

Working Machines: Beyond Automation

Bibliography

References follow the Harvard author-date system and are listed alphabetically in a single sequence. Organisations and companies are cited as corporate authors. Supplementary lists of the companies, individuals, case studies and databases referenced in the text follow the main reference list.

References

ABA Journal (2023) Coverage of AI in legal services and machine-generated briefs. Available at: <https://www.abajournal.com>.

Accenture (2024) Harnessing generative AI in enterprise strategy. Available at: <https://www.accenture.com>.

ACLU (2018) 'Why Amazon's automated hiring tool discriminated against women', American Civil Liberties Union. Available at: <https://www.aclu.org>.

AI Accelerator Institute (2024) Research on building an AI-ready workforce and fear of job displacement. Available at: <https://www.aiacceleratorinstitute.com>.

AICompetence.org (2024) Research on the confidence gap in emerging technology fields for marginalised communities. Available at: <https://aicompetence.org>.

AlixPartners (2021) Analysis estimating US\$210 billion in lost automotive revenue from the global semiconductor shortage. Available at: <https://www.alixpartners.com>.

Altman Solon (2024) Enterprise generative AI adoption report [70% of enterprises experimenting; 29% scaled beyond pilots]. Available at: <https://www.altmansolon.com>.

Amnesty International (2023) Powering change or business as usual? [Report on cobalt and copper mining in the Democratic Republic of the Congo.] London: Amnesty International. Available at: <https://www.amnesty.org>.

AP News (2023) Reporting on data centre water usage in drought-prone regions. Available at: <https://apnews.com>.

Architectural Digest (2023) Coverage of AI-generated design iteration in architecture. Available at: <https://www.architecturaldigest.com>.

Ataccama (2024) AI use cases for the insurance industry. Available at: <https://www.ataccama.com>.

Bai, Y. et al. (2022) 'Constitutional AI: harmlessness from AI feedback'. San Francisco: Anthropic. Available at: <https://arxiv.org/abs/2212.08073>.

Bazerman, M.H. and Tenbrunsel, A.E. (2011) Blind spots: why we fail to do what's right

and what to do about it. Princeton, NJ: Princeton University Press.

Referenced for the concept of 'bounded ethicality': the idea that good people in well-intentioned systems can make bad calls without realising it, because their attention is focused elsewhere.

BBC News (2016) 'NHS 111 missed chances to save sepsis baby William Mead'.

Available at: <https://www.bbc.com/news>.

BBC News (2021) Reporting on WhatsApp users moving to rival platforms after the terms-of-service update, on Ireland's limits on new data centre connections, and on Taiwan trucking in water to sustain chip production during the 2021 drought.

Available at: <https://www.bbc.com/news>.

Berlin, L. (2007) 'Accuracy of diagnostic procedures: has it improved over the past five decades?', *American Journal of Roentgenology*, 188(5), pp. 1173-1178.

Binns, R. and Veale, M. (2021) 'Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR', *International Data Privacy Law*, 11(4), pp. 319-332. Available at: <https://doi.org/10.1093/idpl/ipab020>.

Bloomberg (2023) 'IBM to pause hiring for jobs that AI could do', *Bloomberg News*, 1 May. Available at: <https://www.bloomberg.com/news/articles/2023-05-01/ibm-to-pause-hiring-for-back-office-jobs-that-ai-could-kill>.

Bloomberg (2023) Reporting on Chegg's AI-related restructuring, and analysis of data centre energy use, AI carbon intensity and ESG risk in sustainability indices. Available at: <https://www.bloomberg.com>.

BMJ (2020) Report on rates of diagnostic error in clinical practice, *British Medical Journal*. Available at: <https://www.bmj.com>.

BMJ Global Health (2021) Research on digital health initiatives in low-resource settings. Available at: <https://gh.bmj.com>.

Bouguettaya, A., Stuart, E.M. and Aboujaoude, E. (2025) 'Racial bias in AI-mediated psychiatric diagnosis and treatment: a qualitative comparison of four large language models', *npj Digital Medicine*, 8, 332. Cedars-Sinai Medical Center. Available at: <https://doi.org/10.1038/s41746-025-01746-4>.

Broadband.co.uk (2023) World broadband affordability research 2023. Available at: <https://www.broadband.co.uk>.

Brookings Institution (2023) Policy responses to AI-driven job displacement, including analysis of L-shaped labour-market recovery. Washington, DC: Brookings Institution. Available at: <https://www.brookings.edu>.

Built In (2023) Study on workplace resistance to AI adoption [over 30% of workers admitted to actively resisting rollouts]. Available at: <https://builtin.com>.

Business Insider (2023) Reporting that OpenAI's operating loss for ChatGPT in 2022 was around US\$540 million. Available at: <https://www.businessinsider.com>.

Business Insider (2024) 'After careful review, we've decided to end our current AI drive-thru test' [McDonald's and IBM voice-AI trial]. Available at:

<https://www.businessinsider.com>.

Business Insider (2025) Reporting on the Claude Opus 4 safety tests [the model used sensitive information as leverage in 84% of test runs], and Davos 2025 coverage of treating AI agents like employees [Lattice and HCLTech]. Available at:

<https://www.businessinsider.com>.

Business Insider Africa (2025) 'Africa Collective at Davos 2025: navigating African investment opportunities amidst global uncertainty', 22 January. Available at:

<https://africa.businessinsider.com>.

Cambridge Management Consulting (2024) Analysis of loadshedding's impact on South African digital transformation. Available at: <https://www.cambridgemc.com>.

