

INF1005H1S: Artificial Intelligence Policy

Supplementary Reading List
Last updated January 26, 2024

Recommended Texts

- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.
- Atanasoski, N. & Vora, K. (2019). *Surrogate humanity: Race, robots, and the politics of technological futures*. Duke University Press.
- Christian, B. (2020). *The alignment problem: Machine learning and human values*. W. W. Norton & Company.
- Pasquale, F. (2020). *New laws of robotics: Defending human expertise in the age of AI*. Belknap Press: An Imprint of Harvard University Press.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Renieris, E. M. (2023). *Beyond data: Reclaiming human rights at the dawn of the metaverse*. The MIT Press.
- Agrawal, A., Gans, J., & Goldfarb, A. (2022). *Power and prediction: The disruptive economics of artificial intelligence*. Harvard Business Review Press.
- Iansiti, M. & Lakhani, K. R. (2020). *Competing in the age of AI: Strategy and leadership when algorithms and networks run the world*. Harvard Business Review Press.
- Benanav, A. (2020). *Automation and the future of work*. Verso.
- Dyer-Witheford, N., Kjosen, A. M., & Steinhoff, J. (2019). *Inhuman power: Artificial intelligence and the future of capitalism*. Pluto Press.
- McQuillan, D. (2022). *Resisting AI: An anti-fascist approach to artificial intelligence*. Bristol University Press.
- Vickers, B. & Allado-McDowell, K. (Eds., 2021). *Atlas of anomalous AI*. Ignota Books.
- Halpern, O. & Mitchell, R. (2023). *The smartness mandate*. The MIT Press.

Week 1: Introduction to the Course & AI Systems

- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-360. <https://academic.oup.com/mind/article/LIX/236/433/986238>

- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444. <https://www.nature.com/articles/nature14539>
- Arsanjani, A. (2023). The generative AI life-cycle. *Medium*. <https://dr-arsanjani.medium.com/the-generative-ai-life-cycle-fb2271a70349>
- De Silva, D. & Alahakoon, D. (2022). The artificial intelligence life cycle: From conception to production. *Patterns*, 3, 100489. <https://www.sciencedirect.com/science/article/pii/S2666389922000745>
- Bender, E. M. et al. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. March 2021. 610-623. <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>
- Artificial Intelligence Index (2023). *AI Index Report 2023*, Ch. 2 & 3 (“Technical Performance” and “Technical AI Ethics”): <https://aiindex.stanford.edu/report/#individual-chapters>
- Benaich, N. & Air Street Capital (2023). State of AI report 2023. <https://www.stateof.ai/>
- Lee, T. B. & Trott, S. (2023). A jargon-free explanation of how AI large language models work. *Ars Technica*. <https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/>
- Financial Times (2023). Generative AI exists because of the transformer: This is how it works. <https://ig.ft.com/generative-ai>
- Jones, E. (2023). Explainer: What is a foundation model? *Ada Lovelace Institute*. <https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/>
- Russell, S., Perset, K., & Grobelnik, M. (2023). Updates to the OECD’s definition of an AI system explained. *OECD AI Policy Observatory*. <https://oecd.ai/en/wonk/ai-system-definition-update>

Week 2: Impacts & Ethics of AI Systems

Race, Gender, & Discrimination

- Angwin, J. et al. (2016). Machine bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating systems: Gender, race, and power in AI. *AI Now Institute*.

<https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>

- Keyes, O. (2018). The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2, Issue CSCW, Article No. 88 (1-22).
<https://dl.acm.org/doi/10.1145/3274357>
- Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication, & Society*, 22(7), 900-915. <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2019.1573912>

