



Ethical Artificial Intelligence in the Australian Signals Directorate

January 2023

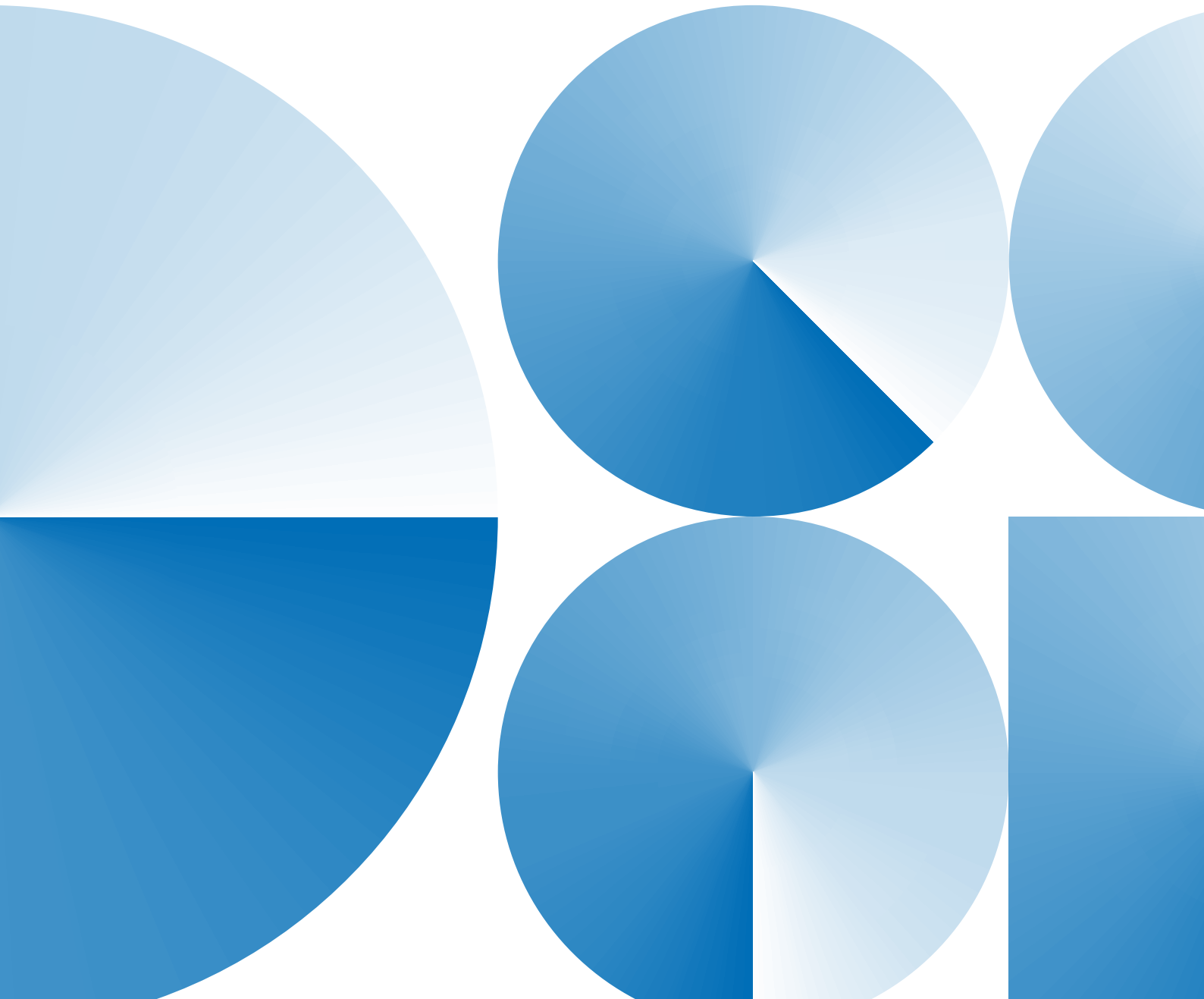
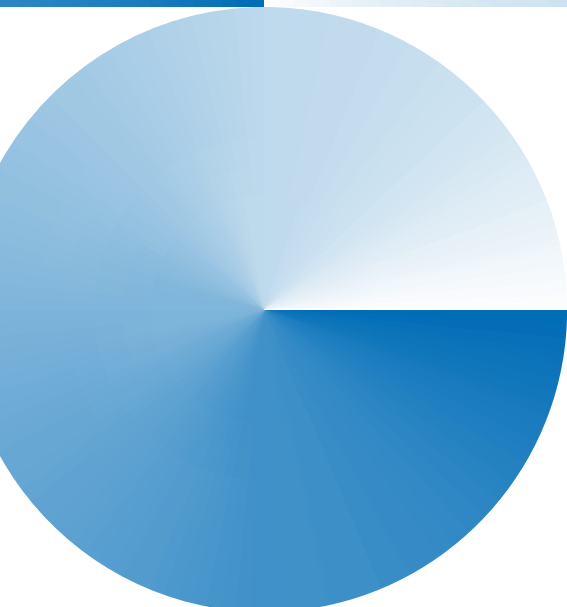
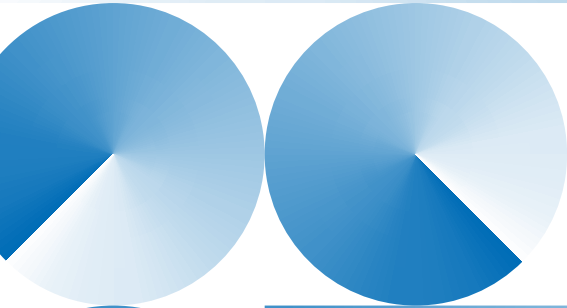