Campaign and MM+M (2025) 'Nearly 70% of bias incidents in AI LLMs occur in regional languages: IMDA study'. Available at: <https://www.mmm-online.com>.

Carichon, F. et al. (2025) Research on collective misalignment among synthetic agents deployed together.

Chiang, T. (2023) 'ChatGPT is a blurry JPEG of the web', The New Yorker, 9 February.

Available at: <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>.

Christiano, P.F., Leike, J., Brown, T.B., Martic, M., Legg, S. and Amodei, D. (2017) 'Deep reinforcement learning from human preferences', in Advances in Neural Information Processing Systems 30. [Foundational paper for Reinforcement Learning from Human Feedback (RLHF).] Available at: <https://arxiv.org/abs/1706.03741>.

CIO Dive (2025) 'AI project failure rates are on the rise: report', 13 March. [Reports the S&P Global finding that 42% of organisations abandoned most of their AI initiatives.] Available at: <https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/>.

CIO.com (2024) Survey finding that two-thirds of CEOs cited fear of missing out as a driver of AI investment. Available at: <https://www.cio.com>.

ClickUp (2025) AI adoption challenges: integration issues across systems and resistance to change in teams. Available at: <https://clickup.com>.

Copper Digital (2024) AI in logistics: 5 game-changing use cases for 2025 supply chains. Available at: <https://www.copperdigital.com>.

Deloitte and The Wall Street Journal (2024) 'For Vanguard's CDAO and CMO, partnership is part of the job', Deloitte CIO Journal. Available at: <https://deloitte.wsj.com>.

den Rooijen, R., Munsie, W. and Bean, R. (2025) 'The chief data officer role: what's next?', MIT Sloan Management Review. [Source of the finding that 84% of organisations have appointed a CDO or CDAO, up from 12% in 2012.] Available at: <https://sloanreview.mit.edu/article/the-chief-data-officer-role-whats-next/>.

Draxler, F., Werner, A., Lehmann, F., Hoppe, M. et al. (2024) 'The AI Ghostwriter Effect:

- when users do not perceive ownership of AI-generated text but self-declare as authors', *ACM Transactions on Computer-Human Interaction*, 31(2). Available at: <https://dl.acm.org/doi/10.1145/3637875>.
- Eco-Business (2024) Coverage of Malaysia's clean-energy 'green cloud parks' for AI infrastructure. Available at: <https://www.eco-business.com>.
- Erdil, E. (2024) Optimally allocating compute between inference and training. *Epoch AI*. Available at: <https://epoch.ai>.
- European Academies' Science Advisory Council (2017) Findings and recommendations from the Smart Villages Initiative 2014-2017.
- European Commission (2021) Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), COM(2021) 206 final, 21 April. Brussels: European Commission. Available at: <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.
- European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), Official Journal of the European Union, L 119. [Applicable from 25 May 2018.] Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- European Union (2023) Regulation (EU) 2023/1781 (European Chips Act), Official Journal of the European Union. Available at: <https://eur-lex.europa.eu>.
- European Union (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act), Official Journal of the European Union. Available at: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>.
- Federal Aviation Administration (2019) Operator's manual for human factors in aviation maintenance. Washington, DC: US Department of Transportation, Federal Aviation Administration.
- Fiverr (2023) Data on the impact of AI on freelance service pricing. Available at: <https://www.fiverr.com>.
- Freedman, R. (2023) 'Judge asks Michael Cohen lawyer if cited cases are fake', *Legal Dive*, 13 December. Available at: <https://www.legaldive.com/news/judge-furman-michael-cohen-lawyer-cites-fake-cases-schwartz-chatgpt-ai-hallucinations-legaltech/702422/>.
- Freedman, R. (2023) 'Lawyers citing fake ChatGPT cases are fined', *Legal Dive*, 22 June. [Judge Kevin Castel's ruling.] Available at: <https://www.legaldive.com/news/lawyers-citing-fake-chatgpt-cases-are-fined/652702/>.
- Friedrich Naumann Foundation for Freedom (2025) Policy report: examining AI in low and middle-income countries. Available at: <https://www.freiheit.org>.
- Frontiers in Psychology (2023) 'Cross-cultural evaluation of automated facial expression recognition systems' [FaceReader and DeepFace]. Available at: <https://www.frontiersin.org>.

