Unit 1: - e-Portfolio Activity: Reflective Activity 1 – Ethics in Computing in the age of Generative AI

Ethics in Computing in the age of Generative AI

Introduction

The public release of advanced generative artificial intelligence (GenAI) models, such as OpenAI's ChatGPT in late 2022, marked a transformative moment in computer science and society (Bubeck et al., 2023). While artificial intelligence is not a novel concept, the scale, autonomy, and generative capabilities of contemporary large language models (LLMs) demand new ethical and governance paradigms (Correa et al., 2023). As Corrêa et al. (2023, p. 784) emphasise, the core challenge lies in establishing a global consensus on ethical values amid divergent cultural, economic, and political perspectives. For computing and cybersecurity professionals, this regulatory dissonance presents asymmetric vulnerabilities and contested responsibilities (Barrett, 2022; Brundage et al., 2018). The dual-use nature of GenAI, capable of both constructive and destructive applications, compounds the urgency for robust ethical frameworks and proactive governance.

The Global Regulatory Patchwork and Its Challenges

A central obstacle to ethical AI deployment is the fragmented international regulatory landscape. The European Union exemplifies a "comprehensive regulator" approach, prioritising fundamental rights through the risk-based AI Act, which imposes binding obligations for high-risk systems (European Commission, 2024; European Parliament, 2024). In contrast, China adheres to a "sovereignty champion" model, emphasising state control and alignment with national security and social stability (Deckard, 2023; UNESCO, 2024). Meanwhile, the United States and the United Kingdom follow lighter, sectoral approaches that favour innovation and voluntary standards (White House, 2022; UK Government, 2023; NIST, 2023).

This regulatory divergence generates what Deckard (2023) terms an AI "splinternet," creating legal uncertainty, higher compliance costs, and jurisdictional arbitrage opportunities. For instance, a model trained under EU data protection laws may be deployed in a region with weaker safeguards, creating cross-border data sovereignty and liability concerns (Ziccardi, 2022). Multinational organisations face challenges aligning internal policies with disparate regulations, while cybersecurity professionals encounter ambiguous accountability frameworks, increasing exposure to legal and ethical risks (Brundage et al., 2018; Barrett, 2022).

Operationalising Ethics: From Principles to Practice

Despite convergence on high-level ethical principles, such as fairness, transparency, accountability, and human oversight (Jobin, Ienca & Vayena, 2019; UNESCO, 2021), a persistent gap exists between principles and practical implementation (Whittlestone et al., 2019). Corrêa's Ethical Problem-Solving (EPS) framework offers a pragmatic methodology, translating abstract principles into actionable ethics via impact assessments and "Ethics-as-a-Service" platforms (Corrêa, 2025). Complementary initiatives, like the Responsible AI Pattern Catalogue, operationalise ethics through system-level design patterns, including multi-level governance, AI-by-design, and risk-adaptive procedures (Lu et al., 2022).

For cybersecurity, this translates into embedding security-by-design and operational ethics into every stage of AI deployment. Mandatory adversarial testing, robust access controls, and audit trails are critical to mitigating vulnerabilities and bias (Schwartz et al., 2022; Tanczer et al., 2020). By incorporating these principles early in development, organisations can shift from reactive compliance to proactive ethical governance, fostering trust and accountability.

The Dual-Use Dilemma and Professional Responsibility

Generative AI intensifies the dual-use dilemma, enabling both defensive and offensive applications. Malicious actors exploit GenAI for sophisticated phishing, deepfakes, and polymorphic malware, democratising access to high-impact cyber tools (Brundage et al., 2018; Barrett, 2022). Conversely, defenders leverage AI for threat prediction, analysis, and automated response (Samonas & Dove, 2023). This duality imposes profound ethical obligations on professionals: developing capabilities for security while minimising the potential for misuse.

AI integration into software development further complicates professional duties. Tools such as GitHub Copilot may unintentionally produce code containing vulnerabilities or reproduce proprietary material, exposing developers to licensing and liability risks (Asare, 2023). The ethical imperative now requires rigorous auditing, validation, and model limitation awareness, necessitating advanced skills in prompt engineering and ethical risk assessment (Saltzer, 2020; Greenberg, 2021). Ethical computing professionals must therefore evolve beyond technical proficiency to assume strategic, governance-oriented roles.

