Editorial: Governance AI ethics

Rebekah Rousi[1]*, Pertti Saariluoma[2] and Mika Nieminen[3]

[1]School of Marketing and Communication, University of Vaasa, Vaasa, Finland, [2]Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland, [3]VTT Technical Research Centre of Finland Ltd, Espoo, Finland

Editorial on the Research Topic
Governance AI ethics

The Special Issue (SI) on "*Governance AI Ethics*" highlights the urgency for systematic consideration of ethically-driven governance models, standards and protocols of artificial intelligent (AI) based technology in society. Featured authors represent a range of disciplines from computer science and computer science to communication studies, social ethics, cultural studies and sociology. Discourse and developments are moving from Society 4.0 (and Industry 4.0), of information and smart tech, to Society 5.0, or intelligent systems, re-formulation and transformation of societies and human roles as a whole. Society 5.0 demands the need to translate, design and implement a blueprint for humanity that will foster basic human values with integrity.

Transitions in the job market are observed (Flynn et al., 2017; Kovacs, 2018) that not only impact lower-skilled and mundane, repetitive work, but also higher-level positions such in administration (i.e., lawyers etc.). AI has long intervened with roles requiring high degrees of accuracy and calculation such as aircraft piloting (McManus and Goodrich, 1989; Bordenkircher, 2020), medical science (Ramesh et al., 2004; Yeasmin, 2019) including the biopharmaceutical industry (Mökander et al.). It should be no surprise that governance could be enabled and facilitated *via* the very technology it is designed to govern (Sharma et al., 2020; Nitzberg and Zysman, 2021; Leikas et al.). Even complex areas of security, cyber or otherwise, will be in the hands or algorithms of the technology itself (Li, 2018; Zhang et al., 2021).

Matters of accountability and responsibility require that researchers and policy-makers connect governance to human sciences. From a cognitive scientific perspective, intentionality (consciousness) plays a crucial role when tackling issues of accountability in moments of crisis (Rousi). Moreover, notions of technology are fluid as the digital is blurred with the physical, and there is no natural way of separating human beings from the technology they create. Interestingly, the pace at which technology is evolving in many ways exceeds that of culture (Murphie and Potts, 2017). AI changes the nature of human-technology relationships. Not only does it change and automate utilitarian processes, but it also transforms social processes. Industrialism brought people to cities to live near factories (urbanization). Its innovations opened pathways to new parts of the globalized economy. Intelligent technology will have consequences of a similar scale that may even witness forms of post-urbanization. It is important to consider these changes holistically. The focus of changes is not in improving technical artifacts but rather the quality of human life.

Applications and use scenarios for AI are already vast. From scenes of AI-enabled cognitive enhancement technology (see e.g., Rousi and Renko, 2020) to stream-lined migration processes (Molnar, 2019) and predictive healthcare (Bohr and Memarzadeh, 2020), each area brings sets of considerations that will require specific forms of framework to ensure the pillars of the transformative practices' reliability (Kurtz and Schrank, 2007; Spremic, 2017; Nwaiwu, 2018). Reliability, consistency and contingency plans (risk management and mitigation) are pre-requisites for human trust (Saariluoma et al., 2018). Particularly human trust in human-made systems. Thus, governance models are imperative for the progress of AI development, implementation and adoption in human societies (Mäntymäki et al., 2022; Viljanen and Parviainen). These in turn, should be seen as part of the technological design itself— from ideation to systemic regulation—any form of AI-based technology should incorporate the broader social, cultural and ethical fabric from the ground up (Bryson and Winfield, 2017; Gasser and Almeida, 2017; Vakkuri et al., 2020).

The main goal of this SI was to outline the holistic nature of AI, its design and other related considerations within efforts of governance and ethical governance model development. The first article to feature in this special issue of *Governance AI Ethics* is Mika Viljanen and Henni Parviainen's, "AI Applications and Regulation: Mapping the Regulatory Strata". This article adopts a critical stance toward classifying AI as immature from a regulatory perspective. The authors argue that AI in its various applications already exist across a broad spectrum of regulation. Already many rules have been established to regulate and guide AI in its development, implementation and sustained use. Two semi-fictional case studies are used to illustrate their argument.

