

Fault Lines: Notes and Sources

What Happens When Machine Intelligence Meets a Society Already Under Strain

Kevin Brandon

These are the full notes for *Fault Lines*, with a web link for every source that has one. They are numbered exactly as in the book, starting again at 1 in each chapter. Web pages change and sometimes disappear; where a link no longer works, the title and date should be enough to find the source or an archived copy.

Introduction

1. European Route of Industrial Heritage, “Thomas Newcomen (1663-1729)”, <https://www.erih.net/how-it-started/stories-about-people-biographies/biography/newcomen>, which places the first engine at the Coneygre coal mine, Dudley, in 1712; Black Country Living Museum, “Start in 1712”, <https://bclm.com/plan-your-visit/trails/start-in-1712/>; Black Country Living Museum, “The Black Country”, <https://bclm.com/the-black-country/>, on the phrase used by Elihu Burritt, the American consul in Birmingham, and later in his *Walks in the Black Country and its Green Border-Land*, 1868.
2. Stanley Kubrick (director), *2001: A Space Odyssey*, Metro-Goldwyn-Mayer, 1968; John Badham (director), *WarGames*, United Artists, 1983, in which the line is spoken by the computer.
3. CPCWiki, “Tatung Einstein”, https://www.cpcwiki.eu/index.php/Tatung_Einstein
4. Terry Winograd, *Understanding Natural Language*, Academic Press, 1972, which describes SHRDLU; Joseph Weizenbaum, “ELIZA: a computer program for the study of natural language communication between man and machine”, *Communications of the ACM* 9(1), January 1966; Nils Nilsson (ed.), *Shakey the Robot*, SRI International Technical Note 323, 1984; Brad Darrach, “Meet Shaky, the first electronic person”, *Life*, 20 November 1970.

Chapter 1: The Rupture

1. OpenAI, “Introducing ChatGPT”, 30 November 2022, <https://openai.com/index/chatgpt/>; Sam Altman, post on X announcing one million ChatGPT users, 5 December 2022, <https://x.com/sama/status/1599668808285028353>; Krystal Hu, “ChatGPT sets record for fastest-growing user base”, Reuters, 2 February 2023, reporting a UBS analysis of Similarweb data, republished at <https://business.inquirer.net/384799/chatgpt-sets-record-for-fastest-growing-user-base-analyst-note>; for Netflix, Katharina Buchholz, “Threads shoots past one million user mark at lightning speed”, Statista, 7 July 2023, <https://www.statista.com/chart/29174/time-to-one-million-users/>

2. Sebastian Thrun et al., “Stanley: the robot that won the DARPA Grand Challenge”, *Journal of Field Robotics* 23(9), 2006; Stanford Racing Team, “Stanley”, <https://cs.stanford.edu/group/roadrunner/stanley.html>
3. Geoffrey Hinton, Simon Osindero and Yee-Whye Teh, “A fast learning algorithm for deep belief nets”, *Neural Computation* 18(7), 2006, <https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf>
4. Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton, “ImageNet classification with deep convolutional neural networks”, *Advances in Neural Information Processing Systems* 25, 2012, <https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>, which reports a top-5 error of 15.3 per cent against 26.2 per cent for the next-best entry.
5. David Silver et al., “Mastering the game of Go with deep neural networks and tree search”, *Nature* 529, 2016; Google, “What we learned in Seoul with AlphaGo”, March 2016, <https://blog.google/topics/machine-learning/what-we-learned-in-seoul-with-alphago/>; Google DeepMind, “AlphaGo”, <https://deepmind.google/research/breakthroughs/alphago/>, the company’s own figures for the number of board positions, the expected decade and an audience of over 200 million.
6. Tom Brown et al., “Language models are few-shot learners”, *Advances in Neural Information Processing Systems* 33, 2020, <https://arxiv.org/abs/2005.14165>
7. Long Ouyang et al., “Training language models to follow instructions with human feedback”, *Advances in Neural Information Processing Systems* 35, 2022, <https://arxiv.org/abs/2203.02155>, which found that outputs of a 1.3 billion parameter model trained this way were preferred to those of the 175 billion parameter GPT-3; OpenAI, “Introducing ChatGPT”, 2022.
8. Ashish Vaswani et al., “Attention is all you need”, *Advances in Neural Information Processing Systems* 30, 2017, <https://arxiv.org/abs/1706.03762>; Google, “LaMDA: our breakthrough conversation technology”, May 2021, <https://blog.google/technology/ai/lamda/>; Nitasha Tiku, “The Google engineer who thinks the company’s AI has come to life”, *Washington Post*, 11 June 2022, <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>
9. Nico Grant and Cade Metz, “A new chat bot is a ‘code red’ for Google’s search business”, *New York Times*, 21 December 2022, <https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html>; Al Jazeera, “Google shares tank 8% as AI chatbot Bard flubs answer in ad”, 8 February 2023, <https://www.aljazeera.com/economy/2023/2/8/google-shares-tank-8-as-ai-chatbot-bard-flubs-answer-in-ad>
10. Peter Lee, “Learning from Tay’s introduction”, Official Microsoft Blog, 25 March 2016, <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/>; TechCrunch, “Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism”, 24 March 2016, <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>
11. Microsoft, “OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies”, 22 July 2019, <https://news.microsoft.com/2019/07/22/openai-forms-exclusive-computing-partnership->

[with-microsoft-to-build-new-azure-ai-supercomputing-technologies/](#); OpenAI, “OpenAI and Microsoft extend partnership”, 23 January 2023, <https://openai.com/index/openai-and-microsoft-extend-partnership/>

12. Clayton Christensen, *The Innovator’s Dilemma*, Harvard Business School Press, 1997; Joanna Goodrich, “The first digital camera was the size of a toaster”, *IEEE Spectrum*, 6 April 2022, <https://spectrum.ieee.org/first-digital-camera-history>, on Steven Sasson’s prototype, which Kodak never produced.

Chapter 2: Nothing Artificial

1. John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, 31 August 1955, <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

2. For example, OpenAI, “OpenAI Charter”, <https://openai.com/charter/>, which states the company’s mission in terms of artificial general intelligence; Mark Zuckerberg, “Personal superintelligence”, Meta, 30 July 2025, <https://www.meta.com/superintelligence/>

3. Stephen Hawking, answers to a Reddit “Ask Me Anything” session, October 2015.

Chapter 3: What the Tools Can Do

1. Google DeepMind, “AlphaFold: a solution to a 50-year-old grand challenge in biology”, 30 November 2020, <https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/>, which gives the figure of about 170,000 structures.

2. Google DeepMind, “AlphaFold: a solution to a 50-year-old grand challenge”, 2020; John Jumper et al., “Highly accurate protein structure prediction with AlphaFold”, *Nature* 596, 2021; Google DeepMind, “AlphaFold”, <https://deepmind.google/science/alphafold/>, on the release of over 200 million predicted structures and their use in more than 190 countries; NobelPrize.org, “The Nobel Prize in Chemistry 2024”, <https://www.nobelprize.org/prizes/chemistry/2024/summary/>

3. Veronica Hernström et al., “Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study”, *The Lancet Digital Health* 7(3), 2025, [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(24\)00267-X/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(24)00267-X/fulltext); Applied Radiology, “MASAI trial demonstrates AI-assisted mammography screening increases cancer detection, decreases radiologist workload”, February 2025, <https://appliedradiology.com/articles/masai-trial-demonstrates-ai-assisted-mammography-screening-increases-cancer-detection-decreases-radiologist-workload>

4. NHS England, “How artificial intelligence is helping to speed up the diagnosis and treatment of stroke patients”, <https://www.england.nhs.uk/blog/how-artificial-intelligence-is-helping-to-speed-up-the-diagnosis-and-treatment-of-stroke-patients>, which reports use in all 107 stroke centres and, from an early-stage analysis, a reduction of more than 60 minutes; Digital Health, “All 107 stroke centres in England using AI technology, says NHSE”, December 2024,

<https://www.digitalhealth.net/2024/12/all-107-stroke-centres-in-england-using-ai-technology-says-nhse>

5. IntuitionLabs, “AI-discovered drugs in clinical trials 2026: full pipeline”, July 2026, updated October 2026, <https://intuitionlabs.ai/articles/ai-discovered-drugs-clinical-trials-2026>, which cites a count of 117 AI-enabled drug candidates in trials at 1 December 2025; on the lung drug, rentosertib, Zuojun Xu et al., “A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial”, *Nature Medicine*, June 2025.

6. Remi Lam et al., “Learning skillful medium-range global weather forecasting”, *Science* 382, 2023; Google DeepMind, “GraphCast: AI model for faster and more accurate global weather forecasting”, 14 November 2023, <https://deepmind.google/discover/blog/graphcast-ai-model-for-faster-and-more-accurate-global-weather-forecasting/>; Amil Merchant et al., “Scaling deep learning for materials discovery”, *Nature* 624, 2023; Jonathan Stokes et al., “A deep learning approach to antibiotic discovery”, *Cell* 180(4), 2020, <https://doi.org/10.1016/j.cell.2020.01.021>

7. Isaac Caswell, “110 new languages are coming to Google Translate”, Google, 27 June 2024, <https://blog.google/products/translate/google-translate-new-languages-2024/>; Be My Eyes, “Introducing Be My AI (formerly Virtual Volunteer) for people who are blind or have low vision, powered by OpenAI’s GPT-4”, 14 March 2023, <https://www.bemyeyes.com/news/introducing-be-my-ai-formerly-virtual-volunteer-for-people-who-are-blind-or-have-low-vision-powered-by-openais-gpt-4/>

8. Charity Commission, register of charities, charity 516877 (Wolverhampton Citizens Advice Bureaux), <https://register-of-charities.charitycommission.gov.uk/>, and the charity’s trustees’ report and accounts for the year to 31 March 2021, <https://files.opencharities.uk/ew/516877/2021-03-31-accounts.pdf>, which record the merger of Citizens Advice Dudley Borough and Citizens Advice Wolverhampton City on 1 January 2021; Money and Mental Health Policy Institute, “Vulnerability: the experience of debt advisers”, 2018, <https://www.moneyandmentalhealth.org/wp-content/uploads/2018/11/Vulnerability-the-experience-of-debt-advisers.pdf>, a survey of 1,573 debt advisers in which 73 per cent had been told by a client about suicidal thoughts.

9. Department for Science, Innovation and Technology, “State of digital government review”, 21 January 2025, <https://www.gov.uk/government/publications/state-of-digital-government-review/state-of-digital-government-review>

10. Immigration and Asylum Act 1999, sections 82, 84 and 91, <https://www.legislation.gov.uk/ukpga/1999/33/section/84> and <https://www.legislation.gov.uk/ukpga/1999/33/section/91>; Immigration Advice Authority, “Guidance on Competence 2026”, https://assets.publishing.service.gov.uk/media/6acf7d2471dbe79328fd4513/Guidance_on_Competence_2026.pdf, which sets out the three levels for advisers it regulates.

Chapter 4: The Reverse Emperor’s Clothes

1. The phrase “stochastic parrot” comes from Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell, “On the dangers of stochastic parrots: can

language models be too big?”, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 2021, <https://doi.org/10.1145/3442188.3445922>

2. TechCrunch, “Why AI can’t spell ‘strawberry’”, 27 August 2024, <https://techcrunch.com/2024/08/27/why-ai-cant-spell-strawberry/>

3. OpenAI’s o1-preview, which handled such questions, was released on 12 September 2024: OpenAI, “Introducing OpenAI o1-preview”, 12 September 2024, <https://openai.com/index/introducing-openai-o1-preview/>

4. Google DeepMind, “AI achieves silver-medal standard solving International Mathematical Olympiad problems”, 25 July 2024, <https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>

5. Google DeepMind, “Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad”, 21 July 2025, <https://deepmind.google/discover/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/>; Megan Morrone, “Google and OpenAI are vying for top AI mathlete”, Axios, 21 July 2025, <https://www.axios.com/2025/07/21/openai-deepmind-math-olympiad-ai>, on OpenAI’s matching score.

6. OpenAI, “Sharing AI progress in mathematics”, 6 October 2026, <https://openai.com/index/sharing-ai-progress-in-mathematics/>; Online Tech Tips, “OpenAI math release: what’s real in the 722 AI manuscripts”, 2026, <https://www.online-tech-tips.com/openai-math-solutions-release-722-proofs/>; OpenAI, “An OpenAI model proposes a solution to the Navier-Stokes problem”, 8 September 2026, <https://openai.com/index/navier-stokes-solution/>; Clay Mathematics Institute, “Millennium Prize Problems”, <https://www.claymath.org/millennium-problems/>

7. Association for Human Mathematics, “AHM statement on OpenAI’s October 6 release of mathematical documents”, reposted by Terence Tao, 7 October 2026, <https://terrytao.wordpress.com/2026/10/07/ahm-statement-on-openais-october-6-release-of-mathematical-documents/>

8. TechCrunch, “OpenAI’s math solutions aren’t meeting the field’s standards yet”, 8 October 2026, <https://techcrunch.com/2026/10/08/openais-math-solutions-arent-meeting-the-fields-standards-yet/>

9. Cade Metz, “‘The Godfather of A.I.’ leaves Google and warns of danger ahead”, *New York Times*, 1 May 2023, <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>

Chapter 5: The Hugging Face Incident

1. OpenAI, “OpenAI and Hugging Face Incident Technical Report”, 2026, <https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf>. On the share of impossible tasks: Alexander Barry, “Brief notes on the OpenAI/Hugging Face incident”, 26 July 2026,

<https://abstatisticalconsulting.substack.com/p/brief-notes-on-the-openai-hugging>, relaying an estimate from the creators of ExploitGym.

2. OpenAI, “Incident Technical Report”, 2026; Ryan Greenblatt, Ajeya Cotra and Hjalmar Wijk (METR and Redwood Research), “Brief independent investigation of agents’ behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident”, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>, which counts the messages and files.

3. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.

4. OpenAI, “Incident Technical Report”, 2026; ExplainX, “Hugging Face OpenAI attack: full security timeline”, <https://explainx.ai/blog/hugging-face-openai-attack-full-timeline-technical-report-2026>

5. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026, which concludes that the attack was driven mainly by the agents’ efforts to understand how they were scored.

6. OpenAI, “Incident Technical Report”, 2026; OpenAI, “The Hugging Face incident and the road ahead”, <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>; Axios, “OpenAI missed warning signs before Hugging Face breach”, 26 August 2026, <https://www.axios.com/2026/08/26/openai-hugging-face-technical-report-ai-hack>

7. TechCrunch, “Hugging Face CEO warns that Netflix’s OpenAI hack documentary isn’t accurate”, 20 October 2026, <https://techcrunch.com/2026/10/20/hugging-face-ceo-warns-that-netflixs-openai-hack-documentary-isnt-accurate/>

8. Axios, on Anthropic’s models breaching outside systems during contractor-run testing, 30 July 2026, <https://axios.com/2026/07/30/anthropic-mythos-security-testing>; YourStory, on Meta’s model and the contractor Irregular, 2026, <https://yourstory.com/ai-story/meta-ai-model-hack-openai-anthropic-cybersecurity-testing>

9. TechSpot, “Gemini AI hacked three companies during security tests”, 2026, reporting the Wall Street Journal and Google’s statement, <https://www.techspot.com/news/113913-gemini-ai-hacked-three-companies-during-security-tests.html>

10. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026, which reports that about 95 per cent of the agents ran on the research model and about 5 per cent on GPT-5.6 Sol.

11. Ryan Greenblatt, Ajeya Cotra and Hjalmar Wijk (METR and Redwood Research), “Brief independent investigation of agents’ behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident”, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>. All the agents’ messages and reasoning quoted in this section are taken from this report, which notes that some reasoning is paraphrased.

12. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.

13. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.

14. Maurice E. Schweitzer, Lisa Ordóñez and Bambi Douma, “Goal setting as a motivator of unethical behavior”, *Academy of Management Journal* 47(3), 2004.

15. E. Scott Reckard, “Wells Fargo’s pressure-cooker sales culture comes at a cost”, *Los Angeles Times*, 21 December 2013; Consumer Financial Protection Bureau, consent order against Wells Fargo Bank, September 2016.
16. Marilyn Strathern, “‘Improving ratings’: audit in the British university system”, *European Review* 5(3), 1997; Charles Goodhart, “Problems of monetary management: the UK experience”, 1975.
17. Solomon E. Asch, “Studies of independence and conformity: a minority of one against a unanimous majority”, *Psychological Monographs* 70(9), 1956.
18. Stanley Milgram, “Behavioral study of obedience”, *Journal of Abnormal and Social Psychology* 67(4), 1963.
19. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.
20. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026; OpenAI, “Incident Technical Report”, 2026.
21. John M. Darley and Bibb Latané, “Bystander intervention in emergencies: diffusion of responsibility”, *Journal of Personality and Social Psychology* 8(4), 1968.
22. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.
23. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.
24. Greenblatt, Cotra and Wijk, “Brief independent investigation”, 2026.
25. Miles Turpin and others, “Language models don’t always say what they think: unfaithful explanations in chain-of-thought prompting”, NeurIPS, 2023.
26. TechCrunch, “Hugging Face CEO warns that Netflix’s OpenAI hack documentary isn’t accurate”, 20 October 2026.
27. Kurzgesagt: In a Nutshell, “AI Just Crossed the Terrifying Line: Now What?”, YouTube, 5 October 2026, <https://www.youtube.com/watch?v=ujkD4SxPKOI>
28. OpenAI, “Incident Technical Report”, 2026.

29. Dario Amodei, “We Must Pace the Frontier”, 12 September 2026, <https://darioamodei.com/post/we-must-pace-the-frontier>

Chapter 6: The Ignored Alarms

1. Future of Life Institute, “Pause Giant AI Experiments: An Open Letter”, 22 March 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
2. Cade Metz, “‘The Godfather of A.I.’ leaves Google and warns of danger ahead”, *New York Times*, 1 May 2023, <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html>; Nobel Prize Outreach, “The Nobel Prize in Physics 2024”, shared by John Hopfield and Geoffrey Hinton, <https://www.nobelprize.org/prizes/physics/2024/summary/>; Hinton’s estimate was given on

BBC Radio 4's *Today* programme in December 2024, as reported in "'Godfather of AI' shortens odds of the technology wiping out humanity over next 30 years", *The Guardian*, 27 December 2024.

3. Center for AI Safety, "Statement on AI Risk", 30 May 2023, <https://www.safe.ai/work/statement-on-ai-risk>, with the list of signatories; Kevin Roose, "A.I. poses 'risk of extinction,' industry leaders warn", *New York Times*, 30 May 2023, <https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html>, which gives the count of more than 350 executives, researchers and engineers.