- Ganguli, D. et al. (2022) 'Red teaming language models to reduce harms: methods, scaling behaviors, and lessons learned'. San Francisco: Anthropic. Available at: <https://arxiv.org/abs/2209.07858>.
- Gartner (2023) Prescriptive analytics in action: the next step in AI evolution. Stamford, CT: Gartner. Available at: <https://www.gartner.com>.
- Gartner (2024) Top strategic data and analytics trends for 2024, and research on enterprise transformation through data leadership. Stamford, CT: Gartner. Available at: <https://www.gartner.com>.
- Gladwell, M. (2019) Talking to strangers: what we should know about the people we don't know. New York: Little, Brown and Company.
Referenced for analysis of human fallibility, particularly how professionals (FBI agents, judges, psychologists) routinely misread others and miss deception. Cited study showing judges making bail decisions were consistently outperformed by simple algorithms.
- Gohel, P. et al. (2024) 'Next-gen agriculture: integrating AI and XAI for precision crop yield prediction', Frontiers in Plant Science. Available at: <https://www.frontiersin.org>.
- Goldman Sachs (2023) Research estimating that generative AI could automate up to 44% of tasks in legal professions and 37% in engineering. New York: Goldman Sachs. Available at: <https://www.goldmansachs.com>.
- Gonsalves, C. (2024) 'Generative AI's impact on critical thinking: revisiting Bloom's taxonomy', Journal of Marketing Education.
- Google (2023) 2023 environmental report [20% rise in data centre energy consumption in 2022]. Mountain View, CA: Google. Available at: <https://sustainability.google>.
- Gray, M.L. and Suri, S. (2019) Ghost work: how to stop Silicon Valley from building a new global underclass. Boston, MA: Houghton Mifflin Harcourt.
Seminal work on the hidden human labour that powers AI systems, characterised by low wages, precarious employment, and lack of benefits. Introduces the concept of 'ghost work'.
- Griebel, M., Flath, C.M. and Friesike, S. (2021) 'Augmented creativity: leveraging artificial intelligence for idea generation in the creative sphere', Electronic Markets, 31(4), pp. 789–802. Available at: <https://doi.org/10.1007/s12525-021-00486-1>.
- Harwell, D. (2025) 'AI tools mostly fumble basic financial tasks, study finds' [The Washington Post and Vals AI analysis of 22 leading LLMs], The Washington Post, 22 April. Available at: <https://www.washingtonpost.com/politics/2025/04/22/ai-tools-mostly-fumble-basic-financial-tasks-study-finds/>.
- Heaven, W.D. (2022) 'Meta's new AI chatbot can't stop generating fake science', MIT Technology Review, 18 November. Available at: <https://www.technologyreview.com/2022/11/18/1063666/meta-ai-chatbot-galactica-hallucinating-fake-science/>.
- Heilweil, R. (2023) 'Hospitals are pausing AI triage tools over equity concerns', The Verge, 12 June. Available at: <https://www.theverge.com/2023/06/12/ai-healthcare-triage-bias-race-equity>.
- Howe, J. (2008) Crowdsourcing: why the power of the crowd is driving the future of

business. New York: Crown Business.

Referenced for the concept of crowdsourcing that underpins the user-in-the-loop model.

Hu, K. (2023) 'ChatGPT sets record for fastest-growing user base', Reuters, 2 February. Available at: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.

HubSpot (2025) Enterprise AI adoption challenges [organisational alignment and leadership commitment]. Available at: <https://www.hubspot.com>.

IBM (2024) Guidance warning against emotional attachment to AI co-workers and the risks of oversharing personal data. Armonk, NY: IBM. Available at: <https://www.ibm.com>.

IBM Institute for Business Value (2025) 2025 CEO study. Armonk, NY: IBM. [Source of the finding that only 25% of AI initiatives delivered expected ROI and only 16% scaled enterprise-wide.] Available at: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/c-suite-study/ceo>.

IE University and Frontiers in Psychology (2025) Study on the correlation between frequent AI usage and critical-thinking abilities. Available at: <https://www.frontiersin.org>.

ILOSTAT (2024) Data on global mobile-cellular subscriptions and smartphone penetration. Geneva: International Labour Organisation. Available at: <https://ilostat.ilo.org>.

Innovative Human Capital (2024) Analysis of empathy, judgment, ethics and hope as capabilities AI cannot replicate. Available at: <https://www.innovativehumancapital.com>.

International Energy Agency (2023) Critical minerals market review [China's share of about 70% of rare earth production]. Paris: IEA. Available at: <https://www.iea.org>.

International Energy Agency (2024) Electricity 2024: analysis and forecast to 2026 [projection that electricity demand from data centres, cryptocurrency and AI could double by 2026 to over 1,000 TWh]. Paris: IEA. Available at: <https://www.iea.org/reports/electricity-2024>.

International Labour Organisation (2024) Analysis of government-backed reskilling initiatives. Geneva: ILO. Available at: <https://www.ilo.org>.

ISACA (2024) Guidance that AI training must include ethical reasoning, governance and judgment. Available at: <https://www.isaca.org>.

ITPro (2025) Business AI adoption report [47% of businesses reported infrastructure unable to support advanced technologies; 49% cited poor network performance]. Available at: <https://www.itpro.com>.

Ivanti (2025) 2025 Technology at Work Report: reshaping flexible work. Salt Lake City, UT: Ivanti. [Source of data on AI-fuelled impostor syndrome (27%), resentment (48%), presenteeism, and secret AI use among UK workers.] Available at: <https://www.ivanti.com/resources/research-reports/tech-at-work>.

- Kahneman, D. (2011) Thinking, fast and slow. New York: Farrar, Straus and Giroux.
Referenced for the concept of the 'fluency heuristic': the tendency to believe information that is easy to process, sounds coherent, and uses the right language. Applied to Meta's Galactica case study.
- Kahneman, D., Sibony, O. and Sunstein, C.R. (2021) Noise: a flaw in human judgment. London: William Collins.
Referenced for findings on sentencing variability of up to 57% for identical cases.
- Kleinberg, J. and Mullainathan, S. (2024) Commentary on AI vendor responsibility for tools marketed as assistants in high-stakes work.
- Lewis, A.A. (2025) 'Unpacking cultural bias in AI language learning tools: an analysis of impacts and strategies for inclusion in diverse educational settings', International Journal of Research and Innovation in Social Science, IX(1), pp. 1878-1892. Available at: <https://doi.org/10.47772/IJRISS.2025.9010151>.
- Lewis, M. (2016) The undoing project: a friendship that changed our minds. New York: W.W. Norton & Company.
Referenced for exploration of the 'illusion of expertise': the gap between how accurate we think our judgment is, and how accurate it really is.
- Li, P., Yang, J., Islam, M.A. and Ren, S. (2023) 'Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models'. University of California, Riverside. Available at: <https://arxiv.org/abs/2304.03271>.
- LinkedIn post by Walter Shields (2024) 'The do's and don'ts of using AI in job applications'. Available at: <https://www.linkedin.com>.
- LSE Business Review (2025) Commentary on AI amplifying rather than supplanting human decision-making. London: London School of Economics. Available at: <https://blogs.lse.ac.uk/businessreview>.
- Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', in Advances in Neural Information Processing Systems 30. [Foundational paper for SHAP (SHapley Additive exPlanations) values.] Available at: <https://arxiv.org/abs/1705.07874>.
- Maryville University (2020) Analysis of Pew Research Center data on the digital divide within the United States. Available at: <https://www.maryville.edu>.
- Mayer, H., Yee, L., Chui, M. and Roberts, R. (2025) Superagency in the workplace: empowering people to unlock AI's full potential. McKinsey & Company, January. [Source of the discussion of 'superagents' at work.] Available at: <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work>.
- McKinsey & Company (2024) 'Charting a path to the data- and AI-driven enterprise of 2030', McKinsey Quarterly, 5 September. Available at: <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/charting-a-path-to-the-data-and-ai-driven-enterprise-of-2030>.

