**Executive Summary**

Artificial Intelligence (AI) has become a transformative force across various sectors, including healthcare, finance, and national security. Although AI has improved efficiency, automation, and data-driven decision-making, there are also moral and legal issues with its growing use. Since malfunctioning in these systems can have detrimental effects on people, communities, and even national security, the hazards connected with high-risk AI applications demand consideration.

Bias is a major ethical issue with AI since many AI models are built on datasets that can contain preconceived notions from the past. If these prejudices are not properly controlled, they may result in discriminatory choices in important domains including healthcare, law enforcement, and employment. Furthermore, accountability and transparency concerns are crucial; AI systems frequently operate as "black boxes," which makes it challenging to audit and understand how they make decisions. Without sufficient transparency, stakeholders and users are unable to ascertain how an AI system came to a specific conclusion.

To address these concerns, two key documents provide insights into AI ethics and regulations. The first, Burton et al. (2017)'s Ethical Considerations in Artificial Intelligence Course, focuses on the moral dilemmas raised by AI, especially those involving accountability, transparency, and fairness. The second is Australia's Government Proposal Paper on Mandatory Guardrails for AI in High-Risk Settings (2024), which lays out guidelines for the responsible use of AI and offers a legal framework for handling high-risk AI applications.

This study critically investigates whether the dangers associated with high-risk AI are adequately captured by the Proposed Principles (PPs) in D2. It evaluates how well they reduce bias, maintain human oversight, and foster inclusivity. It also looks at how flexible they are to control future developments in AI. The main conclusions suggest that while D2's principles offer a solid basis for AI governance, there are some significant gaps, especially about adaptability, bias prevention, and safeguards for underrepresented groups. To guarantee that AI legislation continue to be applicable and efficient as AI technologies advance, this research suggests improvements to these principles.

Introduction

Background and Context

AI is transforming sectors and impacting decision-making on a never-before-seen scale. Law enforcement, national security, financial risk assessments, and medical diagnostics are just a few of the vital industries where artificial intelligence has grown ingrained. The increasing sophistication of AI systems, however, presents substantial ethical challenges, particularly when applied in high-risk settings where poor decision-making can have serious societal and legal repercussions. A framework that guarantees AI governance's accountability, openness, equity, and flexibility is necessary to allay these worries.

Responsible AI use and regulatory control are urgently needed considering the rise of high-risk AI. Some AI applications can change people's lives, including AI-driven medical diagnoses or autonomous decision-making in law enforcement. Inaccurate medical diagnoses, financial discrimination, and unjust arrests could result from unethical AI system performance, which can have disastrous consequences for both individuals and societies.

To navigate these challenges, two key documents help frame AI governance:

1. Burton et al. (2017) - Ethical Considerations in AI: Explores fundamental AI ethics principles, with a focus on fairness, transparency, and responsibility in AI deployment. (D1)
2. Australia’s Government Proposal Paper (2024, D2) - Mandatory Guardrails for AI in High-Risk Settings: Establishes a regulatory framework to control high-risk AI applications and minimize associated risks.

These papers offer a structured approach to evaluating AI ethics and governance frameworks, identifying strengths, limitations, and areas for improvement in current regulatory efforts.

Importance of AI Ethics and Regulations

High-risk AI systems have the potential to be harmful if they are not properly regulated. High-risk AI applications work in delicate fields where mistakes could have permanent repercussions, such healthcare, financial services, public safety, and military operations. The suggested principles (D2) from the Australian Government aim to regulate AI by addressing issues with discriminatory decision-making, bias, a lack of human oversight, and transparency gaps. But an important question still stands: Are these guidelines enough to guarantee moral and responsible AI use?

In order to improve AI governance and reduce the dangers associated with high-risk AI systems, this study intends to assess if D2's principles adequately address the ethical issues raised by AI and ascertain whether any changes, additions, or policy improvements are required.

Research Focus and Key Issues

This research centers around the following core question:

**Do the Proposed Principles (PPs) in D2 Adequately Capture High-Risk AI? Would You Recommend Adding or Removing Any Principles?**

The study evaluates whether the regulatory principles outlined in D2 are sufficient to govern AI responsibly. The analysis is structured around the following key areas:

1. Defining High-Risk AI (D2, p.6-8) – What qualifies as high-risk AI, and why does it require strict regulation? How do we define AI systems that pose a risk to individuals and society?
2. Evaluating the Proposed Principles (D2, p.19-25) – Do the principles outlined in D2 effectively address key ethical risks, societal concerns, and potential AI-related harms?
3. Identifying Gaps and Shortcomings (D1, p.27-31; D2, p.10-12) – Are core ethical considerations such as bias, inclusivity, and transparency adequately accounted for in D2?
4. Protecting Vulnerable Communities (D2, p.26-28) – Do the principles safeguard First Nations people and other marginalized groups? Are AI regulations ensuring equitable protections for all communities?
5. Adaptability to Emerging AI Technologies (D1, p.33-35; D2, p.19-25) – Are the principles flexible enough to regulate General-Purpose AI (GPAI) and Artificial General Intelligence (AGI)?

Research Methodology

This study adopts a multi-faceted research approach to comprehensively assess AI regulations. The methodology includes:

* Case Studies – Examining real-world examples of AI-related failures, risks, and regulatory interventions.
* Government Publications – Reviewing legal reports and official documents concerning AI governance.
* Academic Research – Analyzing peer-reviewed studies on AI ethics, policy frameworks, and regulatory measures.
* Comparative Analysis – Evaluating D2’s principles in relation to global AI regulatory frameworks, including the EU AI Act and the US AI Bill of Rights.

By cross-referencing ethical frameworks outlined in D1 with the regulatory proposals in D2, this research aims to assess the adequacy of D2’s principles, determine their effectiveness in mitigating AI risks, and identify areas that require further refinement.

Scope of Research

The research primarily focuses on:

* Assessing the strengths and weaknesses of Australia’s proposed AI regulations (D2) for high-risk AI applications.
* Identifying gaps and potential areas for improvement within the existing regulatory framework.
* Exploring alternative regulatory models to enhance AI governance, accountability, and ethical oversight.

By contributing to the broader global discussion on AI ethics and responsible governance, this study aims to provide practical insights into refining AI policies and frameworks to align with evolving technological advancements and societal needs.

Findings & Analysis

1. Defining High-Risk AI

What Qualifies as High-Risk AI?

Artificial intelligence (AI) systems classified as high-risk are those that present serious risks to people's health, safety, or basic rights because of their intended use, operational environment, or possible effects. High-risk AI systems are defined by the European Union's Artificial Intelligence Act as those that have the potential to jeopardize safety or fundamental rights. These include applications in critical infrastructures, education, employment, law enforcement, essential private and public services, migration, and border control management.

https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Why Does High-Risk AI Require Strict Regulation?

High-risk AI must be strictly regulated to avoid possible negative effects including discrimination, invasions of privacy, and safety hazards. Inappropriate treatment of particular populations can result from unregulated, high-risk AI systems that reinforce pre-existing biases. For example, prejudices in AI systems employed in healthcare may lead to incorrect diagnoses or unequal access to care for underserved populations. Furthermore, it might be difficult to detect and correct biased or incorrect results when AI decision-making processes lack transparency, which can undermine accountability and confidence.

https://www.nature.com/articles/s41746-025-01503-7

Examples of High Risk AI Applications….

Critique of D2’s approach….