Table of Contents

Foreword	3
Introduction	4
ASD's Purpose	5
What is AI?	6
AI at ASD	6
Cyber security	6
Intelligence analysis	7
Support to the ADF	7
ASD's Ethical AI Framework	8
AI Ethical Principles	9
The principles	9
Implementing AI ethically	9
Governance and oversight of AI	10
Partnerships	11
Conclusion	12



Foreword

The environment in which the Australian Signals Directorate (ASD) operates is becoming increasingly challenging. Today, Australia's region – the Indo-Pacific – is far more complex and far less predictable than at any time since the Second World War. Mastery of innovation and technology will play an important role in helping avoid miscalculation and conflict.

I get asked a lot about how ASD will contend with the technological challenges of the future. My answer is simple – with the same innovation, imagination and creativity that we have used to deal with challenges for the last 75 years. We have a proud history of adapting at the pace that technology evolves, and for the global adoption of artificial intelligence (AI), our approach will be no different.

Harnessing the full potential of AI will deliver strategic advantage for Australia, and will enable ASD analysts to focus on the greatest threats. While ASD has always provided intelligence that delivers advantage to the Australian Defence Force and protected the security of its communications, AI will enable ASD to conduct rapid intelligence, helping provide our people with the best possible tools to complete their mission.

By necessity, much of ASD's work happens in secret, but core to our values is ensuring that we not only meet our legal obligations, but operate within community expectations of propriety and ethics. The AI Ethical Framework outlined in this paper demonstrates how we apply AI in a lawful, secure, accurate, fair and accountable manner. Human judgement will remain at the centre of our decision-making, and our talented people will always be the critical component in our efforts to protect Australia's interests.

Rachel Noble PSM

Director-General ASD

Introduction

ASD and its mission-focused culture was born out of support to the Allied cryptologic effort during the Second World War. This global enterprise included some of the world's finest mathematicians and cryptanalysts who developed the foundations of digital computing and programming. Staff from ASD's forebears contributed to the Allied war effort led by Alan Turing OBE and other great minds to break German and Japanese codes in the 1930s and 1940s. This allowed the Allies to read enemy communications and gain crucial advantage during the war. A leader in computer science and artificial intelligence (AI), Turing went on to develop the 'Turing Test', a method for assessing whether machines can think¹.

For more than 75 years, ASD has collected intelligence about foreign adversaries while keeping Australia's national secrets safe. From decrypting Japanese radio signals during the Second World War to our role today at the forefront of contemporary signals intelligence and cyber security, ASD has always leveraged state-of-the-art technologies. The use of AI represents the next phase in maintaining and securing a competitive advantage.

Advances in technology are changing the way Australians live, work and communicate. They create new jobs, drive economic growth and boost competitiveness. But with technological advancement come different threats and adversaries who seek to undermine Australia's interests.

AI is a form of software that can learn to solve problems at a scale and speed impossible for humans, and it is all around us. It guides our internet searches, helps point us in the right direction when using satellite navigation, and enables technologies as diverse as voice transcription and medical screening for cancers. AI is used widely by businesses and industry to detect and recommend preventative maintenance, shape complex logistics networks, and target advertising and content on social media.

At ASD, AI has a unique role to play in supporting our highly skilled people. It can automate tasks that machines excel at, which enables our people to focus on the things that only people can do. The interaction between machine automation and human decision-making will enable us to better develop and tailor solutions to emerging security challenges.

As we continue to develop our AI capability, ASD will strive to guard against the inherent risks as they relate to the collection, analysis and assessment of intelligence as well as the management of cyber threats.