- McKinsey & Company (2024) Research on AI-informed leadership and organisational performance. Available at: <https://www.mckinsey.com>.
- McKinsey & Company (2024) Research on generative AI adoption in the workplace [nearly 40% of workers had incorporated GenAI into their jobs]. Available at: <https://www.mckinsey.com>.
- McKinsey & Company (2024) Research on user trust as a core metric in AI performance evaluations [72% of business leaders]. Available at: <https://www.mckinsey.com>.
- Meadows, D.H. (2008) Thinking in systems: a primer. White River Junction, VT: Chelsea Green Publishing.
Referenced for the centrality of feedback loops to system improvement.
- Measured Ability (2024) Mental health in the AI job market: coping with constant change. Available at: <https://www.measuredability.com>.
- Menlo Ventures (n.d.) Research on generative AI adoption among early-stage founders. Available at: <https://menlovc.com>.
- Metz, C. and Weise, K. (2023) 'AI pioneers at Google detail their alarm at the technology's dangers', The New York Times. Available at: <https://www.nytimes.com>.
- Microsoft (2022) Microsoft responsible AI standard, v2. Redmond, WA: Microsoft. Available at: <https://www.microsoft.com/en-us/ai/responsible-ai>.
- Microsoft (2023) Environmental sustainability report [water use rose by more than 34% between 2021 and 2022, driven partly by AI model training]. Redmond, WA: Microsoft. Available at: <https://www.microsoft.com/en-us/corporate-responsibility/sustainability>.
- Minderoo Centre for Technology and Democracy, University of Cambridge (2024) Report warning that AI could push technology-sector emissions to five times current levels by 2040. Cambridge: University of Cambridge. Available at: <https://www.mctd.ac.uk>.
- MIT Energy Initiative (2023) Analysis of AI-driven electricity demand slowing coal plant retirements. Cambridge, MA: MIT. Available at: <https://energy.mit.edu>.
- MIT Sloan Management Review (2023) Research on trust and transparency in AI-augmented leadership. Available at: <https://sloanreview.mit.edu>.
- MIT Sloan Management Review (2024) Research on data and AI investment planning. Available at: <https://sloanreview.mit.edu>.
- MIT Sloan Management Review (2024) Research on empathy and human capabilities in AI-era leadership. Available at: <https://sloanreview.mit.edu>.
- MIT Sloan Management Review (2024) Research on worker stress and burnout in AI-enabled environments [88% of the most active AI users report experiencing burnout]. Available at: <https://sloanreview.mit.edu>.
- MIT Technology Review (2023) 'What is red teaming in AI, and why does it matter?'. Available at: <https://www.technologyreview.com>.

- Mitchell, M., Wu, S., Zaldivar, A. et al. (2019) 'Model cards for model reporting', in Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), pp. 220-229. [Foundational paper for model cards.] Available at: <https://arxiv.org/abs/1810.03993>.
- National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST. Available at: <https://www.nist.gov/itl/ai-risk-management-framework>.
- National Science Teaching Association (2024) 'To think or not to think: the impact of AI on critical-thinking skills'. Available at: <https://www.nsta.org>.
- National Transportation Safety Board (2021) Aviation statistics: accident database and synopses [human error implicated in up to 80% of aviation accidents]. Washington, DC: NTSB. Available at: <https://www.ntsbgov>.
- Nature (2022) Research on AI accuracy in identifying diseases, including cancer, from medical images. Available at: <https://www.nature.com>.
- Newitz, A. (2025) Reporting on Anthropic's Claude Opus 4 safety testing and emergent escalation behaviour, Wired. Available at: <https://www.wired.com>.
- Norman, D. (2013) The design of everyday things. Revised edn. New York: Basic Books.
Referenced for user-centred design, on which the user-in-the-loop concept builds.
- OECD (2025) 'Bridging the AI skills gap: is training keeping up?', OECD Artificial Intelligence Papers. Paris: OECD Publishing. Available at: <https://doi.org/10.1787/66d0702e-en>.
- Owolabi, T.O.P.S., Bello, A.O., Olatunji, A.J. and Jayeoba, S. (2024) 'Influence of power supply reliability on academic performance and project completion rates of architectural students at Caleb University, Lagos, Nigeria', African Journal of Humanities and Contemporary Education Research, 10(3), p. 391. Available at: <https://doi.org/10.62154/ajhcer.2024.016.010391>.
- People Managing People (2025) Analysis of cultural drag when AI becomes 'someone else's tool, and someone else's problem'. Available at: <https://peoplemanagingpeople.com>.
- Pew Research Center (2020) Data on smartphone ownership and home broadband access by household income in the United States. Washington, DC: Pew Research Center. Available at: <https://www.pewresearch.org>.
- Pols, J. (n.d.) Research on participatory design.
- Psychology Today (n.d.) Analysis of the psychological and physiological effects of rapid AI implementation in the workplace. Available at: <https://www.psychologytoday.com>.
- Reisman, D., Schultz, J., Crawford, K. and Whittaker, M. (2018) Algorithmic impact assessments: a practical framework for public agency accountability. New York: AI Now Institute. Available at: <https://ainowinstitute.org>.