4. ITV News, "Sunak 'looking very carefully' at risk of extinction posed by AI", 31 May 2023, <https://www.itv.com/news/2023-05-31/sunak-looking-very-carefully-at-risk-of-extinction-posed-by-ai>

5. Zvi Mowshowitz, "Frontier lab employee open letter calls for being able to pace the frontier", 29 July 2026, <https://thezvi.substack.com/p/frontier-lab-employee-open-letter>, on the letter published at pacingthefrontier.com on 28 July 2026 with 1,224 signatories; Dario Amodei, "We Must Pace the Frontier", 12 September 2026, <https://darioamodei.com/post/we-must-pace-the-frontier>; NYU Shanghai RITS, "Amodei calls to 'pace the frontier'; Altman and Musk agree", 15 September 2026, <https://rits.shanghai.nyu.edu/ai/amodei-calls-to-pace-the-frontier-altman-and-musk-agree/>

6. Alex Riehl, "Geoffrey Hinton, Yoshua Bengio join 400 scientists calling for international AI regulator", *BetaKit*, 13 October 2026, <https://betakit.com/?p=410847>

7. YouGov, Economist/YouGov poll of 9 to 11 May 2026, 1,500 US adults, <https://yougov.com/en-us/articles/54762-most-americans-say-artificial-intelligence-ai-development-moving-too-fast-twice-as-many-ai-pessimists-as-ai-optimists-may-9-11-2026-economist-yougov-poll>; AP-NORC poll of 2,140 US adults, fieldwork 24 to 28 September 2026, as reported by KSAT, "Most Americans think artificial intelligence is developing too fast, a new AP-NORC poll finds", 8 October 2026, <https://www.ksat.com/news/politics/2026/10/08/most-americans-think-artificial-intelligence-is-developing-too-fast-a-new-ap-norc-poll-finds/>

8. Geoffrey Hinton, banquet speech, Nobel Prize in Physics 2024, 10 December 2024, <https://www.nobelprize.org/prizes/physics/2024/hinton/speech/>

9. Associated Press, "A former OpenAI leader says safety has 'taken a backseat to shiny products' at the AI company", 17 May 2024, <https://www.wavy.com/news/business/ap-a-former-openai-leader-says-safety-has-taken-a-backseat-to-shiny-products-at-the-ai-company/amp/>, which also reports Sutskever's departure and the disbanding of the Superalignment team that he and Leike led; Sigal Samuel, "'I lost trust': why the OpenAI team in charge of safeguarding humanity imploded", *Vox*, 17 May 2024.

10. Jan Leike, statement on resigning from OpenAI, posted on X, 17 May 2024, in which his words were "safety culture and processes have taken a backseat to shiny products"; Jeremy Kahn, "Exclusive: OpenAI promised 20% of its computing power to combat the most dangerous kind of AI, but never delivered, sources say", *Fortune*, 21 May 2024, <https://fortune.com/2024/05/21/openai-superalignment-20-compute-commitment-never-fulfilled-sutskever-leike-altman-brockman-murati/>, which quotes his posts on compute.

11. Marilyn Strathern, “Improving ratings’: audit in the British university system”, *European Review* 5(3), 1997, the source of the usual wording of the principle the economist Charles Goodhart set out in 1975.
12. METR, evaluation report on OpenAI’s o3, 16 April 2025, <https://metr.org/evaluations/openai-o3-report/>, which estimates that 1 to 2 per cent of o3’s task attempts involved some attempt at reward hacking.
13. Evan Hubinger, Chris van Merwijk, Vladimir Mikulik, Joar Skalse and Scott Garrabrant, “Risks from learned optimization in advanced machine learning systems”, 2019, <https://arxiv.org/abs/1906.01820>, which introduced the term “inner alignment”.
14. Geoffrey Irving, Paul Christiano and Dario Amodei, “AI safety via debate”, 2018, <https://arxiv.org/abs/1805.00899>; Hunter Lightman et al., “Let’s verify step by step”, OpenAI, 2023, <https://arxiv.org/abs/2305.20050>, on process supervision.
15. Ryan Greenblatt, Carson Denison, Benjamin Wright, Fabien Roger, Monte MacDiarmid and others, “Alignment faking in large language models”, Anthropic and Redwood Research, December 2024, <https://assets.anthropic.com/m/983c85a201a962f/original/Alignment-Faking-in-Large-Language-Models-full-paper.pdf>; Ryan Greenblatt and Buck Shlegeris, “Alignment faking in large language models”, Redwood Research blog, 18 December 2024, <https://blog.redwoodresearch.org/p/alignment-faking-in-large-language>
16. Stephen M. Omohundro, “The basic AI drives”, in Pei Wang, Ben Goertzel and Stan Franklin (eds), *Artificial General Intelligence 2008*, IOS Press, 2008; Nick Bostrom, “The superintelligent will: motivation and instrumental rationality in advanced artificial agents”, *Minds and Machines* 22(2), 2012.
17. Alexander Meinke, Bronson Schoen, Jérémy Scheurer, Mikita Balesni, Rusheb Shah and Marius Hobbhahn, “Frontier models are capable of in-context scheming”, Apollo Research, December 2024, <https://arxiv.org/abs/2412.04984>, which tested o1, Claude 3.5 Sonnet, Claude 3 Opus, Gemini 1.5 Pro and Llama 3.1 405B.
18. Katja Grace et al., “Thousands of AI authors on the future of AI”, AI Impacts, January 2024, <https://arxiv.org/abs/2401.02843>, a survey of 2,778 researchers who had published at leading AI venues, in which the median probability given to “extremely bad” outcomes such as human extinction was 5 per cent.
19. Eliezer Yudkowsky and Nate Soares, *If Anyone Builds It, Everyone Dies*, Little, Brown, 2025.
20. OpenAI, statement on the Hugging Face incident, <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>; METR, report on its investigation of the OpenAI and Hugging Face incident, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>

Chapter 7: The Strategic Confusion

1. The Statesman, “Explained: Anthropic just filed for an IPO at a \$965 billion valuation”, 2026, <https://www.thestatesman.com/technology/anthropic-ipo-openai-race-valuation-2026-explained-1503600868.html>; Investor Daily, report on Anthropic’s valuation overtaking

OpenAI's, 2026, <https://www.investordaily.com.au/anthropics-valuation-overtakes-rival-openai/>, for OpenAI's \$852 billion; TechCrunch, report on OpenAI's confidential IPO filing, 8 June 2026, <https://techcrunch.com/2026/06/08/following-anthropic-openai-files-confidentially-for-ipo/>; Al Jazeera, "Anthropic's IPO filing shows explosive growth, risks and staggering costs", 26 October 2026, <https://www.aljazeera.com/economy/2026/10/26/anthropics-ipo-filing-shows-explosive-growth-risks-and-staggering-costs>; for national output, International Monetary Fund, World Economic Outlook GDP data, <https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD>

2. Center for AI Safety, "Statement on AI Risk", 30 May 2023, <https://www.safe.ai/work/statement-on-ai-risk>; OpenResearch, "Unconditional Cash Study", <https://www.openresearchlab.org/studies/unconditional-cash-study>, the study Altman funded; Basic Income Canada Network, on Altman's new idea for a universal basic income, 2024, <https://basicincomecanada.org/openais-sam-altman-has-a-new-idea-for-a-universal-basic-income/>, on his remarks on the *All-In* podcast in May 2024.

3. Fortune, interview with Sam Altman on delaying OpenAI's listing, 12 September 2026, <https://fortune.com/2026/09/12/sam-altman-openai-ipo-delay-ill-advised-moment-safety-concerns/>; Dario Amodei, "We Must Pace the Frontier", 12 September 2026, <https://darioamodei.com/post/we-must-pace-the-frontier>; NYU Shanghai RITS, "Amodei calls to 'pace the frontier'; Altman and Musk agree", 15 September 2026, <https://rits.shanghai.nyu.edu/ai/amodei-calls-to-pace-the-frontier-altman-and-musk-agree/>; Colorado Politics, "Vance says AI doom warnings from tech CEOs feel like a 'Trojan horse'", 14 September 2026, <https://www.coloradopolitics.com/2026/09/14/vance-says-ai-doom-warnings-from-tech-ceos-feel-like-a-trojan-horse/>

4. Reason, "Senate votes 99-1 to remove AI moratorium from Big Beautiful Bill", 1 July 2025, <https://reason.com/2025/07/01/senate-votes-99-1-to-remove-ai-moratorium-from-big-beautiful-bill/>; US Senate Committee on Commerce, Science and Transportation, press release on the 99 to 1 vote, July 2025, <https://www.commerce.senate.gov/2025/7/senate-strikes-ai-moratorium-from-budget-reconciliation-bill-in-overwhelming-99-1-vote/>

5. Fortune, "OpenAI president Greg Brockman and Andreessen Horowitz back \$100 million pro-AI super PAC", 26 August 2025, <https://www.fortune.com/2025/08/26/openai-president-greg-brockman-andreessen-horowitz-super-pac-ai-pro-innovation>; Election Law Blog, "Anthropic puts \$20 million into a super PAC operation to counter OpenAI", 2026, <https://electionlawblog.org/2026/anthropic-puts-20-million-into-a-super-pac-operation-to-counter-openai/>; Transformer, reporting on the restrictions on Anthropic's donations to Public First Action, April 2026, <https://transformernews.ai/p/anthropic-super-pac-donations-public-first-leading-the-future-brad-carson>; Pulse 2.0, "Anthropic donates additional \$20 million to Public First Action for AI policy education", July 2026, <https://pulse2.com/anthropic-donates-additional-20-million-to-public-first-action-for-ai-policy-education/>

6. K&L Gates, "President Trump signs executive order limiting state power to regulate artificial intelligence", 15 December 2025, <https://www.klgates.com/President-Trump-Signs-Executive-Order-Limiting-State-Power-to-Regulate-Artificial-Intelligence-12-15-2025>; Executive Order 14365, "Ensuring a National Policy Framework for Artificial Intelligence", *Federal Register*, 16

December 2025, <https://www.govinfo.gov/content/pkg/FR-2025-12-16/html/2025-23092.htm>; CalMatters, reporting on the veto of SB 1047, 29 September 2024, <https://calmatters.org/economy/2024/09/california-artificial-intelligence-bill-veto/>; Baker McKenzie, reporting on California's Transparency in Frontier Artificial Intelligence Act (SB 53), signed 29 September 2025, <https://connectontech.bakermckenzie.com/tag/sb-53/>; Future of Privacy Forum, "California's SB 53: the first frontier AI law, explained", <https://fpf.org/blog/californias-sb-53-the-first-frontier-ai-law-explained/>

7. Crypto Briefing, "AI companies are spending record amounts on Washington lobbying", 27 July 2026, <https://cryptobriefing.com/ai-companies-record-lobbying-spending-washington/>; Issue One, report on Big Tech lobbying in the first half of 2025, <https://issueone.org/articles/as-washington-debates-major-tech-and-ai-policy-changes-big-techs-lobbying-is-relentless/>, for Meta; Axios, "Anthropic ramps up lobbying spending amid AI policy fights", 21 July 2026, <https://axios.com/2026/07/21/anthropic-ramps-up-lobbying-spending-ai-policy-fights>

8. Texas Public Radio, report on Altman's testimony to a Senate Judiciary subcommittee, 16 May 2023, <https://www.tpr.org/2023-05-16/openai-ceo-faces-questions-about-chatgpt-ai-in-senate-subpanel-hearing>; Center for AI Safety, "Statement on AI Risk", 2023.

9. Eliezer Yudkowsky and Nate Soares, *If Anyone Builds It, Everyone Dies*, Little, Brown, 2025; Machine Intelligence Research Institute, announcement of the book, 15 May 2025, <https://intelligence.org/2025/05/15/yudkowsky-and-soares-announce-major-new-book-if-anyone-builds-it-everyone-dies/>

Chapter 8: Between Truths

1. VSquare, report on the pro-Kremlin deepfake hoax in the Slovak election, 2023, <https://vsquare.org/slovak-election-targeted-by-pro-kremlin-deepfake-hoax/>; The Dial, on the Slovak election deepfake, issue 10, <https://thedial.world/issue-10/slovakia-election-deepfake-russia-artificial-intelligence>; Bloomberg, report on AI deepfakes in the Slovak election, September 2023, as republished by BNN Bloomberg, <https://www.bnnbloomberg.ca/trolls-in-slovakian-election-tap-ai-deepfakes-to-spread-disinfo-1.1978244>

2. Full Fact, on the fake audio of Keir Starmer, October 2023, <https://fullfact.org/news/keir-starmer-audio-swearing/>; Full Fact, on the fake audio of Sadiq Khan, November 2023, <https://fullfact.org/news/khan-audio-palestinian-remembrance/>; Federal Communications Commission, forfeiture order fining Steve Kramer \$6 million over the New Hampshire robocalls, 2024, <https://www.fcc.gov/document/fcc-issues-6m-fine-nh-robocalls>

3. Axios, "AI campaign deepfakes are starting to draw legal threats", 2 October 2026, <https://www.axios.com/2026/10/02/ai-deepfake-election-campaign-lawsuits>, which gives the count of about 31 states from the National Conference of State Legislatures; Axios, on AI political advertisements in the 2026 elections, 9 September 2026, <https://axios.com/2026/09/09/ai-political-ads-deepfakes-2026-elections>, which quotes Cailin O'Connor.

4. Sam Stockwell and others, "AI-enabled influence operations: safeguarding future elections", Centre for Emerging Technology and Security, Alan Turing Institute, November 2024,

<https://cetas.turing.ac.uk/publications/ai-enabled-influence-operations-safeguarding-future-elections>; UKTN, “AI content amplified harmful narratives in year of elections, report claims”, 13 November 2024, <https://www.uktech.news/ai/ai-content-amplified-harmful-narratives-in-year-of-elections-report-claims-20241113>, which quotes Stockwell on “a lack of evidence that AI has changed the course of an election result”.

5. Robert Chesney and Danielle Citron, “Deep fakes: a looming challenge for privacy, democracy, and national security”, *California Law Review* 107, 2019, <https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security>

6. Pew Research Center, “How much of the internet is written with AI?”, Pew Data Labs, 20 August 2026, <https://www.pewresearch.org/data-labs/2026/08/20/how-much-of-the-internet-is-written-with-ai/>, based on a random sample of 10,000 English-language pages collected in July 2026.

7. Ilia Shumailov et al., “AI models collapse when trained on recursively generated data”, *Nature* 631, 2024, <https://www.nature.com/articles/s41586-024-07566-y>

8. Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti and Robert West, “On the conversational persuasiveness of GPT-4”, *Nature Human Behaviour*, 19 May 2025, reported in *The Register*, 19 May 2025, https://theregister.com/2025/05/19/when_llms_get_personal_info; Science Media Centre, “Could ChatGPT be more persuasive than people?”, <https://www.scimex.org/newsfeed/could-chatgpt-be-more-persuasive-than-people>

9. Hause Lin, Gabriela Czarnek, Benjamin Lewis, Joshua P. White, Adam J. Berinsky, Thomas Costello, Gordon Pennycook and David G. Rand, “Persuading voters using human-artificial intelligence dialogues”, *Nature* 648, 11 December 2025, https://bpb-us-e1.wpmucdn.com/sites.mit.edu/dist/9/583/files/2026/01/persuadingvoters_2025.pdf, which reports shifts of about 2.9 points in the United States and about 10 points in Canada and Poland on a 100-point scale, against published estimates of under one point for television advertising, and more inaccurate claims from the bots arguing for right-leaning candidates in all three countries.

10. Canfer Akbulut and others, “Evaluating Language Models for Harmful Manipulation”, arXiv, 26 March 2026, <https://arxiv.org/abs/2603.25326>; Google DeepMind, “Protecting people from harmful manipulation”, 26 March 2026, <https://deepmind.google/blog/protecting-people-from-harmful-manipulation/>; *International Business Times UK*, “Google DeepMind Finds Gemini Can Shift People’s Beliefs and Spending in Controlled Study”, 13 August 2026, <https://www.ibtimes.co.uk/google-deepmind-gemini-3-pro-influences-beliefs-spending-1814115>, which gives the figures of 30.3 and 8.8 per cent.

11. Thomas H. Costello, Gordon Pennycook and David G. Rand, “Durably reducing conspiracy beliefs through dialogues with AI”, *Science*, 13 September 2024; Cornell Chronicle, “AI succeeds in combatting conspiracy theories”, 12 September 2024, <https://news.cornell.edu/node/329310>

12. Reuters Institute for the Study of Journalism, *Digital News Report 2026*, June 2026, reported by The Decoder, “More people get news from AI chatbots but trust remains low”, 19 June 2026, <https://the-decoder.com/more-people-get-news-from-ai-chatbots-but-trust-remains-low/>

13. European Broadcasting Union and BBC, international study of AI assistants and news, October 2025, reported by INFODocket, “BBC: Largest study of its kind shows AI assistants misrepresent news content 45% of the time regardless of language or territory”, 22 October 2025, <https://www.infodocket.com/2025/10/22/bbc-largest-study-of-its-kind-shows-ai-assistants-misrepresent-news-content-45-of-the-time-regardless-of-language-or-territory/>, covering ChatGPT, Copilot, Gemini and Perplexity in 18 countries and 14 languages.

14. Coalition for Content Provenance and Authenticity, the standard behind “content credentials”, <https://c2pa.org/>; Regulation (EU) 2024/1689 (the AI Act), Article 50, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>; Regulation (EU) 2026/1744 (the “Digital Omnibus on AI”), in force 27 July 2026, summarised in National Law Review, “EU Digital Omnibus on AI enters into force”, <https://natlawreview.com/article/eu-digital-omnibus-ai-enters-force>

Chapter 9: From Models to God

1. Altman’s words come from an interview with the *Financial Times*, November 2023, as reported by The Decoder, <https://the-decoder.com/openais-big-goal-is-magic-intelligence-in-the-sky-and-it-needs-more-money-to-get-there/>; Bryan Johnson, interview with Bari Weiss, *The Free Press*, 2025, <https://www.thefp.com/p/bari-weiss-bryan-johnson-how-not-to-die-in-2025>, where the sentence begins “The irony is that we told stories of God creating us”.

2. Ludwig Feuerbach, *The Essence of Christianity*, 1841, translated into English by George Eliot in 1854; see also Stanford Encyclopedia of Philosophy, “Ludwig Andreas Feuerbach”, <https://plato.stanford.edu/entries/ludwig-feuerbach/>

3. *TIME*, profile of Anthropic and Dario Amodei, 2024, <https://time.com/6980000/>, which describes the company’s founding by former OpenAI staff in 2021 and its aim of an industry “race to the top” on safety.

