Unit 1 - Ethics in Computing in the age of Generative AI:

Artificial Intelligence (AI) has emerged as a fast-growing research domain, revolutionising industries and society. However, research on the ethical dimensions of AI has struggled to keep pace with advancements in the field. To address this gap, Correa et al. (2023) conducted an extensive analysis of 200 AI governance documents from 37 countries across six continents. Their study sheds light on the diversity of perspectives on AI ethics but also highlights critical limitations and challenges that must be addressed. This reflection expands on their findings, incorporates additional insights and outlines actionable steps to foster a globally coherent yet locally adaptable framework for AI ethics.

Correa et al. (2023) highlight that the analysed documents come from diverse cultural and professional contexts, originating in equal proportions from governmental bodies and companies, where corporate contributions might prioritise profit-driven considerations over public interest. For instance, industry actors often advocate for AI guardrails that align with their business models, potentially skewing the ethical discourse. Moreover, the study points out an underrepresentation of documents from China and India, countries that are important to AI development through their significant contributions in research and talent. The timeline of the documents analysed is also noteworthy. A majority were published between 2017 and 2019, a period marked by heightened interest in data ethics following high-profile incidents such as the Cambridge Analytica scandal in 2018. This event exposed the misuse of data-driven algorithms in political processes, spurring global demand for robust ethical guidelines in AI. Despite this surge in interest, the study reveals a major limitation: only 2% of the analysed documents emphasise practical implementation of

ethical principles, while most discussions on AI ethics remain theoretical, focusing on normative principles rather than actionable frameworks. Fjeld et al. (2020) similarly underscore a significant gap between the articulation of normative principles and their application in real-world scenarios. They argue that while many governance documents define principles such as transparency, accountability, and fairness, they often fail to provide mechanisms for operationalising these values in AI systems. For instance, voluntary compliance, which is a common feature in government documents, as noted by Correa et al. (2023) accounts for 91.6% of the analysed policies. This reliance on non-binding guidelines undermines the enforcement of ethical standards, leaving room for inconsistencies and potential misuse.

Despite these challenges, there are promising signs of progress in AI governance. Zhang et al. (2022) report that the number of AI-related laws increased from just one in 2016 to 18 in 2021, reflecting growing legislative attention to the field. However, they caution that only a minority of these laws address AI security, leaving critical gaps in safeguarding against misuse and ensuring system reliability. This trend underscores the need for a more comprehensive regulatory approach that prioritises both ethical principles and technical robustness. One of the most significant barriers to global AI governance is the divergence in ethical priorities across regions. Correa et al. (2023) note that Europe and North America largely converge on the five normative principles they perceive as most important, including transparency and accountability. In contrast, Asia emphasises different principles, reflecting variations in cultural and societal values. Furthermore, even when similar principles are identified, their definitions often differ across regions. For example, the concept of privacy may be interpreted more stringently in Europe under the General Data

Protection Regulation (GDPR) than in other parts of the world (European Parliament, 2016). These discrepancies complicate efforts to establish a unified global standard for AI ethics, as they require reconciling not only diverse priorities but also differing interpretations of shared values.

To address these challenges, several steps are necessary. Building on initiatives such as the OECD AI Principles (OECD, 2024) and UNESCO's Recommendation on the Ethics of AI (UNESCO, 2021), a global framework should define baseline standards for ethical AI development. This framework must prioritise universal values such as transparency, accountability, and inclusivity while allowing for regional adaptations to reflect local cultural and legal contexts. As noted by Jobin et al. (2019), global alignment on core ethical principles can foster trust and cooperation among stakeholders. Correa et al. (2023) emphasise the need for tools to catalogue and compare AI governance documents. A centralised repository would facilitate knowledge-sharing, enabling policymakers, researchers, and industry stakeholders to identify areas of convergence and divergence. Such a resource could also highlight best practices for practical implementation, bridging the gap between normative principles and real-world applications. Regional initiatives, such as the European Union's AI Act (European Parliament, 2024) or the African Union's AI blueprint (Smart Africa, 2021), should be encouraged to address local challenges while aligning with global standards. Partnerships among governments, academia, industry, and civil society are essential to ensure diverse perspectives are incorporated into AI governance. As argued by Mittelstadt et al. (2016), inclusive collaboration can help mitigate biases and ensure that ethical frameworks are equitable and representative. To move beyond voluntary compliance, it is crucial to

develop binding regulations that enforce ethical principles. For example, accountability mechanisms could include third-party audits of AI systems, mandatory reporting of algorithmic decisions, and penalties for non-compliance. Binding laws would ensure that ethical standards are consistently applied, reducing the risk of harm from AI technologies.