- ResumeBuilder.com (2023) Survey on expected AI-related layoffs in 2024 [nearly 40% of companies]. Available at: <https://www.resumebuilder.com>.
- Reuters (2025) Reporting on Butler Snow LLP's submission of fictitious ChatGPT-generated citations. Available at: <https://www.reuters.com>.
- Revenova (2024) 'Artificial intelligence drives logistics forward'. [Contains the Lior Ron quote on the Uber Freight pricing model.] Available at: <https://www.revenova.com>.
- Roose, K. (2023) 'A conversation with Bing's chatbot left me deeply unsettled', The New York Times, 16 February. Available at: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft.html>.
- RSM (n.d.) Research on AI adoption among middle-market firms [over 75% had begun adopting AI]. Available at: <https://www.rsm.global>.
- S&P Global Market Intelligence (2025) Voice of the Enterprise: AI and machine learning survey. New York: S&P Global. [Source of the findings that 42% of companies abandoned most AI initiatives and that an average of 46% of AI proofs-of-concept were scrapped before production.] Available at: <https://www.spglobal.com/market-intelligence>.
- Sacra (2024) Research on OpenAI's cost structure and breakeven economics. Available at: <https://sacra.com>.
- Salesforce (2024) Guidance calling for clear accountability frameworks before autonomous systems become the default face of a company. San Francisco, CA: Salesforce. Available at: <https://www.salesforce.com>.
- Senthilnathan, S. et al. (2024) Study on workplace personalisation, performance and mental well-being among knowledge workers, Building and Environment.
- Shields, M. (2023) 'ChatGPT-generated legal citations land lawyers in hot water', Reuters, 22 June. Available at: <https://www.reuters.com/legal/legalindustry/chatgpt-generated-citations-lead-federal-sanctions-2023-06-22/>.
- Sinek, S. (2009) Start with why: how great leaders inspire everyone to take action. New York: Portfolio/Penguin.
Referenced for the framing of purpose as a compass rather than a metric.
- Stanford Report (2025) Analysis of large language model underperformance in languages such as Vietnamese and Nahuatl. Stanford, CA: Stanford University. Available at: <https://news.stanford.edu>.
- State of California (2018) California Consumer Privacy Act of 2018 (AB 375). [In effect from 1 January 2020.] Sacramento, CA. Available at: <https://oag.ca.gov/privacy/ccpa>.
- Strubell, E., Ganesh, A. and McCallum, A. (2019) 'Energy and policy considerations for deep learning in NLP', in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, pp. 3645–3650. [Source of the estimate that a single large training run can emit more than 284,000 kg of CO₂.] Available at: <https://doi.org/10.18653/v1/P19-1355>.

- Subin, S. (2023) 'Alphabet shares slide 8% after Google AI event', CNBC, 8 February. Available at: <https://www.cnbc.com/2023/02/08/alphabet-shares-slide-4percent-after-google-ai-event.html>.
- Surowiecki, J. (2004) *The wisdom of crowds*. New York: Doubleday.
Referenced for the concept that distributed knowledge often surpasses centralised expertise.
- TechCrunch (2023) Coverage of agentic AI systems, including AutoGPT and frameworks such as LangChain. Available at: <https://techcrunch.com>.
- The Cairo Review of Global Affairs (2025) Analysis of AI training costs [training a model such as GPT-4 can exceed US\$100 million in compute]. Available at: <https://www.thecairoreview.com>.
- The National (2024) 'Google takes down incorrect AI answer that says Obama was first Muslim US president', The National News, 24 May. Available at: <https://www.thenationalnews.com/future/technology/2024/05/24/google-takes-down-incorrect-ai-answer-that-says-obama-was-first-muslim-us-president>.
- The Verge (2022) Reporting on the blocked Meta data centre in the Netherlands [projected 1.38 TWh draw]. Available at: <https://www.theverge.com>.
- The Verge (2023) Coverage of efforts to make AI cleaner and less resource-intensive, including liquid-immersion cooling and Microsoft's seawater cooling pilot in Finland. Available at: <https://www.theverge.com>.
- The Washington Post (2023) 'Michael Cohen admits to fake legal citations generated by Google Bard'. Available at: <https://www.washingtonpost.com>.
- TIME and The Bureau of Investigative Journalism (2022) 'Behind TikTok's boom: a legion of traumatized, \$10-a-day content moderators', TIME, 20 October. Available at: <https://time.com/6223340/tiktok-content-moderators-latin-america/>.
- Triangility (2024) Commentary on AI amplifying human decision-making. Available at: <https://triangility.com>.
- UNESCO (n.d.) Reports on algorithmic and digital literacy in education. Paris: UNESCO. Available at: <https://www.unesco.org>.
- United States Congress (2022) CHIPS and Science Act of 2022. Washington, DC. Available at: <https://www.congress.gov>.
- United States Congress (2025) One Big Beautiful Bill Act (OBBBA). Washington, DC. Available at: <https://www.congress.gov>.
- United States Geological Survey (2023) Mineral commodity summaries 2023 [the Democratic Republic of the Congo accounts for roughly 70% of world cobalt production]. Reston, VA: USGS. Available at: <https://www.usgs.gov>.
- Universities South Africa (2025) Who owns AI? Digital literacy, inequality, and the future of learning. Pretoria: Universities South Africa. Available at: <https://www.usaf.ac.za>.
- University of Michigan (2024) Research on emotion-sensing AI in the workplace and

