Stef Abolade

230985072

Submission Date: 10/05/2024

Word Count: 3141

ECS7025P - Ethics, Regulation and Law in Advanced Digital Information Processing and Decision Making

Data Science and Artificial Intelligence

Individual assignment

Table of Contents

[List of abbreviations 2](#_Toc166263197)

[Introduction 2](#_Toc166263198)

[Section A: Reflective Account 2](#_Toc166263199)

[Section B: Case Study Analysis of AI/Data Ethics Breach 4](#_Toc166263200)

[Section C: Proposal for an Ethical Framework 6](#_Toc166263201)

[References 8](#_Toc166263202)

[Appendix 8](#_Toc166263203)

## List of abbreviations

AI - Artificial Intelligence

MSc - Master of Science

HART - Harm Assessment Risk Tool

OECD - Organisation for Economic Co-operation and Development

UNESCO - United Nations Educational, Scientific and Cultural Organization

GDPR - General Data Protection Regulation

## Introduction

In this report, we begin with Section A, where I reflect on the profound growth in my understanding of AI ethics, influenced by frameworks and module insights. In Section B, I present a case study analysis of an unethical AI application, examining it through various ethical lenses and stakeholder perspectives to offer recommendations. Finally, in Section C, I propose a tailored ethical framework for AI use in organisations, outlining its structure, implementation strategies, and the roles of continuous training and monitoring in fostering an ethically responsible AI environment.

## Section A: Reflective Account

Introduction

Reflecting on my journey through this MSc, I recognise how my understanding of artificial intelligence (AI) and ethics has changed. Before beginning this module in the Ethics, Regulation and Law in Advanced Digital Information Processing and Decision Making, my perspective was quite naive. I had understood ethics in a broad sense, moral principles like stealing is wrong. I hadn't grasped the ways these principles merge with AI technology and its impact on society. My background in a biological lab set the foundation of my ethical understanding within the scope of 'morals.' I worked in a company developing a robotic consumer product. This experience stimulated my curiosity and marked the beginning of a shift in my comprehension of AI ethics.

Initial Understanding and Expectations

My main goal entering this module was to bridge the gaps in my understanding of how AI, an autonomous and objective tool, links with ethical issues. This objective came from my experience in the industry, where the ethical dimensions of AI were often oversimplified or overlooked. My initial thoughts about AI ethics were centred around avoiding harm in a general sense, without a deep understanding of the ethical dilemmas presented by AI technologies.

Transformative Moments and Insights

The course discussions, particularly those from the first week, were eye-opening. Geoffrey Hinton's warning about the "Existential Threat" of AI and Eric Horvitz's message in "AI and the Future of Human Flourishing" presented two contrasting yet complementary views. These discussions brought key questions about AI's integration into our future and the influences shaping its learning processes.

A transformative moment for me was the realisation that there are currently no comprehensive laws governing AI. This was concerning, given the multitude of potential risks AI poses. This sparked my interest in how ethical frameworks can act as supplements for legal structures, guiding the development and deployment of AI in the absence of formal legislation. Among the various frameworks we studied, the UNESCO framework stood out to me. It helped bridge my understanding of how terms like sustainability, transparency, and safety are interconnected with ethics in AI. This provided a structured approach to linking theoretical principles with practical implications.

Ethical Theories and Practical Integration

Regarding ethical theories, I find that no single theory fits every scenario in AI. However, Consequentialism, which focuses on the consequences of actions, seems like an ideal compromise, though it is subject to biases based on differing views of what constitutes a positive outcome. Realising this it shows the importance of a balanced approach, considering multiple ethical perspectives to navigate the complexities of AI development. In my future role as an AI practitioner possibly as an AI engineer, I aim to integrate these insights into the development of AI models. My goal is to ensure that the AI systems we create are not only fair but also adhere to the ethical guidelines and legislation established by my company. I plan to use my knowledge to offer fresh insights and contribute to shaping policies that prioritise ethical considerations in AI development.

Challenges and Personal Development

One of the main challenges I anticipate in applying these ethical principles as touched on earlier is the subjectivity of human perspectives and the presence of unconscious biases. This introduces complications in aligning AI development with ethical ideals as well as maintaining vigilance over ethical ideals and the practical realities of AI development. Recognising these challenges has made me more aware of the need to continuously examine our biases and strive for objectivity in AI applications.

This module has broadened my perspective, revealing my own limitations and ethical considerations I need to think about. I now feel better prepared to identify potential ethical breaches and to advocate for ethical practices in AI development, having learned to spot the red flags that might signal underlying issues.