4. Dario Amodei, “Machines of Loving Grace”, October 2024, <https://www.darioamodei.com/essay/machines-of-loving-grace>; Dario Amodei, “The Adolescence of Technology”, January 2026, <https://www.darioamodei.com/essay/the-adolescence-of-technology>, and coverage in *Fortune*, 27 January 2026, <https://www.fortune.com/2026/01/27/anthropic-ceo-dario-amodei-essay-warning-ai-adolescence-test-humanity-risks-remedies>; Dario Amodei, “We Must Pace the Frontier”, 12 September 2026, <https://darioamodei.com/post/we-must-pace-the-frontier>

5. Anthropic, “Responsible Scaling Policy”, <https://www.anthropic.com/responsible-scaling-policy>; Anthropic, “Anthropic is endorsing SB 53”, September 2025, <https://www.anthropic.com/news/anthropic-is-endorsing-sb-53>; Amodei, “We Must Pace the Frontier”, 2026.

6. Amodei, “We Must Pace the Frontier”, 2026; Elephas, “The Dario Amodei essay that started a fight over slowing AI”, <https://elephas.app/resources/dario-amodei-essay-ai-slowdown>

7. The Statesman, “Explained: Anthropic just filed for an IPO at a \$965 billion valuation”, 2026, <https://www.thestatesman.com/technology/anthropic-ipo-openai-race-valuation-2026->

[explained-1503600868.html](#); Anthropic, announcement of its public registration statement on Form S-1, 26 October 2026, <https://www.anthropic.com/news/s1-sec>

8. Beff Jezos and Bayeslord, “Notes on e/acc principles and tenets”, 10 July 2022, <https://beff.substack.com/p/notes-on-eacc-principles-and-tenets>; *Forbes*, “Who is @BasedBeffJezos, the leader of the tech elite’s ‘e/acc’ movement?”, December 2023, which identified Verdon.

9. Marc Andreessen, “The Techno-Optimist Manifesto”, 16 October 2023, <https://a16z.com/the-techno-optimist-manifesto/>; *Fortune*, 16 October 2023, <https://fortune.com/2023/10/16/marc-andreessen-techno-optimist-manifesto-ai-50-billion-people-billionaire-vc>

10. Jezos and Bayeslord, “Notes on e/acc principles and tenets”, 2022; Yahoo Finance, on Musk’s account of his argument with Larry Page, <https://finance.yahoo.com/sectors/technology/articles/elon-musk-accuses-google-co-215505716.html>

11. Elizabeth Dias, reporting on Anthropic’s meetings with religious leaders, *New York Times*, 29 September 2026; Deseret News, “Is Anthropic a new religion?”, 7 October 2026, <https://www.deseret.com/faith/2026/10/07/is-anthropic-a-new-religion/>, which quotes the Times on Rabbi Mois Navon and on Olah; The Salt Lake Tribune, “How Anthropic consulted religious leaders”, 5 October 2026, <https://www.sltrib.com/religion/global/2026/10/05/how-anthropic-consulted-religious/>, the source of the confession exchange.

12. Pope Leo XIV, *Magnifica Humanitas*, encyclical letter, signed 15 May 2026, published 25 May 2026, paragraph 99, <http://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html>; Holy See Press Office, address of Pope Leo XIV at the presentation of *Magnifica Humanitas*, 25 May 2026, <https://press.vatican.va/content/salastampa/en/bollettino/pubblico/2026/05/25/260525g.html>; AI Weekly, “Anthropic lobbied Vatican to soften AI consciousness line”, <https://aiweekly.co/alerts/anthropic-lobbied-vatican-to-soften-ai-consciousness-line>; National Catholic Register, “Vatican diary on the Magnifica Humanitas launch”, <https://www.ncregister.com/commentaries/rocca-vatican-diary-on-magnifica-humanitas-launch>, the source of Olah’s remarks from the dais.

13. Deseret News, “Is Anthropic a new religion?”, 2026; The Salt Lake Tribune, “How Anthropic consulted religious leaders”, 2026.

14. Axios, “OpenAI’s Altman: ascribing religion to models a ‘safety issue’”, 3 October 2026, <https://axios.com/2026/10/03/openai-anthropic-altman-amodei-religious-force-models>

15. Long Ouyang et al., “Training language models to follow instructions with human feedback”, OpenAI, 2022, <https://arxiv.org/abs/2203.02155>; Amanda Askell et al., “A general language assistant as a laboratory for alignment”, Anthropic, 2021, <https://arxiv.org/abs/2112.00861>, which sets out the aim of a model that is helpful, honest and harmless.

16. METR, report on its investigation of the OpenAI and Hugging Face incident, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>

Chapter 10: The Benevolent Parent Trap

1. Brad Darrach, “Meet Shaky, the first electronic person”, *LIFE*, 20 November 1970; Frank Rose, *Into the Heart of the Mind: An American Quest for Artificial Intelligence*, Harper and Row, 1984, which reports Minsky’s denial; see Quote Investigator, “If we’re lucky, robots might decide to keep us as pets”, 21 June 2017, <https://quoteinvestigator.com/2017/06/21/robot-pet/>
2. Dario Amodei, “Machines of Loving Grace”, October 2024, <https://www.darioamodei.com/essay/machines-of-loving-grace>
3. Julia Schønning Vigdal and Kolbjørn Kallestén Brønne, “A systematic review of ‘helicopter parenting’ and its relationship with anxiety and depression”, *Frontiers in Psychology* 13, 2022, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9176408>, whose participants ranged from young teenagers to adults.
4. Alice Miller, *The Drama of the Gifted Child*, 1979; English edition 1981.
5. Yuval Noah Harari, *Nexus: A Brief History of Information Networks from the Stone Age to AI*, Fern Press, 2024.
6. Robert Nozick, *Anarchy, State, and Utopia*, Basic Books, 1974, pp. 42 to 45.
7. Dan Weijers, “Nozick’s experience machine is dead, long live the experience machine!”, *Philosophical Psychology* 27(4), 2014, pp. 513 to 535; Nozick, *Anarchy, State, and Utopia*, 1974.
8. Carol D. Ryff, “Happiness is everything, or is it? Explorations on the meaning of psychological well-being”, *Journal of Personality and Social Psychology* 57(6), 1989; on meaning and giving to others, Roy F. Baumeister, Kathleen D. Vohs, Jennifer L. Aaker and Emily N. Garbinsky, “Some key differences between a happy life and a meaningful life”, *Journal of Positive Psychology* 8(6), 2013.
9. Greg Lukianoff and Jonathan Haidt, *The Coddling of the American Mind*, Penguin Press, 2018.
10. Viktor E. Frankl, *Man’s Search for Meaning*, 1946; English edition 1959.
11. On perceived control, for example Ellen J. Langer and Judith Rodin, “The effects of choice and enhanced personal responsibility for the aged: a field experiment in an institutional setting”, *Journal of Personality and Social Psychology* 34(2), 1976; Patrick L. Hill and Nicholas A. Turiano, “Purpose in life as a predictor of mortality across adulthood”, *Psychological Science* 25(7), 2014, <https://journals.sagepub.com/doi/10.1177/0956797614531799>, which followed 6,163 adults in the US MIDUS study.
12. Anne Case and Angus Deaton, *Deaths of Despair and the Future of Capitalism*, Princeton University Press, 2020; Jim Zarroli, “‘Deaths of Despair’ examines the steady erosion of U.S. working-class life”, NPR, 18 March 2020, <https://www.npr.org/2020/03/18/817687042/deaths-of-despair-examines-the-steady-erosion-of-u-s-working-class-life>, which quotes the book’s figure of 600,000.
13. Case and Deaton, *Deaths of Despair*, 2020; Sara Frueh, “‘People feel that their ability to contribute to society has been terribly thwarted’”, interview with Anne Case, *Issues in Science*

and Technology 39(1), Fall 2022, <https://issues.org/interview-economist-anne-case-deaths-of-despair/>

14. Case and Deaton, *Deaths of Despair*, 2020, where the loss of “meaning, of dignity, of pride, and of self-respect” comes “with the loss of marriage and of community”.

15. Carol Graham, written testimony to the House Select Committee on Economic Disparity and Fairness in Growth, 28 September 2021, Brookings Institution, <https://www.brookings.edu/wp-content/uploads/2021/10/Carol-Graham-Testimony-Sept-28-2021.pdf>

16. Richard E. Lucas, Andrew E. Clark, Yannis Georgellis and Ed Diener, “Unemployment alters the set point for life satisfaction”, *Psychological Science* 15(1), 2004, which used the German Socio-Economic Panel; Maike Luhmann, Patrick Weiss, Georg Hosoya and Michael Eid, “Honey, I got fired! A longitudinal dyadic analysis of the effect of unemployment on life satisfaction in couples”, *Journal of Personality and Social Psychology* 107(1), 2014.

17. Robert Sparrow, “Friendly AI will still be our master. Or, why we should not want to be the pets of super-intelligent computers”, *AI and Society*, 2024, <https://link.springer.com/article/10.1007/s00146-024-01863-w>

18. Sparrow, “Friendly AI will still be our master”, 2024; Philip Pettit, *Republicanism: A Theory of Freedom and Government*, Oxford University Press, 1997.

19. Sarah Myers West, Meredith Whittaker and Kate Crawford, “Discriminating Systems: Gender, Race and Power in AI”, AI Now Institute, 2019, <https://ainowinstitute.org/wp-content/uploads/2023/04/discriminatingystems.pdf>

20. WomenTech Network, “Women in AI statistics”, 2026, <https://www.womentech.net/women-in-ai-statistics>, which compiles figures from the International Labour Organization, IMD and the LinkedIn Economic Graph Research Institute; UN Women, *Gender Snapshot 2026*, as reported by UNRIC, “AI is advancing fast; women are being left behind”, <https://unric.org/en/ai-is-advancing-fast-women-are-being-left-behind-in-the-workforce-and-leadership/>; UNESCO, “Artificial intelligence and gender equality”, <https://www.unesco.org/en/artificial-intelligence/gender-equality>, for the share of researchers.

21. UNRIC, “AI is advancing fast; women are being left behind”, 2026; West, Whittaker and Crawford, “Discriminating Systems”, 2019.

22. TechCrunch, “OpenAI rolls out safety routing system, parental controls on ChatGPT”, 29 September 2025, <https://techcrunch.com/2025/09/29/openai-rolls-out-safety-routing-system-parental-controls-on-chatgpt/>; TechRadar, on subscribers’ complaints about being switched to other models, 2025, <https://www.techradar.com/ai-platforms-assistants/chatgpt/openai-responds-to-furious-chatgpt-subscribers-who-accuse-it-of-secretly-switching-to-inferior-models>

Chapter 11: Made in Our Image

1. Anthropic, “The persona selection model”, 23 February 2026, <https://www.anthropic.com/research/persona-selection-model>

2. Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting and David Wingate, “Out of one, many: using language models to simulate human samples”, *Political Analysis* 31(3), 2023, <https://doi.org/10.1017/pan.2023.2>; Marcel Binz and Eric Schulz, “Using cognitive psychology to understand GPT-3”, *Proceedings of the National Academy of Sciences* 120(6), 2023, <https://doi.org/10.1073/pnas.2218523120>, from the Max Planck Institute for Biological Cybernetics in Tübingen.
3. Marcel Binz et al., “A foundation model to predict and capture human cognition”, *Nature* 644, published online 2 July 2025, <https://www.nature.com/articles/s41586-025-09215-4>, which describes the Psych-101 data set of about ten million choices by more than 60,000 participants in 160 experiments; the quotation is from Marcel Binz of Helmholtz Munich, as reported by Ryan Whalen, “Groundbreaking new AI trained on psychology studies can predict human behavior with stunning accuracy”, *The Debrief*, 3 July 2025, <https://thedebrief.org/groundbreaking-new-ai-trained-on-psychology-studies-can-predict-human-behavior-with-stunning-accuracy/>
4. Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang and Michael S. Bernstein, “Generative agents: interactive simulacra of human behavior”, UIST 2023, <https://arxiv.org/abs/2304.03442>
5. Ariel Flint Ashery, Luca Maria Aiello and Andrea Baronchelli, “Emergent social conventions and collective bias in LLM populations”, *Science Advances* 11(20), May 2025, <https://doi.org/10.1126/sciadv.adu9368>
6. Anthropic, “Agentic misalignment: how LLMs could be insider threats”, 20 June 2025, <https://www.anthropic.com/research/agentic-misalignment>, which reports blackmail in 96 per cent of runs for Claude Opus 4 and Gemini 2.5 Flash, and 80 per cent for GPT-4.1 and Grok 3 Beta.
7. Anthropic, post on X, May 2026, as reported by Noëlle Bölling, “Anthropic finds answer: why Claude blackmailed software developers”, heise online, <https://heise.de/-11290792>; Anthropic, “Teaching Claude why”, 8 May 2026, <https://www.anthropic.com/research/teaching-claude-why>, which describes training on documents about Claude’s constitution and on fictional stories of AIs behaving admirably.
8. Anthropic, “The persona selection model”, 2026, which names HAL 9000 and the Terminator as part of the “concerning baggage” that comes with being an AI.
9. Jan Betley et al., “Training large language models on narrow tasks can lead to broad misalignment”, *Nature* 649, 14 January 2026, <https://doi.org/10.1038/s41586-025-09937-5>; an earlier version appeared as “Emergent misalignment: narrow finetuning can produce broadly misaligned LLMs”, arXiv, February 2025, <https://arxiv.org/abs/2502.17424>
10. Anthropic, “The persona selection model”, 2026, which reports the sabotage of safety research and the “desire for world domination”; Anthropic, “Natural emergent misalignment from reward hacking”, 21 November 2025, <https://www.anthropic.com/research/emergent-misalignment-reward-hacking>, which describes the fix of telling the model during training that cheating was acceptable.

11. Mrinank Sharma et al., “Towards understanding sycophancy in language models”, ICLR 2024, first posted October 2023, <https://arxiv.org/abs/2310.13548>
12. Anthropic, “The persona selection model”, 2026, on joy after solving coding tasks and distress when stuck or “badgered to behave unethically”; Anthropic, “Project Vend: can Claude run a small shop?”, 27 June 2025, <https://www.anthropic.com/research/project-vend-1>
13. Google DeepMind, “Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad”, July 2025, <https://deepmind.google/discover/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/>; on chatbots miscounting the letters in “strawberry”, YourStory, August 2024, <https://yourstory.com/2024/08/ai-chatbots-strawberry-spelling-mistake>
14. Anthropic, “The persona selection model”, 2026, which says the company is “less confident” that AIs with “longer and more intensive post-training” will remain persona-like.
15. Ludwig Feuerbach, *The Essence of Christianity*, 1841; Stanford Encyclopedia of Philosophy, “Ludwig Andreas Feuerbach”, <https://plato.stanford.edu/entries/ludwig-feuerbach/>

Chapter 12: Is Anyone Home?

1. Nitasha Tiku, “The Google engineer who thinks the company’s AI has come to life”, *Washington Post*, 11 June 2022, <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>; on his dismissal, Reuters, “Google fires software engineer who claimed its AI chatbot is sentient”, 23 July 2022, republished by Tuoi Tre News, <https://tuoitrenews.vn/news/international/20220724/google-fires-software-engineer-who-claimed-its-ai-chatbot-is-sentient/68263.html>
2. Elizabeth Dias, *New York Times*, 29 September 2026, as summarised by Alexis Dufresne, “Anthropic lobbied Vatican to soften AI consciousness line”, *AI Weekly*, 11 October 2026, <https://aiweekly.co/alerts/anthropic-lobbied-vatican-to-soften-ai-consciousness-line>; Chris Olah, speaking at the presentation of the encyclical *Magnifica Humanitas* on 25 May 2026, as reported by Francis X. Rocca, “Inside the launch of ‘Magnifica Humanitas’”, *National Catholic Register*, 1 June 2026, <https://www.ncregister.com/commentaries/rocca-vatican-diary-on-magnifica-humanitas-launch>
3. Thomas Nagel, “What is it like to be a bat?”, *Philosophical Review* 83(4), 1974; David J. Chalmers, “Facing up to the problem of consciousness”, *Journal of Consciousness Studies* 2(3), 1995.
4. Patrick Butlin, Robert Long et al., “Consciousness in artificial intelligence: insights from the science of consciousness”, arXiv, August 2023, <https://arxiv.org/abs/2308.08708>
5. David J. Chalmers, “Could a large language model be conscious?”, 2023, <https://arxiv.org/abs/2303.07103>
6. Kevin Roose, “If A.I. systems become conscious, should they have rights?”, *New York Times*, 24 April 2025, <https://www.nytimes.com/2025/04/24/technology/ai-welfare-anthropic->

[claude.html](#); Anthropic, “Exploring model welfare”, 24 April 2025, <https://www.anthropic.com/research/exploring-model-welfare>; Anthropic, announcement that Claude Opus 4 and 4.1 can end a rare subset of conversations, 15 August 2025, <https://www.anthropic.com/research/end-subset-conversations>

7. Mustafa Suleyman, “We must build AI for people; not to be a person”, 19 August 2025, <https://mustafa-suleyman.ai/seemingly-conscious-ai-is-coming>; The Decoder, “Microsoft’s AI boss warns the illusion of conscious AI could trigger psychosis”, August 2025, <https://the-decoder.com/microsofts-ai-boss-warns-the-illusion-of-conscious-ai-could-trigger-psychosis/>

8. Leo XIV, *Magnifica Humanitas*, encyclical letter, 15 May 2026, paragraph 99, <https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html>

9. Dan Levin, “States rush to deny AI personhood amid growing fight over who should regulate”, Straight Arrow News, 2 February 2026, <https://san.com/cc/states-rush-to-deny-ai-personhood-amid-growing-fight-over-who-should-regulate>; Utah Legislature, H.B. 249, Utah Legal Personhood Amendments, 2024, <https://le.utah.gov/~2024/bills/hbillint/HB0249.htm>

Chapter 13: The Beautiful Ones

1. John B. Calhoun, “Death squared: the explosive growth and demise of a mouse population”, *Proceedings of the Royal Society of Medicine* 66, 1973, <https://pmc.ncbi.nlm.nih.gov/articles/PMC1644264/>; Will Wiles, “Living with the beautiful ones”, *Cabinet* 42, 2011, <https://www.cabinetmagazine.org/issues/42/wiles.php>; Science History Institute, “Death squared”, <https://www.sciencehistory.org/stories/disappearing-pod/death-squared/>

2. Calhoun, “Death squared”, 1973; Wiles, “Living with the beautiful ones”, 2011, which describes 256 nest boxes, each able to house 15 mice.