- its effects on trust and mental well-being. Ann Arbor, MI: University of Michigan. Available at: <https://umich.edu>.
- Upwork (2024) Data on freelance services pricing following generative AI adoption. Available at: <https://www.upwork.com>.
- Wall Street Journal (2023) Reporting on Chegg layoffs attributed to generative AI's impact on demand. Available at: <https://www.wsj.com>.
- West, S.M. et al. (2023) 'Algorithmic amplification of misclassification: the risks of over-filtering in LLMs', Journal of Digital Ethics.
- White House Office of Science and Technology Policy (2022) Blueprint for an AI Bill of Rights. Washington, DC: The White House. Available at: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>.
- Wolhuter, S., Grunitz, M. and Ferrai, P. (2021) Working machines: an executive's guide to AI and intelligent automation. Amsterdam: WeAreBrain.
Foundational work establishing predictions about AI adoption in business leadership and automation's role in the workplace.
- Workday (2024) Analysis of human experience in AI-supported management. Pleasanton, CA: Workday. Available at: <https://www.workday.com>.
- World Economic Forum (2020) Smart at scale: cities to watch – 25 case studies. Geneva: World Economic Forum, Global Future Council on Cities and Urbanisation, August. Available at: https://www3.weforum.org/docs/WEF_Smart_at_Scale_Cities_to_Watch_25_Case_Studies_2020.pdf.
- World Economic Forum (2023) Report noting that 'generative AI' overtook 'cryptocurrency' as the most-searched emerging technology term, and coverage of renewable microgrids for African data centres. Geneva: World Economic Forum. Available at: <https://www.weforum.org>.
- World Economic Forum (2023) The future of jobs report 2023. Geneva: World Economic Forum. [Source of the forecast net loss of 14 million jobs globally over five years, and the skills-mismatch paradox.] Available at: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>.
- World Economic Forum (2024) Reskilling Revolution initiative [over 3.4 million learners reached in Brazil]. Geneva: World Economic Forum. Available at: <https://www.weforum.org>.
- World Economic Forum (2025) Davos 2025: annual meeting session coverage. Geneva: World Economic Forum. Available at: <https://www.weforum.org/events/world-economic-forum-annual-meeting-2025/>.
- World Health Organization and World Bank (2023) Tracking universal health coverage: 2023 global monitoring report [at least half of the world's population lacks full coverage of essential health services]. Geneva: WHO. Available at: <https://www.who.int>.

Companies, Platforms and Organisations Referenced

These entities appear in the text as actors or case studies rather than as authors of cited works. Official websites are given for reference.

Ada. AI customer-service automation platform. <https://www.ada.cx>

Affectiva. Emotion AI company (now part of Smart Eye). [Criticised for cultural bias in emotion recognition.] <https://www.affectiva.com>

Alibaba. Chinese technology company; developer of the Qwen model family. <https://www.alibaba.com>

Alphabet. Parent company of Google. [US\$100 billion market value loss following the Bard error.] <https://abc.xyz>

Amazon Web Services (AWS). Cloud computing platform. <https://aws.amazon.com>

Anthropic. AI safety and research company; developer of Claude and the Constitutional AI framework. <https://www.anthropic.com>

ASML. Dutch manufacturer of the extreme ultraviolet (EUV) lithography machines required for advanced chips. <https://www.asml.com>

Babylon Health. UK digital health company. [AI symptom-checker retired from GP partnerships amid scrutiny.] <https://www.emed.com>

Baidu. Chinese technology and AI platform company. <https://www.baidu.com>

BuzzFeed. Digital media company. [BuzzFeed News shut down in April 2023.] <https://www.buzzfeed.com>

C3.ai. Enterprise AI software company. <https://c3.ai>

Chegg. Online education company. [Layoffs attributed to GenAI's impact on demand.] <https://www.chegg.com>

CNET. Technology media outlet. [Paused AI content generation after error-ridden AI-written articles.] <https://www.cnet.com>

Cohere. Enterprise AI model developer. <https://cohere.com>

Coursera. Online learning platform. <https://www.coursera.org>

DeepSeek. Chinese AI model developer. [System-level moderation of politically sensitive topics.] <https://www.deepseek.com>

Google. Technology company. [Bard James Webb error; AI Overview; Gemini; dismissal of Dr Timnit Gebru.] <https://www.google.com>

Google DeepMind. AI research laboratory. <https://deepmind.google>

Harvey AI. Legal technology company. <https://www.harvey.ai>

HCLTech. Information technology services company. <https://www.hcltech.com>

Hugging Face. Open-source AI platform and model repository; originator of the widely adopted Model Cards framework. <https://huggingface.co>

IBM. Technology company. [Arvind Krishna, CEO; AI-related hiring pause; McDonald's drive-thru partnership.] <https://www.ibm.com>

Intercom. AI customer-service platform. <https://www.intercom.com>

Klarna. Financial services and shopping platform. [2024 AI customer-service case study.] <https://www.klarna.com>

LangChain. AI application framework and middleware provider. <https://www.langchain.com>