Fairness, Accountability, Transparency, & Safety

- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).
<https://journals.sagepub.com/doi/full/10.1177/2053951716679679>
- Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns*, 2(2). [https://www.cell.com/patterns/pdf/S2666-3899\(21\)00015-5.pdf](https://www.cell.com/patterns/pdf/S2666-3899(21)00015-5.pdf)
- Miceli, M., Posada, J., & Yang, T. (2022). Studying up machine learning data: Why talk about bias when we mean power? *Proceedings of the ACM on Human-Computer Interaction*. Volume 6, Issue GROUP, Article 34.
<https://dl.acm.org/doi/10.1145/3492853>
- Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. *FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1186-1197.
<https://dl.acm.org/doi/10.1145/3593013.3594073>
- Attard-Frost, B., De los Ríos, A., & Walters, D. R. (2022). The ethics of AI business practices: A review of 47 AI ethics guidelines. *AI and Ethics*, 3, 389-406.
<https://link.springer.com/article/10.1007/s43681-022-00156-6>
- Raji, I. D. et al. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 33-44.
<https://dl.acm.org/doi/10.1145/3351095.3372873>
- Adadi, A. & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138-52160.
<https://ieeexplore.ieee.org/document/8466590>
- Brundage, M. et al. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv:1802.07228.
<https://arxiv.org/pdf/1802.07228.pdf>
- Anthropic (2023). Core views on AI safety: When, why, what, and how.
<https://www.anthropic.com/index/core-views-on-ai-safety>
- Lazar, S. & Nelson, A. (2023). AI safety on whose terms? *Science*, 381(6654), 138. <https://www.science.org/doi/10.1126/science.adi8982>

Political Economy & Ecology of AI Systems

- Williams, A., Miceli, M., & Gebru, T. (2022). The exploited labor behind artificial intelligence. *Noema*.
<https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/>
- Whittaker, M. (2021). The steep cost of capture. *Interactions of the ACM*, 26(6), 50-55. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4135581
- Widder, D. G., Whittaker, M., & West, S. M. (2023). Open (for business): Big tech, concentrated power, and the political economy of open AI. *SSRN*.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4543807
- Ahmed, N., Wahed, M., & Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, 375(6635), 884-886.
https://ide.mit.edu/wp-content/uploads/2023/03/0303PolicyForum_Ai_FF-2.pdf
- Gmyrek, P., Berg, J., & Bescond, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. *International Labour Organization*. ILO Working Paper 96.
https://www.ilo.org/wcmsp5/groups/public/---dgreports/---inst/documents/publication/wcms_890761.pdf
- Gupta, A. (2021). The imperative for sustainable AI systems. *The Gradient*.
<https://thegradient.pub/sustainable-ai/>
- Kaack, L. H. et al. (2020). Artificial intelligence and climate change: Opportunities, considerations, and policy levers to align AI with climate change goals. *Heinrich-Böll-Stiftung*.
https://eu.boell.org/sites/default/files/2021-04/Artificial%20Intelligence%20and%20Climate%20Change_FINAL_14042021.pdf
- Li, P. et al. (2023). Making AI less ‘thirsty’: Uncovering and addressing the secret water footprint of AI models. arXiv:2304.03271.
<https://arxiv.org/pdf/2304.03271.pdf>
- Solaiman, I. et al. (2023). Evaluating the social impact of generative AI systems in systems and society. arXiv:2306.05949. <https://arxiv.org/abs/2306.05949>
- Vipra, J. & West, S. M. (2023). Computational power and AI. *AI Now Institute*.
<https://ainowinstitute.org/publication/policy/compute-and-ai>
- Lewis, J. et al. (2018). Making kin with the machines. *Journal of Design and Science*. <https://jods.mitpress.mit.edu/pub/lewis-arista-pechawis-kite/release/1>

Sectoral Challenges

- Chui et al. (2018). Applying AI for social good. *McKinsey*.
<https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good>
- Awad, E. et al. (2018). The Moral Machine experiment. *Nature*, 563, 59-64.
<https://www.nature.com/articles/s41586-018-0637-6>

- Davis, J. et al. (2022). Five ethical challenges facing data-driven policing. *AI and Ethics*, 2, 185-198. <https://link.springer.com/article/10.1007/s43681-021-00105-9>
- CIGI (2023). The ethics of automated warfare and artificial intelligence. <https://www.cigionline.org/the-ethics-of-automated-warfare-and-artificial-intelligence/>
- Akgun, S. & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2, 431-440. <https://link.springer.com/article/10.1007/s43681-021-00096-7>
- Murphy, K. et al. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Medical Ethics*, 22, Article 14. <https://bmcmethics.biomedcentral.com/articles/10.1186/s12910-021-00577-8>
- Brintrup, A. et al. (2023). Trustworthy, responsible, ethical AI in manufacturing and supply chains: Synthesis and emerging research questions. arXiv: 2305.11581. <https://arxiv.org/pdf/2305.11581.pdf>
- Jiang, H. et al. (2023). AI art and its impact on artists. *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 363-374. <https://dl.acm.org/doi/10.1145/3600211.3604681>

