Tech Ethics Resources:

ORGANISATIONS

Centre for Long term resilience https://www.longtermresilience.org/

https://www.just-ai.net/methodology

Responsible Innovation Labs: https://www.rilabs.org/report

https://www.worldbenchmarkingalliance.org/mission/

https://investorsforhumanrights.org/

https://cset.georgetown.edu/

Beijing Academy of Artificial Intelligence https://www.baai.ac.cn/english.html

Turing Institute

Ada Lovelace

ETHICAL FRAMEWORKS:

https://www.rilabs.org/responsible-ai Responsible AI framework

The Vatican: https://www.romecall.org/the-call/

ICO AI and Data protection:
https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/ai-and-data-protection-risk-toolkit/

IEEE standard: https://xplorestaging.ieee.org/browse/standards/get-program/page/series?id=93

US AI Risk framework https://www.nist.gov/itl/ai-risk-management-framework

OECD AI principles https://oecd.ai/en/ai-principles

https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Blueprint of a Bill of AI rights https://www.whitehouse.gov/ostp/ai-bill-of-rights/

U.S. Equal Employment Opportunity Commission

https://www.federalregister.gov/documents/2023/01/10/2023-00283/draft-strategic-enforcement-plan#print

Joint initiative between EU and US on Trustworthy AI

https://digital-strategy.ec.europa.eu/en/library/ttc-joint-roadmap-trustworthy-ai-and-risk-management?utm_source=POLITICO.EU&utm_campaign=3e89ce51a0-EMAIL_CAMPAIGN_2023_01_19_12_30&utm_medium=email&utm_term=0_10959edeb5-3e89ce51a0-%5BLIST_EMAIL_ID%5D

AI accountability in the US government

https://www.actiac.org/system/files/2023-01/AI%20Accountability.pdf?utm_source=POLITICO.EU&utm_campaign=3e89ce51a0-EMAIL_CAMPAIGN_2023_01_19_12_30&utm_medium=email&utm_term=0_10959edeb5-3e89ce51a0-%5BLIST_EMAIL_ID%5D

REGULATION

EU AI Act: https://artificialintelligenceact.eu/the-act/

China regulating deep fake imagery

https://www.wsj.com/articles/china-a-pioneer-in-regulating-algorithms-turns-its-focus-to-deepfakes-11673149283

AI measurement and methodology - US National Institute of Standards and Technology

https://www.nist.gov/system/files/documents/2021/06/16/AIME_at_NIST-DRAFT-20210614.pdf?utm_source=POLITICO.EU&utm_campaign=3e89ce51a0-EMAIL_CAMPAIGN_2023_01_19_12_30&utm_medium=email&utm_term=0_10959edeb5-3e89ce51a0-%5BLIST_EMAIL_ID%5D

G20 statement in support of Trustworthy AI and OECD principles
https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf?utm_source=POLITICO.EU&utm_campaign=3e89ce51a0-EMAIL_CAMPAIGN_2023_01_19_12_30&utm_medium=email&utm_term=0_10959edeb5-3e89ce51a0-%5BLIST_EMAIL_ID%5D

https://www.ft.com/content/45857293-366a-4347-96dd-f5a37fbc6385?desktop=true&segmentId=7c8f09b9-9b61-4fbb-9430-9208a9e233c8#myft:notification:daily-email:content

UK Government National AI strategy
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020405/The_UK_s_National_AI_Strategy_-_Flow_diagram.pdf

Whitehouse on Crypto
https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/16/fact-sheet-white-house-releases-first-ever-comprehensive-framework-for-responsible-development-of-digital-assets/

REPORTS

Stanford AI Index (Human Centred AI)
https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf
https://aiindex.stanford.edu/report/ (slides)

McKinsey
https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#talent

https://creatingfutureus.org/inaugural-report-from-investors-for-a-sustainable-digital-economy-highlights-risks-from-data-use-and-investors-role-in-mitigating-consequences/

https://www.turing.ac.uk/sites/default/files/2020-10/understanding_bias_in_facial_recognition_technology.pdf