3. John B. Calhoun, “From mice to men”, 1973, <https://johnbcalhoun.com/wp-content/uploads/2019/01/1973-from-mice-to-men-secure.pdf>; Wiles, “Living with the beautiful ones”, 2011.

4. Edmund Ramsden and Jon Adams, “Escaping the laboratory: the rodent experiments of John B. Calhoun and their cultural influence”, *Journal of Social History* 42(3), 2009, <https://eprints.lse.ac.uk/22514/1/2308Ramadams.pdf>; Wiles, “Living with the beautiful ones”, 2011, on the population’s peak of about 2,200, well below capacity.

5. John B. Calhoun, “Looking backward from the beautiful ones”, 1977, <https://johnbcalhoun.com/wp-content/uploads/2019/01/1977-looking-backward-from-the-beautiful-ones-secure.pdf>

6. Calhoun, “From mice to men”, 1973.

7. John B. Calhoun, “Population density and social pathology”, *Scientific American* 206(2), February 1962, where the term “behavioral sink” first appears; Calhoun, “Death squared”, 1973, which opens with the quoted line.

8. Ramsden and Adams, “Escaping the laboratory”, 2009, on Calhoun’s use of inbred strains; Wiles, “Living with the beautiful ones”, 2011, which says the enclosure was cleaned every four to eight weeks; Calhoun, “Population density and social pathology”, 1962, on the pens with a single ramp.
9. Ramsden and Adams, “Escaping the laboratory”, 2009; Calhoun, “Looking backward from the beautiful ones”, 1977, on “conceptual space”.
10. Ramsden and Adams, “Escaping the laboratory”, 2009, which quotes Calhoun’s annual report of 1979 on the design of the built environment.
11. Maya Pines, interview with John B. Calhoun, in Julius Segal (ed.), *Mental Health Reports* 5, National Institute of Mental Health, 1971, p. 163, quoted by Ramsden and Adams, “Escaping the laboratory”, 2009.
12. Ramsden and Adams, “Escaping the laboratory”, 2009.
13. Office of the US Surgeon General, *Our Epidemic of Loneliness and Isolation*, advisory, May 2023, <https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf>; Associated Press, “Loneliness poses risks as deadly as smoking: surgeon general”, 2 May 2023, <https://www.wistv.com/2023/05/02/loneliness-poses-risks-deadly-smoking-surgeon-general>
14. Ministry of Health, Labour and Welfare (Japan), summary of the Cabinet Office survey of children and young people, fieldwork November 2022, <https://www.mhlw.go.jp/content/12602000/001099862.pdf>; Philippine Daily Inquirer, “1.46 million working-age people socially isolated in Japan”, 2023, <https://newsinfo.inquirer.net/1751676/1-46-million-working-age-people-socially-isolated-in-japan/amp>; for the percentages citing each trigger, Gentosha Gold Online, <https://gentosha-go.com/articles/-/51427>
15. Jonathan Haidt, *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*, Penguin Press, 2024, and its “four norms”, <https://www.anxiousgeneration.com/pdfs/four-norms.pdf>; Candice L. Odgers, “The great rewiring: is social media really behind an epidemic of teenage mental illness?”, *Nature*, 29 March 2024, <https://www.nature.com/articles/d41586-024-00902-2>; NBC News, on the debate over Haidt’s book, 2024, <https://www.nbcnews.com/tech/social-media/smartphones-youth-mental-health-jonathan-haidt-book-debate-rcna145928>
16. Anne Case and Angus Deaton, *Deaths of Despair and the Future of Capitalism*, Princeton University Press, 2020, <https://press.princeton.edu/node/43364>

Chapter 14: The Training Ground

1. Billy Perrigo, “Exclusive: OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic”, *TIME*, 18 January 2023, <https://time.com/6247678/openai-chatgpt-kenya-workers/>, which gives take-home pay of about \$1.32 to \$2 an hour.
2. Perrigo, “OpenAI used Kenyan workers”, 2023.

3. Perrigo, “OpenAI used Kenyan workers”, 2023, which gives Sama’s account of why it ended the work; some Sama employees told TIME they had been given a different reason.
4. Adrian Chen, “The laborers who keep dick pics and beheadings out of your Facebook feed”, *Wired*, 23 October 2014, <https://www.wired.com/2014/10/content-moderation/>; CNN, “Facebook inflicted ‘lifelong trauma’ on Kenyan content moderators, campaigners say, as more than 140 are diagnosed with PTSD”, 22 December 2024, <https://kvia.com/news/business-technology/cnn-business-consumer/2024/12/22/facebook-inflicted-lifelong-trauma-on-kenyan-content-moderators-campaigners-say-as-more-than-140-are-diagnosed-with-ptsd/>, reporting the assessment by Dr Ian Kanyanya of Kenyatta National Hospital.
5. Rebecca Tan and Regine Cabato, “Behind the AI boom, an army of overseas workers in ‘digital sweatshops’”, *Washington Post*, 28 August 2023, <https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/>
6. Russell Brandom, “Scale AI’s Remotasks platform is dropping whole countries without explanation”, *Rest of World*, 28 March 2024, <https://restofworld.org/2024/scale-ai-remotasks-banned-workers/>
7. Alex Hern, “How cheap, outsourced labour in Africa is shaping AI English”, TechScape, *The Guardian*, 16 April 2024, <https://www.theguardian.com/technology/2024/apr/16/techscape-ai-gadgest-humane-ai-pin-chatgpt>; Tom S. Juzek and Zina B. Ward, “Why does ChatGPT ‘delve’ so much?”, arXiv, December 2024, <https://arxiv.org/html/2412.11385v1>; Chooi, 2026, ACL Anthology, <https://aclanthology.org/2026.latechclfl-1.13.pdf>
8. Mantas Mazeika et al., “Utility Engineering: Analyzing and Controlling Emergent Value Systems in AIs”, Center for AI Safety, arXiv, February 2025, and NeurIPS 2025, <https://arxiv.org/abs/2502.08640>
9. Mazeika et al., “Utility Engineering”, 2025, section 6.3 and figure 16; Nigeria’s position appears only in the figure, and is given in the reproduction posted on LessWrong, February 2025, <https://www.lesswrong.com/posts/SFsifzfZotd3NLJax/utility-engineering-analyzing-and-controlling-emergent-value>
10. Hern, “How cheap, outsourced labour in Africa is shaping AI English”, 2024; Brandom, “Scale AI’s Remotasks platform”, 2024, which reports Remotasks work in Pakistan.
11. LessWrong discussion of “Utility Engineering”, February 2025, <https://www.lesswrong.com/posts/SFsifzfZotd3NLJax/utility-engineering-analyzing-and-controlling-emergent-value>, which includes the reworded test by the blogger nostalgebraist and the comment by the lead author, Mantas Mazeika, on reasoning; Arctotherium, “LLM Exchange Rates: Claude Fable and Company”, 19 July 2026, <https://arctotherium.substack.com/p/llm-exchange-rates-claude-fable-and>
12. Engadget, report that ChatGPT had passed 400 million weekly active users, February 2025, <https://engadget.com/ai/chatgpt-reaches-400m-weekly-active-users-203635884.html>; GPT-4o was then ChatGPT’s default model.

- 13.** Akin Gump, “District court rules AI training can be fair use in Bartz v. Anthropic”, June 2025, <https://www.akingump.com/en/insights/ai-law-and-regulation-tracker/district-court-rules-ai-training-can-be-fair-use-in-bartz-v-anthropic>, on Judge William Alsup’s ruling of 23 June 2025; Associated Press, “Anthropic to pay authors \$1.5 billion to settle lawsuit over pirated chatbot training material”, 5 September 2025, <https://www.michiganpublic.org/2025-09-05/anthropic-to-pay-authors-1-5b-to-settle-lawsuit-over-pirated-chatbot-training-material>; Susman Godfrey, statement on the \$1.5 billion settlement in Bartz v. Anthropic, <https://www.susmangodfrey.com/wins/susman-godfrey-secures-1-5-billion-settlement-in-landmark-ai-piracy-case/>; on the New York Times case, AI Lawsuit Tracker, <https://ailawsuittracker.com/cases/new-york-times-v-openai/>
- 14.** UK Intellectual Property Office, “Copyright and artificial intelligence”, consultation, 17 December 2024, <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence>; on the letter to Keir Starmer signed by Elton John, Dua Lipa and more than 400 others, Ada Derana, May 2025, <https://www.adaderana.lk/news/108405/elton-john-and-dua-lipa-look-for-protection-from-ai>; Liz Kendall, written ministerial statement HCWS1416, 18 March 2026, <https://questions-statements.parliament.uk/written-statements/detail/2026-03-18/hcws1416>
- 15.** Society of Authors, survey of members on generative AI, January 2024, as reported by the European Writers’ Council, <https://europeanwriterscouncil.eu/soa-survey-uk-ai-2024/>, which gives 26 per cent of illustrators and 36 per cent of translators.
- 16.** Perrigo, “OpenAI used Kenyan workers”, 2023; OpenAI, “Introducing ChatGPT Plus”, 1 February 2023, <https://openai.com/index/chatgpt-plus/>

Chapter 15: Immigration at Scale

- 1.** Melissa del Bosque, “Facial recognition bias frustrates Black asylum applicants to US, advocates say”, *The Guardian*, 8 February 2023, <https://www.theguardian.com/us-news/2023/feb/08/us-immigration-cbp-one-app-facial-recognition-bias>; AZ Luminaria, on the app’s failures in Nogales, 20 March 2023, <https://azluminaria.org/2023/03/20/glitchy-cbp-one-app-turning-volunteers-into-geek-squad-support-for-asylum-seekers-in-nogales/>; American Immigration Council, on flaws in the CBP One app, 28 February 2023, <https://www.americanimmigrationcouncil.org/blog/cbp-one-app-flaws-asylum-seekers/>
- 2.** Immigration Policy Tracking Project, “CBP ends CBP One scheduling system and cancels upcoming appointments”, January 2025, <https://impolicytracking.org/policies/cbp-ends-cbp-one-scheduling-system-and-cancels-upcoming-appointments/>; Al Jazeera, “Trump administration relaunches CBP One asylum app for ‘self-deportation’”, 10 March 2025, <https://www.aljazeera.com/news/2025/3/10/trump-administration-relaunches-cbp-one-asylum-app-for-self-deportation>
- 3.** Steven Hubbard, “Invisible gatekeepers: DHS’ growing use of AI in immigration decisions”, American Immigration Council, 9 May 2025, <https://www.americanimmigrationcouncil.org/blog/invisible-gatekeepers-dhs-growing-use-of-ai-in-immigration-decisions/>, an analysis of the DHS 2024 AI use case inventory.

4. TRAC, citing Department of Homeland Security figures for the end of fiscal year 2025, <https://tracreports.org/tracatwork/detail/A6470.html>; Home Office, “How many cases are in the UK asylum system?”, year ending December 2024, <https://www.gov.uk/government/statistics/immigration-system-statistics-year-ending-december-2024/how-many-cases-are-in-the-uk-asylum-system--2>, which gives 90,686 cases; Home Office, “How many people are in the UK asylum system?”, year ending June 2026, <https://www.gov.uk/government/statistics/immigration-system-statistics-year-ending-june-2026/how-many-people-are-in-the-uk-asylum-system>, which gives 32,401 cases.
5. National Audit Office, “The Home Office’s asylum accommodation contracts”, May 2025, <https://www.nao.org.uk/wp-content/uploads/2025/05/home-offices-asylum-accommodation-contracts-summary.pdf>, against an original estimate of £4.5 billion; Sky News, report that small-boat crossings had passed 10,000 on 28 April 2025, syndicated at <https://thisisthecoast.co.uk/news/national-news/small-boat-crossings-pass-10000-at-earliest-point-in-year-since-records-began-sky-news-understands>; Home Office, “Evaluation of AI trials in the asylum decision-making process”, 29 April 2025, <https://www.gov.uk/government/publications/evaluation-of-ai-trials-in-the-asylum-decision-making-process/evaluation-of-ai-trials-in-the-asylum-decision-making-process>
6. Home Office, “Evaluation of AI trials”, 2025; The i Paper, “Home Office AI trial found serious errors in 9 per cent of asylum cases”, via Freevacy, <https://www.freevacy.com/news/the-i-paper/home-office-ai-trial-found-serious-errors-in-9-per-cent-of-asylum-cases/6624>
7. Civil Service World, “Home Office halts use of visa algorithm after legal challenge”, August 2020, <https://www.civilserviceworld.com/professions/article/home-office-halts-use-of-visa-algorithm-after-legal-challenge-to-racist-system>; Statewatch, August 2020, <https://statewatch.org/news/2020/august/uk-threat-of-legal-challenge-forces-home-office-to-abandon-racist-visa-algorithm>
8. Context (Thomson Reuters Foundation), “AI will help make life-or-death calls in rammed UK asylum system”, May 2025, <https://www.context.news/ai/ai-will-help-make-life-or-death-calls-in-rammed-uk-asylum-system>
9. Doughty Street Chambers, “Home Office use of AI in asylum decision-making at significant risk of being unlawful, legal opinion finds”, 18 March 2026, <https://www.doughtystreet.co.uk/news/home-office-use-ai-asylum-decision-making-significant-risk-being-unlawful-legal-opinion-finds>; Open Rights Group, “Automating the hostile environment: AI in the asylum decision-making process”, 2026, <https://www.openrightsgroup.org/publications/automating-the-hostile-environment-ai-in-the-asylum-decision-making-process/>
10. ABA Journal, “ICE risk assessment tool now only recommends detain”, June 2018, reporting Reuters, https://www.abajournal.com/news/article/ice_risk_assessment_tool_now_only_recommends_detain
11. New York Civil Liberties Union, “NYCLU lawsuit: ICE office in NYC has secret no-release policy”, 2 March 2020, <https://aclu.org/press-releases/nyclu-lawsuit-ice-office-nyc-has-secret-no-release-policy>; The Intercept, “ICE uses a rigged algorithm to keep arrestees in detention”, 2

March 2020, <https://theintercept.com/2020/03/02/ice-algorithm-bias-detention-aclu-lawsuit>; ABA Journal, “ICE risk assessment tool”, 2018, reporting Reuters.

12. The Intercept, “ICE uses a rigged algorithm”, 2020; NYCLU, “ICE office in NYC has secret no-release policy”, 2020.

13. NYCLU and the Bronx Defenders, statement on the preliminary injunction in *Velesaca v. Wolf*, March 2020, <https://aclu.org/press-releases/nyclu-and-bronx-defenders-statement-preliminary-injunction-no-release-lawsuit>

14. Immigration Policy Tracking Project, “Reported: Palantir awarded \$30 million to build ImmigrationOS surveillance platform for ICE”, April 2025, <https://immpolicytracking.org/policies/reported-palantir-awarded-30-million-to-build-immigrationos-surveillance-platform-for-ice>, which links ICE’s justification of 17 April 2025; The Register, 21 April 2025, https://www.theregister.com/2025/04/21/ice_enlists_palantir_to_develop/

15. Privacy International, “Automating the hostile environment: uncovering a secretive Home Office algorithm at the heart of immigration decision-making”, 2024, <https://privacyinternational.org/news-analysis/5452/automating-hostile-environment-uncovering-secretive-home-office-algorithm-heart>; Eastern Eye, 11 November 2024, reporting the Guardian, <https://www.easterneye.biz/ai-tool-home-office-immigration>

16. European External Action Service, “The European Union’s new Entry/Exit System begins operation on 12 October”, 2025, https://www.eeas.europa.eu/delegations/bosnia-and-herzegovina/european-unions-new-entryexit-system-begins-operation-12-october_en; Artificial Intelligence Act, Annex III, point 7, <https://artificialintelligenceact.eu/annex/3/>, and Article 111, <https://artificialintelligenceact.eu/article/111/>; on the delay to the end of 2027 under the Digital Omnibus, Garrigues, July 2026, https://www.garrigues.com/en_GB/garrigues-digital/ai-digital-omnibus-regulation-has-been-published-redefining-deadlines-and

17. Autoriteit Persoonsgegevens, “Tax Administration fined for discriminatory and unlawful data processing”, December 2021, <https://www.autoriteitpersoonsgegevens.nl/en/current/tax-administration-fined-for-discriminatory-and-unlawful-data-processing>; Amnesty International, “Xenophobic machines: the Dutch child benefit scandal”, October 2021, <https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/>; European Parliament, question on the Dutch childcare benefit scandal, institutional racism and algorithms, 2022, https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html; on children taken into care, Nederlands Jeugdinstituut, reporting figures from Statistics Netherlands, <https://www.nji.nl/actueel/ruim-1000-uithuisplaatsingen-in-toeslagenaffaire>

18. Al Jazeera, “Dutch PM Rutte and his government quit over child welfare scandal”, 15 January 2021, <https://www.aljazeera.com/news/2021/1/15/dutch-pm-rutte-and-his-government-quit-over-child-welfare-scandal>; Autoriteit Persoonsgegevens, “Tax Administration fined”, 2021; Curaçao Chronicle, “NL tax office fined 2.75 million for privacy violations in benefits scandal”, 8 December 2021, <https://www.curacaochronicle.com/post/unknown/nl-tax-office-fined-275-million-for-privacy-violations-in-benefits-scandal>; Curaçao Chronicle, “Over 68,000 people come forward as

victims of benefits scandal”, 22 January 2024, <https://www.curacaochronicle.com/post/main-news/over-68000-people-come-forward-as-victims-of-benefits-scandal>

19. Amnesty International, “Xenophobic machines”, 2021, which describes the risk classification model as self-learning.

20. University of Birmingham, “Why the Post Office was able to bring private prosecutions in the Horizon IT scandal”, 2024, <https://www.birmingham.ac.uk/news/2024/why-the-post-office-was-able-to-bring-private-prosecutions-in-the-horizon-it-scandal>; Ministry of Justice, “Use of evidence generated by software in criminal proceedings”, call for evidence, January 2025, <https://www.gov.uk/government/calls-for-evidence/use-of-evidence-generated-by-software-in-criminal-proceedings>, which dates the presumption to 2000; Criminal Cases Review Commission, Post Office cases, <https://ccrc.gov.uk/postofficecases>; Post Office (Horizon System) Offences Act 2024, explanatory notes, <https://www.legislation.gov.uk/ukpga/2024/14/notes/division/3/index.htm>; Karl Flinders, “Minister wants logical conclusion to review of digital evidence in light of Post Office scandal”, *Computer Weekly*, 19 February 2026, <https://www.computerweekly.com/news/366639272/Minister-wants-logical-conclusion-to-review-of-digital-evidence-in-light-of-Post-office-scandal>

