

Bibliography

1. Science-Oriented AI

a. Mandatory:

- Castelfranchi, C., 2020, September. For a Science-oriented, Socially Responsible, and Self-aware AI: beyond ethical issues. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)* pp. 1-4. IEEE.
- Turing, A. M., 2022. "Computing Machinery and Intelligence." *Mind*, vol. 59, no. 236, 1950, pp. 433–60. *JSTOR*, <http://www.jstor.org/stable/2251299>. Accessed 19 July.
- Winfield, A.F., Michael, K., Pitt, J. and Evers, V., 2019. Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]. *Proceedings of the IEEE*, 107(3), pp.509-517.

b. Optional:

- Ricci, A., Tummolini, L. and Castelfranchi, C., 2019. Augmented societies with mirror worlds. *Ai & Society*, 34(4), pp.745-752.
- SHERPA Project, 2019. Briefings Released on the Ethical Implications of Smart Information Systems.

2. Introduction to Ethics

a. Mandatory:

- Floridi, L. and Cowls, J., 2022. A unified framework of five principles for AI in society. *Machine Learning and the City: Applications in Architecture and Urban Design*, pp.535-545.
- Shafer-Landau, R., 2020. The Kantian Perspective: Fairness and Justice. In *The fundamentals of ethics* (Fifth Edition). Oxford: Oxford University Press.
- Shafer-Landau, R., 2020. Virtue Ethics. In *The fundamentals of ethics* (Fifth Edition). Oxford: Oxford University Press.
- Shafer-Landau, R., 2021, What is Morality? *Living Ethics. An Introduction with Readings*, II Edition, Oxford: Oxford University Press.
- Shafer-Landau, R., 2021, Consequentialism *Living Ethics. An Introduction with Readings*, II Edition, Oxford: Oxford University Press.

b. Optional:

- Allen, C., Smit, I. and Wallach, W., 2005. Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and information technology*, 7(3), pp.149-155.
- Bostrom, N. and Yudkowsky, E., 2018. The ethics of artificial intelligence. In *Artificial intelligence safety and security*, pp. 57-69. Chapman and Hall/CRC.
- Taddeo, M., 2016. The moral value of information and information ethics. In *The Routledge Handbook of Philosophy of Information*, pp. 377-390. Routledge.

- Shafer-Landau, R., 2020. The Social Contract Tradition. In *The fundamentals of ethics* (Fifth Edition). Oxford: Oxford University Press.

3. Do Artifacts have politics?

a. Mandatory

- Winner, L., 1980. Do artifacts have politics?. In *Modern Technology: Problem or Opportunity?*, (109, 1). MIT Press. pp. 121-136

b. Optional:

- Contissa, G., 2017. Automation and Liability: an analysis in the context of socio-technical systems. In *i-lex*, (11, 1) pp. 15-45.
- Vermaas, P., Kroes, P., van de Poel, I., Franssen, M., Houkes, W. 2011. Ethics and Designing. In: *A Philosophy of Technology. Synthesis Lectures on Engineers, Technology, & Society*. Springer, Cham.

4. Introduction to Ethics and Digital Humanism

a. Mandatory:

- Floridi, L. and Taddeo, M., 2016. What is data ethics? *Philosophical Transactions of the Royal Society A*, 374.

5. Value Alignment

a. Mandatory:

- Russell, S., Dewey, D. and Tegmark, M., 2015. Research priorities for robust and beneficial artificial intelligence. *Ai Magazine*, 36(4), pp.105-114.

b. Optional:

- Rossi, F. and Mattei, N., 2019, July. Building ethically bounded AI. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), pp. 9785-9789.
- Loreggia, A., Mattei, N., Rossi, F. and Venable, K.B., 2018, July. A notion of distance between cp-nets. In *Proceedings of AAMAS*, pp. 955-963.
- Burton, E., Goldsmith, J. and Mattei, N., 2018. How to teach computer ethics through science fiction. *Communications of the ACM*, 61(8), pp.54-64.
- Rossi, F. and Loreggia, A., 2019, May. Preferences and Ethical Priorities: Thinking Fast and Slow in AI. In *Proceedings of AAMAS*, pp. 3-4.