Swiss Re pieces on AI in insurance
https://www.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/opportunities-ai-insurance.html?utm_campaign=cpn-1788_aiininsurancepart2a_group_global_2023&utm_medium=socialMedia_organic&utm_source=linkedin&utm_content=digitalisation_bylineArticle_en&utm_term=part2antonio

Citi report on AI regulation
https://www.citivelocity.com/rendition/digidoc/eppublic/public/pdfRenderContainer.do?data=ZG9jaWQ9MzAxNTY1MDQmdGl0bGU9UVVrZ1ZHbHRaU0E2SUZkb1IYUWdlVzkxSUc1bFpXUWdkRzhnYTI1dmR5QmhZbTkxZENCQlNTQnlaV2QxYkdGMGFXOXUmZGV2aWNlVHlwZT1ERVNLVE9QJnBkZkZpbGVOYW1lPSZwbGF0Zm9ybT0xMDkmc291cmNlX2lkPTAmdmlld09wdGlvbj1wZGZPbmx5JmNoYW5uZWw9Y2l0aXZlbG9jaXR5JnN1YkNoYW5uZWw9V2Vi

COURSES/ Skills/ Networks

https://aiethicscourse.org/index.html

https://alltechishuman.org/community-slack

https://learning.edx.org/course/course-v1:EdinburghX+DEI01x+2T2022/home

Coverage:

Trends
https://www.generalist.com/briefing/what-to-watch-in-ai?utm_source=substack&utm_medium=email

https://www.technologyreview.com/2023/01/09/1066394/10-breakthrough-technologies-2023?utm_campaign=mb&utm_medium=newsletter&utm_source=morning_brew

Using Algorithms to shortlist job candidates
https://www.ethicallyalignedai.com/post/responsible-innovation-would-you-let-an-algorithm-shortlist-you-for-the-job

Collusion of algorithmic pricing
https://www.technologyreview.com/2021/10/27/1038835/podcast-how-pricing-algorithms-learn-to-collude/

Investor action:
https://capitalmonitor.ai/factor/social/engaging-with-facebook-and-twitter-is-like-talking-to-kids

US tech lobbying refuses Americans protections that they are providing to Europeans
https://www.brookings.edu/blog/techtank/2022/12/28/big-tech-giving-european-consumers-what-they-deny-americans/?utm_source=POLITICO.EU&utm_campaign=3c4efdb629-EMAIL_CAMPAIGN_2023_01_12_12_30&utm_medium=email&utm_term=0_10959edeb5-3c4efdb629-%5BLIST_EMAIL_ID%5D

2023 is the year of AI governance
https://finance.yahoo.com/news/ces-2023-a-is-at-a-really-important-moment-says-sonys-ai-ethics-expert-200046533.html?guccounter=1

AI Legal work is growing
https://oecd.ai/en/wonk/increasing-legal-cases?utm_source=POLITICO.EU&utm_campaign=3c4efdb629-EMAIL_CAMPAIGN_2023_01_12_12_30&utm_medium=email&utm_term=0_10959edeb5-3c4efdb629-%5BLIST_EMAIL_ID%5D

Emotional Recognition Technology
https://www.wired.co.uk/article/empathy-artificial-intelligence
https://iapp.org/news/a/exploring-the-privacy-ethical-issues-with-emotion-detection-tech/#
Based on flawed science and categorising of emotion without understanding its cause/ context can result in problematic conclusions. Plus less reliable with non-white faces, women 'mask' more and the neurodivergent

https://iapp.org/news/a/exploring-the-privacy-ethical-issues-with-emotion-detection-tech/#
(privacy & ethical issues with ERT)

Wellbeing apps in schools
https://www.technologyreview.com/2023/04/17/1071137/denmark-teachers-apps-student-mood-audit-software/?utm_source=engagement_email&utm_medium=email&utm_campaign=wklysun&utm_content=04.23.23.nonsubs_eng&mc_cid=9eac6bde65&mc_eid=d5afc044cc

Facial Recognition technology

https://bigbrotherwatch.org.uk/wp-content/uploads/2022/02/Whos-Watching-You_The-dominance-of-Chinese-state-owned-CCTV-in-the-UK-17746.pdf
https://www.codastory.com/authoritarian-tech/uk-supermarket-biometric-cameras/