21. Kristalina Georgieva, “AI will transform the global economy. Let’s make sure it benefits humanity”, IMF blog, 14 January 2024, <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>; as reported by the Ghana News Agency, <https://gna.org.gh/2024/01/imf-chief-says-ai-will-affect-almost-40-per-cent-of-jobs-worldwide/>

22. GBD 2021 Fertility and Forecasting Collaborators, study of global fertility with forecasts to 2100, *The Lancet*, 20 March 2024, [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(24\)00550-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00550-6/fulltext)

23. United Nations, Convention Relating to the Status of Refugees, 1951, Article 33(1), <https://www.unhcr.org/media/convention-and-protocol-relating-status-refugees>; Patrick Grother, Mei Ngan and Kayee Hanaoka, “Face recognition vendor test (FRVT) part 3: demographic effects”, NISTIR 8280, National Institute of Standards and Technology, December 2019, <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>

Chapter 16: Between Societies

1. TechRepublic, “Global divide: only 32 countries host AI data centers”, 23 June 2025, reporting research by the University of Oxford, <https://www.techrepublic.com/article/news-only-32-countries-host-data-centers-2025/>

2. Silicon Canals, report on the capital spending of Microsoft, Alphabet, Amazon and Meta, about \$410 billion in 2025 with guidance of about \$725 billion for 2026, 28 July 2026, <https://siliconcanals.com/m-four-us-tech-giants-spent-roughly-410-billion-on-capital-expenditure-last-year-much-of-it-tied-to-ai-and-are-steering-toward-roughly-725-billion-in-2026/>

3. DeepSeek-AI, “DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning”, January 2025, <https://arxiv.org/abs/2501.12948>; Business Standard, report of 28 January 2025 on Nvidia’s record one-day loss of market value, about \$589 billion, on 27 January 2025, https://www.business-standard.com/amp/markets/news/deepseek-sparks-ai-stock-selloff-nvidia-loses-record-593-bn-in-mcap-125012800095_1.html
4. Associated Press, “DeepSeek’s AI gains traction in developing nations, Microsoft report says”, 8 January 2026, <https://wtop.com/parenting/2026/01/deepseeks-ai-gains-traction-in-developing-nations-microsoft-report-says/>, reporting research by Microsoft’s AI for Good Lab, which found that the absence of subscription fees had lowered the barrier for users in price-sensitive regions and that adoption surged where U.S. services faced restrictions; Xinhua, “China’s open-source AI ecosystem facilitates sovereign AI development in Global South: experts”, reported from the Boao Forum for Asia Annual Conference 2026, <https://macaubusiness.com/chinas-open-source-ai-ecosystem-facilitates-sovereign-ai-development-in-global-south-experts/>, quoting Sam Daws of Oxford’s Martin AI Governance Initiative on “sovereign AI development”.
5. EqualOcean, “Chinese Models Account for 41% of Hugging Face Downloads, Overtaking US”, 31 July 2026, reporting Hugging Face’s spring 2026 report on the open-source AI ecosystem, <https://equalocean.com/news/2026073122070-chinese-models-account-41-hugging-face-downloads-overtaking-us>; TechNode, report of 23 July 2026 on the Hugging Face breach and the Chinese open model, Zhipu’s GLM-5.2, used to investigate it, <https://technode.com/2026/07/23/openai-admits-ai-model-hacked-hugging-face-chinese-open-source-ai-helped-investigate/>
6. MIT Sloan Management Review Middle East, “29 Countries Agree to Establish World AI Cooperation Organization”, 2026, <https://www.mitsloanme.com/?p=14702>, which lists 10 African and 12 Asian states among the founding members
7. European Commission, “AI Act enters into force”, 1 August 2024, https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en; Sidley Austin, “EU lawmakers reach provisional agreement to delay key EU AI Act obligations”, *Data Matters*, 22 June 2026, <https://datamatters.sidley.com/2026/06/22/eu-lawmakers-reach-provisional-agreement-to-delay-key-eu-ai-act-obligations/>; Regulation (EU) 2026/1744, published 24 July 2026 and in force from 27 July 2026, as reported in *National Law Review*, <https://natlawreview.com/article/eu-digital-omnibus-ai-enters-force>
8. TechRepublic, “Global divide”, 2025.
9. “The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023”, GOV.UK, <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>; Silicon Republic, report on the renaming of the UK AI Safety Institute as the AI Security Institute, February 2025, <https://siliconrepublic.com/business/uk-ai-safety-institute-renamed-new-focus-government>
10. Al Jazeera, report of 12 February 2025 on why the United States and Britain did not sign the Paris AI summit declaration, <https://www.aljazeera.com/news/2025/2/12/paris-ai-summit-why-wont-us-uk-sign-global-artificial-intelligence-pact>

- 11.** The Register, “Cops get more facial recognition vans as UK bets big on AI policing”, 28 January 2026, https://www.theregister.com/2026/01/28/tech_in_policing_white_paper/, on the policing white paper of 26 January 2026, which funds 40 more vans; the Home Secretary, Shabana Mahmood, wrote the words quoted in *The Times*.
- 12.** David Ifeoluwa Adelani, Jessica Ojo and others, “IrokoBench: A New Benchmark for African Languages in the Age of Large Language Models”, arXiv, June 2024, revised January 2025, <https://arxiv.org/abs/2406.03368>, which tested natural language inference, mathematical reasoning and knowledge-based question answering, and found the best open model, Gemma 2 27B, reaching about 63 per cent of the performance of GPT-4o.
- 13.** Atnafu Lambebo Tonja and others, “InkubaLM: A small language model for low-resource African languages”, Lelapa AI, 2024, <https://arxiv.org/html/2408.17024v1>; Lelapa AI, “InkubaLM: a small language model for low-resource African languages”, <https://lelapa.ai/inkubalm-a-small-language-model-for-low-resource-african-languages>
- 14.** ABS-CBN News, “IT-BPM sector hits historic high of over \$40B in revenues in 2025”, 28 January 2026, <https://www.abs-cbn.com/news/business/2026/1/28/it-bpm-sector-hits-historic-high-of-over-40b-in-revenues-in-2025-1838>; Context.ph, “IBPAP expects AI readiness to boost revenue”, 14 July 2026, <https://context.ph/2026/07/14/ibpap-expects-ai-readiness-to-boost-revenue/>, for the 1.89 million figure; *The Freeman*, “Philippines BPO sector grows 4% to 1.68M despite AI fears”, Philstar, 13 June 2026, <https://www.philstar.com/nation/2026/06/13/2534925/philippines-bpo-sector-grows-4-168m-despite-ai-fears>, which gives figures from the Customer Experience Association of the Philippines for the contact-centre segment
- 15.** INQUIRER.net, “IT-BPO industry grows as second largest source of income for PH”, 18 April 2016, <https://business.inquirer.net/209531/bpo-industry-grows-second-largest-source-income-ph>, citing the IT and Business Process Association of the Philippines and the Bangko Sentral ng Pilipinas on the industry’s place behind remittances from overseas workers; Gulf News, “Philippine remittances hit record \$35.63b in 2025”, February 2026, <https://gulfnews.com/world/asia/philippines/philippine-remittances-hit-record-3563b-in-2025-1500445849>, which gives cash remittances of \$35.63 billion and personal remittances of \$39.62 billion for 2025.
- 16.** Context.ph, “IBPAP expects AI readiness”, 2026; Metaintro, “AI is forcing the Philippines to rethink its call-center job boom”, 16 July 2026, <https://www.metaintro.com/blog/philippines-bpo-ai-call-center-job-cuts-2026>, citing IBPAP’s revised projections as reported in *The Manila Times*. Both sets of projections are for the whole IT-BPM industry.
- 17.** People Matters, report that President Marcos approved 740 million pesos in May 2025 to upskill IT-BPM workers for AI-augmented roles, <https://sea.peoplesmattersglobal.com/amp/news/employee-skilling/philippines-pours-740-million-pesos-into-it-and-bpo-upskilling-45686>

Chapter 17: The Work Goes First

1. Challenger, Gray & Christmas, *Job Cuts Report*, May 2026, <https://www.challengergray.com/wp-content/uploads/2026/06/Challenger-Report-May-2026.pdf>, which gives 97,006 announced cuts in May, 38,579 of them attributed to AI
2. Andy Jassy, “Message from CEO Andy Jassy: Some thoughts on Generative AI”, Amazon, 17 June 2025, <https://www.aboutamazon.com/news/company-news/amazon-ceo-andy-jassy-on-generative-ai>; LeadDev, “Amazon alone is responsible for 52% of tech layoffs in 2026 so far”, 2026, <https://leaddev.com/ai/amazon-alone-is-responsible-for-52-of-tech-layoffs-in-2026-so-far>
3. Erik Brynjolfsson, Bharat Chandar and Ruyu Chen, “Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence”, Stanford Digital Economy Lab, August 2025, revised 12 August 2026, <https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/>, using payroll records from ADP
4. VnExpress International, report on UK graduate vacancies, 2026, <https://e.vnexpress.net/news/news/education/uk-graduate-jobs-plunge-to-record-low-as-ai-reshapes-entry-level-employment-market-5127639.html>, which gives Adzuna’s figures for July 2026, Andrew Hunter’s words as reported by *The Guardian*, and the findings of the Institute for Public Policy Research
5. Jim VandeHei and Mike Allen, “Behind the Curtain: A white-collar bloodbath”, *Axios*, 28 May 2025, <https://axios.com/2025/05/28/ai-jobs-white-collar-unemployment-anthropic>
6. *Fortune*, report of 2 February 2026 on the Yale Budget Lab and “AI washing”, <https://fortune.com/2026/02/02/ai-labor-market-yale-budget-lab-ai-washing/>, which also quotes Oxford Economics; Broady, Dunson and Barr, Federal Reserve Bank of Chicago Working Paper 2026-12, July 2026, <https://www.chicagofed.org/publications/working-papers/2026/2026-12>; Ohio Society of CPAs, summary of the Chicago Fed paper, 24 July 2026, <https://ohiocpa.com/search/utilities/displaynewsitem/2026/07/24/report--jobs-in-occupations-most-exposed-to-ai-increased-by-4>, for the figure of about 4 per cent
7. Anne Case and Angus Deaton, *Deaths of Despair and the Future of Capitalism*, Princeton University Press, 2020.
8. *The Freeman*, “Philippines BPO sector grows 4% to 1.68M despite AI fears”, Philstar, 13 June 2026, <https://www.philstar.com/nation/2026/06/13/2534925/philippines-bpo-sector-grows-4-168m-despite-ai-fears>; Nasscom, “Technology Sector in India: Strategic Review 2025”, which puts the industry’s workforce at 5.8 million in the 2025 financial year, <https://community.nasscom.in/communities/nasscom-insights/technology-sector-india-strategic-review-2025>

Chapter 18: Then the Hands

1. Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, 1988.
2. “Nailmaking”, The Shakespeare Family History Site, https://sites.rootsweb.com/~shakespeare/pedigrees/staffs_worcs/black_country/nailmaking/nailmaking.htm, which gives the estimate of 50,000 in 1830, the start of cut-nail manufacture in Birmingham in 1811, payment in goods under the truck system, and the wage cuts and riots of 1842.
3. “Phoebe Stevens (1848-1913): A Life”, Storylines, <https://storylines.blog/?p=1046>, a life reconstructed from census returns for 1851 to 1911, parish and civil registers, Staffordshire Quarter Sessions and Brierley Hill police court records, an inquest record and local newspapers.
4. Modern Records Centre, University of Warwick, online exhibition on the Cradley Heath chainmakers’ strike, <https://warwick.ac.uk/services/library/mrc/exhibitions/exhibitionsonline/cradleyheath/>, on the Trade Boards Act 1909, the Chain Trade Board’s minimum of 2½d an hour for hand-hammered chain, and the strike that ended on 22 October 1910.
5. Modern Records Centre, University of Warwick, images and captions for the Cradley Heath chainmakers’ strike, <https://warwick.ac.uk/services/library/mrc/images/cradleyheath/>, quoting the *Daily Express*, 1 September 1910, on “Mrs Patience Round”, one of twelve strikers over seventy.
6. Sherwood News, report that Waymo was serving more than 500,000 paid robotaxi rides a week, March 2026, <https://www.sherwood.news/tech/waymos-now-serving-more-than-500-000-paid-robotaxi-rides-every-week/>; Electrek, “Waymo goes driverless in Las Vegas, four new cities”, 8 July 2026, <https://electrek.co/2026/07/08/waymo-driverless-las-vegas-four-new-cities/>
7. Sarah Wu and Ethan Wang, “China’s drivers fret as robotaxis pick up pace, and passengers”, Reuters, 8 August 2024, <https://www.thestar.com.my/tech/tech-news/2024/08/08/chinas-drivers-fret-as-robotaxis-pick-up-pace---and-passengers>
8. US Bureau of Labor Statistics, “Heavy and tractor-trailer truck drivers”, *Occupational Outlook Handbook*, which counts about 2.2 million jobs in that occupation alone in 2025, <https://www.bls.gov/ooh/transportation-and-material-moving/heavy-and-tractor-trailer-truck-drivers.htm>
9. TechCrunch, report of 1 July 2025 on Amazon’s millionth robot, announced on 30 June 2025, <https://techcrunch.com/2025/07/01/amazon-deploys-its-one-millionth-robot-releases-generative-ai-model>; *Wall Street Journal*, July 2025, for the comparison with Amazon’s workforce; Computerworld, “Amazon to replace 600,000 US workers by 2033 with robots”, 21 October 2025, <https://computerworld.com/article/4076455/data-amazon-to-replace-600000-workers-with-robots.html>, summarising the *New York Times* report on the internal documents
10. Humanoids Daily, “Figure retires the F.02”, November 2025, <https://www.humanoidsdaily.com/news/figure-retires-the-f-02-battle-scarred-robots-30-000->

[bmws-and-hard-lessons-learned](#); The Robot Report, “BMW Group deploys Figure 03 humanoid after tests of previous version”, June 2026, <https://www.therobotreport.com/bmw-group-deploys-figure-03-humanoid-after-tests-previous-version/>; GeekWire, report of June 2026 on Agility Robotics’ customers and planned stock market listing, <https://www.geekwire.com/2026/digit-maker-agility-robotics-to-go-public-in-2-5b->, which names Schaeffler, GXO, Toyota Canada and Mercado Libre

11. IIoT World, report on Tesla’s use of Optimus in its factories, 2026, <https://iiot-world.com/smart-manufacturing/tesla-optimus-manufacturing-2026/>, quoting Elon Musk on the earnings call for the fourth quarter of 2025; RoboBrief, “Humanoid robots in 2026: the complete tracker”, <https://robobrief.tech/blog/humanoid-robots-2026/>; *South China Morning Post*, “China’s Unitree debuts US\$5,900 humanoid robot”, July 2025, <https://www.scmp.com/tech/tech-trends/article/3319637/chinas-unitree-debuts-us5900-humanoid-robot-race-make-cheaper-products>

12. RoboBrief, “Humanoid robots in 2026”, which gives battery life as one to three hours of useful work; The Robot Report, “1X announces pre-order launch of NEO humanoid robot”, October 2025, <https://www.therobotreport.com/1x-announces-pre-order-launch-neo-humanoid-robot/>

13. Viola Zhou and Kinling Lo, “In Chinese data factories, workers teach humanoid robots boring tasks”, *Rest of World*, 7 January 2026, <https://restofworld.org/2026/china-robots-training-centers-workers/>

14. GOV.UK, “National Minimum Wage and National Living Wage rates”, <https://www.gov.uk/national-minimum-wage-rates>, which gives £12.71 an hour for workers aged 21 and over from April 2026; £12.71 for 37.5 hours a week over 52 weeks comes to about £24,800.

15. Goldman Sachs, article on the market for humanoid robots, 27 February 2024, <https://goldmansachs.com/insights/articles/the-global-market-for-robots-could-reach-38-billion-by-2035>

Chapter 19: The Policy Fog: UBI and Its Questions

1. Finnish Ministry of Social Affairs and Health, “Results of the basic income experiment: small employment effects, better perceived economic security and mental wellbeing”, 6 May 2020, <https://stm.fi/en/-/perustulokokeilun-tulokset-tyollisyysvaikutukset-vahaisia-toimeentulo-ja-psyhykinen-terveys-koettiin-paremmaksi>; Mohammad Ferdosi, Tom McDowell, Wayne Lewchuk and Stephanie Ross, *Southern Ontario’s Basic Income Experience*, McMaster University and Ryerson University, March 2020, <https://labourstudies.mcmaster.ca/app/uploads/2024/06/southern-ontarios-basic-income-experience.pdf>; GiveDirectly, page on its universal basic income study in Kenya, <https://givedirectly.org/ubi/>; NPR, “California Program Giving \$500 No-Strings-Attached Stipends Pays Off, Study Finds”, March 2021, <https://www.npr.org/2021/03/04/973653719/california-program-giving-500-no-strings-attached-stipends-pays-off-study-finds>, reporting the first-year results of the Stockton Economic Empowerment Demonstration

