Rail Delivery Group

National Rail

ARTIFICIAL INTELLIGENCE (AI) ETHICAL STANDARDS

CONTENTS

EXECUTIVE SUMMARY

Artificial Intelligence (AI) is revolutionising various sectors including rail, providing remarkable new capabilities and efficiencies that were once unimaginable. However, this power brings a significant responsibility to ensure AI is used ethically and responsibly. It is critical to prevent the misuse of AI, as well as to guard against discrimination and breaches of privacy.

This document serves as a comprehensive guide to AI Ethical Standards, with a particular focus on data privacy. One of the core principles outlined is the strict prohibition against using Personally Identifiable Information (PII) in data collection processes. PII includes any data that could potentially identify a specific individual, such as names, addresses, and social security numbers. The use of such data poses significant privacy risks and ethical concerns.

In addition to the prohibition on collecting PII, these standards extend to the processing and analysis stages of data handling. During data processing, it is essential to implement robust methods to anonymise or de-identify data, ensuring that individual identities cannot be reconstructed. This involves techniques such as data masking, encryption, and aggregation to prevent any potential re-identification of individuals within the data sets.

When it comes to data analysis, it is equally important to adhere to ethical guidelines. Analysts must ensure that the data used in AI models does not inadvertently reintroduce PII or other sensitive information. This requires continuous monitoring and auditing of data sets to identify and mitigate any privacy risks. Moreover, analytical methods should be designed to minimise biases and prevent discriminatory outcomes, thereby ensuring that AI systems are fair and equitable.

It is also mandated that all data must be thoroughly sanitised before being ingested into AI systems or Large Language Models (LLMs). Data sanitisation involves the process of removing or obfuscating any sensitive information to ensure that privacy is maintained. Additionally, sensitive information encompasses business confidential information, which includes trade secrets, proprietary data, and any internal business information that could harm the organisation if disclosed.

This step is crucial to prevent any unintended exposure of personal or business-sensitive information and to uphold the ethical standards of AI usage. By thoroughly sanitising data, we can ensure that AI systems operate responsibly, protecting both individual privacy and the confidentiality of business information. This practice is essential for maintaining trust, ensuring compliance with legal and ethical standards, and supporting the sustainable and fair development of AI technologies.

Furthermore, ensuring fairness in AI systems is critical. By eliminating the use of PII and sanitising data, we can help prevent biases and discriminatory practices that may arise from improperly handled data. This commitment to fairness is fundamental to the ethical deployment of AI.

Any output from AI systems must always be analysed and reviewed before publishing to ensure that AI biases do not exist in the output. This step is essential to maintain fairness and accuracy in AI-driven decisions and predictions. By carefully examining the results generated by AI, we can identify and mitigate any potential biases that may have been introduced during data processing or model training. This thorough review process helps to uphold ethical standards, ensuring that the AI outputs are reliable, unbiased, and equitable before they are made public or used in decision-making.

Alan Cain

Chief Information Security Officer

Lastly, maintaining the integrity of AI systems is paramount. By following these ethical standards, we ensure that AI technologies operate in a manner that is consistent with our values and principles. This not only enhances the credibility of AI systems but also promotes their sustainable and responsible development.

In conclusion, using AI responsibly is not just a technical need but a moral duty. By following the AI Ethical Standards described in this document, especially regarding data privacy and cleaning, we can use AI in a way that is both creative and responsible. This method will help us create a future where AI helps everyone while protecting people's basic rights and privacy.

Alan Cain
Chief Information Security Officer

AI ETHICAL STANDARDS

1. TRANSPARENCY AND ACCOUNTABILITY

1.1. Transparency in the use of AI, ML and LLM's is essential. Organisations must clearly document and communicate the methodologies, datasets, and algorithms used in AI systems. This ensures that stakeholders understand how decisions are made and can trust the system's integrity.

1.2. Implementation: Develop detailed documentation for each AI project, including data sources, model architecture, and decision-making processes. Establish clear communication channels for stakeholders to understand AI operations. Use tools and platforms that support explainable AI (XAI) to make the decision-making process transparent. Regularly update documentation to reflect any changes in the AI system.

2. DATA PRIVACY AND SECURITY

2.1. AI systems must prioritise the protection of personal data. No Personally Identifiable Information (PII) should be utilised in data collection, processing, and analysis processes. Ensuring data privacy is not just a legal obligation but a fundamental ethical requirement.

2.2. Implementation: Before any data is ingested into AI systems, it must undergo a thorough sanitisation process to remove or anonymise PII. Employ techniques such as data masking, encryption, and pseudonymisation. Conduct regular audits to ensure compliance with data privacy standards. Implement data governance frameworks that define clear policies for data handling and access control. Use privacy-preserving technologies like differential privacy to further enhance data protection.