Week 3: Governance of AI Systems

Government perspectives on AI governance

- G7 (2018). Charlevoix common vision for the future of artificial intelligence. https://www.international.gc.ca/world-monde/assets/pdfs/international_relations-internationales/g7/2018-06-09-artificial-intelligence-artificielle-en.pdf
- G7 (2023). G7 Hiroshima Process on Generative Artificial Intelligence (AI). https://read.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en#page1
- Government of Canada (2022). Pan-Canadian Artificial Intelligence Strategy. <https://ised-isde.canada.ca/site/ai-strategy/en>
- UK Office for Artificial Intelligence (2023). A pro-innovation approach to AI regulation. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>
- UK Parliament (2023). The governance of artificial intelligence: Interim report – Report Summary. <https://publications.parliament.uk/pa/cm5803/cmselect/cmsctech/1769/summary.html>
- Trump White House (2020). Artificial intelligence for the American people. <https://trumpwhitehouse.archives.gov/ai/>
- Chopra, R. et al. (2023). Joint statement on enforcement efforts against discrimination and bias in automated systems.

https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf

- European Commission (2018). Communication from the Commission: Artificial intelligence for Europe. SWD(2018) 137 final.
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN>
- HM Government (2021). National AI strategy [United Kingdom].
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf
- Roberts, H. et al. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*, 36, 59-77.
<https://link.springer.com/article/10.1007/s00146-020-00992-2>
- NITI Aayog (2018). National strategy for artificial intelligence [India].
<https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf>

Industry perspectives on AI governance

- Google (2023). Google AI policy perspectives.
<https://ai.google/responsibility/public-policy-perspectives/>
- Microsoft (2023). Microsoft responsible AI.
<https://www.microsoft.com/en-us/ai/responsible-ai>
- Smith, B. (2023). How do we best govern AI? *Microsoft*.
<https://blogs.microsoft.com/on-the-issues/2023/05/25/how-do-we-best-govern-ai>
- Hagemann, R. & Leclerc, J.-M. (2020). Precision regulation for artificial intelligence. *IBM Policy Lab*. <https://www.ibm.com/policy/ai-precision-regulation/>
- Anthropic (2023). Charting a path to AI accountability.
<https://www.anthropic.com/index/charting-a-path-to-ai-accountability>
- Responsible AI Institute (2023). How we help.
<https://www.responsible.ai/how-we-help>
- Holistic AI (2023). <https://www.holisticai.com/>

Civil society perspectives on AI governance

- Gillespie, N. et al. (2023). Trust in artificial intelligence: A global study. *KPMG and the University of Queensland*.
<https://assets.kpmg.com/content/dam/kpmg/au/pdf/2023/trust-in-ai-global-insights-2023.pdf>
- Maddaus, G. (2023). How the WGA decided to harness – but not ban – artificial intelligence. *Variety*.
<https://variety.com/2023/biz/news/wga-ai-writers-strike-technology-ban-1235610076/>

- Dalton, A. & The Associated Press (2023). Writers' strike: Writers strike: Why A.I. is such a hot-button issue in Hollywood's labor battle with SAG-AFTRA. *Fortune*. <https://fortune.com/2023/07/24/sag-aftra-writers-strike-explained-artificial-intelligence/>
- Writers Guild of Canada (2023). Re: Artificial intelligence and the role of Canadian creators. <https://www.wgc.ca/sites/default/files/2023-06/WGC%20Letter%20to%20PCH-ISED%20on%20AI%20June%202023.pdf>
- Canadian Civil Liberties Association (2023). Category: Artificial intelligence. <https://ccla.org/category/privacy/surveillance-technology/artificial-intelligence/>
- Delgado, F. et al. (2023). The participatory turn in AI design: Theoretical foundations and the current state of practice. *EAAMO '23, October 30–November 1, 2023*. <https://arxiv.org/pdf/2310.00907.pdf>
- Birhane, A. et al. (2022). Power to the people? Opportunities and challenges for participatory AI. *EAAMO '22: Equity and Access in Algorithms, Mechanisms, and Optimization*, 6. <https://dl.acm.org/doi/pdf/10.1145/3551624.3555290>

