Transcript

Task 'Individual Reflection Presentation on a Tutor-Specified Question'

Unit 12 Part B

Evaluating the Future of Computing – Ethical and Security Challenges

Slide 1: Introduction and Scope (2 mins)

Slide Content:

- AI: a powerful tool for enterprise innovation
- Risks: Bias, privacy, Security, and accountability
- Technical foundations: risks, governance, real-world cases, and best practices

AI is transforming many industries helping automate processes, enabling cost-effective prediction, and drives innovation. Though the use of AI also brings ethical and Security concerns. Today, I will critically explore how enterprises can balance AI innovation with ethical and Security responsibilities, moving from technical to the strategic level.

The challenge lies in balancing the use of innovation with reducing bias, improving privacy protection, and ensuring accountability. Relying solely on technical measures is not enough. A combination of top-down governance, best practices, and learning from experience is crucial to ensure the controlled and responsible use of new technologies.

Slide 2: Technical Foundations of AI (2 mins)

Slide Content:

- AI types: machine Learning, deep Learning, generative AI
- Needs: Data, computer power, cloud/edge infrastructure
- Challenges: Drift, overfitting, adversarial attacks

While AI is utilized primarily in machine learning, encompassing supervised and unsupervised methods, as well as reinforcement learning. In addition, deep learning, which is used in models like GPT and DALL-E and relies on large neural networks that are trained on immense datasets and while these systems maybe generalize well, they can be vulnerable to model drift, overfitting, and adversarial inputs.

From a technical and environmental perspective, there are issues around computing and power consumption. The use of data for general purposes can be enough. This is when it comes to more critical decisions. Furthermore, new vulnerabilities are introduced by different AI models, which must be addressed.

(Pelekis, Koutroubas, Blika, 2025)

Slide 3: Real-World Applications (2 mins)

Slide Content:

- Finance: Credit scoring, fraud detection
- Healthcare: Diagnostics, patient Triage

- Retail: Recommendations, personalization
- Cybersecurity: Intrusion detection and prevention

AI innovation is being utilized in many sectors, such as finance, where it helps automate credit risk analysis and detect fraud. In healthcare, AI aids in diagnosis and clinical conclusions. In the retail sector, AI is used for shopping personalization, and in cybersecurity, teams leverage anomaly detection to identify and stop breaches. However, these benefits often have ethical and Security trade-offs.

There is significant potential in advanced technology that helps improve many practices, quality of life for individuals and positively influences society. Nevertheless, the use of such advancements must be evaluated from multiple perspectives, with logic and critical thinking.

(Leenes, Martin, 2021)

Slide 4: Ethical and Security Risks (2 mins)

Slide Content:

- Ethical: Bias, transparency, privacy breaches
- Security: Data poisoning, adversarial attacks
- Dual use: Deepfakes, automated hacking

A significant concern in AI systems is bias that arises from models trained on unvalidated data. This could lead to reinforcement discrimination. In addition, black-

box models reduce transparency and limit accountability. From a security perspective, AI could be susceptible to poisoning attacks, model theft, and adversarial inputs that manipulate outputs, AI tools can even be used maliciously, such as for creating deepfakes or automating cyberattacks.

Bias comes in different forms and can cause damage, including discrimination, loss of reputation and data. Organizations must understand the technologies they use, including the sources of their components, code, vulnerabilities and design elements.

(Balasubramanian, Liyana, Sankaran, 2025)

Slide 5: Governance and Regulatory Frameworks (2 mins)

Slide Content:

- GDPR: Data rights, automated decisions
- EU AI Act: Risk-based classifications
- NIST AI RMF: Risk management guidance
- Organizational models: Ethics boards, audits

The GDPR provides individuals which are referred to as "Data Subjects", with progressive rights over decisions regarding their personal data. The EU AI Act classifies AI systems by risk, banning unacceptable uses and regulating high-risk systems. The US NIST AI framework helps enterprises evaluate and control AI risks. Internally, organizations may implement AI ethics boards, oversight committees, and continuous audit practices.

Lawmakers and the industry are taking active action to regulate fast-evolving technology. To avoid loss of revenue, reputation and penalties for compliance breaches, organizations must follow up and keep risk management, privacy and ethical practices up to date and continuously evolve with trends and regulations.

(Tabassi, 2023; Bolgouras, Zarras, Leka, 2025)

Slide 6: Case Studies (2 mins)

Slide Content:

- Estonia: Ethics not consistently embedded in public AI
- Healthcare hackathon: Gaps in explainability and consent
- Cybersecurity: IoT systems vulnerable to AI attacks
- Defense: Governance embedded in autonomous platforms

Some states have realized the social, ethical, and privacy risks AI brings. Students at a healthcare hackathon struggled with EU ethical AI guidelines, for instance, issues related to explainability and informed consent. AI-powered IoT solutions bring improvements in cybersecurity, although they have vulnerabilities such as poisoning. Ethics must be embedded in autonomous systems, with layered governance and human oversight.

Case studies have highlighted the ethical and privacy challenges associated with emerging technologies such as AI. Some political unions, for instance the EU, have imposed strict regulations, namely the GDPR, which threaten big tech companies.