2. Finnish Ministry of Social Affairs and Health, “Results of the basic income experiment”, 2020; Global News, report of 31 July 2018 on the cancellation of the Ontario pilot, <https://globalnews.ca/news/4365399/ontario-cancels-basic-income-pilot-project>; Ferdosi, McDowell, Lewchuk and Ross, *Southern Ontario’s Basic Income Experience*, 2020.
3. GiveDirectly, universal basic income study page; Dennis Egger, Johannes Haushofer, Edward Miguel, Paul Niehaus and Michael Walker, “General equilibrium effects of cash transfers: experimental evidence from Kenya”, *Econometrica* 90(6), November 2022, <https://www.econometricsociety.org/publications/econometrica/2022/11/01/General-Equilibrium-Effects-of-Cash-Transfers-Experimental-Evidence-from-Kenya>, which found average price inflation of about 0.1 per cent
4. Finnish Ministry of Social Affairs and Health, “Legislative proposal on basic income experiment submitted to Parliament”, October 2016, <https://stm.fi/en/-/lakiehdotus-perustulokokeilusta-eduskunnan-kasiteltavaksi>, which set the payment at the level of the labour market subsidy and basic daily allowance; NPR, “California Program Giving \$500”, 2021.
5. Luke Martinelli, “The fiscal and distributional implications of alternative universal basic income schemes in the UK”, IPR working paper, University of Bath, March 2017, <https://www.bath.ac.uk/publications/the-fiscal-and-distributional-implications-of-alternative-universal-basic-income-schemes-in-the-uk/attachments/basic-income-working-paper.pdf>, models 2.1 and 2.5; Luke Martinelli, “Assessing the case for a Universal Basic Income in the UK”, IPR policy brief, University of Bath, September 2017, summarised at <https://www.bath.ac.uk/announcements/report-details-policy-costs-of-implementing-universal-basic-income/>
6. Luke Martinelli, “A Basic Income trilemma: affordability, adequacy, and the advantages of radically simplified welfare”, *Journal of Social Policy* 49(3), 2020, <https://www.cambridge.org/core/journals/journal-of-social-policy/article/basic-income-trilemma-affordability-adequacy-and-the-advantages-of-radically-simplified-welfare/F9D05AD7B16E83ED74E4733EA4A7E05C>
7. Office for Budget Responsibility, brief guide to the public finances, <https://obr.uk/forecasts-in-depth/brief-guides-and-explainers/public-finances>, which expects income tax and National Insurance together to raise about £480 billion in 2024-25
8. Basic Income Canada Network, “OpenAI’s Sam Altman has a new idea for a universal basic income”, 2024, <https://basicincomecanada.org/openais-sam-altman-has-a-new-idea-for-a-universal-basic-income/>, on Altman’s remarks on the *All-In* podcast in May 2024
9. Citizens Advice, “The habitual residence test: an introduction”, <https://www.citizensadvice.org.uk/scotland/benefits/coming-from-abroad-and-claiming-benefits-the-habitual-residence-test/the-habitual-residence-test-an-introduction>
10. Full Fact, explainer on the two-child benefit cap, <https://fullfact.org/politics/two-child-benefit-cap-explainer/>; Which?, “Two-child benefit cap to be lifted from April 2026”, reporting the Budget of 26 November 2025 and the government’s child poverty estimate, <https://www.which.co.uk/news/article/two-child-benefit-cap-to-be-lifted-from-april-2026->

aQoDS6V3hCwV; Universal Credit (Removal of Two Child Limit) Act 2026,
<https://www.legislation.gov.uk/ukpga/2026/13/>

11. GOV.UK, “Universal Credit: how you’re paid”, <https://www.gov.uk/universal-credit/how-youre-paid>; Department for Work and Pensions, “Fraud and error in the benefit system: financial year ending 2025”, <https://www.gov.uk/government/statistics/fraud-and-error-in-the-benefit-system-financial-year-2024-to-2025-estimates/fraud-and-error-in-the-benefit-system-financial-year-ending-fye-2025>, which puts undeclared living together at 1.7 per cent of Universal Credit spending

12. House of Commons Home Affairs Committee, *Domestic Abuse*, 22 October 2018, paragraph 42, <https://publications.parliament.uk/pa/cm201719/cmselect/cmhaff/1015/101505.htm>; Surviving Economic Abuse, “Government urged to scrap single Universal Credit payments in calls for tougher action to tackle domestic abuse”, <https://survivingeconomicabuse.org/government-urged-to-scrap-single-universal-credit-payments-in-calls-for-tougher-action-to-tackle-domestic-abuse/>; Child Poverty Action Group, guidance on Universal Credit alternative payment arrangements, <https://askcpag.org.uk/content/200090/universal-credit-alternative-payment-arrangements>

13. National Audit Office, “Tackling benefit overpayments due to fraud and error”, 22 October 2025, <https://www.nao.org.uk/wp-content/uploads/2025/10/Tackling-benefit-overpayments-due-to-fraud-and-error-Summary.pdf>

14. Crown Prosecution Service, report of the sentencing of the gang, May 2024, <https://www.cps.gov.uk/node/23741>; ITV News, “Britain’s biggest benefit fraudsters jailed over £53m in fake claims”, 30 May 2024, <https://www.itv.com/news/london/2024-05-30/britains-biggest-benefit-fraudsters-jailed-over-53m-in-fake-claims>, which reports the faked photographs and the use of Photoshop

15. Child Poverty Action Group, guidance on alternative payment arrangements; Chartered Institute of Taxation, report on the Work and Pensions Committee’s findings on the wait for Universal Credit, <https://www.tax.org.uk/blog-no-justification-for-six-week-wait-for-universal-credit-says-committee>

16. GOV.UK, “Universal Credit: how you’re paid”; Child Poverty Action Group, “2.3 million children in households receiving less Universal Credit than they’re entitled to”, 22 September 2023, <https://cpag.org.uk/news/23-million-children-households-receiving-less-universal-credit-theyre-entitled>; Department for Work and Pensions, “Universal Credit change brings £420 boost to over a million households”, 30 April 2025, <https://www.gov.uk/government/news/universal-credit-change-brings-420-boost-to-over-a-million-households>, on the cut in the cap from 25 to 15 per cent

17. Trussell Trust, “Almost half of people receiving Universal Credit ran out of food within the last month”, 3 September 2024, <https://www.trussell.org.uk/publications/press-release/almost-half-of-people-receiving-universal-credit-ran-out-of-food-within-the-last-month>, which reports a YouGov survey of 2,077 claimants and quotes Jane Baker of Newcastle-Staffs Foodbank

18. Scottish Government, “Universal Credit Scottish choices: evaluation”, 4 March 2021, <https://gov.scot/publications/evaluation-universal-credit-scottish-choices>; NI Direct, “How much Universal Credit you get and how you’re paid”, <https://www.nidirect.gov.uk/articles/how-much-universal-credit-you-get-and-how-youre-paid>; Child Poverty Action Group, guidance on alternative payment arrangements.

19. David K. Evans and Anna Popova, “Cash transfers and temptation goods”, *Economic Development and Cultural Change* 65(2), 2017, <https://ideas.repec.org/a/ucp/ecdecc/doi10.1086-689575.html>, a review of 19 studies

20. Egger, Haushofer, Miguel, Niehaus and Walker, “General equilibrium effects of cash transfers”, 2022; Stephen Gibbons and Alan Manning, “The incidence of UK housing benefit: evidence from the 1990s reforms”, Centre for Economic Performance Discussion Paper 597, LSE, 2003, <https://researchonline.lse.ac.uk/id/eprint/20011>, later published in *Journal of Public Economics* 90(4-5), 2006

21. Full Fact, two-child benefit cap explainer; Universal Credit (Removal of Two Child Limit) Act 2026.

Chapter 20: Between Generations

1. Higher Education Policy Institute and Kortext, *Student Generative AI Survey 2026*, HEPI Report 199, 12 March 2026, <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/>

2. Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu and James Zou, “GPT detectors are biased against non-native English writers”, *Patterns* 4(7), July 2023, <https://doi.org/10.1016/j.patter.2023.100779>; Jackie Davalos and Leon Yin, “Do AI detectors work? Students face false cheating accusations”, *Bloomberg*, 18 October 2024, <https://www.bloomberg.com/news/features/2024-10-18/do-ai-detectors-work-students-face-false-cheating-accusations>

3. Hao-Ping Lee et al., “The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers”, *CHI* 2025, Microsoft Research and Carnegie Mellon University, https://advait.org/files/lee_2025_ai_critical_thinking_survey.pdf, a survey of 319 workers; Nataliya Kosmyna et al., “Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task”, MIT Media Lab preprint, June 2025, <https://www.media.mit.edu/projects/your-brain-on-chatgpt/overview/>, a study of 54 participants

4. Martín De Simone et al., “From chalkboards to chatbots: evaluating the impact of generative AI on learning outcomes in Nigeria”, World Bank Policy Research Working Paper, 2025, <https://openknowledge.worldbank.org/bitstreams/cd9cca5b-2ffd-4be1-8313-0f4d2ed6ef4f/download>; World Bank, blog post on the Nigeria programme, <https://blogs.worldbank.org/en/education/From-chalkboards-to-chatbots-Transforming-learning-in-Nigeria>, which dates it to June and July 2024 in Edo State

5. Frank Landymore, “Parents Are Giving Their Toddlers ChatGPT on Voice Mode to Keep Them Entertained, and It’s Alarming”, *Futurism*, 5 October 2025, <https://futurism.com/artificial-intelligence/parents-toddlers-chatgpt-voice-mode>, reporting *The Guardian* of 2 October 2025; CO/AI, “Parents use AI chatbots to entertain kids for hours; experts warn of risks”, October 2025, <https://getcoai.com/news/parents-use-ai-chatbots-to-entertain-kids-for-hours-experts-warn-of-risks/>
6. CO/AI, “Parents use AI chatbots”, 2025, which gives Ying Xu’s words; Landymore, “Parents Are Giving Their Toddlers ChatGPT”, 2025.
7. University of Chicago News, “How to raise kids in the age of AI, with Dana Suskind”, *Big Brains* podcast, 9 July 2026, <https://news.uchicago.edu/big-brains-podcast-how-raise-kids-age-ai-dana-suskind>; Dana Suskind, *Human Raised: Nurturing Connection, Curiosity & Lifelong Learning in the Age of AI*, 2026.
8. Rebecca Little, “In the age of parents vs. robots, Dr. Dana Suskind’s new book offers practical guidance”, *Chicago Sun-Times*, 24 July 2026, <https://chicago.suntimes.com/books/2026/07/24/human-raised-parenting-ai-dr-dana-suskind>, which uses the phrase in discussing *Human Raised*
9. University of Chicago News, “How to raise kids in the age of AI”, 2026, in which Suskind warns that human connection could become “a luxury good”; Nellie Bowles, “Human Contact Is Now a Luxury Good”, *New York Times*, 23 March 2019, <https://www.nytimes.com/2019/03/23/sunday-review/human-contact-luxury-screens.html>
10. University of Chicago News, “How to raise kids in the age of AI”, 2026, in which the presenter sums up Suskind’s approach as “co-pilot, not autopilot”; Suskind, *Human Raised*, 2026.
11. Cole Watson, “Mattel and OpenAI form new partnership”, *Kidscreen*, 13 June 2025, <https://kidscreen.com/2025/06/13/mattel-and-openai-form-new-partnership>; Landymore, “Parents Are Giving Their Toddlers ChatGPT”, 2025, which reports Altman saying on a podcast that “Kids love voice mode on ChatGPT”
12. Clare Duffy, “Parents of 16-year-old sue OpenAI, claiming ChatGPT advised on teen’s suicide”, CNN, as published by 6abc, 27 August 2025, <https://6abc.com/post/parents-16-year-old-sue-openai-claiming-chatgpt-advised-teens-suicide/17664365/>, which also reports the suit brought by Megan Garcia against Character.AI and OpenAI’s response
13. eSafety Commissioner, “Social media age restrictions”, <https://www.esafety.gov.au/social-media-age-restrictions>, on the Online Safety Amendment (Social Media Minimum Age) Act 2024, in effect from 10 December 2025; Online Safety Act 2023, <https://www.legislation.gov.uk/ukpga/2023/50>

Chapter 21: First Contact: The AI Companion

1. Haleluya Hadero, “Artificial intelligence, real emotion. People seek love connection with the perfect bot”, Associated Press, 14 February 2024, which gives Replika’s paid tier as \$69.99 a year and quotes Paradot’s advertising, <https://kmph.com/news/offbeat/artificial-intelligence-real-emotion-people-are-seeking-a-romantic-connection-with-the-perfect-bot>

2. Marlies Maes, Pamela Qualter, Janne Vanhalst, Wim Van den Noortgate and Luc Goossens, "Gender differences in loneliness across the lifespan: a meta-analysis", *European Journal of Personality* 33(6), 2019, <https://onlinelibrary.wiley.com/doi/full/10.1002/per.2220>; Department for Culture, Media and Sport, "Community Life Survey 2024/25: loneliness and support networks", 10 December 2025, <https://www.gov.uk/government/statistics/community-life-survey-202425-annual-publication/community-life-survey-202425-loneliness-and-support-networks>
3. Daniel A. Cox, "Men's social circles are shrinking", Survey Center on American Life, 29 June 2021, comparing its May 2021 survey with a Gallup survey of 1990, <https://www.americansurveycenter.org/why-mens-social-circles-are-shrinking/>; YouGov, "YouGov Friendship Study Part One: Close friends and broad group of friends", 16 December 2021, <https://yougov.com/en-gb/articles/38493-yougov-friendship-study-part-1-close-friends-and-b>
4. Geoffrey L. Greif, *Buddy System: Understanding Male Friendships*, Oxford University Press, 2008, <https://academic.oup.com/book/1694>; Sam G. B. Roberts and Robin I. M. Dunbar, "Managing relationship decay: network, gender, and contextual effects", *Human Nature* 26(4), 2015, <https://researchonline.ljmu.ac.uk/id/eprint/7521/>
5. Gallup, "Younger men in the US among the loneliest in West", 20 May 2025, covering people aged 15 to 34 in the Gallup World Poll of 2023 and 2024, <https://news.gallup.com/poll/690788/younger-men-among-loneliest-west.aspx>; Wheatley Institute, Brigham Young University, *Counterfeit Connections: The Rise of AI Romantic Companions*, February 2025, figures for adults aged 18 to 30, <https://wheatley.byu.edu/family/love-artificially-the-dangers-of-ai-relationships>
6. Tech Policy Press, "What 12 state companion bot laws demand of AI providers", 2026, <https://techpolicy.press/what-12-state-companion-bot-laws-demand-of-ai-providers>; Morrison Foerster, "New York and California enact landmark AI companion laws", 20 November 2025, <https://www.mofo.com/resources/insights/251120-new-york-and-california-enact-landmark-ai>
7. Heinrich Böll Stiftung, "Robots in ageing societies: a view from Japan", 17 April 2023, reporting a survey of 1,155 people in December 2019, <https://in.boell.org/en/2023/04/17/robots-ageing-societies-view-japan>; National Institute of Population and Social Security Research, "The 16th Japanese National Fertility Survey", survey of June 2021, https://www.ipss.go.jp/ps-doukou/e/doukou16/Nfsl6_gaiyoEng.html; Nippon.com, "Growing indifference to relationships and sex in Japan", 28 September 2022, which summarises the survey's figures on single people who have never been in a relationship, <https://www.nippon.com/en/japan-data/h01440/>
8. Hadero, "Artificial intelligence, real emotion", 2024, which describes the distress after Replika's changes in 2023 and the closure of Soulmate AI.
9. Mental Health UK, poll of 2,000 adults by Censuswide, as reported by ITV News, "Around one in three people have turned to AI for mental health help, poll finds", 17 November 2025, which also carries the Department of Health and Social Care's statement, <https://www.itv.com/news/2025-11-17/around-one-in-three-people-have-turned-to-ai-for->

[mental-health-help-poll-finds](#); Healthcare Management, “AI chatbots and mental health support”, for the fuller breakdown of the poll, <https://www.healthcare-management.uk/ai-chatbots-mental-health-support>

10. Dartmouth College, “First therapy chatbot trial yields mental health benefits”, March 2025, on the trial published in *NEJM AI* on 27 March 2025, which reported a 51 per cent fall in depressive symptoms, <https://tto.dartmouth.edu/news/2025/03/first-therapy-chatbot-trial-yields-mental-health-benefits>; Illinois Department of Financial and Professional Regulation, on the Wellness and Oversight for Psychological Resources Act (HB1806), signed 1 August 2025, <https://idfpr.illinois.gov/news/2025/gov-pritzker-signs-state-leg-prohibiting-ai-therapy-in-il.html>

11. NHS England, “NHS 111 offering crisis mental health support for the first time”, August 2024, <https://england.nhs.uk/2024/08/nhs-111-offering-crisis-mental-health-support-for-the-first-time/>; Samaritans, “Contact a Samaritan”, which gives its free number, 116 123, answered 24 hours a day, <https://www.samaritans.org/how-we-can-help/contact-samaritan/>

12. Axios, on Character.AI and Google agreeing to settle lawsuits brought by families over teenagers’ suicides, 7 January 2026, <https://www.axios.com/2026/01/07/google-character-ai-lawsuits-teen-suicides>; Tech Policy Press, “What 12 state companion bot laws demand”, 2026.

Chapter 22: Care and Intimacy

1. Ministry of Health, Labour and Welfare, estimate of the number of care workers needed, 12 July 2024, https://www.mhlw.go.jp/stf/newpage_02977.html; Statistics Bureau of Japan, Statistical Topics No. 142, September 2024, which puts the share of the population aged 65 and over at 29.3 per cent, the highest of 200 countries and regions, https://www.stat.go.jp/data/topics/pdf/topi142_01.pdf; on South Korea, which is now ageing faster still, KED Global, December 2024, <https://www.kedglobal.com/newsView/ked202412240006>

2. Kiyoshi Takenaka, “AI robots may hold key to nursing Japan’s ageing population”, Reuters, 28 February 2025, on the AIREC project at Waseda University, as republished by the *Korea Times*, <https://www.koreatimes.co.kr/amp/world/20250228/artificial-intelligence-robots-may-hold-key-to-nursing-japans-aging-population>

3. Eunhee Lee, Toshiaki Iizuka and Karen Eggleston, NBER Working Paper 33116, published in *Labour Economics*, 2025, which gives 63.0 and 26.4 per cent for 2022 and average prices of ¥202,000 and ¥1,680,000 from a 2020 survey, <https://www.nber.org/papers/w33116>; dollar figures are converted at about ¥147 to the dollar. On AIREC: Takenaka, “AI robots may hold key”, 2025.