3. FAIRNESS AND NON-DISCRIMINATION

3.1. AI technologies must be used in such a way to promote fairness and avoid discrimination. Algorithms should be scrutinised for biases that could lead to unfair treatment of individuals or groups. Fairness involves ensuring that AI systems do not perpetuate existing biases or create new forms of discrimination.

3.2. Implementation: Implement bias detection and mitigation techniques throughout the AI lifecycle. Use diverse datasets to train models and perform regular fairness assessments. Establish review boards to monitor and address any issues of bias or discrimination. Incorporate fairness metrics into the development and evaluation process of AI models. Regularly engage with diverse stakeholder groups to understand the impact of AI systems on different communities and make necessary adjustments.

Alan Cain
Chief Information Security Officer

4. SAFETY AND RELIABILITY

4.1. AI systems must be safe and reliable, performing as intended without causing harm. This involves rigorous testing and validation to ensure accuracy and robustness. Safety also includes the ability to handle unexpected inputs or situations gracefully.

4.2. Implementation: Conduct extensive testing in varied scenarios to validate AI performance. Implement fail-safes and redundancy measures to handle unexpected failures. Regularly update systems based on feedback and new findings. Establish robust monitoring and logging mechanisms to detect and respond to issues in real-time. Perform security assessments to identify and mitigate potential vulnerabilities.

5. HUMAN OVERSIGHT

5.1. Despite advancements in AI, human oversight remains crucial. Decisions made by AI should be transparent and explainable, allowing for human intervention when necessary. All decisions/actions taken by AI must be logged and wherever possible reversible. Human oversight ensures accountability and helps mitigate risks associated with automated decision-making.

5.2. Implementation: Develop interfaces that enable human operators to monitor and intervene in AI operations. Ensure AI decisions are explainable through clear, understandable logic and documentation. Create protocols for human-in-the-loop (HITL) systems where humans can review and override AI decisions. Train staff to understand AI systems and their limitations, ensuring they are prepared to intervene effectively when needed. Develop decision making guard-rails that pause the processing of any decision that does not appear to be normal behaviour, until after human intervention.

6. ENVIRONMENTAL SUSTAINABILITY

6.1. The development and deployment of AI systems should consider environmental impacts. Efforts should be made to minimise the carbon footprint and promote sustainability. This involves choosing energy-efficient technologies and practices that reduce environmental harm.

6.2. Implementation: Opt for energy-efficient hardware and algorithms. Implement measures to reduce energy consumption and promote the recycling of AI infrastructure components. Regularly assess and optimise the environmental impact of AI projects. Use cloud-based solutions that rely on renewable energy sources where possible. Develop AI models that are computationally efficient to reduce energy usage during training and deployment.

7. ETHICAL USE OF AI

7.1. AI should be used for beneficial purposes and not to cause harm. Ethical guidelines should be established to govern the deployment and use of AI technologies. This involves considering the broader social implications of AI systems and striving to use AI for the greater good.

7.2. Implementation: Develop a code of ethics for AI use within the organisation. Conduct regular ethics training for employees and stakeholders. Establish an ethics committee to review and guide AI projects. Encourage a culture of ethical awareness and responsibility. Regularly review AI applications to ensure they align with ethical guidelines and make necessary adjustments based on feedback and evolving ethical standards.

8. COMPLIANCE WITH LEGAL AND REGULATORY STANDARDS

8.1 Systems must comply with all relevant legal and regulatory requirements, including data protection laws and industry-specific regulations. Compliance ensures that AI practices are lawful and adhere to established standards.

8.2 Implementation: Stay informed about legal developments and ensure AI practices comply with current laws. Engage with legal experts to review AI systems and practices. Maintain thorough records of compliance efforts. Develop internal policies and procedures to ensure ongoing compliance with legal requirements. Conduct regular audits and assessments to identify and address potential compliance issues.

8.3 Non-Compliance

All staff are required to adhere to the guidance detailed within this standard. Staff should be aware that use of AI in a way that is deemed to breach these guidelines, may lead to disciplinary action under the RDG's disciplinary procedure. Serious breaches of these guidelines may constitute gross misconduct and may lead to action under the disciplinary procedure up to and including dismissal.

CONCLUSION

Adhering to these AI Ethical Standards is critical for fostering trust, ensuring fairness, and safeguarding the privacy and rights of individuals. The prohibition of PII in data collection and the mandate for data sanitisation are fundamental to achieving these objectives. By implementing these standards, organisations can harness the benefits of AI responsibly and ethically.

Version	Date	Author	Comments
V0.1	01/03/2024	Alan Cain	Initial draft
V0.1.1	30/05/2024	Mark Hemsley	Initial reviewer
V0.1.2	30/05/2024	Susan Alake	Initial reviewer