Future Aspirations and Global Perspectives

My career aspirations have evolved through this module; I am now driven to be part of or establish an organisation that prioritises people over products. I envision a role where I can say that my contributions make a positive difference in the world, ensuring that ethical considerations are at the forefront of AI development. This ambition is fuelled by the knowledge I've gained, which has reshaped my approach to job hunting.

Being originally from Nigeria, I realised the importance of considering diverse global perspectives in AI ethics. The module encouraged us to reflect on our countries of origin and research the current situations there. This approach was a reminder that ethical considerations in AI must include the differences unique to each area, furthering our understanding that our perspectives should not be confined to our immediate environments.

Learning about international conversations and initiatives like the G20 AI Principles has been enlightening. These global efforts have increased awareness which is vital for creating a more inclusive and ethically aware AI community.

Diverse Stakeholder Views

The varying objectives and perspectives of different stakeholders from developers and policymakers to users demonstrate that ethics relate differently to each stakeholder in the AI ecosystem. This diversity affects ethical decision-making in AI, with each stakeholder group bringing its unique set of concerns and priorities.

For instance, developers may prioritise innovation and efficiency, sometimes at the expense of transparency or fairness, while policymakers might focus on establishing guidelines that protect users without stifling technological advancements. Users often seek transparency and fairness but may not fully understand the complexities involved in achieving these in AI systems.

Conclusion

This module has significantly enhanced my understanding of AI ethics. From my initial perspective, I have a more nuanced and comprehensive view that considers a range of ethical theories, frameworks, and stakeholder perspectives. My future endeavours in AI will be deeply influenced by this enriched understanding, guiding me to contribute to the development of AI technologies that are not only innovative but also ethically responsible and aligned with global ethical standards.

I am committed to being a proactive advocate for ethical AI, using the insights gained from this module to inform my practices and policies, and to encourage a culture of ethical awareness in every AI project I engage with. This approach will be crucial in navigating the complex ethical landscape of AI, ensuring that we leverage this powerful technology to enhance human flourishing while mitigating its potential risks.

## Section B: Case Study Analysis of AI/Data Ethics Breach

Introduction

In the age of advanced computing and data-driven decision-making, the intersection of artificial intelligence (AI) and ethical considerations has become increasingly important, especially in the field of law enforcement. A case showing this intersection is the Harm Assessment Risk Tool (HART), developed by Durham Constabulary in collaboration with the University of Cambridge. HART was designed to predict the risk of individuals reoffending and influencing police decisions about possible interventions. While HART aims to enhance resource allocation and decision-making efficiency, its deployment raises substantial ethical concerns, particularly regarding transparency, fairness, and potential bias. This analysis explores the multifaceted ethical issues surrounding HART, how the tool aligns with or deviates from established ethical guidelines and regulatory frameworks and examines various stakeholder perspectives.

The HART Case Overview

HART is an algorithmic risk-assessment tool used by UK police to assess reoffending risk (Oswald et al., 2018). Developed with machine learning, particularly the random forest algorithm, HART categorises individuals into risk groups based on data like age, gender, criminal history, and geography. Its goal is to improve decision-making by predicting future offenses and suggesting preventative interventions (Oswald et al., 2018). However, HART's application has sparked intense debate over using algorithmic predictions in criminal justice. Concerns focus on issues of transparency, as the algorithms are not open for public scrutiny or external evaluation, raising questions about accountability and fairness.

Ethical Guidelines and Regulatory Frameworks

The deployment of HART breached several ethical principles, including:

* The algorithms behind HART were not accessible to the public, making it hard for third parties to evaluate their fairness or identify biases (Oswald et al., 2018). This lack of transparency directly challenges the Organisation for Economic Co-operation and Development (OECD) AI Principles, which advocates for clear and responsible disclosure of AI methodologies. HART’s algorithms prevent independent validation undermining the tool's accountability.
* HART's use of demographic data and criminal history raised concerns about perpetuating existing biases, potentially leading to discriminatory outcomes. This conflicts with UNESCO's Recommendations on the Ethics of AI, which emphasises AI systems' need to promote fairness and avoid bias (Pedro et al., 2019).
* The individuals assessed by HART had limited options to challenge or understand their assessments due to the system's inscrutability. This contradicts the Asilomar AI Principles, which highlight the importance of human oversight and the right for individuals to appeal against automated decisions.

Regulatory Frameworks

* Data Protection Act 2018:

This act emphasises individuals' rights over their data, requiring transparency and consent for data usage. HART's operation, which involves extensive personal data without clear consent, suggests non-compliance and the need for enhanced privacy notices and explicit consent mechanisms.