However, despite the benefits of using AI, there is a need for proactive and innovative strategies to prevent data exposure and protect against threat actors.

(Hinton 2023; Roberson, Bornstein, Liivoja, Ng, Scholz, Devitt, 2022; Pourzolfaghar, Alfano, Helfert, 2023; Yang, He, Wang, Qu, Zhang, 2023)

Slide 7: Balancing Innovation with Responsibility (2 mins)

Slide Content:

- Ethics and Security by Design
- Human oversight (HITL)
- Transparency and monitoring
- Regulatory alignment

To ensure balanced innovation and ethical integrity, organizations must embed privacy and security by design thinking into their company processes. For instance, human-in-the-loop (HITL) oversight can be utilized to help ensure that critical decisions remain accountable where regular audits and monitoring help manage drift and Bias, and governance must evolve in response to regulatory updates and stakeholder expectations.

As demonstrated by some of the case studies, it is challenging to control and manage the negative effects of AI. More importantly, it is even more challenging to determine who is accountable in cases of bias that leads to discrimination, intellectual property violations and privacy breaches. Uncontrolled use of AI brings many uncertainties, which makes effective risk management a challenge.

(Ricciardi, and Zomaya, 2025)

Slide 8: Conclusion (2 mins)

Slide Content:

- AI is an opportunity and a responsibility
- Risks require proactive governance
- A multi-layered approach is key

AI can help drive enormous value, but it also presents challenges, such as ethical concerns, security threats, and regulatory constraints, which must be addressed very early. A layered governance model spanning technical, organizational, and legal aspects could offer a practical way forward.

Senior leadership is ultimately accountable for negligence relating to violations. Failure to fulfil duties can have a negative impact on the business due to penalties imposed by laws, regulations and compliance requirements. To prevent such occurrences, a combination of well-tailored strategic, tactical and operational measures must be put in place. These measures should be adopted to suit each organization's business, culture and tolerance level.

(Papagiannidis, Mikalef, Conboy, 2025; Malka, 2025)

References

- Balasubramanian, P., Liyana, S., and Sankaran, H. (2025). Generative AI for cyber threat intelligence: Applications, challenges, and analysis of real-world case studies. *Artif Intell* Rev, 58, 336. Available at: https://doi.org/10.1007/s10462-025-11338-z [Accessed 5 October 2025].
- Bolgouras, V., Zarras, A., and Leka, C. (n.d.). The EU regulatory ecosystem for ethical AI. *AI Ethics* 5, 5063–5080 (2025). Available at: https://doi.org/10.1007/s43681-025-00749-x [Accessed 7 October 2025].
- Hinton, C., 2023. The state of ethical AI in practice: A multiple case study of Estonian public service organizations. International Journal of Technoethics, 14(1), pp.1–17. Available at: https://www.igi-global.com/pdf.aspx?tid=322017andptid=310218andctid=4andoa=trueandisxn=9781668479568 [Accessed 8 October 2025].
- Leenes, R., Martin, A. (2021). Technology and regulation (2020). Open Press TiU. Available at: https://jstor.org/stable/community.34023115 [Accessed 5 October 2025].
- Malka, A. (2025). T Governance Framework Analysis. University of **********. Available at: ********** [Accessed 16 October 2025].
- Papagiannidis, E., Mikalef, P. and Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), p.101885.
- Pelekis, S., Koutroubas, T., Blika, A. Adversarial machine learning: a review of methods, tools, and critical industry sectors. Artificial Intelligence Review, Rev

58, 226 (2025). Available at: https://doi.org/10.1007/s10462-025-11147-4 [Accessed 6 October 2025].

- Pourzolfaghar, Z., Alfano, M. and Helfert, M., (2023). Application of ethical AI requirements to an AI solution use case in the healthcare domain. *American Journal of Business*, 38(3), pp.112–128.
- Ricciardi Celsi, L. and Zomaya, A.Y., (2025). Perspectives on Managing AI Ethics in the Digital Age. Information, 16(4), p.318.
- Roberson, T., Bornstein, S., Liivoja, R., Ng, S., Scholz, J., and Devitt, K. (2022). A method for ethical AI in defence: A case study on developing trustworthy autonomous systems. Journal of Responsible Technology, 11, p.100036. Available at: https://doi.org/10.1016/j.jrt.2022.100036 [Accessed 6 October 2025].
- Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0). NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD Available at: https://doi.org/10.6028/NIST.AI.100-1 [Accessed 7 October 2025].
- Xiang, Q., Zi, L., Cong, X., and Wang, Y. (2023). Concept drift adaptation methods under the deep learning framework: A literature review. Applied Sciences, 13(11), p.6515. Available at: https://doi.org/10.3390/app13116515 [Accessed 5 October 2025].
- Yang, R., He, H., Wang, Y., Qu, Y. and Zhang, W., (2023). Dependable federated learning for IoT intrusion detection against poisoning attacks. Computers and Security, 132, p.103381.

This document has been written solely for educational purposes. All references, names, and trademarks mentioned here remain the property of their respective owners and are used here strictly for the educational context. Grammarly was used exclusively for proofreading and enhancing the clarity and language of the text. ChatGPT was consulted for general research. All academic writing, analysis, argumentation, and conclusions are entirely the original work of the author.