6. AI and Human Rights

a. Mandatory:

- Floridi, L., Cows, J., Beltrametti, M. et al., 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, 28, pp. 689–707.
- European Digital Rights (EDRi), 2021. Open letter: Civil society call for the introduction of red lines in the upcoming European Commission proposal on Artificial Intelligence.

- Council of Europe Commissioner for Human Rights, 2019. Unboxing artificial intelligence: 10 steps to protect human rights.
- Sartor, G., 2017. Human rights and information technologies. *The Oxford Handbook of Law, Regulation and Technology*, pp.424-458.

b. Optional:

- Sen, A., 2017. Elements of a theory of human rights. In *Justice and the capabilities approach*, pp. 221-262. Routledge.
- Internet Rights & Principles Coalition, 2014. The charter of human rights and principles for the internet. Internet Governance Forum, United Nations.

7. Logic Programming

a. Mandatory:

b. Optional:

- Apt, K. R. 2005., The logic programming paradigm and Prolog. In Mitchell, J. C., editor, *Concepts in Programming Languages*, chapter 15, pp. 475–508. Cambridge University Press, Cambridge, UK.
- Baroni, P. and Giacomin, M. 2007. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10) pp.675–700.
- Baroni, P. and Giacomin, M., 2009. Semantics of Abstract Argument Systems, pp. 25–44. Springer US, Boston, MA.
- Borning, A., Maher, M. J., Martindale, A., and Wilson, M., 1989. Constraint hierarchies and logic programming. In Levi, G. and Martelli, M., editors, 6th *International Conference on Logic Programming*, volume 89, pp. 149–164, Lisbon, Portugal. MIT Press.
- Calegari, R., Ciatto, G., Denti, E., and Omicini, A. 2020. Logic-based technologies for intelligent systems: State of the art and perspectives. *Information*, 11(3), pp.1–29.
- Calegari, R., Denti, E., Dovier, A., and Omicini, A. 2018a. Extending logic programming with labelled variables: Model and semantics. *Fundamenta Informaticae*, 161(1-2), pp.53–74.
- Calegari, R., Denti, E., Mariani, S., and Omicini, A. 2018b. Logic programming as a service. *Theory and Practice of Logic Programming*, 18(3-4), pp.1–28.
- Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2), pp.321–357.
- Dyckhoff, R., Herre, H., and Schroeder-Heister, P., 1996. Extensions of Logic Programming, *5th International Workshop, ELP'96*, volume 1050 of LNCS, Leipzig, Germany. Springer.
- Levesque, H. J. 1989. A knowledge-level account of abduction. In *IJCAI*, pp. 1061–1067.
- Omicini, A., Ricci, A., and Viroli, M. 2008. Artifacts in the A&A meta-model for multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 17(3), pp.432–456.
- Poole, D. 1993. Logic programming, abduction and probability. *New Generation Computing*, 11(3–4), p.377.

- Riveret, R., Oren, N., and Sartor, G. 2020. A probabilistic deontic argumentation framework. *International Journal of Approximate Reasoning*, 126, pp.249–271.
- Saptawijaya, A. and Pereira, L. M. 2019. From logic programming to machine ethics. In Bendel, O., editor, *Handbuch Maschinenethik*, pp. 209–227. Springer VS, Wiesbaden.

8. Data Protection

a. Mandatory:

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
- Sartor, G., and Lagioia, F., 2020. Study: The impact of the General Data Protection Regulation on artificial intelligence. European Parliament.

b. Optional:

- Zarsky, T.Z., 2016. Incompatible: The GDPR in the age of big data. *Seton Hall L. Rev.*, 47, pp.995-1020.

9. A Framework for Ethical Principles

a. Mandatory:

- Tsamados, A., Aggarwal, N., Cows, J., Morley, J., Roberts, H., Taddeo, M., and Floridi, L., 2021. The ethics of algorithms: Key problems and solutions. In Floridi, L., editor, *Ethics, Governance, and Policies in Artificial Intelligence*. Springer.

10. Fairness in Automated Decisions

a. Mandatory:

- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A., 2021. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), pp. 3-44.
- Lagioia, F., Rovatti, R., & Sartor, G., 2022. Algorithmic fairness through group parities? The case of COMPAS-SAPMOC. *AI & SOCIETY*, pp. 1-20.