* Equality Act 2010:
* This act prevents discrimination based on protected characteristics. If HART's algorithms lead to biased outcomes, it risks indirect discrimination under the act. Ensuring HART’s fairness is vital to comply with the Equality Act 2010, necessitating a review of data inputs and algorithmic decision-making to eliminate biases.

The UK's regulatory frameworks reveal gaps in AI policing. There's no comprehensive framework for the ethical use of AI in policing, and existing laws don't mandate full explainability of AI systems (Metcalfe et al., 2021). This inhibits individuals’ ability to challenge AI decisions. The lack of regulations on AI decision-making processes and the evolving definitions of fairness under the Equality Act 2010 require updated guidelines. Balancing AI benefits with ethical considerations remains a fundamental challenge, needing a blend of legal, technical, and ethical knowledge. AI development spanning multiple countries requires international cooperation for consistent ethical standards, currently lacking. These gaps highlight the need for more specific regulations and international standards.

Stakeholder Perspectives

The deployment of HART shows diverse perspectives on AI's ethical use in law enforcement.

Individuals assessed by HART face potential stigmatisation due to unclear, algorithm-driven judgments. The lack of transparency and appeal routes prevents them from challenging assessments, creating a demand for clearer explanations and accessible appeal processes.

Police using HART to improve efficiency face ethical complexities. Balancing algorithmic insights with professional judgment without fully understanding the algorithms calls for more transparent AI systems to support informed decisions and maintain community trust.

Policy makers and legal experts work to balance creating legislation against AI misuse while supporting innovation. They bridge the gap between technological advances and legal frameworks, advocating for responsible AI deployment in policing with mechanisms to correct biases.

Developers creating algorithms are at the forefront of ensuring these tools are both effective and ethically sound. They address biases and incorporate ethical considerations, including stakeholder feedback to refine AI tools and meet ethical standards.

The public and community groups, especially those representing marginalised communities are influenced by the deployment of tools like HART. Likely to be skeptical of AI in law enforcement, especially regarding surveillance and potential discrimination. They push for strict ethical standards and community-involved decision-making to ensure AI's fairness and benefits for all.

Strategies for Fostering Collaboration Among Stakeholders

Collaboration is vital to address AI concerns like HART. Strategies include:

Establish Ethical Oversight Committees with representatives from all stakeholder groups to oversee AI projects' ethical aspects from design to implementation.

Facilitating Transparent Conversations to demystify AI operations and foster mutual understanding.

Implementing Co-Operative Design by involving end-users, especially from vulnerable groups, in design and testing for fair solutions.

Educating law enforcement and policymakers on AI capabilities and limitations to enhance decision-making.

Developing Standardised Ethical Frameworks to guide AI development and usage, ensuring consistency and fairness.

Ethical Decision-Making in AI Development

The development of HART highlights the need for ethical considerations throughout the AI development process, ensuring systems are efficient and ethically sound. Developers must balance ethical considerations with technological advancements, focusing on fairness and transparency. Incorporating stakeholder feedback and implementing rigorous testing are key to maintaining ethical standards and enhancing AI development.

Conclusion

The case of exemplifies the ethical challenges in deploying AI in law enforcement. Adhering to ethical guidelines and involving stakeholders ensures effective and equitable AI implementations. As AI evolves, maintaining ethical practices is crucial to ensuring these technologies are used transparently and beneficially. Learning from HART can guide future AI developments toward more responsible and ethical outcomes, with developers playing a crucial role in balancing innovation with ethical considerations to enhance public safety without compromising standards. This commitment, combined with stakeholder engagement and oversight, is essential to harness AI's benefits while minimising risks.Top of Form

## Section C: Proposal for an Ethical Framework

Introduction

As artificial intelligence (AI) continues to evolve, the need for ethical frameworks and ethics committees has become necessary, especially in the context of safeguarding marginalised groups often underrepresented in data sets. Our commitment to ethical AI aims to prevent these biases, ensuring fairness and equity in all our technological applications.

Committee Objective and Framework Proposal Rationale

The primary objective of the ethics committee is to guide the responsible development and use of AI technologies within the organisation. This committee will act as the overseer to ensure that AI systems are designed and operated in a way that upholds ethical principles promoting responsible AI use across all operations.

Importance of Ethical Framework and Committee:

1. Preventing Bias: Marginalized groups are often underrepresented in AI training sets, leading to biases in AI applications. Our ethical framework aims to identify and mitigate these biases, ensuring AI fairness.
2. Stakeholder Trust: By demonstrating a commitment to ethical AI, we foster trust among users and stakeholders, demonstrating a commitment to ethical standards and transparency.
3. Aligning with Core Values: Developing an ethical framework aligns with the organisation’s commitment to social responsibility and inclusivity. It ensures that AI technologies are not just efficient but are also supports our broader goal of unbiassed service, ensuring our technologies serve diverse communities effectively.
4. Regulatory Compliance: With increasing scrutiny from regulatory bodies, having a proactive ethical framework helps the organisation stay ahead of legal requirements and avoid penalties related to unethical AI use. Forming this committee is a proactive step to prevent issues within our organisation.