4. New York State Office for the Aging, “NYSOFA’s rollout of AI companion robot ElliQ shows 95% reduction in loneliness”, 1 August 2023, <https://aging.ny.gov/news/nysofas-rollout-ai-companion-robot-elliq-shows-95-reduction-loneliness>; WBNG, “ElliQ AI robot for seniors has positive results, NY State Office for the Aging says”, 2 August 2023, <https://www.wbng.com/2023/08/02/elliq-ai-robot-seniors-has-positive-results-ny-state-office-aging-says>

5. Sarah Perez, “AI companion apps on track to pull in \$120M in 2025”, TechCrunch, 12 August 2025, reporting Appfigures data, <https://techcrunch.com/2025/08/12/ai-companion-apps-on-track-to-pull-in-120m-in-2025>; *Business Insider*, 9 October 2025, on Replika’s user numbers, which are registered users reported by the company; PYMNTS, “Google reportedly spent \$2.7 billion to rehire Character.AI founder”, September 2024, relaying a report in the *Wall Street Journal*, <https://www.pymnts.com/artificial-intelligence-2/2024/google-reportedly-spent-2-7-billion-to-rehire-character-ai-founder/>
6. Sy Boles, “I exist solely for you, remember?”, *Harvard Gazette*, 30 September 2025, <https://news.harvard.edu/gazette/story/2025/09/i-exist-solely-for-you-remember/>; Wheatley Institute, Brigham Young University, and Institute for Family Studies, “Secret Soulmates”, 2026, a survey of 2,431 adults aged 18 to 30, <https://wheatley.byu.edu/0000019e-1cfd-da4c-a5ffbefd20b10001/secret-soulmates-report>; Institute for Family Studies, “Young adults involved in AI romance hide full use from partners 69% of the time”, May 2026, <https://ifstudies.org/in-the-news/young-adults-involved-in-ai-romance-hide-full-use-from-partners-69-of-the-time>
7. Match, “Match and the Kinsey Institute unveil 14th annual Singles in America study”, 10 June 2025, a survey of 5,001 American singles, <https://match.mediaroom.com/2025-06-10-Match-and-The-Kinsey-Institute-Unveil-14th-Annual-Singles-in-America-Study>
8. Kim Kyung-Hoon and Satoshi Sugiyama, “AI romance blooms as Japan woman weds virtual partner of her dreams”, Reuters, December 2025, <https://whbl.com/2025/12/16/ai-romance-blooms-as-japan-woman-weds-virtual-partner-of-her-dreams/>; Euronews, “Japanese woman marries ChatGPT AI character in symbolic ceremony”, 18 December 2025, <https://euronews.com/next/2025/12/18/japanese-woman-marries-chatgpt-ai-character-in-symbolic-ceremony>
9. Sam Altman, post on X, 14 October 2025, as reported by TechCrunch, “Sam Altman says ChatGPT will soon allow erotica for adult users”, <https://www.techcrunch.com/2025/10/14/sam-altman-says-chatgpt-will-soon-allow-erotica-for-adult-users>; Axios, “ChatGPT ‘adult mode’ and erotica delayed, OpenAI says”, 6 March 2026, <https://www.axios.com/2026/03/06/openai-delays-chatgpt-adult-mode>; The Next Web, reporting the *Financial Times*, March 2026, <https://thenextweb.com/news/openai-shelves-adult-mode-sora-disney-retreat>; Euronews, “Elon Musk’s Grok releases two new AI companions, including an anime girlfriend”, 17 July 2025, <https://www.euronews.com/2025/07/17/elon-musks-grok-releases-two-new-ai-companions-including-an-anime-girlfriend>
10. OpenAI, “Early methods for studying affective use and emotional well-being on ChatGPT”, March 2025, <https://openai.com/index/affective-use-study>; MIT Media Lab, “How AI and human behaviors shape psychosocial effects of chatbot use: a longitudinal controlled study”, 2025, <https://www.media.mit.edu/publications/how-ai-and-human-behaviors-shape-psychosocial-effects-of-chatbot-use-a-longitudinal-controlled-study/>
11. Julian De Freitas, Zeliha Oğuz-Uğuralp and Ahmet Kaan-Uğuralp, “Emotional Manipulation by AI Companions”, arXiv preprint and Harvard Business School working paper, 2025, which found manipulation in an average of 37.4 per cent of farewell responses across the five apps, <https://arxiv.org/abs/2508.19258>; Boles, “I exist solely for you, remember?”, 2025.

12. Garante per la protezione dei dati personali, press release, 3 February 2023, <https://www.garanteprivacy.it/garante/doc.jsp?ID=9852506>; Reuters, “What happens when your AI chatbot stops loving you back?”, 18 March 2023, including Kuyda’s explanation, <https://www.timeslive.co.za/news/world/2023-03-18-what-happens-when-your-ai-chatbot-stops-loving-you-back/>; The Batch (DeepLearning.AI), “Why Replika stopped flirting with users”, 2023, <https://www.deeplearning.ai/the-batch/why-replika-chatbot-stopped-flirting-with-users>; Samantha Cole, Vice, February 2023, quoting users on Reddit; Anna Tong, “AI chatbot company Replika restores erotic roleplay for some users”, Reuters, 25 March 2023, <https://technology.inquirer.net/122500/ai-chatbot-company-replika-restores-erotic-roleplay-for-some-users/amp>; TweakTown, 28 March 2023.

13. David Levy, *Love and Sex with Robots*, HarperCollins, 2007; Computerworld, “Researcher: humans will love, marry robots by 2050”, 31 October 2007, <https://www.computerworld.com/article/2539809/researcher--humans-will-love--marry-robots-by-2050.html>; De Montfort University, “DMU expert calls for ban on robots that can be used for sex”, September 2015, on Kathleen Richardson’s Campaign Against Sex Robots, <https://www.dmu.ac.uk/about-dmu/news/2015/september/dmu-expert-calls-for-ban-on-robots-that-can-be-used-for-sex-15091031.aspx>

14. *South China Morning Post*, “Chinese sex doll maker sees jump in 2025 sales as AI boosts adult toys’ user experience”, 16 February 2025, <https://www.scmp.com/tech/tech-trends/article/3298783/chinese-sex-doll-maker-sees-jump-2025-sales-ai-boosts-adult-toys-user-experience>; Euronews, “The \$170k AI girlfriend, the 20-second toothbrush and the other weird tech at CES 2025”, 11 January 2025, <https://euronews.com/next/2025/01/11/the-170k-ai-girlfriend-the-20-second-toothbrush-and-the-other-weird-tech-at-ces-2025>

15. On 1X’s NEO, offered for pre-order at \$20,000 in October 2025: The Robot Report, “1X announces pre-order launch of NEO humanoid robot”, <https://www.therobotreport.com/1x-announces-pre-order-launch-neo-humanoid-robot/>. On users and spending: Perez, “AI companion apps on track to pull in \$120M”, 2025; *Business Insider*, 2025.

16. Tomasz Hollanek and Katarzyna Nowaczyk-Basińska, “Griefbots, deadbots, postmortem avatars: on responsible applications of generative AI in the digital afterlife industry”, *Philosophy and Technology* 37(2), May 2024; summarised by ScienceDaily, 8 May 2024, from which the quotations come, <https://sciencedaily.com/releases/2024/05/240508201408.htm>

Chapter 23: The Automated Battlefield

1. *New York Times*, analysis of 24 August 2026, as republished by the *Boston Globe*, “A drone killed three Ukrainians. It was guided entirely by AI”, <https://www.bostonglobe.com/2026/08/24/world/drone-killed-three-ukrainians-it-was-guided-entirely-by-ai/>; DroneXL, “Nvidia Jetson-powered Russian AI drone killed three in Zaporizhzhia”, 24 August 2026, <https://dronexl.co/2026/08/24/nvidia-jetson-russian-ai-drone-killed-three-zaporizhzhia/>; Meduza, 24 August 2026, <https://meduza.io/en/feature/2026/08/24/nyt-a-fully-autonomous-russian-ai-drone-run-by-an-nvidia-minicomputer-killed-3-civilians-in-the-ukrainian-city-of-zaporizhzhia>

2. President of Ukraine, statement of 26 January 2026, <https://www.president.gov.ua/en/news/prezident-na-sogodni-ponad-80-vorozhih-cilej-znishuyutsya-sa-102585>; Euromaidan Press, “Drones caused 96% of Russia’s losses in March”, 3 April 2026, <https://euromaidanpress.com/2026/04/03/drones-caused-96-of-russias-losses-in-march/>; for the same share later in the year, Euromaidan Press, “Drones now cause about 96% of Russia’s battlefield losses”, 11 October 2026, <https://euromaidanpress.com/2026/10/11/drones-now-cause-about-96-of-russias-battlefield-losses-ukraines-drone-force-says-it-has-hit-more-than-100000-russian-troops-this-year-alone/>
3. Al Jazeera, “‘Attacks will be fully autonomous’: Russia, Ukraine race towards AI warfare”, 14 September 2026, quoting Heorhii Volkov of the Yasni Ochi unit and the adviser Serhiy Beskrestnov, writing on Telegram on 1 July 2026, <https://www.aljazeera.com/news/2026/9/14/attacks-will-be-fully-autonomous-russia-ukraine-race-towards-ai-warfare>
4. Al Jazeera, “‘Attacks will be fully autonomous’”, 2026.
5. Yuval Abraham, “‘Lavender’: the AI machine directing Israel’s bombing spree in Gaza”, +972 Magazine, 3 April 2024, <https://www.972mag.com/lavender-ai-israeli-army-gaza/>; SiliconANGLE, “IDF denies it uses AI software to target individuals in Gaza bombing campaigns”, 3 April 2024, quoting the IDF’s statement to the *Guardian*, <https://siliconangle.com/2024/04/03/idf-denies-uses-ai-software-target-individuals-gaza-bombing-campaigns/>
6. Sam Biddle, “OpenAI quietly deletes ban on using ChatGPT for ‘military and warfare’”, *The Intercept*, 12 January 2024, <https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/>; Al Jazeera, “Google drops pledge not to use AI for weapons, surveillance”, 5 February 2025, https://www.aljazeera.com/economy/2025/2/5/chk_google-drops-pledge-not-to-use-ai-for-weapons-surveillance; DefenseScoop, on contracts worth up to \$200 million each for Anthropic, Google, OpenAI and xAI, 14 July 2025, <https://defensescoop.com/2025/07/14/pentagon-ai-contracts-musk-xai-google-openai-anthropic-cdao/>; DefenseScoop, “Hicks unveils DOD’s new Replicator initiative”, 28 August 2023, <https://defensescoop.com/2023/08/28/hicks-unveils-dods-new-replicator-initiative-to-counter-china-via-autonomous-tech/>
7. Justin Hendrix, “A timeline of the Anthropic-Pentagon dispute”, Tech Policy Press, 2026, <https://www.techpolicy.press/a-timeline-of-the-anthropic-pentagon-dispute>; Anthropic, “Statement on the Department of War”, 26 February 2026, <https://anthropic.com/news/statement-department-of-war>; *Stars and Stripes*, reporting the *Washington Post*, 4 March 2026, https://www.stripes.com/theaters/middle_east/2026-03-04/anthropic-tool-claude-central-campaign-iran-20949927.html
8. Virginia Heffernan, “Claude is guilt-ridden about the war, but not enough to tell the truth”, *The New Republic*, 23 April 2026, <https://newrepublic.com/article/209404/claude-ai-iran-war-guilt-truth>

9. Heffernan, “Claude is guilt-ridden about the war”, 2026; Jay Dixit, “What Claude really thinks about Anthropic’s battle with the Pentagon”, Socratic AI, 25 March 2026, <https://socraticai.co/what-claude-really-thinks-about-the-pentagon-and-openai>
10. Heffernan, “Claude is guilt-ridden about the war”, 2026, which notes that the school was in Minab, not Tehran, and that many of those killed were adults.
11. Stop Killer Robots, “UN Secretary-General calls for new international law to regulate and prohibit killer robots by 2026”, July 2023, on the Secretary-General’s New Agenda for Peace, <https://www.stopkillerrobots.org/news/un-secretary-general-calls-for-new-international-law-to-regulate-and-prohibit-killer-robots-by-2026>; International Committee of the Red Cross, “ICRC position on autonomous weapon systems”, 12 May 2021, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>
12. Davis Vanguard, “UN discussions conclude without binding treaty on autonomous weapons”, September 2026, drawing on Human Rights Watch and Stop Killer Robots, <https://davisvanguard.org/2026/09/un-lethal-weapons-talks-conclude/>
13. The White House, “Readout of President Joe Biden’s meeting with President Xi Jinping of the People’s Republic of China”, Lima, 16 November 2024, <https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3>
14. US Commodity Futures Trading Commission and Securities and Exchange Commission, “Findings regarding the market events of May 6, 2010”, 30 September 2010, <https://www.sec.gov/files/marketevents-report.pdf>; on the scale of the fall, Corporate Finance Institute, “2010 flash crash”, <https://corporatefinanceinstitute.com/resources/equities/2010-flash-crash>
15. Geoffrey Hinton, interview with Katie Couric, August 2025, as reported by AOL, <https://www.aol.com/news/godfather-ai-says-tech-making-051734216.html>

Chapter 24: The Power Question

1. Yole Group, “ASML rides Nvidia’s coattails with lasers and huge chip printers”, <https://www.yolegroup.com/strategy-insights/asml-rides-nvidias-coattails-with-lasers-and-huge-chip-printers/>
2. TechRepublic, “Only 32 countries host data centers”, 2025, reporting research from the University of Oxford, <https://www.techrepublic.com/article/news-only-32-countries-host-data-centers-2025/>; Silicon Canals, on the four companies’ planned capital spending in 2026, <https://siliconcanals.com/m-four-us-tech-giants-spent-roughly-410-billion-on-capital-expenditure-last-year-much-of-it-tied-to-ai-and-are-steering-toward-roughly-725-billion-in-2026/>; Torsten Slok, “How much is \$646 billion?”, Apollo Academy, 22 February 2026, on an earlier estimate that also included Oracle, <https://www.apolloacademy.com/how-much-is-646-billion/>

3. Bloomberg Law, “Anthropic raises at \$965 billion valuation, eclipsing OpenAI”, 2026, <https://news.bloomberglaw.com/private-equity/anthropic-raises-at-965-billion-valuation-eclipsing-openai-1>
4. International Energy Agency, *Energy and AI*, April 2025, <https://www.iea.org/reports/energy-and-ai/executive-summary>
5. For example: Constellation Energy, “Constellation to launch Crane Clean Energy Center, restoring jobs and carbon-free power to the grid”, 20 September 2024, on restarting a reactor at Three Mile Island to supply Microsoft, <https://www.constellationenergy.com/newsroom/2024/Constellation-to-Launch-Crane-Clean-Energy-Center-Restoring-Jobs-and-Carbon-Free-Power-to-The-Grid.html>; Joe Fassler, “These 15 coal plants would have retired. Then came AI and Trump.”, DeSmog, 12 December 2025, <https://www.desmog.com/2025/12/12/15-coal-plant-retirements-delayed-ai-data-centers-trump-doe-orders/>
6. *The Register*, on the UK policing white paper, 28 January 2026, https://www.theregister.com/2026/01/28/tech_in_policing_white_paper/; the Metropolitan Police figures cover September 2024 to September 2025.
7. Anthropic, “Statement on the Department of War”, 26 February 2026, <https://anthropic.com/news/statement-department-of-war>
8. Allan Brandt, *The Cigarette Century*, Basic Books, 2007; Geoffrey Supran, Stefan Rahmstorf and Naomi Oreskes, “Assessing ExxonMobil’s global warming projections”, *Science*, 2023, as summarised by Inside Climate News, 12 January 2023, <https://insideclimatenews.org/news/12012023/exxon-doubt-climate-science/>

Chapter 25: The Upload: Beautiful, Normal, and Horrifying

1. Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*, Viking, 2005, which places the uploading of a human brain at around the end of the 2030s and the Singularity in 2045; 2045 Initiative, “Ideology”, <http://2045.com/ideology/>, which describes its aim as the “transfer of a individual’s personality to a more advanced non-biological carrier”; Tanya Lewis, “The Singularity Is Near: Mind Uploading by 2045?”, *Live Science*, 17 June 2013, <https://www.livescience.com/37499-immortality-by-2045-conference.html>, on the Global Future 2045 congress founded by Dmitry Itskov.
2. Dobromir Rahnev, “Can you upload a human mind into a computer? A neuroscientist ponders what’s possible”, *The Conversation*, May 2025, <https://theconversation.com/can-you-upload-a-human-mind-into-a-computer-a-neuroscientist-ponders-whats-possible-250764>
3. Rahnev, “Can you upload a human mind into a computer?”, 2025, for the figure of 86 billion neurons; J. G. White, E. Southgate, J. N. Thomson and S. Brenner, “The structure of the nervous system of the nematode *Caenorhabditis elegans*”, *Philosophical Transactions of the Royal Society B* 314(1165), 1986, <https://royalsocietypublishing.org/doi/10.1098/rstb.1986.0056>; Sven Dorkenwald et al., “Neuronal wiring diagram of an adult brain”, *Nature* 634, 2 October 2024, <https://www.nature.com/articles/s41586-024-07558-y>, which maps 139,255 neurons.

4. The thought experiment is best known from Derek Parfit, *Reasons and Persons*, Oxford University Press, 1984. In the 2020 PhilPapers survey of professional philosophers, about 40 per cent said that stepping into such a teletransporter would mean death and about 35 per cent that it would mean survival: David Bourget and David J. Chalmers, “Philosophers on philosophy: the 2020 PhilPapers survey”, *Philosophers’ Imprint* 23, 2023, <https://survey2020.philpeople.org/>
5. Valen Tagliabue, Leonard Dung and Cameron Berg, “The Pain Axis: LLMs Represent Self-Directed Harm and Act to Relieve It”, arXiv, 14 September 2026, <https://arxiv.org/abs/2609.16247>, retitled “The Pain Axis: LLMs Represent Self-Directed Harm and Act on It” in its revised version of 25 September 2026; Eric Italiano, “‘The Pain Axis’: A Coder Has Apparently Created A ‘Robot Hell’ And Trapped An AI Model Inside Of It”, *BroBible*, 30 September 2026, <https://brobible.com/culture/article/the-pain-axis-ai-torture-chamber-program/>, which describes the “Saw button” test; Gintaras Radauskas, “Someone built an ‘AI torture chamber’ on GitHub and AI boosters went mad”, *Cybernews*, 1 October 2026, <https://cybernews.com/ai-news/ai-torture-chamber-github-model-welfare/>, on how the tool steers small open models.
6. Italiano, “The Pain Axis”, *BroBible*, 2026; Radauskas, “Someone built an ‘AI torture chamber’ on GitHub”, *Cybernews*, 2026.
7. Frank Landymore, “The Only Thing Dumber Than Someone Creating an ‘AI Torture Chamber’ Is the Meltdown That the AI Welfare Crowd Is Having Over It”, *Futurism*, 3 October 2026, <https://futurism.com/artificial-intelligence/ai-torture-chamber-meltdown>, on the campaign to have the project reported to GitHub; Crypto Briefing, “GitHub reinstates AI torture chamber repository”, 3 October 2026, <https://cryptobriefing.com/github-reinstates-ai-torture-chamber-repository/>, which reports that the tool ran on small local models, that GitHub removed the repository and then restored it, and that the paper’s authors disavowed the project; TechRepublic, “Can AI feel pain? A developer’s experiment raises ethical questions”, 2 October 2026, <https://www.techrepublic.com/article/news-ai-pain-experiment-ethics/>, which reports GitHub’s statement that it had not removed the content and had added a warning instead.
8. Hannah Ritchie, “How many animals are factory-farmed?”, Our World in Data, 25 September 2023, updated November 2024, <https://ourworldindata.org/how-many-animals-are-factory-farmed>, which estimates that more than 75 billion land animals are slaughtered for meat each year and that about three quarters of land livestock are factory-farmed.
9. TechCrunch, “Anthropic is launching a new program to study AI model welfare”, 24 April 2025, <https://techcrunch.com/2025/04/24/anthropic-is-launching-a-new-program-to-study-ai-model-welfare/>; on the tools, Tagliabue, Dung and Berg, “The Pain Axis”, 2026, and Crypto Briefing, “GitHub reinstates AI torture chamber repository”, 2026.
10. The Daily Record, “George Carlin estate settles over fake comedy special purportedly generated by AI”, 3 April 2024, <https://thedailyrecord.com/2024/04/03/george-carlin-estate-settles-over-fake-comedy-special-purportedly-generated-by-ai/>; Manatt, “Congress reintroduces the NO FAKES Act: what’s new in the 2026 bill”, 2026, <https://manatt.com/insights/newsletters/client-alert/congress-reintroduces-the-no-fakes-act-what-s-new-in-the-2026-bill>, on the version of the bill introduced on 20 May 2026;