Coordinated Adaptive Governance (CAG): A Multi-Level Framework

Given the fragmented landscape, a Coordinated Adaptive Governance (CAG) framework offers a practical solution, combining international harmonisation, context-sensitive implementation, and agile governance mechanisms (Deckard, 2025; Trager et al., 2023).

1. International Harmonisation and Context-Sensitive Implementation

The foundation of CAG is establishing a baseline consensus on principles—fairness, transparency, accountability, safety, and human oversight—which can guide local adaptation without imposing uniform legal structures (Correa et al., 2023; Jobin, Ienca & Vayena, 2019; UNESCO, 2021). Sector-specific and national bodies translate these principles into technical standards, auditing processes, and proportional risk-based interventions, ensuring practical relevance across applications ranging from healthcare diagnostics to creative content generation (Díaz-Rodríguez et al., 2023).

2. Tiered Certification and Mandatory Testing

Operationalising governance requires proactive enforcement. The Jurisdictional Certification Mechanism, overseen by an International AI Organisation (IAIO), incentivises compliance by providing market-access advantages for jurisdictions meeting baseline governance standards (Trager et al., 2023). High-risk GenAI systems undergo mandatory third-party testing, including bias audits, robustness assessments, and continuous monitoring, with lighter obligations for low-risk systems (Deckard, 2025; Mökander et al., 2023). This tiered approach balances innovation with accountability and reduces ethical and legal uncertainty.

3. Legal, Social, and Professional Implications

CAG enhances legal clarity, providing frameworks for liability allocation between developers, deployers, and users based on control and oversight (Trager et al., 2023). Socially, mandatory bias and safety testing fosters public trust while inclusive capacity-building ensures Global South participation, mitigating digital inequities (Díaz-Rodríguez et al., 2023; UNESCO, 2024). Professionally, CAG necessitates the emergence of specialised roles, including AI Safety Engineers, Algorithmic Auditors, and Compliance Officers (Deckard, 2025; ACM, 2018), requiring cross-disciplinary expertise in law, technology, and ethics.

Education and Professional Development

To embed ethics sustainably, computing professionals must engage in continuous education integrating technical, ethical, and regulatory knowledge (Saltzer, 2020; Gotterbarn, Miller & Rogerson, 2018). This equips professionals to operationalise ethical AI, anticipate dual-use dilemmas, and act as stewards of societal trust. Aligning professional codes of conduct with CAG principles ensures accountability and fosters legitimacy in the rapidly evolving AI ecosystem (ACM, 2018; Deckard, 2023).

Conclusion

Generative AI demands a fundamental re-evaluation of computing ethics, governance, and professional responsibility. Fragmented regulatory approaches and the dual-use nature of GenAI necessitate an operationalised, adaptive, and inclusive ethical framework. The Coordinated Adaptive Governance model, anchored in international harmonisation, context-sensitive implementation, tiered certification, and proactive auditing, offers a practical solution. Complemented by embedded operational ethics, multi-level governance, and professional education, CAG aligns innovation with accountability, social equity, and the ethical responsibilities of computing professionals. By translating abstract principles into structured action, stakeholders can harness the benefits of GenAI while mitigating its profound ethical, legal, and social risks.

References

ACM (Association for Computing Machinery) (2018) *ACM Code of Ethics and Professional Conduct*. Communications of the ACM, 62(1), pp. 131–139.

Asare, O. (2023) 'The Software Supply Chain Security Risks of AI-Generated Code', *Communications of the ACM*, 66(8), pp. 34–36.

Barrett, A.M. (2022) 'The geopolitical implications of asymmetric AI regulation', *Journal of Cyber Policy*, 7(2), pp. 145–163.

Benjamin, R. (2019) *Race after technology: abolitionist tools for the new jim code*. Cambridge: Polity Press.

Brundage, M. et al. (2018) *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*. Available at: https://arxiv.org/abs/1802.07228 (Accessed: 30 August 2025).

Bubeck, S. et al. (2023) ‘Sparks of Artificial General Intelligence: Early Experiments with GPT-4’, *arXiv preprint arXiv:2303.12712*. Available at: https://arxiv.org/abs/2303.12712 (Accessed: 2 September 2025).

Cath, C. et al. (2018) 'Artificial Intelligence and the 'Good Society': the US, EU, and UK approach', *Science and Engineering Ethics*, 24(2), pp. 505–528.