These actions provide various benefit. Harmonising global standards would reduce regulatory fragmentation, enabling smoother cross-border collaboration and trade (Financial Stability Board, 2024). However, achieving this requires reconciling conflicting national laws and addressing geopolitical tensions which might be challenging. Binding regulations could also provide clarity on liability issues, ensuring accountability for AI-related harm. A globally coherent framework would promote equity and inclusivity by ensuring that marginalised voices are represented in AI governance (Cheong, 2024). Transparent and accountable AI systems can also build public trust, addressing societal fears about misuse and bias. However, effective communication and education are necessary to ensure public understanding of AI's benefits and risks (Ateeq et al., 2024). Computing professionals must navigate ethical dilemmas, balancing innovation with societal risks. Adherence to global standards would provide clear benchmarks for ethical decision-making. Furthermore, professionals must stay informed about evolving regulations, emphasising the importance of continuous learning and adaptability.

Concluding, the generative AI revolution presents unprecedented opportunities and challenges, requiring coordinated global efforts to address its ethical, legal, and social implications. By establishing a global framework, creating a centralised

repository, and promoting regional collaborations, the complexities of AI governance can be managed (OECD, 2024; UNESCO, 2021). The proposed approach balances the need for global coherence with respect for local autonomy, ensuring that AI technologies are developed and deployed responsibly (Mittelstadt et al., 2016). Achieving global convergence on AI guardrails is essential to harnessing the transformative potential of AI while safeguarding societal well-being (Correa et al., 2023).

List of References:

Ateeq, A., Milhem, M., Alzoraiki, M., Dawwas, M. I. F., Ali, S. A. and Al Astal, A. Y. (2024) 'The impact of AI as a mediator on effective communication: enhancing interaction in the digital age', *Frontiers in Human Dynamics* 6. Available at: https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2024.1467384/full

Cheong, B. C. (2024) 'Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making', *Frontiers in Human Dynamics* 6. Available at: https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2024.1421273/full

Correa, N. et al. (2023) 'Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance', *Patterns* 4 (10). Available at: https://doi.org/10.1016/j.patter.2023.100857

European Parliament (2024) AI Act. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (Accessed: 28 January 2025)

European Parliament (2016) General Data Protection Regulation. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng (Accessed: 28 January 2025)

Financial Stability Board (2024) Recommendations to Promote Alignment and Interoperability Across Data Frameworks Related to Cross-border Payments. Available at: https://www.fsb.org/uploads/P121224-1.pdf (Accessed: 28 January 2025)

Fjeld, J., et al. (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Research Publication: 2020 - 2021*. Available at: https://cyber.harvard.edu/publication2020/principled-ai (Accessed: 28 January 2025)

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence* 1 (9), pp. 389-399. Available at: https://doi.org/10.1038/s42256-019-0088-2

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) 'The Ethics of Algorithms: Mapping the Debate', *Big Data & Society* 3(2), pp. 1-21. Available at: https://doi.org/10.1177/2053951716679679

OECD (2024) AI principles. Available at: https://www.oecd.org/en/topics/ai-principles.html#:~:text=The%20OECD%20AI%20Principles%20are%20the%20first%20intergovernmental,AI%20that%20respects%20human%20rights%20and%20democratic%20values. (Accessed: 28 January 2025)

Smart Africa (2021) Blueprint Artificial Intelligence for Africa. Available at: https://smartafrica.org/wp-content/uploads/2023/11/70029-eng_ai-for-africa-blueprint-min.pdf (Accessed: 28 January 2025)

UNESCO (2021) UNESCO Recommendation on the ethics of Artificial Intelligence. Available at: https://unesco.org.uk/resources/unesco-recommendation-on-the-ethics-of-artificial-intelligence-key-facts#:~:text=The%20Recommendation%20emphasises%20who%20should%20be%20in%20control,and%20ensure%20they%20contribute%20to%20the%20public%20good. (Accessed: 28 January 2025)

Zhang, B. et al. (2021) The AI index 2021 annual report, *arXiv preprint*. Available at: https://arxiv.org/abs/2103.06312