Chang, Welton et al. (2016). "Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments". In: *Judgment and Decision making* 11.5, pp. 509–526.
Cheng, Ruijia et al. (2024). ""It would work for me too": How Online Communities Shape Software Developers' Trust in AI-Powered Code Generation Tools". In: *ACM Transactions on Interactive Intelligent Systems* 14.2, pp. 1–39.
Chiang, Wei-Lin et al. (2024). *Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference*. arXiv: 2403.04132 [cs.AI].
Choi, Jonathan H and Daniel Schwarcz (2024). "AI Assistance in Legal Analysis: An Empirical Study". In: *Journal of Legal Education* 73. Forthcoming.
Dardaman, Emily and Abhishek Gupta (2023). "Asking Better Questions: The Art and Science of Forecasting". In: *CHI 2023 Designing Technology and Policy Simultaneously: Towards A Research Agenda and New Practice Workshop*. Hamburg, Germany: ACM.
Davis-Stober, Clintin P. et al. (2014). "When is a Crowd Wise?" In: *Decision* 1.2, p. 79.
Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski (2023). "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination". In: *Review of Economics and Statistics*, pp. 1–47.
Dell'Acqua, Fabrizio et al. (2023). "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality". In: *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013.
Depounti, Iliana, Paula Saukko, and Simone Natale (2023). "Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend". In: *Media, Culture & Society* 45.4, pp. 720–736.
Doshi, Anil R and Oliver Hauser (2023). *Generative artificial intelligence enhances creativity*. URL: https://ssrn.com/abstract=4535536.
Douglas, Benjamin D, Patrick J Ewell, and Markus Brauer (2023). "Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA". In: *Plos one* 18.3, e0279720.
Firat, Mehmet and Saniye Kuleli (2023). "What if GPT4 became autonomous: The Auto-GPT project and use cases". In: *Journal of Emerging Computer Technologies* 3.1, pp. 1–6.
Fraiwan, Mohammad and Natheer Khasawneh (2023). *A Review of ChatGPT Applications in Education, Marketing, Software Engineering, and Healthcare: Benefits, Drawbacks, and Research Directions*. arXiv: 2305.00237 [cs.CY].
Gao, Jie et al. (2024). "A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration". In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–11.
George, A Shaji and T Baskar (2023). "The Impact of AI Language Models on the Future of White-Collar Jobs: A Comparative Study of Job Projections in Developed and Developing Countries". In: *Partners Universal International Research Journal* 2.2, pp. 117–135.
Goyal, Tanya, Junyi Jessy Li, and Greg Durrett (2023). *News Summarization and Evaluation in the Era of GPT-3*. arXiv: 2209.12356 [cs.CL].
Grove, Nathaniel P and Stacey Lowery Bretz (2012). "A Continuum of Learning: From Rote Memorization to Meaningful Learning in Organic Chemistry". In: *Chemistry Education Research and Practice* 13.3, pp. 201–208.
Gruver, Nate et al. (2023). *Large Language Models Are Zero-Shot Time Series Forecasters*. arXiv: 2310.07820 [cs.LG].
Guo, Yi et al. (2024). "Talk2data: A natural language interface for exploratory visual analysis via question decomposition". In: *ACM Transactions on Interactive Intelligent Systems* 14.2, pp. 1–24.
Halawi, Danny et al. (2024). "Approaching Human-Level Forecasting with Language Models". In: *arXiv preprint arXiv:2402.18563*.
Hazell, Julian (2023). *Spear Phishing With Large Language Models*. arXiv: 2305.06972 [cs.CY].
Heiding, Fredrik et al. (2023). "Devising and detecting phishing: Large language models vs. smaller human models". In: *arXiv preprint arXiv:2308.12287*.
Himmelstein, Mark, David V Budescu, and Ying Han (2023). "The Wisdom of Timely Crowds". In: *Judgment in Predictive Analytics*. Springer, pp. 215–242.
Himmelstein, Mark et al. (Jan. 2024). *The Forecasting Proficiency Test: A Practical Forecaster Evaluation Tool*. Conference Presentation. Helsinki, Finland.
Jiao, Wenxiang et al. (2023). *Is ChatGPT a Good Translator? Yes with GPT-4 as the Engine*. arXiv: 2301.08745 [cs.CL].

Jin, Ming et al. (2023). "Time-llm: Time series forecasting by reprogramming large language models". In: *arXiv preprint arXiv:2310.01728*.
Jones, Cameron R and Benjamin K Bergen (2024). "People cannot distinguish GPT-4 from a human in a Turing test". In: *arXiv preprint arXiv:2405.08007*.