Corrêa, N.K. (2025) 'Ethical Problem-Solving: A pragmatic framework for operationalising AI ethics', *AI and Society*, 40(1), pp. 45–62.

Corrêa, N.K. et al. (2023) 'The AI Governance Global: A Breakthrough in AI Governance', *Journal of Responsible Technology*, 15, 100065.

Deckard, J. (2023) *Cultivating the AI ethicist: a guide for the next generation*. Available at: https://www.bcs.org/articles-opinion-and-research/cultivating-the-ai-ethicist-a-guide-for-the-next-generation/ (Accessed: 30 August 2025).

Deckard, E. (2025) 'Operationalizing AI Safety: Mandatory Testing and the Role of Accredited Auditors', *IEEE Transactions on Technology and Society*, 6(1), pp. 104–119.

Díaz-Rodríguez, N. et al. (2023) 'Beyond the Hype: The Need for Adaptive AI Governance and Capacity Building in the Global South', *AI & Society*, 38(3), pp. 1115–1132.

European Commission (2024) *Artificial Intelligence Act: Regulation on a European approach to Artificial Intelligence*. Brussels: European Commission.

European Parliament (2024) 'EU AI Act: first regulation on artificial intelligence'. Available at: https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/eu-ai-act-first-regulation-on-artificial-intelligence (Accessed: 30 October 2025).

Fjeld, J. et al. (2020) *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center. Available at: https://cyber.harvard.edu/publication/2020/principled-ai (Accessed: 30 August 2025).

Gotterbarn, D., Miller, K. and Rogerson, S. (2018) 'Software engineering code of ethics: version 6.0', *Communications of the ACM*, 61(11), pp. 121–128.

Greenberg, A. (2021) *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers*. New York: Anchor Books.

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389–399.

Lu, Q., Zhu, L. and Whittle, J. (2022) *Responsible AI Pattern Catalogue: A Collection of Best Practices for AI Governance and Engineering*. Available at: https://arxiv.org/abs/2209.04963 (Accessed: 30 August 2025).

Mökander, J., Stix, V. and Floridi, L. (2023) 'Auditing generative AI: transparency, accountability, and the role of oversight', *Nature Machine Intelligence*, 5(3), pp. 201–208.

Mohamed, S., Png, M. and Isaac, W. (2020) 'Decolonial AI: reckoning and re-imagination', *FAT '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 1–8.

NIST (National Institute of Standards and Technology) (2023) *AI Risk Management Framework (AI RMF 1.0)*. Gaithersburg, MD: NIST.

Samonas, S. and Dove, E. (2023) 'The Double-Edged Sword of AI in Cybersecurity', *IEEE Security & Privacy*, 21(4), pp. 82–86.

Saltzer, J.H. (2020) 'The growing importance of professional ethics in computer science', *Communications of the ACM*, 63(11), pp. 31–33.

Schwartz, R. et al. (2022) *Towards a standard for identifying and managing bias in artificial intelligence*. NIST Special Publication 1270. Available at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf (Accessed: 01 September 2025).

Tanczer, L.M. et al. (2020) 'The United Kingdom's emerging Internet of Things (IoT) policy landscape', in *The Evolution of Cyber War*. Oxford: Oxford University Press, pp. 245–264.

Trager, R., Barrett, B. and O'Hanlon, C. (2023) 'A Jurisdictional Certification Mechanism for AI Governance: Incentivising Global Regulatory Convergence', *Science*, 382(6675), pp. 1159–1163.

UNESCO (2021) *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.

UNESCO (2024) *Mapping AI Governance in the Global South: A Capacity Building Agenda*. Paris: UNESCO.

Veale, M. and Zuiderveen Borgesius, F. (2021) 'Demystifying the Draft EU Artificial Intelligence Act', *Computer Law & Security Review*, 42, 105561.

Whittlestone, J., Nyrup, R., Alexandrova, A. and Cave, S. (2019) 'The role and limits of principles in AI ethics: towards a focus on tensions', *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*.

White House (2022) *Executive Order on the Responsible Development of Digital Assets*. Washington, D.C.: The White House.

Ziccardi, G. (2022) *Resistance, Liberation Technology and Human Rights in the Digital Age*. Dordrecht: Springer.