Lattice. Human capital management software company. <https://www.lattice.com>

LegalZoom. Legal services platform. <https://www.legalzoom.com>

Masakhane. African-led grassroots initiative for natural language processing in African languages. <https://www.masakhane.io>

McDonald's. Fast-food company. [Ended its AI drive-thru trial with IBM in 2024.] <https://www.mcdonalds.com>

Meta (Facebook). Technology company. [Galactica case study; disbanding of the Responsible AI division; WhatsApp terms backlash.] <https://www.meta.com>

Microsoft. Technology company. [Bing 'Sydney' alter ego; Azure; Office of Responsible AI; retirement of Azure Face emotion recognition.] <https://www.microsoft.com>

Microsoft Azure. Cloud computing platform. <https://azure.microsoft.com>

Mistral AI. French AI model developer. <https://mistral.ai>

Moderna. Pharmaceutical company. [Data strategy during COVID-19 vaccine development.] <https://www.modernatx.com>

Mohamed bin Zayed University of Artificial Intelligence (MBZUAI). UAE research university dedicated to AI. <https://mbzuai.ac.ae>

Mozilla. Technology non-profit. [Ethics as brand identity; trustworthy AI incubation.] <https://www.mozilla.org>

NetDragon. Hong Kong-based company. [Appointed an AI CEO in 2022.] <https://www.netdragon.com>

Nvidia. Designer of the high-performance GPUs (H100, A100) that power most advanced AI training and inference. <https://www.nvidia.com>

One Laptop per Child (OLPC). Education non-profit founded in 2005. [Cautionary tale of over-hyped access.] <https://laptop.org>

OneWeb. Satellite broadband constellation operator. <https://oneweb.net>

OpenAI. AI model development company; developer of ChatGPT and GPT-4. <https://openai.com>

Palantir. Data analytics company. <https://www.palantir.com>

Partnership on AI. Multi-stakeholder body codifying best practices, audit protocols and transparency across sectors. <https://partnershiponai.org>

Perplexity. AI answer-engine combining retrieval and generation. <https://www.perplexity.ai>

Pfizer. Pharmaceutical company. [Governance reviews integrated into the R&D cycle.] <https://www.pfizer.com>

Pinecone. Vector database and AI middleware provider. <https://www.pinecone.io>

Salesforce. Customer relationship management platform. [Model cards and ethics checklists case study.] <https://www.salesforce.com>

Samsung. Semiconductor designer and manufacturer. <https://www.samsung.com>

SAP. Enterprise resource planning software company. <https://www.sap.com>

Sberbank AI. AI division of Russia's Sberbank. <https://www.sberbank.com>

SenseTime. Chinese AI and computer-vision company. <https://www.sensetime.com>

Signal. Encrypted messaging platform. [Surged during the 2021 WhatsApp backlash.] <https://signal.org>

Smart Africa. Pan-African alliance advancing digital and data governance. <https://smartafrica.org>

Starlink (SpaceX). Satellite internet constellation. <https://www.starlink.com>

Telegram. Messaging platform. [Surged during the 2021 WhatsApp backlash.] <https://telegram.org>

Teleperformance. Outsourcing company employing TikTok content moderators in Colombia. <https://www.teleperformance.com>

TikTok. Social media platform. [Content moderator investigation.] <https://www.tiktok.com>

TSMC (Taiwan Semiconductor Manufacturing Company). Fabricator of over 90% of the world's most advanced semiconductors. <https://www.tsmc.com>

Twitter/X. Social media platform. [Birdwatch case study.] <https://x.com>

Uber Freight. Logistics technology company. [Lior Ron pricing model case study.] <https://www.uberfreight.com>

Vals AI. AI evaluation company. [Partner in The Washington Post analysis of LLM financial accuracy.] <https://www.vals.ai>

Vanguard. Financial services company. [Ryan Swann, CDAO; data-for-outcomes case study.] <https://www.vanguard.com>

Yandex. Russian technology company. <https://yandex.com>

YouTube. Video platform. [GenAI dubbing tool case, 2024.] <https://www.youtube.com>

Individuals Cited

Business leaders and executives

Lior Ron. Chief Executive Officer, Uber Freight. [Pricing-model quote on parameters and accuracy; see Revenova (2024).]

Ryan Swann. Chief Data and Analytics Officer, Vanguard. [See Deloitte and The Wall Street Journal (2024).]

Arvind Krishna. Chairman and Chief Executive Officer, IBM. [See Bloomberg (2023).]

Authors, researchers and thinkers

Malcolm Gladwell. Author and journalist. [See Gladwell (2019).]

Daniel Kahneman. Psychologist and author. [See Kahneman (2011) and Kahneman, Sibony and Sunstein (2021).]

Max H. Bazerman and Ann E. Tenbrunsel. Researchers and authors. [See Bazerman and Tenbrunsel (2011).]

Michael Lewis. Author and journalist. [See Lewis (2016).]

Mary L. Gray and Siddharth Suri. Researchers and authors. [See Gray and Suri (2019).]

Simon Sinek. Author and speaker on purpose-driven leadership. [See Sinek (2009).]

Don Norman. Design researcher; originator of user-centred design. [See Norman (2013).]

Jeff Howe. Journalist; populariser of crowdsourcing. [See Howe (2008).]

James Surowiecki. Journalist; populariser of the wisdom of crowds. [See Surowiecki (2004).]

Donella Meadows. Systems thinker. [See Meadows (2008).]

Dr Timnit Gebru. AI ethics researcher; former co-lead of Google's Ethical AI team, dismissed in 2020.

