**Old Version:**

# Introduction

Artificial Intelligence (AI) is often described as one of the most powerful and important technologies. It is expected to contribute over USD 13 trillion to the global economy by 2030, creating more industries and opportunities [1]. Although the development of AI is beneficial for us, it also brings serious risks, including biased algorithms that discriminate against minority groups as well as generative ability, which can be misused to produce deepfakes and harmful and misleading content [1]. Researchers such as Burton et al. [2] have also emphasised that these risks are wide-ranging, it should not only focus on current concerns like fairness, accountability, and privacy, but also on longer-term challenges, including autonomy, the misuse of general-purpose systems and even the possibility of superintelligence. These show that the risks surrounding AI are not merely technical but also social, ethical and political issues. As a result, governments worldwide are under increasing pressure to establish principles and guardrails that can ensure AI is developed and used correctly.

According to the paper published by the Australian Government, Department of Industry, Science and Resources in 2024 [3], Australia has responded to the AI risk challenge by setting out six proposed principles (PPs) to guide the designation of AI systems as "high-risk" [3]. These principles mainly focus on risks to human rights, physical and mental health, legal status, cultural and collective rights, broader societal and environmental impacts and the overall severity of such harms. The intention is to ensure that AI systems with the potential to cause serious harm are placed under stricter control. In contrast, Malaysia has also developed its own National Guidelines on AI Governance and Ethics (AIGE) with the support of the United Nations Educational, Scientific and Cultural Organisation (UNESCO) to address AI risk challenges with a different emphasis. The guidelines outline seven core principles, which are fairness, reliability and safety, privacy and security, inclusiveness, transparency, accountability, as well as the pursuit of human benefit and happiness [1]. Malaysia's approach is distinct from Australia's because it highlights inclusiveness and human well-being, expressed through the theme "AI for Malaysia, AI for All", which shows that, in addition to aligning with international standards, the guidelines also reflect Malaysia's own culture and society, where AI is meant to support different communities fairly.

As mentioned above, Malaysia’s approach to addressing AI risk places a stronger emphasis on inclusiveness and human well-being, whereas Australia’s proposed principles take a more regulatory and compliance-focused direction. This contrast highlights how different countries have adopted different approaches to AI governance, shaped by their own cultural, social, and political contexts. While Australia’s six proposed principles provide a useful starting point, some of them are very broad and can be difficult to interpret with precision. At the same time, they do not explicitly reflect values such as inclusiveness and human well-being, which play a central role in Malaysia’s framework. For this reason, this report will further evaluate the adequacy of Australia’s proposed principles, compare them with Malaysia’s guidelines, and explore whether the existing principles should remain unchanged, be revised, removed, or whether new principles should be introduced.

# Comment:

The approach for introduction too content based (It should be avoided)

It should be briefer and more focused on the AI-risk first, don’t make the comparison at this moment, it is too early to mention about the pps contents, it may make the introduction too packed.

Put in the title such as Methodology, Scope

What is high risk AI?

Why we need to capture high-risk AI?

The body to answer the topic chosen.

**New version:**

# Introduction

Artificial Intelligence (AI) has rapidly evolved from an emerging research field into one of the most transformative technologies of the 21st century. It is expected to contribute over USD 13 trillion to the global economy by 2030, creating new industries [1]. However, alongside these benefits, AI also brings significant risks including biased algorithms, accountability gaps, the misuse of generative systems, and even the potential emergence of superintelligence [2]. Such concerns highlight the importance of establishing robust principles to capture these high-risk AI.

A central concern in current discussions of AI governance is the concept of high-risk AI. High-risk AI refers to systems that, if misused or poorly regulated, have the potential to cause serious harm to individuals or society in terms of human rights, public safety, or mental health [3]. Clearly identifying and categorising high-risk AI is critical because it enables governments and organisations to apply stricter oversight and ensure that AIs are designed, deployed as well as monitored in ways that minimise potential harms. Without appropriate principles, the rapid expansion of AI could easily outpace the capacity of regulators and institutions to protect society.

The need to capture and regulate high-risk AI is urgent for several reasons. First, AI has already become part of everyday life: people make use of it for learning, assessment, and decision-making, where mistakes or bias in it can harm vulnerable individuals. Second, the scalability of AI means that even small design flaws can produce widespread harm when deployed at scale. Third, because AI is developed and adopted globally, differences in national standards and regulations can create inconsistencies and accountability gaps. Together, these challenges underscore the importance of assessing whether existing governance frameworks are sufficient or whether new principles are needed.

# Scope

This report is limited to the examination of governance principles for high-risk AI. The primary focus will be on evaluating the six principles proposed by Australian Government, Department of Industry, Science and Resources [3], with a comparative analysis of Malaysia’s National Guidelines on AI Governance and Ethics [1]. The report will not attempt to review all global frameworks but will instead narrow its focus to principles that specifically address the risks associated with high-risk AI. Special attention will be given to whether these principles sufficiently address issues of human rights, societal well-being as well as long-term impacts.

# Methodology

This research adopts a qualitative approach, drawing primarily on policy document analysis and comparison with academic literature. The Australian and Malaysian governance frameworks will be reviewed in light of ethical considerations highlighted by Burton et al. [2], as well as other relevant scholarly and policy sources. The aim is to evaluate whether the proposed principles are adequate for mitigating the challenges of high-risk AI, to determine if revisions are necessary, and to consider whether additional or removal principles should be introduced.

# Reference

1. Ministry of Science, Technology and Innovation (MOSTI), The National Guidelines on AI Governance & Ethics. Putrajaya, Malaysia: MASTIC, 2024.
2. E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh, “Ethical considerations in artificial intelligence courses,” AI Magazine, vol. 38, no. 2, pp. 22–34, 2017.
3. Department of Industry, Science and Resources, Safe and Responsible AI in Australia: Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings. Canberra, Australia: Commonwealth of Australia, 2024.