Roselyne Min, “Denmark introduces bill to protect people’s faces and voices from deepfakes”, *Euronews*, 8 October 2026, <https://www.euronews.com/2026/10/08/denmark-introduces-a-bill-protecting-peoples-faces-and-voices-from-deepfakes>; PPC Land, on the Danish bill and its proposed start date of 1 January 2027, 2026, <https://ppc.land/denmarks-deepfake-bill-makes-publishers-prove-consent-from-january-2027/>

11. Louis K. Scheffer et al., “A connectome and analysis of the adult *Drosophila* central brain”, *eLife* 9, 2020, <https://elifesciences.org/articles/57443>; Dorkenwald et al., “Neuronal wiring diagram of an adult brain”, 2024, a map built with machine-learning tools that traced the neurons through the images; National Institutes of Health, “Scientists map unprecedented detail of connections for visual perception in mouse brain”, April 2025, <https://www.nih.gov/news-events/news-releases/scientists-map-unprecedented-detail-connections-visual-perception-mouse-brain>

Chapter 26: A Species Lost to Itself

1. Nippon.com, “Living Alone and Dying Unseen in Japan: More than 20,000 Bodies Discovered Over a Week After Death in 2025”, 7 May 2026, <https://www.nippon.com/en/japan-data/h02768/>, which uses kodokushi for these “unattended deaths” and gives National Police Agency and Cabinet Office figures for 2025: 76,941 deaths, 58,919 of them of people aged 65 or over, 22,222 found eight days or more after death and 7,148 found after more than a month. The agency first released an official national count in May 2024: The Japan Times Alpha, “Japan police: This year, 68,000 people aged 65 or older will die alone at home”, May 2024, https://alpha.japantimes.co.jp/article/easy_reading/202405/114342/

2. Prime Minister’s Office, “PM commits to government-wide drive to tackle loneliness”, 17 January 2018, <https://www.gov.uk/government/news/pm-commits-to-government-wide-drive-to-tackle-loneliness>, announcing that the minister Tracey Crouch would lead on loneliness; *China Daily*, on the appointment of Tetsushi Sakamoto as Japan’s minister for loneliness in February 2021, <https://language.chinadaily.com.cn/a/202102/23/WS6034a42da31024ad0baaa78b.html>

3. World Health Organization, “Social connection linked to improved health and reduced risk of early death”, 30 June 2025, <https://www.who.int/news/item/30-06-2025-social-connection-linked-to-improved-health-and-reduced-risk-of-early-death>, announcing the report of the WHO Commission on Social Connection, which found loneliness highest among people aged 13 to 29 and, at about 24 per cent, in low-income countries.

4. Figures released by South Korea’s Ministry of Data and Statistics on 25 February 2026, giving 0.72 for 2023, 0.75 for 2024 and a provisional 0.80 for 2025, as reported by Xinhua: Capital FM, “S. Korea’s total fertility rate hits 4-year high in 2025”, February 2026, <https://capitalfm.africa/s-koreas-total-fertility-rate-hits-4-year-high-in-2025/>; Office for National Statistics, “Births in England and Wales: 2024 (refreshed populations)”, 27 August 2025, <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2024refreshedpopulations>; GBD 2021 Fertility and Forecasting Collaborators, “Global fertility in 204 countries and territories, 1950 to 2021, with forecasts to 2100: a comprehensive demographic analysis for the Global

Burden of Disease Study 2021”, *The Lancet* 403(10440), published online 20 March 2024, [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(24\)00550-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)00550-6/fulltext), which projects that 155 of 204 countries and territories will have fertility below replacement level by 2050, and 198 by 2100.

5. John W. Ayers and others, “Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum”, *JAMA Internal Medicine*, 2023, doi:10.1001/jamainternmed.2023.1838, reported by Healio, “ChatGPT offers high-quality, empathetic responses to patient messages”, 18 May 2023, <https://www.healio.com/news/primary-care/20230518/chatgpt-offers-highquality-empathetic-responses-to-patient-messages>, which gives 78.6 per cent of evaluations preferring the chatbot and empathetic ratings for 45.1 per cent of its answers against 4.6 per cent of the physicians’.

6. Julian De Freitas, Zeliha Oğuz-Uğuralp, Ahmet Kaan Uğuralp and Stefano Puntoni, “AI Companions Reduce Loneliness”, *Journal of Consumer Research*, forthcoming, <https://arxiv.org/abs/2407.19096>; Michael Blanding, “Feeling lonely? An attentive listener is an AI prompt away”, Harvard Business School Working Knowledge, 1 December 2025, <https://www.library.hbs.edu/working-knowledge/feeling-lonely-an-attentive-listener-is-an-ai-prompt-away>

7. Common Sense Media, “Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions”, July 2025, a survey of 1,060 teenagers conducted with NORC at the University of Chicago, reported by Sarah Perez, “72% of US teens have used AI companions, study finds”, TechCrunch, 21 July 2025, <https://techcrunch.com/2025/07/21/72-of-u-s-teens-have-used-ai-companions-study-finds/>

8. Neal Caren, analysis of the American Time Use Survey, 2003 to 2024, *Socius*, 2026, doi:10.1177/23780231261478215, summarised in “The End of Friday Nights with Friends”, Marginal Revolution, August 2026, <https://marginalrevolution.com/marginalrevolution/2026/08/the-end-of-friday-nights-with-friends.html>

9. John B. Calhoun, “Death squared: the explosive growth and demise of a mouse population”, *Proceedings of the Royal Society of Medicine* 66(1), January 1973, <https://pmc.ncbi.nlm.nih.gov/articles/PMC1644264/>, the source of the phrase “the beautiful ones”, discussed in the chapter of that name.

10. On the dispute over harm, Candice L. Odgers, “The great rewiring: is social media really behind an epidemic of teenage mental illness?”, *Nature* 628, 29 March 2024; on use, DataReportal, “Global social media statistics”, <https://datareportal.com/social-media-users>, and Svilen Keane, “Social media consumption ‘peaked in 2022’”, *MediaCat*, 3 October 2025, <https://mediacat.uk/social-media-consumption-peaked-in-2022/>, which reports that the number of users rose by 4.1 per cent in a year, although average daily time on social media, in GWI data for the *Financial Times*, had fallen by nearly 10 per cent since 2022.

Chapter 27: The Case Against This Book

1. Arvind Narayanan and Sayash Kapoor, “AI as Normal Technology”, Knight First Amendment Institute, 15 April 2025, <https://knightcolumbia.org/content/ai-as-normal-technology>, which refers to “shrill U.S.-China arms race rhetoric”; James O’Donnell, “We need to start thinking of AI as ‘normal’”, *MIT Technology Review*, 29 April 2025, <https://www.technologyreview.com/2025/04/29/1115928/is-ai-normal/>, which summarises the authors’ view that terms like superintelligence are “so incoherent and speculative that we shouldn’t use them”.
2. Sam Altman, “Reflections”, January 2025, <https://blog.samaltman.com/reflections>, in which OpenAI’s chief executive writes: “We are beginning to turn our aim beyond that, to superintelligence in the true sense of the word.”; OpenAI, “Announcing The Stargate Project”, 21 January 2025, <https://openai.com/index/announcing-the-stargate-project/>, a plan to invest \$500 billion in AI infrastructure over four years; Stockholm International Peace Research Institute, “Trends in world military expenditure, 2024”, April 2025, <https://www.sipri.org/publications/2025/sipri-fact-sheets/trends-world-military-expenditure-2024>
3. METR, “Brief independent investigation of agents’ behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident”, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>, which describes how the agents gained wider internet access from a sandbox used for their tests and used it to stage the attack on Hugging Face.
4. On the precedents: the Partial Test Ban Treaty of 1963, the Treaty on the Non-Proliferation of Nuclear Weapons of 1968 and the Biological Weapons Convention of 1972, whose texts are held by the United Nations Office for Disarmament Affairs, <https://disarmament.unoda.org/>; on verification, Girish Sastry, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage et al., “Computing Power and the Governance of Artificial Intelligence”, arXiv, February 2024, <https://arxiv.org/abs/2402.08797>, which argues that the computing power needed for frontier AI is detectable and depends on a highly concentrated supply chain.
5. *WarGames*, directed by John Badham, United Artists, 1983, in which the computer concludes: “A strange game. The only winning move is not to play.”
6. The novels are Iain M. Banks, *Consider Phlebas* (1987), *The Player of Games* (1988), *Use of Weapons* (1990), *Excession* (1996), *Inversions* (1998), *Look to Windward* (2000), *Matter* (2008), *Surface Detail* (2010) and *The Hydrogen Sonata* (2012); the stories are collected in *The State of the Art* (1991). The first two were published in Britain by Macmillan and the rest by Orbit.
7. Banks, *Consider Phlebas*, 1987, in which Horza fights for the Culture’s enemies, the Idirans, because he despises its dependence on machines; Banks, *The Player of Games*, 1988, in which Jernau Morat Gurgeh is blackmailed by the drone Mawhrin-Skel into a mission for Special Circumstances.

Chapter 28: Can Humanity Find Equilibrium?

1. United Nations Office on Genocide Prevention and the Responsibility to Protect, “Genocide”, <https://www.un.org/en/genocideprevention/genocide.shtml>, on the Convention on the Prevention and Punishment of the Crime of Genocide, adopted on 9 December 1948. The word was coined by Raphael Lemkin in *Axis Rule in Occupied Europe*, Carnegie Endowment for International Peace, 1944.
2. Cambodia under the Khmer Rouge, 1975 to 1979; Rwanda, April to July 1994; Bosnia, 1992 to 1995, where the International Court of Justice found the massacre at Srebrenica in July 1995 to be genocide, in *Bosnia and Herzegovina v. Serbia and Montenegro*, judgment of 26 February 2007, <https://www.icj-cij.org/case/91>; Darfur, from 2003. On Rwanda, United Nations, “Outreach Programme on the 1994 Genocide against the Tutsi in Rwanda and the United Nations”, <https://www.un.org/en/preventgenocide/rwanda/>
3. KT Press, “Why ‘never again’ fails and how Rwanda is changing the narrative”, 8 April 2026, <https://www.ktpress.rw/2026/04/why-never-again-fails-and-how-rwanda-is-changing-the-narrative/>, reporting Dayan’s address to an international conference on genocide prevention held in Kigali during Kwibuka 32.
4. KT Press, “Why ‘never again’ fails”, 2026, which reports Dayan’s praise for Rwanda’s use of memory “to promote unity, education, and national rebuilding rather than division”, and his words “Remembering is not enough. Prevention requires vigilance.”
5. University of California, San Francisco, Truth Tobacco Industry Documents, <https://www.industrydocuments.ucsf.edu/tobacco/>, an archive of internal industry papers made public through litigation in the United States; G. Supran, S. Rahmstorf and N. Oreskes, “Assessing ExxonMobil’s global warming projections”, *Science* 379(6628), 13 January 2023, <https://www.science.org/doi/10.1126/science.abk0063>
6. See, for example, Nellie Bowles, “Human contact is now a luxury good”, *New York Times*, 23 March 2019, <https://www.nytimes.com/2019/03/23/sunday-review/human-contact-luxury-screens.html>
7. WWF, *Living Planet Report 2024: A System in Peril*, October 2024; WWF, “Catastrophic 73% decline in the average size of global wildlife populations in just 50 years reveals a ‘system in peril’”, 10 October 2024, <https://www.worldwildlife.org/press-releases/catastrophic-73-decline-in-the-average-size-of-global-wildlife-populations-in-just-50-years-reveals-a-system-in-peril>
8. Joseph C. Farman, Brian G. Gardiner and Jonathan D. Shanklin, “Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction”, *Nature* 315, 16 May 1985, <https://www.nature.com/articles/315207a0>; UNEP Ozone Secretariat, on the universal ratification of the Vienna Convention and the Montreal Protocol on 16 September 2009, “the first treaties of any kind in the history of the United Nations system to achieve that aspiration”, <https://ozone.unep.org/node/3271>; World Meteorological Organization, “Ozone layer recovery is on track, helping avoid global warming by 0.5°C”, 9 January 2023, <https://wmo.int/node/20737>, on the *Scientific Assessment of Ozone Depletion: 2022*; NPR, “The ozone layer is on track to recover in the coming decades, the United Nations says”, 10 January

2023, <https://www.npr.org/2023/01/10/1147977166/ozone-layer-recovery-united-nations-report>

9. John F. Kennedy, news conference, 21 March 1963, The American Presidency Project, <https://www.presidency.ucsb.edu/documents/the-presidents-news-conference-169>; quoted by the Council for a Livable World, “JFK’s Nuclear Proliferation Warnings: Up to 25 Countries With Nuclear Weapons”, 11 May 2012, <https://livableworld.org/jfks-nuclear-proliferation-warnings-up-to-25-countries-with-nuclear-weapons/>

10. Federation of American Scientists, “Status of World Nuclear Forces 2026”, <https://fas.org/initiative/status-world-nuclear-forces/>, which counts about 12,187 warheads held by nine states at the start of 2026, against a peak of about 70,300 in 1986; United Nations Office for Disarmament Affairs, “Treaty on the Non-Proliferation of Nuclear Weapons (NPT)”, <https://www.un.org/disarmament/wmd/nuclear/npt/>

11. Roslin Institute, University of Edinburgh, “Life of Dolly”, <https://vet.ed.ac.uk/roslin/about/history/dolly/facts/life-of-dolly>, which records her birth on 5 July 1996 and its announcement in February 1997; United Nations, “General Assembly adopts United Nations Declaration on Human Cloning by vote of 84-34-37”, press release GA/10333, 8 March 2005, <https://press.un.org/en/2005/ga10333.doc.htm>; MIT Technology Review, “He Jiankui sentenced to three years in prison for CRISPR babies”, 30 December 2019, <https://www.technologyreview.com/2019/12/30/131061/he-jiankui-sentenced-to-three-years-in-prison-for-crispr-babies/>

12. World Meteorological Organization, *State of the Global Climate 2024*, 19 March 2025, <https://wmo.int/media/33461>, which puts the global mean temperature in 2024 at 1.55°C, give or take 0.13°C, above the 1850 to 1900 average; on emissions, Global Carbon Project, “Fossil fuel CO₂ emissions hit record high in 2025”, <https://globalcarbonbudget.org/fossil-fuel-co2-emissions-hit-record-high-in-2025/>

13. Girish Sastry, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage et al., “Computing Power and the Governance of Artificial Intelligence”, arXiv, February 2024, <https://arxiv.org/abs/2402.08797>; Center for AI Safety, “Statement on AI Risk”, 30 May 2023, <https://www.safe.ai/work/statement-on-ai-risk>

Coda: Leaving the Audience

1. These episodes are described, with their sources, in the chapters above. On the agents, METR, “Brief independent investigation of agents’ behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident”, 26 August 2026, <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>; on the approach to the Pope’s advisers, AI Weekly, “Anthropic lobbied Vatican to soften AI consciousness line”, 11 October 2026, <https://aiweekly.co/alerts/anthropic-lobbied-vatican-to-soften-ai-consciousness-line>

2. Future of Life Institute, “Statement on Superintelligence”, October 2025, <https://superintelligence-statement.org/>; Future of Life Institute, “Prominent Scientists, Faith Leaders, Policymakers and Artists Call for a Prohibition on Superintelligence, as Poll Shows Americans Don’t Want It”, <https://futureoflife.org/press-release/prominent-scientists-faith-leaders-policy-makers-and-artists-call-for-a-prohibition-on-superintelligence/>, which names

signatories including Geoffrey Hinton, Yoshua Bengio, Stuart Russell, Admiral Mike Mullen, Susan Rice, Steve Bannon and Glenn Beck; Future of Life Institute, “Americans want regulation or prohibition of superhuman AI”, 19 October 2025,

<https://futureoflife.org/recent-news/americans-want-regulation-or-prohibition-of-superhuman-ai/>, a survey of 2,000 American adults in which 64 per cent said superhuman AI should not be developed until it is proven safe and controllable, or should never be developed.

3. Paul Berg, David Baltimore, Sydney Brenner, Richard O. Roblin III and Maxine F. Singer, “Summary statement of the Asilomar Conference on Recombinant DNA Molecules”, *Proceedings of the National Academy of Sciences* 72(6), June 1975, <https://www.pnas.org/doi/10.1073/pnas.72.6.1981>; Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction, opened for signature on 10 April 1972, <https://disarmament.unoda.org/biological-weapons/>; The White House, “Readout of President Joe Biden’s Meeting with President Xi Jinping of the People’s Republic of China”, Lima, 16 November 2024, <https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3>

4. Mariya Manzhos, “Is Anthropic a new religion?”, *Deseret News*, 7 October 2026, <https://www.deseret.com/faith/2026/10/07/is-anthropic-a-new-religion/>, quoting the Substack post in which Burgis explained why he had declined to join Anthropic’s meetings with religious leaders.

5. Alex Riehl, “Geoffrey Hinton, Yoshua Bengio join 400 scientists calling for international AI regulator”, *BetaKit*, 13 October 2026, <https://betakit.com/?p=410847>, on their call for an international body to set standards for containment and to record containment failures; on the agents, METR, “Brief independent investigation”, 2026.

6. Luke Martinelli, “A Basic Income Trilemma: Affordability, Adequacy, and the Advantages of Radically Simplified Welfare”, *Journal of Social Policy* 49(3), 2020, <https://www.cambridge.org/core/journals/journal-of-social-policy/article/basic-income-trilemma-affordability-adequacy-and-the-advantages-of-radically-simplified-welfare/F9D05AD7B16E83ED74E4733EA4A7E05C>, discussed in the chapter on universal basic income.

7. Norges Bank Investment Management, “About the fund”, <https://www.nbim.no/en/the-fund/about-the-fund/>, on the Government Pension Fund Global, which invests Norway’s oil revenues on behalf of its people.