Karger, Ezra, Pavel D. Atanasov, and Philip Tetlock (2022). "Improving judgments of existential risk: Better forecasts, questions, explanations, policies". In: *Questions, Explanations, Policies (January 17, 2022)*.
Karvetski, Christopher W et al. (2022). "What do Forecasting Rationales Reveal about Thinking Patterns of Top Geopolitical Forecasters?" In: *International Journal of Forecasting* 38.2, pp. 688–704.
Kasparov, Garry (2010). "The chess master and the computer". In: *The New York Review of Books* 57.2, pp. 16–19.
Katz, Daniel Martin et al. (2023). "GPT-4 Passes the Bar Exam". In: *SSRN*. URL: https://ssrn.com/abstract=4389233.
Kinniment, Megan et al. (2023). "Evaluating language-model agents on realistic autonomous tasks". In: *arXiv preprint arXiv:2312.11671*.
Kjærland, Frode et al. (2018). "An analysis of bitcoin's price dynamics". In: *Journal of Risk and Financial Management* 11.4, p. 63.
Li, Daliang et al. (2023). "Large Language Models with Controllable Working Memory". In: *Findings of the Association for Computational Linguistics: ACL 2023*. Toronto, Canada: Association for Computational Linguistics, pp. 1774–1793. DOI: 10.18653/v1/2023.findings-acl.112. URL: https://aclanthology.org/2023.findings-acl.112.
Lichtenstein, Sarah and Baruch Fischhoff (1977). "Do those who know more also know more about how much they know?" In: *Organizational behavior and human performance* 20.2, pp. 159–183.
Magar, Inbal and Roy Schwartz (May 2022). "Data Contamination: From Memorization to Exploitation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 157–165. DOI: 10.18653/v1/2022.acl-short.18. URL: https://aclanthology.org/2022.acl-short.18.
Mannes, Albert E., Jack B. Soll, and Richard P. Larrick (2014). "The Wisdom of Select Crowds". In: *Journal of Personality and Social Psychology* 107.2, p. 276.
McAndrew, Thomas et al. (2022a). "Chimeric Forecasting: Combining Probabilistic Predictions from Computational Models and Human Judgment". In: *BMC Infectious Diseases* 22.1, p. 833.
McAndrew, Thomas et al. (2022b). "Early Human Judgment Forecasts of Human Monkeypox, May 2022". In: *The Lancet Digital Health* 4.8, e569–e571.
McAndrew, Thomas et al. (2024). "Assessing Human Judgment Forecasts in the Rapid Spread of the Mpox Outbreak: Insights and Challenges for Pandemic Preparedness". In: *arXiv preprint arXiv:2404.14686*.
McIntosh, Timothy R et al. (2024). "A Reasoning and Value Alignment Test to Assess Advanced GPT Reasoning". In: *ACM Transactions on Interactive Intelligent Systems*.
Mellers, Barbara et al. (2015a). "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions". In: *Perspectives on Psychological Science* 10.3, pp. 267–281.
Mellers, Barbara et al. (2015b). "The psychology of intelligence analysis: Drivers of prediction accuracy in world politics." In: *Journal of Experimental Psychology: Applied* 21.1, p. 1.
Meow, Jordy (2024). *AI Engine*. https://wordpress.org/plugins/ai-engine/. WordPress Plugin. (Visited on 07/24/2024).
Metaculus (2023). *Quarterly Cup*. URL: https://www.metaculus.com/tournament/quarterly-cup-2023q3/.
Moore, Don A and Paul J Healy (2008). "The trouble with overconfidence." In: *Psychological review* 115.2, p. 502.
Naveed, Humza et al. (2023). *A Comprehensive Overview of Large Language Models*. https://github.com/humza909/LLM_Survey.git.
Ngo, Richard, Lawrence Chan, and Sören Mindermann (2023). *The Alignment Problem from a Deep Learning Perspective*. arXiv: 2209.00626 [cs.AI].
Nori, Harsha et al. (2023). *Capabilities of GPT-4 on Medical Challenge Problems*. arXiv: 2303.13375 [cs.CL].
Noy, Shakked and Whitney Zhang (2023). "Experimental evidence on the productivity effects of generative artificial intelligence". In: *SSRN*. URL: https://ssrn.com/abstract=4375283.
OpenAI (2018). *OpenAI Charter*. OpenAI. URL: https://openai.com/charter.

OpenAI (2023a). *GPT-4 Technical Report*. arXiv: 2303.08774 [cs.CL].