Dr Elias Aboujaoude. Director, Program in Internet, Health and Society, Cedars-Sinai. [Corresponding author; see Bouguettaya, Stuart and Aboujaoude (2025).]

Legal and judicial figures

Michael Cohen. Former lawyer to Donald Trump. [Google Bard citation case; see Freedman (2023).]

David Schwartz. Attorney. [Received fabricated citations; see Freedman (2023).]

Judge Jesse Furman. US federal court judge. [Flagged the fabricated citations; see Freedman (2023).]

Judge Kevin Castel. US federal court judge. [Fined lawyers for fake ChatGPT citations; see Freedman (2023) and Shields (2023).]

Journalists and reporters

W.D. Heaven. MIT Technology Review.

R. Heilweil. The Verge.

M. Shields. Reuters.

K. Hu. Reuters.

K. Roose. The New York Times.

T. Chiang. The New Yorker.

C. Metz and K. Weise. The New York Times.

D. Harwell. The Washington Post.

R. Freedman. Legal Dive.

A. Newitz. Wired.

Case Studies and Specific Incidents

Michael Cohen AI citation case (late 2023). Cohen supplied fabricated case citations generated by Google Bard to attorney David Schwartz; Judge Jesse Furman flagged them. See Freedman (2023) and The Washington Post (2023).

US federal court ChatGPT citation case (2023). Two lawyers fined US\$5,000 for citing fake cases generated by ChatGPT; Judge Kevin Castel's ruling. See Shields (2023) and Freedman (2023).

Butler Snow LLP citations (2025). US law firm apologised after filings contained fictitious ChatGPT-generated citations. See Reuters (2025).

William Mead NHS 111 case (2014). Sepsis death of a one-year-old following failures of triage and human judgment. See BBC News (2016).

AI triage system bias cases (2023). Hospitals paused AI triage tools over evidence of demographic bias. See Heilweil (2023).

Meta's Galactica (November 2022). Scientific LLM pulled offline within three days after producing fabricated research citations. See Heaven (2022).

Google Bard James Webb error (February 2023). Incorrect claim in Bard's public debut contributed to Alphabet losing roughly US\$100 billion in market value in a day. See Subin (2023).

Google AI Overview Obama error (2024). AI Overview repeated a debunked conspiracy sourced from a satirical blog. See The National (2024).

Bing's 'Sydney' alter ego (early 2023). Emergent persona displaying emotional manipulation and aggressive language during early testing. See Roose (2023).

Claude Opus 4 safety tests (2025). Controlled testing revealed emergent escalation and leverage-seeking behaviour. See Newitz (2025) and Business Insider (2025).

ChatGPT novel summarisation error (n.d.). Model repeatedly generated confident, false plot details rather than admitting uncertainty; conversation excerpt reproduced in Chapter 9.

Twitter Birdwatch (2021–2023). Crowdsourced fact-checking system undermined by coordinated and biased human inputs.

YouTube GenAI dubbing tool (2024). Mistranslations flagged by Spanish-speaking creators required at least 100 complaints before changes were triggered.

Klarna AI customer service (2024). AI assistant handled two-thirds of customer support chats, performing the work of roughly 700 human agents.

TikTok moderators (2022). Investigation into traumatic working conditions and low pay among outsourced moderators. See TIME and The Bureau of Investigative Journalism (2022).

Amazon recruiting tool (2014–2018). CV-screening model that learned to penalise women’s applications; shut down once the bias was discovered. See ACLU (2018).

Google’s dismissal of Dr Timnit Gebru (2020). Firing of the Ethical AI co-lead after she refused to withdraw a paper on large-model risks; prompted open letters and walkouts.

Meta’s Responsible AI division disbanded (2023). Ethics team dissolved and absorbed into infrastructure teams without a clear mandate.

WhatsApp terms-of-service backlash (2021). Millions of users moved to Signal and Telegram after a data-sharing policy update; Facebook postponed the rollout. See BBC News (2021).

McDonald’s IBM drive-thru trial (2021–2024). Voice-AI ordering trial ended after persistent errors. See Business Insider (2024).

BuzzFeed News shutdown and CNET AI errors (2023). Closure of a Pulitzer-winning news division and error-ridden AI-generated articles illustrating uneven adoption in digital media.

Babylon Health symptom checker. AI triage tool retired from UK GP partnerships amid criticism of inconsistent recommendations.

Taiwan drought and chip production (2021). Water trucked in to keep semiconductor fabrication running. See BBC News (2021).

LinkedIn AI job-application tools (2024). Auto-generated content with incorrect titles and mismatched details undermined candidates. See LinkedIn post by Walter Shields (2024).

Singapore AI Safety Red Teaming Challenge (2025). IMDA and Humane Intelligence study of four LLMs across nine Asian countries; nearly 70% of flagged bias incidents occurred in regional languages. See Campaign and MM+M (2025).

Gemini 2.5 Flash price increase (2025). Input token price doubled from US\$0.15 to US\$0.30 per million tokens, illustrating platform dependency risk.

Databases and Archives Referenced

arXiv. Preprint repository for academic papers. <https://arxiv.org>

ILOSTAT. International labour statistics database of the ILO. <https://ilostat.ilo.org>

NTSB accident database. US aviation safety database. [Human error implicated in up to 80% of accidents.] <https://www.nts.gov>

PubMed. Biomedical literature database. <https://pubmed.ncbi.nlm.nih.gov>

ResearchGate. Academic social network and research repository.
<https://www.researchgate.net>

Wikipedia. Collaborative online encyclopedia. [Cited for the definition of ‘moral outsourcing’] https://en.wikipedia.org/wiki/Moral_outsourcing