Surveillance at work and its perverse outcomes
https://psyche.co/ideas/why-ai-surveillance-at-work-leads-to-perverse-outcomes?utm_source=Psyche+Magazine&utm_campaign=34bc9de1dc-EMAIL_CAMPAIGN_2023_01_24_11_54&utm_medium=email&utm_term=0_-34bc9de1dc-%5BLIST_EMAIL_ID%5D

Cyborg cells to treat cancer
https://interestingengineering.com/health/semi-living-cyborg-cells-treat-cancer?utm_campaign=etb&utm_medium=newsletter&utm_source=morning_brew

Why AI is unlikely to suffer Crypto's fate
https://luttig.substack.com/p/is-ai-the-new-crypto

Cognitive privacy and AI reading your thoughts
https://www.theguardian.com/science/2023/mar/26/nita-farahany-the-battle-for-your-brain-neurotechnology?CMP=Share_iOSApp_Other

Shortlisting by hiring algorithms
https://www.ethicallyalignedai.com/post/responsible-innovation-would-you-let-an-algorithm-shortlist-you-for-the-job

How AI will change science (by Eric Schmidt)
https://www.technologyreview.com/2023/07/05/1075865/eric-schmidt-ai-will-transform-science/?utm_source=engagement_email&utm_medium=email&utm_campaign=sun-special-offer&utm_term=07.09.23.nonsubs_25off&utm_content=JA23ISS.25Off.ENG&mc_cid=10123864c4&mc_eid=d5afc044cc

The rise of AI
https://www.emergingtechbrew.com/stories/2023/02/03/amid-wave-of-layoffs-some-tech-giants-are-elevating-ai?cid=30450724.149517&mid=79f808bc294a9aa75dc732bf63dc6616&utm_campaign=etb&utm_medium=newsletter&utm_source=morning_brew

Facebook privacy new antitrust rules
https://www.ft.com/content/d9a9af52-4cb5-4bfb-92ed-55e301343ff5

How social media hacks our brains
https://www.humanetech.com/brain-science

Ledger of harms
https://ledger.humanetech.com/

Online harms bill & ofcom hires
https://www.ft.com/content/19e9da57-5da4-40bc-989d-9820bf3d2aff?desktop=true&segmentId=7c8f09b9-9b61-4fbb-9430-9208a9e233c8#myft:notification:daily-email:content

COURSES

https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en?utm_source=POLITICO.EU&utm_campaign=3e89ce51a0-EMAIL_CAMPAIGN_2023_01_19_12_30&utm_medium=email&utm_term=0_10959edeb5-3e89ce51a0-%5BLIST_EMAIL_ID%5D

Generative AI - speed over responsibility
https://www.nytimes.com/2023/04/07/technology/ai-chatbots-google-microsoft.html

CAUTIONARY CORPORATE TALES

Hireview
https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/

GPT 4 (and why its better than 3)
https://www.technologyreview.com/2023/03/14/1069823/gpt-4-is-bigger-and-better-chatgpt-openai/?utm_source=engagement_email&utm_medium=email&utm_campaign=wklysun&utm_content=03.19.23.nonsubs_eng&mc_cid=d42ec9a242&mc_eid=d5afc044cc

Image Bias
https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/?utm_source=engagement_email&utm_medium=email&utm_campaign=wklysun&utm_content=04.02.23.nonsubs_eng&mc_cid=e5d3e0eca0&mc_eid=d5afc044cc

Impact on productivity
https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionize-economy-decide-what-looks-like/?utm_source=Stanford+HAI&utm_campaign=9009a37862-Mailchimp_HAI_Newsletter_March+2023_2_COPY_01&utm_medium=email&utm_term=0_aaf04f4a4b-9009a37862-214034162

https://www.theguardian.com/business/2023/sep/06/cars-collect-extensive-personal-data-on-drivers-study-warns

Case Studies of Ethics Fails:

Amazon's hiring algorithm

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Phone surveillance by the courts

https://www.wired.com/story/anti-porn-covenant-eyes-bond-revoked/?bxid=6276599b28cfb4543b074f74&cndid=69678237&esrc=growl2-regGate-0321&mbid=mbid%3DCRMWIR012019%0A%0A&source=Email_0_EDT_WIR_NEWSLETTER_0_DAILY_ZZ&utm_brand=wired&utm_campaign=aud-dev&utm_content=WIR_Daily_061223&utm_mailing=WIR_Daily_061223&utm_medium=email&utm_source=nl&utm_term=P1

Data Ethics Resources:

Counterfactuals
https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf

AI Blindspots
https://aiblindspot.media.mit.edu/

Consequence scanning
https://doteveryone.org.uk/project/consequence-scanning/

Synthetic data
https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic_Data_Survey-24.pdf

Centre for Data Ethics and Innovation (UK Government)
https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation

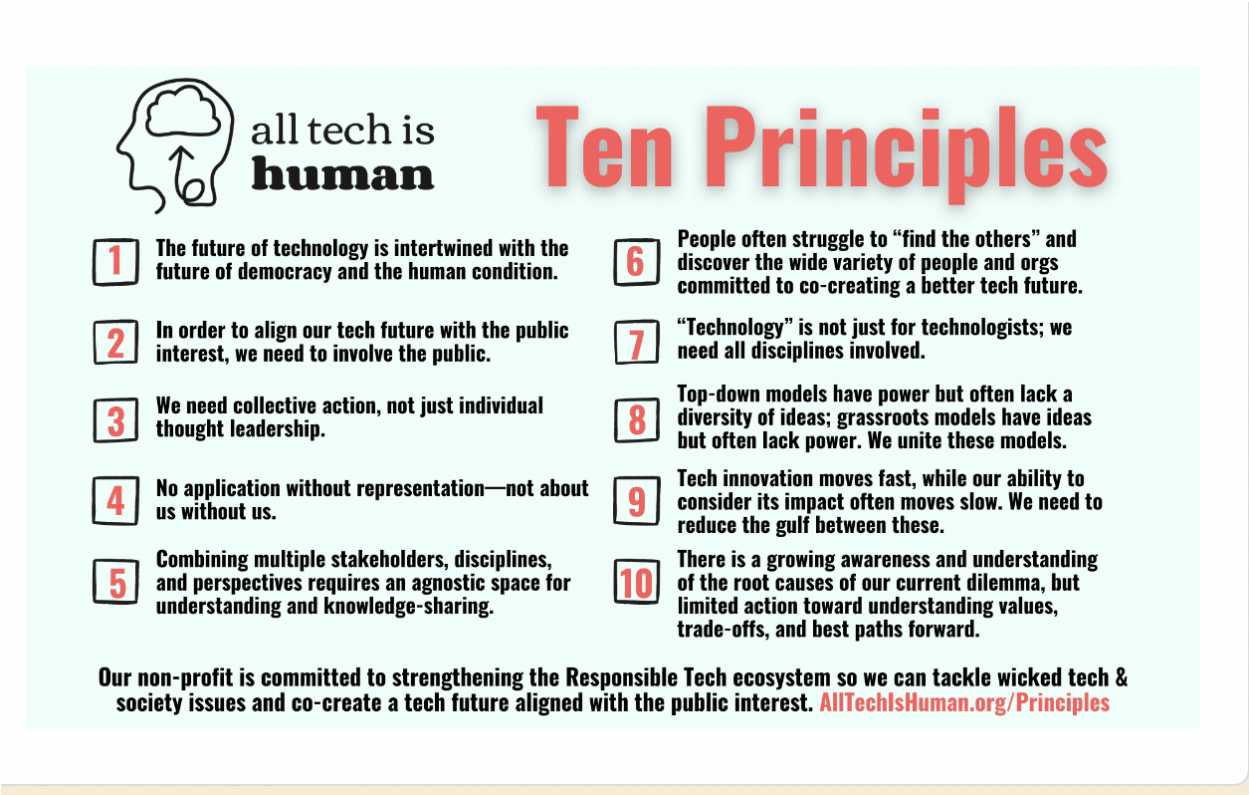
Copyright and artists (class action)
https://www.wired.com/story/artificial-intelligence-copyright-law/?bxid=6276599b28cfb4543b074f74&cndid=69678237&esrc=growl2-regGate-0321&mbid=mbid%3DCRMWIR012019%0A%0A&source=Email_0_EDT_WIR_NEWSLETTER_0_DAILY_ZZ&utm_brand=wired&utm_campaign=aud-dev&utm_content=WIR_Daily_072523&utm_mailing=WIR_Daily_072523&utm_medium=email&utm_source=nl&utm_term=P1