Customised Ethical AI Framework Development

Core Principles and Values:

1. Fairness and Equity: We commit to developing AI systems that provide equitable outcomes for all.
2. Transparency and Explainability: We will ensure that decisions made by AI systems are transparent and explainable. Stakeholders should be able to understand how and why decisions are made.
3. Accountability: We hold ourselves accountable for the impact of our AI systems. This means implementing mechanisms to monitor outcomes and taking corrective actions if our systems cause harm.
4. Privacy and Data Protection: Safeguarding user data, especially sensitive information is key. We will adhere to the highest standards of data protection, ensuring that data used to train AI systems is handled responsibly.
5. Inclusivity and Diversity: Our AI solutions will be developed with input from diverse groups. This approach ensures that a wide range of perspectives is considered, leading to more inclusive AI systems.

These principles are chosen because they directly address the core areas where AI has the potential to impact ethical issues, particularly bias. By focusing on fairness, transparency, accountability, privacy, and inclusivity, we can develop AI technologies that are not only innovative but also equitable and supportive of all users.

Ethical Impact Assessment Development

To ensure our AI initiatives align with our ethical framework, we will develop an ethical impact assessment process.

* Assess how AI models and datasets handle biases, with tools to identify and rectify biases related to race and other sensitive attributes.
* Include perspectives from marginalized groups to understand potential impacts and concerns.
* Evaluate the clarity and accessibility of explanations provided by AI systems regarding their decision-making processes.
* Check compliance with privacy laws and guidelines, ensuring that data protection measures are robust.
* Use metrics that specifically measure fairness and equity, alongside traditional performance indicators.

Implementation of the Assessment Framework:

The framework will be applied at multiple stages of the AI development lifecycle, from initial design to post-deployment. It will involve regular reviews and updates based on feedback from stakeholders and continuous monitoring of AI systems. To integrate the ethical framework into existing and future AI initiatives, the organisation will:

Incorporate the ethical impact assessment as a standard part of the development and review process for all AI projects.

Revise internal policies and documentation to reflect the ethical framework principles, ensuring that all team members are aware of these standards.

Ensure that the ethical framework is consistent with current regulatory requirements, such as the GDPR in Europe and other relevant local laws, enhancing compliance and reducing regulatory risks.

Training, Monitoring, and Continuous Improvement

To ensure the effective implementation of our ethics committee and framework, we will:

* Develop comprehensive training for all AI team members on the ethical framework, with specific modules focusing on identifying and mitigating biases.
* Regular workshops and seminars will be held to keep the team updated on the latest ethical AI practices and regulatory developments.
* Establish a monitoring system to regularly assess AI systems against the ethical framework, with a mechanism for reporting issues and non-compliance.
* Use a dashboard to track ethical performance metrics, providing visibility into how AI systems are performing concerning ethical standards.
* Create a feedback loop involving stakeholders, including marginalized racial groups, to continually refine AI systems and the ethical framework.
* Regularly review and update the ethical framework based on new research, changes in societal values, and stakeholder feedback to ensure it remains relevant and effective.

Conclusion

By establishing this ethical framework and ethics committee, we are committing to the responsible use of AI ensuring that AI technologies are developed and used in a manner that is transparent, fair, inclusive, and aligned with the organisation’s core values. By prioritising these ethical considerations, the organization can harness the benefits of AI while minimising risks, particularly those affecting marginalised groups. This commitment to ethical AI development, combined with proactive stakeholder engagement and thorough oversight, will position the organisation as a leader in responsible AI use, this is not just about fulfilling ethical obligations it's about leading by example in the development and use of AI for the good of all people.

## References

Maclure, J. (2021) ‘Ai, explainability and public reason: The argument from the limitations of the human mind’, *Minds and Machines*, 31(3), pp. 421–438. doi:10.1007/s11023-021-09570-x.

Oswald, M. *et al.* (2018) ‘Algorithmic risk assessment policing models: Lessons from the durham hart model and “experimental” proportionality’, *Information &amp; Communications Technology Law*, 27(2), pp. 223–250. doi:10.1080/13600834.2018.1458455.

Pedro, F., Subosa, M., Rivas, A. and Valverde, P. (2019). ‘Artificial intelligence in education: Challenges and opportunities for sustainable development.’

## Appendix

A white sheet with blue text

Description automatically generated