OpenAI (2023b). *New models and developer products announced at DevDay*. https://help.openai.com/en/articles/8555510-gpt-4-turbo.
OpenAI (2024). *Models - OpenAI API*. https://platform.openai.com/docs/models. Accessed on July 25, 2024. URL: https://platform.openai.com/docs/models.
Park, Peter S. (2022). "The evolution of cognitive biases in human learning". In: *Journal of Theoretical Biology* 541, p. 111031.
Park, Peter S., Philipp Schoenegger, and Chongyang Zhu (2024). "Diminished diversity-of-thought in a standard large language model". In: *Behavior Research Methods*, pp. 1–17.
Park, Peter S.. and Max Tegmark (2023). *Divide-and-Conquer Dynamics in AI-Driven Disempowerment*. arXiv: 2310.06009 [cs.CY].
Park, Peter S. et al. (2023). *AI Deception: A Survey of Examples, Risks, and Potential Solutions*. arXiv: 2308.14752 [cs.CY].
Peeperkorn, Max et al. (2024). "Is temperature the creativity parameter of large language models?" In: *arXiv preprint arXiv:2405.00492*.
Peng, Sida et al. (2023). "The impact of ai on developer productivity: Evidence from github copilot". In: *arXiv preprint arXiv:2302.06590*.
Petropoulos, Fotios et al. (2022). "Forecasting: Theory and Practice". In: *International Journal of Forecasting* 38.3, pp. 705–871.
Sallam, Malik et al. (2023). "ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations". In: *Narra J* 3.1, e103–e103.
Schoemaker, Paul JH and Philip E Tetlock (2016). "Superforecasting: How to upgrade your company's judgment". In: *Harvard Business Review* 94.5, pp. 73–78.
Schoenegger, Philipp and Peter S. Park (2023). *Large Language Model Prediction Capabilities: Evidence from a Real-World Forecasting Tournament*. arXiv: 2310.13014 [cs.CY].
Schoenegger, Philipp et al. (2024a). "Can AI Understand Human Personality?–Comparing Human Experts and AI Systems at Predicting Personality Correlations". In: *arXiv preprint arXiv:2406.08170*.
Schoenegger, Philipp et al. (2024b). "Wisdom of the silicon crowd: Llm ensemble prediction capabilities match human crowd accuracy". In: *arXiv preprint arXiv:2402.19379*.
Shen, Zhiqiang et al. (2023). "SlimPajama-DC: Understanding Data Combinations for LLM Training". In: *arXiv preprint arXiv:2309.10818*.
Shiller, Robert J (2015). "Irrational exuberance". In: *Irrational exuberance*. Princeton university press.
Solaiyappan, Siddharth et al. (2023). "Utilizing Machine Learning Algorithms Trained on AI-generated Synthetic Participant Recent Music-Listening Activity in Predicting Big Five Personality Traits". In.
Steyvers, Mark and Aakriti Kumar (2023). "Three challenges for AI-assisted decision-making". In: *Perspectives on Psychological Science*, p. 17456916231181102.
Summers, Lawrence H and Steve Rattner (2023). *Larry Summers on who could be replaced by AI [Interviewed by Bloomberg TV's David Westin]*. URL: https://www.youtube.com/watch?v=8Epl9yAu0gk.
Sutton, Rich (2023). *AI succession [Youtube video of talk]*. World Artificial Intelligence Conference in Shanghai. URL: https://www.youtube.com/watch?v=NgHFMolXs3U.
Tetlock, Philip E. and Dan Gardner (2016). *Superforecasting: The Art and Science of Prediction*. Random House.
Tetlock, Philip E., Barbara A Mellers, and J Peter Scoblic (2017). "Bringing Probability Judgments into Policy Debates via Forecasting Tournaments". In: *Science* 355.6324, pp. 481–483.
Tetlock, Philip E. et al. (2014). "Forecasting Tournaments: Tools for Increasing Transparency and Improving the Quality of Debate". In: *Current Directions in Psychological Science* 23.4, pp. 290–295.
Vaswani, Ashish et al. (2017). "Attention is All You Need". In: *Advances in Neural Information Processing Systems* 30.
Vectara (2024). *Leaderboard Comparing LLM Performance at Producing Hallucinations when Summarizing Short Documents*. https://github.com/vectara/hallucination-leaderboard. Accessed: 2024-07-24.
Vemprala, Sai et al. (2023). "Chatgpt for robotics: Design principles and model abilities". In: *Microsoft Auton. Syst. Robot. Res* 2, p. 20.