Toolkits:

Responsible Tech tools
https://ethicalexplorer.org/
https://ethicalexplorer.org/introducing-the-ethical-explorer-pack/
From:
https://omidyar.com/omidyar-networks-journey/

Trust in Machine Learning
https://news.mit.edu/2023/strengthening-trust-machine-learning-models-0328

All Tech is Human principles



Why UK should lead the charge to AI regulation
https://www.washingtonpost.com/business/2023/06/16/let-the-british-government-write-ai-s-rules/92dd563a-0bfe-11ee-8132-a84600f3bb9b_story.html

https://corporatedigitalresponsibility.net

Joanna Bryson article on AI regulation
https://journals.sas.ac.uk/amicus/article/view/5612/5286

https://journals.sas.ac.uk/amicus/article/view/5612

AI & ESG
https://www.technologyreview.com/2023/06/20/1074244/climate-vulnerability-digital-divide/?utm_source=engagement_email&utm_medium=email&utm_campaign=wklysun&utm_term=06.25.23.nonsubs_slashprice&utm_content=JA23ISS.SlashPrice.ENG&mc_cid=b0b9f58d7a&mc_eid=d5afc044cc

AI & Catastrophic risks
https://arxiv.org/abs/2306.12001

Open AI for tech (nick clegg)
https://www.ft.com/content/ac3b585a-ce50-43d1-b71d-14dfe6dce999?desktop=true&segmentId=7c8f09b9-9b61-4fbb-9430-9208a9e233c8#myft:notification:daily-email:content

Tony Blair Institute on innovation and the future.
https://www.institute.global/insights/politics-and-governance/new-national-purpose-innovation-can-power-future-britain#conclusion

AI a real worry now not existential crazies :)
https://sloanreview.mit.edu/article/what-me-worry/

Ada Lovelace governance
https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/

EU AI act
https://www.europarl.europa.eu/news/en/press-room/20240212IPR17618/artificial-intelligence-act-committees-confirm-landmark-agreement

World Economic Forum
https://www.weforum.org/events/ai-governance-summit-2023/

Democratising AI alignment - claude
https://www.maginative.com/article/anthropic-explores-democratizing-ai-alignment-using-public-input/

Tech Uk video on future of work
https://www.youtube.com/watch?v=qhAOr8wngUs

Brian's blog on the certificate
https://www.zyen.com/community/the-ethical-ai-thought-space/putting-ethical-ai-into-practice/

Turing on AI & kids
https://www.turing.ac.uk/research/research-projects/children-and-ai

Chartered institute of IT - need professional ethics

https://www.bcs.org/articles-opinion-and-research/living-with-ai-and-emerging-technologies-meeting-ethical-challenges-through-professional-standards/

Digital leaders
https://digileaders.com

AI now - generally
https://ainowinstitute.org/2023-landscape

Safeguards and alignment
https://www.iaps.ai/research/safeguarding-the-safeguards

Tweet from Yann debunking the whole thang
https://twitter.com/ylecun/status/1802679017402757162?cxt=HBwW1MC7xf3As4QyAAAA&cn=ZmxleGlibGVfcmVjcw%3D%3D&refsrc=email

Adeba tweet on synthetic data
https://twitter.com/Abebab/status/1797574190087966858?cxt=HBwWlMLUmfaMovIxAAAA&cn=ZmxleGlibGVfcmVjcw%3D%3D&refsrc=email

Adeba tweet of Bloomberg piece on data centres vs renewable energy
https://twitter.com/Abebab/status/1804877672486404458?cxt=HBwW1IWwucarm4wyAAAA&cn=ZmxleGlibGVfcmVjcw%3D%3D&refsrc=email