative AI Risk (MTAIR) - LessWrong." July 28, 2021. <https://www.lesswrong.com/s/aERZoriyHfCqvWkzg>.
- Maslej N., Fattorini I., Perrault R., Gil Y., Parli V., Kariuki N., Capstick E., Reuel A., Brynjolfsson E., Etchemendy J., Ligett K., Lyons T., Manyika J., Niebles J., Shoham Y., Wald R., Walsh T., Hamrah A.,

- Santarlaschi L., Betts Lotufo J., Rome A., Shi A., Oak S. “The AI Index 2025 Annual Report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025.
- McCaslin, T., Rosenberg J., Karger E., Morris A., Hickman M., Glover S., Jacobs Z., Tetlock P. 2024. “Conditional Trees: A Method for Generating Informative Questions about Complex Topics.” Forecasting Research Institute. <https://static1.squarespace.com/static/635693acf15a3e2a14a56a4a/t/66ba37a144f1d6095de467df/1723479995772/AIConditionalTrees.pdf>.
- Metropolitansky D., Larson J. 2025. “Towards Effective Extraction and Evaluation of Factual Claims.” February 15, 2025. <https://doi.org/10.48550/arXiv.2502.10855>.
- Miotti, Andrea, Tolga Bilge, Dave Kasten, and James Newport. 2024. “A Narrow Path.” <https://www.narrowpath.co/>.
- Nelson, Roger B. 2006. *An Introduction to Copulas*. Springer Series in Statistics. New York, NY: Springer New York. <https://doi.org/10.1007/0-387-28678-0>.
- Ngajie, Berty Nsolly, Yan Li, Dawit Tibebe Tiruneh, and Mengmeng Cheng. 2020. “Investigating the Effects of a Systematic and Model-Based Design of Computer-Supported Argument Visualization on Critical Thinking.” *Thinking Skills and Creativity* 38: 100742. <https://www.sciencedirect.com/science/article/pii/S1871187120302169>.
- Paul, M. (2024, September 15) The elephAInt – Are we all like the six blind men when it comes to AI? PRISMAGuard <https://www.prismaguard.com/the-elephaint-are-we-all-like-the-six-blind-men-when-it-comes-to-ai/>
- Pearl, J. 2014. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Elsevier. [https://books.google.ca/books?hl=en&lr=&id=mn2jBQAAQBAJ&oi=fnd&pg=PP1&dq=Pearl,+J.+\(1988\).+Probabilistic+Reasoning+in+Intelligent+Systems&ots=4tEX2A4Ha8&sig=lgUs\\_RCoEXEuGwM5xMEoyJy4HI](https://books.google.ca/books?hl=en&lr=&id=mn2jBQAAQBAJ&oi=fnd&pg=PP1&dq=Pearl,+J.+(1988).+Probabilistic+Reasoning+in+Intelligent+Systems&ots=4tEX2A4Ha8&sig=lgUs_RCoEXEuGwM5xMEoyJy4HI).
- Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K. ; New York: Cambridge University Press.
- Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. 2nd ed. Cambridge University Press.
- Pollock, John L. 1995. *Cognitive Carpentry: A Blueprint for How to Build a Person*. Mit Press. [https://books.google.ca/books?hl=en&lr=&id=JAfHrHTqswAC&oi=fnd&pg=PA1&dq=Pollock,+J.+\(1995\).+Cognitive+Carpentry&ots=rq-qSCBcxV&sig=aAfHGSGUosxl\\_1-JuxIEA7C2QO4](https://books.google.ca/books?hl=en&lr=&id=JAfHrHTqswAC&oi=fnd&pg=PA1&dq=Pollock,+J.+(1995).+Cognitive+Carpentry&ots=rq-qSCBcxV&sig=aAfHGSGUosxl_1-JuxIEA7C2QO4).
- Prokudin, D. E., E. N. Lisanyuk, and I. R. Baymuratov. 2024. “Visualization Functions in Argumentation Representation Software.” *Scientific Visualization* 16 (3). <https://sv-journal.org/2024-3/11/en.pdf>.
- Rehman, I. 2025. “The Battle for Brilliant Minds: From the Nuclear Age to AI.” *War on the Rocks*. January 13, 2025. <https://warontherocks.com/2025/01/the-battle-for-brilliant-minds-from-the-nuclear-age-to-ai/>.
- Samborska V. (2025) - “Scaling up: how increasing inputs has made artificial intelligence more capable” Published online at OurWorldinData.org. Retrieved from: ‘<https://ourworldindata.org/scaling-up-ai>’ [Online Resource]
- Samuel, S. 2023. “AI Is a ‘Tragedy of the Commons.’ We’ve Got Solutions for That.” *Vox*. July 7, 2023. <https://www.vox.com/future-perfect/2023/7/7/23787011/ai-arms-race-tragedy-commons-risk-safety>.
- Schelling, T. 1960. “1960. *The Strategy of Conflict*.” Cambridge, Mass.
- Scheuer, Oliver, Frank Loll, Niels Pinkwart, and Bruce M. McLaren. 2010. “Computer-Supported Argumentation: A Review of the State of the Art.” *International Journal of Computer-Supported Collaborative Learning* 5 (1): 43–102. <https://doi.org/10.1007/s11412-009-9080-x>.
- Solomon, Jill. 2020. *Corporate Governance and Accountability*. John Wiley & Sons. [https://books.google.ca/books?hl=en&lr=&id=JAX9DwAAQBAJ&oi=fnd&pg=PR1&dq=review+of+the+effects+of+liability+frameworks+on+corporate+governance+&ots=ny23\\_vd-U0&sig=3LuNNhvSWXriEeg-ipAdDIQGAgO](https://books.google.ca/books?hl=en&lr=&id=JAX9DwAAQBAJ&oi=fnd&pg=PR1&dq=review+of+the+effects+of+liability+frameworks+on+corporate+governance+&ots=ny23_vd-U0&sig=3LuNNhvSWXriEeg-ipAdDIQGAgO).

- Squires, Chandler, and Caroline Uhler. 2023. “Causal Structure Learning: A Combinatorial Perspective.” *Foundations of Computational Mathematics* 23 (5): 1781–1815. <https://doi.org/10.1007/s10208-022-09581-9>.
- Tegmark, M. 2024. “Asilomar AI Principles.” Future of Life Institute. 2024. <https://futureoflife.org/open-letter/ai-principles/>.
- Tetlock, Phil. 2022. “Conditional Trees: AI Risk.” 2022. <https://www.metaculus.com/tournament/3508/>.
- Voigt, Christian. (2014) 2025. “Christianvoigt/Argdown.” <https://github.com/christianvoigt/argdown>.
- Walton, Douglas. 2009. “Argument Visualization Tools for Corroborative Evidence.” In *Proc. Of the 2nd International Conference on Evidence Law and Forensic Science*, 32–49. <https://www.academia.edu/download/37718171/09ArguVis.pdf>.
- Yang, Jie, Soyeon Caren Han, and Josiah Poon. 2022. “A Survey on Extraction of Causal Relations from Natural Language Text.” *Knowledge and Information Systems* 64 (5): 1161–86. <https://doi.org/10.1007/s10115-022-01665-w>.
- Yudkowsky, Eliezer. 2008. “Artificial Intelligence as a Positive and Negative Factor in Global Risk.” In *Global Catastrophic Risks*, by Eliezer Yudkowsky. Oxford University Press. <https://doi.org/10.1093/oso/9780198570509.003.0021>.