Hallamaa and Kallikoski's article, "AI Ethics as Applied Ethics" uses bioethics to illustrate a weighing of metaethical and methodological approaches utilized in AI ethics. Moreover, the authors stress the need to embed AI ethics within the realm of applied ethics through drawing on other domains (i.e., safety research and impact assessment) to solidify theory within actionable contexts. Leikas et al. present a study on the co-development of a large-scale AI project, AuroraAI, that is aimed at providing Finnish citizens with tailored and timely pubic services. The case is used to highlight practical challenges in utilizing AI within administration. Similarly, Sigfrids et al. present a systematic review of literature concerning AI governance for public administration. As a result, they propose the Comprehensive, Inclusive, Institutionalized and Actionable framework (CIIA) as a comprehensive AI governance model. In "Governance of Responsible AI" Gianni et al. investigate the potential of translating ethical guidelines to cooperative policies. They draw our attention toward discourse on the roles and limitations of AI and how theory can be implemented in practice.

Cañas takes a human-AI collaborative perspective to discussing ethics. Supervision is seen as the driver in delineating shared responsibilities and is the lens through which performance is measured pertaining to human and AI actors. In "With clear intention", Rebekah Rousi proposes a responsibility model for robot governance that not only looks at AI or robotics as co-workers, but autonomous learning agents in their own right. The article problematizes accountability relationships in a time when humans do not program AI, and the choice between wrong and right is within the autonomous artifact and system executing action. Finally, from the biopharmaceutical industry our special issue features a study of COVID-19 vaccine AstraZenica's organizational AI governance model in which Mökander et al. present the challenges and progress of their organizational case.

Ethics as a field, is concerned with what is good or right, what is justified and legal, and what should be done (e.g., Monahan and Loftus, 1982; Kent, 2000; Bunge, 2012). Ethics has a pivotal position in well-guided social changes. Social changes often mean alterations in ethical rules (Rousi, 2021). Not only may sub-conscious human values and biases materialize in algorithmic logic, but AI as a tool (or weapon) may be used with ill-will and malicious intention (Brundage et al., 2018; Horowitz, 2019; Pantserev, 2020). Governance is a type of human action that provides parameters and protocol (Humphrey and Schmitz, 2001). Designing AI actions for governance includes many open ethical issues. This SI shed light on these issues and evolving developments to address these for future years to come.

Author contributions

RR was the main author of this Editorial. The structure and significant contributions were provided by PS, with additional insight provided by MN. All authors contributed to the article and approved the submitted version.

Acknowledgments

Conflict of interest

MN was employed by VTT Technical Research Centre of Finland Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Bohr, A., and Memarzadeh, K. (2020). "The rise of artificial intelligence in healthcare applications," in *Artificial Intelligence in healthcare*. Cambridge, MA: Academic Press. p. 25–60. doi: 10.1016/B978-0-12-818438-7.00002-2

Bordenkircher, B. A. (2020). The Unintended Consequences of Automation and Artificial Intelligence: Are Pilots Losing Their Edge? in *Issues Aviation Law and Policy*. Chicago, IL: International Aviation Law Institute p. 19.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B.,., and Amodei, D. (2018). The malicious use of artificial intelligence: forecasting, prevention, and mitigation. *arXiv*. Available online at: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Bryson, J., and Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*. 50, 116–119. doi: 10.1109/MC.2017.154

Bunge, M. (2012). *Treatise on Basic Philosophy: Ethics: The Good and The Right*. Berlin, Germany: Springer Science and Business Media.

Flynn, J., Dance, S., and Schaefer, D. (2017). Industry 4.0 and its potential impact on employment demographics in the UK. *Adv. Prod. Eng. Manag.* 6, 239–244. Available online at: https://core.ac.uk/download/pdf/131168407.pdf

Gasser, U., and Almeida, V. A. (2017). A layered model for AI governance. *IEEE Int. Comput.* 21, 58–62. doi: 10.1109/MIC.2017.4180835

Horowitz, M. C. (2019). When speed kills: Lethal autonomous weapon systems, deterrence and stability. *J. Strategic Stud.* 42, 764–788. doi: 10.1080/01402390.2019.1621174

Humphrey, J., and Schmitz, H. (2001). Governance in global value chains. *IDS Bull.* 32, 19–29. doi: 10.1111/j.1759-5436.2001.mp32003003.x