11. Autonomous Vehicles

a. Mandatory:

- Contissa, G., Lagioia, F., & Sartor, G., 2017. The Ethical Knob: ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law*, 25(3), 365-378.
- Nyholm, S., 2018. The ethics of crashes with self-driving cars: A roadmap, II. *Philosophy Compass*, 13(7).
- Nyholm, S., 2018. The ethics of crashes with self-driving cars: A roadmap, I. *Philosophy Compass*, 13(7).

b. Optional:

- Etzioni, A., & Etzioni, O., 2017. Incorporating ethics into artificial intelligence. *The Journal of Ethics*, 21(4), pp. 403-418.
- Gentzel, M., 2020. Classical liberalism, discrimination, and the problem of autonomous cars. *Science and Engineering Ethics*, 26(2), pp. 931-946.
- Nyholm, S., 2018. Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. *Science and engineering ethics*, 24(4), pp. 1201-1219.

12. Intelligent Weapons

a. Mandatory:

- House, P. B., 2015. Autonomous weapons systems: five key human rights issues for consideration. *Amnesty International Publications. London, 28*.
- Arkin, R., 2018. Lethal autonomous systems and the plight of the non-combatant. In *The political economy of robots*, pp. 317-326. Palgrave Macmillan, Cham.
- Arkin, R., et al., 2019. Autonomous Weapon Systems: A Roadmapping Exercise, *Policy workshop organized by Max Tegmark, Emilia Javorsky and Meia Chita-Tegmark*.
- Scharre, P., Horowitz, M. C., 2015. An Introduction to AUTONOMY in WEAPON SYSTEMS, *CNAS Working Papers*.
- Walzer, M., 2015. Just and unjust wars: A moral argument with historical illustrations. Hachette UK, chapters 8,9.

b. Optional:

- Arkin, R., 2009. Governing lethal behavior in autonomous robots. Chapman and Hall/CRC, chapter 10.
- Sartor, G., & Omicini, A., 2016. The autonomy of technological systems and responsibilities for their use. In N. Bhuta, S. Beck, R. Geiß, H. Liu, & C. Kreß (Eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (pp. 39-74). Cambridge: Cambridge University Press.
- Sharkey, N., 2014. Towards a principle for the human supervisory control of robot weapons. *Politica & societa*, 3(2), pp. 305-324.
- Sharkey, A., 2019. Autonomous weapons systems, killer robots and human dignity. *Ethics and Information Technology*, 21(2), pp. 75-87.

13. Ethics of Filtering

a. Mandatory:

- Sartor, G., and Loreggia, A., Study: The impact of algorithms for online content filtering or moderation (“upload filters”). European Parliament, 2020.

14. AI and Unfairness in ToS and PPs

a. Mandatory:

- Drawzeski, K., Galassi, A., Jablonowska, A., Lagioia, F., Lippi, M., Micklitz, H.W., Sartor, G., Tagiuri, G. and Torroni, P., 2021, November. A Corpus for Multilingual Analysis of Online Terms of Service. In *Proceedings of the Natural Legal Language Processing Workshop 2021*, pp. 1-8.

- Contissa, G., Docter, K., Lagioia, F., Lippi, M., Micklitz, H.W., Pałka, P., Sartor, G. and Torroni, P., 2018. Claudette meets gdpr: Automating the evaluation of privacy policies using artificial intelligence. Study Report, Funded by The European Consumer Organisation (BEUC).
- Lippi, M., Pałka, P., Contissa, G., Lagioia, F., Micklitz, H.W., Sartor, G. and Torroni, P., 2019. CLAUDETTE: an automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law*, 27(2), pp.117-139.
- Ruggeri, F., Lagioia, F., Lippi, M. and Torroni, P., 2022. Detecting and explaining unfairness in consumer contracts through memory networks. *Artificial Intelligence and Law*, 30(1), pp.59-92.
- Lippi, M., Contissa, G., Jablonowska, A., Lagioia, F., Micklitz, H.W., Palka, P., Sartor, G. and Torroni, P., 2020. The force awakens: artificial intelligence for consumer law. *Journal of artificial intelligence research*, 67, pp.169-190.

b. Optional:

- Lagioia, F., Jabłonowska, A., Liepina, R. and Drazewski, K., 2022. AI in Search of Unfairness in Consumer Contracts: The Terms of Service Landscape. *Journal of Consumer Policy*, pp.1-56.