Week 4: Regulation, Standards, & Guidance

Legislation & Regulation

- Parliament of Canada (2023). Bill C-27 [Part 3: Artificial Intelligence and Data Act]. <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>
- Minister of Innovation, Science and Industry (2023). [Letter to Chair of the House of Commons Standing Committee on Industry and Technology, November 28, 2023]. <https://www.ourcommons.ca/content/Committee/441/INDU/WebDoc/WD12751351/12751351/MinisterOfInnovationScienceAndIndustry-2023-11-28-Combined-e.pdf>
- European Parliament (2023, June). Artificial intelligence act. Briefing: EU legislation in progress. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)
- Hilliard, A. & Gulley, A. (2023). What is the EU AI Act? *Holistic AI*. <https://www.holisticai.com/blog/eu-ai-act>
- Toronto Police Services Board (2022). Use of artificial intelligence technology. <https://tpsb.ca/policies-by-laws/board-policies/195-use-of-artificial-intelligence-technology>

Standards & Certification

- NIST (2023). AI Risk Management Framework (AI RMF 1.0). <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>
- Responsible Artificial Intelligence Institute (2023). The Responsible AI Certification: Program guidebook. https://www.responsible.ai/files/ugd/f70099_77bb74c4c23a45219e35520b66922b7a.pdf
- Microsoft (2022, June). Microsoft Responsible AI Standard, v2. <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf>
- Auld, G. et al. (2022). Governing AI through ethical standards: Learning from the experiences of other private governance initiatives. *Journal of European Public Policy*, 29(11), 1822-1844. <https://www.tandfonline.com/doi/full/10.1080/13501763.2022.2099449>

Guidance Documents

- Government of Canada (2023). Canadian Guardrails for Generative AI – Code of Practice. <https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>
- The White House (2022). Blueprint for an AI Bill of Rights. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- Barrett, A. (2023). Standards around generative AI. *AP*. <https://blog.ap.org/standards-around-generative-ai>
- UNESCO (2023). Guidance for generative AI in education and research. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- OpenAI (2023). Teaching with AI. <https://openai.com/blog/teaching-with-ai>
- OpenAI (2022). Best practices for deploying language models. <https://openai.com/blog/best-practices-for-deploying-language-models>
- Leslie, D. (2020). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. *The Alan Turing Institute*. https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf
- Office of the Privacy Commissioner of Canada (2023). Principles for responsible, trustworthy and privacy-protective generative AI technologies. https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

- University of Toronto (2023). Use of generative AI in the classroom. <https://teaching.utoronto.ca/resources/generative-artificial-intelligence-in-the-classroom/>
- University of Toronto (2023). Guidance on the appropriate use of generative artificial intelligence in graduate theses. <https://www.sgs.utoronto.ca/about/guidance-on-the-use-of-generative-artificial-intelligence/>

AI Audits & Audit Tools

- Faveri, B. & Auld, G. (2023). Informing possible futures for the use of third-party audits in AI regulations. https://repository.library.carleton.ca/concern/research_works/2z10wr54f?locale=en
- Gebru, T. et al. (2018). Datasheets for datasets. *Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning*. <https://www.microsoft.com/en-us/research/uploads/prod/2019/01/1803.09010.pdf>
- Mitchell, M. et al. (2019). Model cards for model reporting. *FAT* '19*. <https://arxiv.org/pdf/1810.03993.pdf>
- Raji, I. D. et al. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *FAT* '20*. <https://dl.acm.org/doi/pdf/10.1145/3351095.3372873>
- Raji, I. D. et al. (2022). Outsider oversight: Designing a third party audit ecosystem for AI governance. *AIES '22*. <https://arxiv.org/pdf/2206.04737.pdf>

Emerging Topics: Participatory AI & Frontier Regulation

- Delgado, F. et al. (2023). The participatory turn in AI design: Theoretical foundations and the current state of practice. *EAAMO '23, October 30–November 1, 2023*. <https://arxiv.org/pdf/2310.00907.pdf>
- Government of Canada (2023). Learning together for responsible artificial intelligence: Report of the Public Awareness Working Group. <https://ised-isde.canada.ca/site/advisory-council-artificial-intelligence/en/public-awareness-working-group/learning-together-responsible-artificial-intelligence>
- Anderljung, M. et al. (2023). Frontier AI regulation: Managing emerging risks to public safety. arXiv:2307.03718. <https://arxiv.org/pdf/2307.03718.pdf>