Kent, G. (2000). *Ethical Principles. Research Training for Social Scientists*. London: Sage. p. 61–7 doi: 10.4135/9780857028051.d11

Kovacs, O. (2018). The dark corners of industry 4.0–Grounding economic governance 2.0. *Technol. Soc.* 55, 140–145. doi: 10.1016/j.techsoc.2018.07.009

Kurtz, M. J., and Schrank, A. (2007). Growth and governance: models, measures, and mechanisms. *J. Polit.* 69, 538–554. doi: 10.1111/j.1468-2508.2007.00549.x

Li, J. H. (2018). Cyber security meets artificial intelligence: a survey. *Front. Inf. Technol. Electron.* 19, 1462–1474. doi: 10.1631/FITEE.1800573

Mäntymäki, M., Minkkinen, M., Birkstedt, T., and Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*. 1–7. doi: 10.1007/s43681-022-00143-x

McManus, J., and Goodrich, K. (1989). Application of artificial intelligence (AI) programming techniques to tactical guidance for fighter aircraft. *In Guidance, Navigation and Control Conference (p. 3525)*. doi: 10.2514/6.1989-3525

Molnar, P. (2019). Technology on the margins: AI and global migration management from a human rights perspective. *Camb. Int. Law J.* 8, 305–330. doi: 10.4337/cilj.2019.02.07

Monahan, J., and Loftus, E. F. (1982). The psychology of law. *Annu. Rev. Psychol.* 33, 441–475. doi: 10.1146/annurev.ps.33.020182.002301

Murphie, A., and Potts, J. (2017). *Culture and Technology*. London: Macmillan International Higher Education.

Nitzberg, M., and Zysman, J. (2021). Algorithms, data, and platforms: the diverse challenges of governing AI. *J. Eur. Public Policy*. 26. doi: 10.2139/ssrn.3802088

Nwaiwu, F. (2018). Review and comparison of conceptual frameworks on digital business transformation. *J. Competitiveness*. doi: 10.7441/joc.2018.03.06

Pantserev, K. A. (2020). "The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability," in *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*. Cham: Springer. p. 37–55. doi: 10.1007/978-3-030-35746-7_3

Ramesh, A. N., Kambhampati, C., Monson, J. R., and Drew, P. J. (2004). Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* 86, 334. doi: 10.1308/147870804290

Rousi, R. (2021). "Ethical stance and evolving technosexual culture–a case for human-computer interaction," in *International Conference on Human-Computer Interaction*. Cham: Springer. p. 295–310. doi: 10.1007/978-3-030-77431-8_19

Rousi, R., and Renko, R. (2020). Emotions toward cognitive enhancement technologies and the body–attitudes and willingness to use. *Int. J. Hum. Comput. Stud.* 143, 102472. doi: 10.1016/j.ijhcs.2020.102472

Saariluoma, P., Karvonen, H., and Rousi, R. (2018). "Techno-trust and rational trust in technology–A conceptual investigation," in *IFIP Working Conference on Human Work Interaction Design*. Cham: Springer. p. 283–293. doi: 10.1007/978-3-030-05297-3_19

Sharma, G. D., Yadav, A., and Chopra, R. (2020). Artificial intelligence and effective governance: a review, critique and research agenda. *Sustainable Fut.* 2, 100004. doi: 10.1016/j.sftr.2019.100004

Spremic, M. (2017). Governing digital technology–how mature IT governance can help in digital transformation?. *Int. J. Econ. Manag[[Inline Image]]*. 2. Available online at: https://www.iaras.org/iaras/filedownloads/ijems/2017/007-0029(2017).pdf

Vakkuri, V., Kemell, K. K., Kultanen, J., and Abrahamsson, P. (2020). The current state of industrial practice in artificial intelligence ethics. *IEEE Software*. 37, 50–57. doi: 10.1109/MS.2020.2985621

Yeasmin, S. (2019). "Benefits of artificial intelligence in medicine," in *2019 2nd International Conference on Computer Applications and Information Security (ICCAIS)*. IEEE. p. 1–6. doi: 10.1109/CAIS.2019.8769557

Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J.,., and Choo, K. K. R. (2021). Artificial intelligence in cyber security: research advances, challenges, and opportunities. *Artif. Int. Rev.* 55, 1–25. doi: 10.1007/s10462-021-09976-0