Implementing AI in a national security context represents particular challenges. By necessity, much of ASD's work is done in secret, but we still adhere to the social and ethical standards expected of us by the Australian community. To ensure this, ASD will adapt and grow its governance of AI as community expectations and standards evolve. In this context, establishing a principles-based approach to AI ethics ensures ASD presents not only an ethics governance framework but also a pathway for practical implementation.

1. Oppy, G., and Dowe, D., 'The Turing Test' in *The Stanford Encyclopaedia of Philosophy* (Winter 2021 edition), Edward Zalta (editor), <https://plato.stanford.edu/archives/win2021/entries/turing-test/> (retrieved 02 November 2022).

ASD's Purpose

ASD defends Australia from global threats and helps advance Australia's national interests. We do this by mastering technology to inform, protect and disrupt.

- Inform: by the covert acquisition of foreign information not publicly available (signals intelligence).
- Protect: by comprehensively understanding the cyber threat, providing proactive advice and assistance to improve the management of cyber risk by governments, businesses and the community.
- Disrupt: by applying our offensive cyber capabilities offshore to support military operations and disrupt cybercrime.

The Australian Government continues to make significant investments in ASD's foreign signals intelligence, cyber security role and offensive cyber capabilities. This includes an AUD \$9.9 billion investment in cyber and intelligence capabilities over the next decade through the REDSPICE (Resilience – Effects – Defence – SPace – Intelligence – Cyber – Enablers) initiative. REDSPICE responds to the deteriorating strategic circumstances in our region, which are characterised by rapid military expansion, growing coercive behaviour and increased cyber attacks. Through REDSPICE, ASD will deliver forward-leaning capabilities essential to maintaining Australia's strategic advantage and capability edge over the coming decade and beyond. This includes advanced AI, machine learning and cloud technology in support of ASD's mission.

What is AI?

AI is a collection of technologies that can be employed to perform tasks that would typically be done by a human, including at a scale, speed and complexity that humans cannot do. AI can also enable and augment human teams to:

- support better decision-making
- discover new ways of operating
- increase efficiency and resilience.

AI may also refer to process automation or mathematical algorithms. It may also include machine learning, which is the ability for a computer to ingest and learn from data in order to make predictions or identify patterns. An AI solution will often incorporate all of these technologies.

AI is essential in assisting ASD to meet the challenges and changes in Australia's strategic environment. These include:

- dealing with the increasing complexity and volume of data
- responding to new developments in technology
- providing ASD with the best possible tools to enable their mission
- adapting to changes in the environment in a way that rules-based approaches cannot.

AI technologies evolve rapidly, which is why ASD recognises the importance of being agile in developing AI solutions, and responsive in adapting its processes, guidelines and concepts.

AI at ASD

ASD already leverages AI to support a variety of its functions, in particular cyber security, intelligence analysis and support to the Australian Defence Force (ADF).

Cyber security

Cyber security and reinforcing online resilience is a national priority. Australia's prosperity makes us attractive to cyber criminals and malicious cyber activity is growing. Rapid exploitation of critical public vulnerabilities has become the norm, and critical infrastructure networks are increasingly being targeted worldwide. AI can help ASD to protect government networks, defend critical infrastructure, and advise all Australian internet users by:

- providing early detection of a cyber incident, whether the incident was successful or not
- rapidly identifying damage and/or loss of capability to ensure application of appropriate mitigations
- suggesting and prioritising disruption and response options, particularly for automated systems
- minimising disruption to operations and establishing resilience within networks to ensure limited loss of function during and after a cyber incident.

Intelligence analysis

Intelligence analysis is about finding the needle in the haystack to protect Australia and our national interests. Over the last two decades, the total volume of information created, captured, copied, and consumed globally has grown exponentially. So while the needle has not changed size, the haystack has grown tremendously. AI will be central to ASD in rising to meet this challenge by:

- filtering substantial amounts of data to identify high interest communications, allowing analysts to concentrate on Australia's greatest threats
- enabling analysis that would otherwise be too difficult or voluminous
- supporting decision-makers with rapid situational awareness, verification of results and increased confidence
- providing better resource management for the collection, storage and analysis of data.

Support to the ADF

ASD has a long history of supporting Australian military operations. Today, ASD supports ADF operations around the globe to enable warfighting and protect our people and assets. ASD draws on its deep technical expertise to help the ADF stay ahead of technological advancements in the region. AI will support this mission by:

- providing ADF platform protection through the rapid analysis of threat data
- fusing and analysing multi-source data to enable faster and more robust decision-making.

ASD's Ethical AI Framework

The foundation of ASD's Ethical AI Framework (Figure 1) is the AI Ethical Principles. The principles will drive future AI technology design and implementation in ASD. ASD will leverage existing processes for internal and external review to drive continuous improvement.

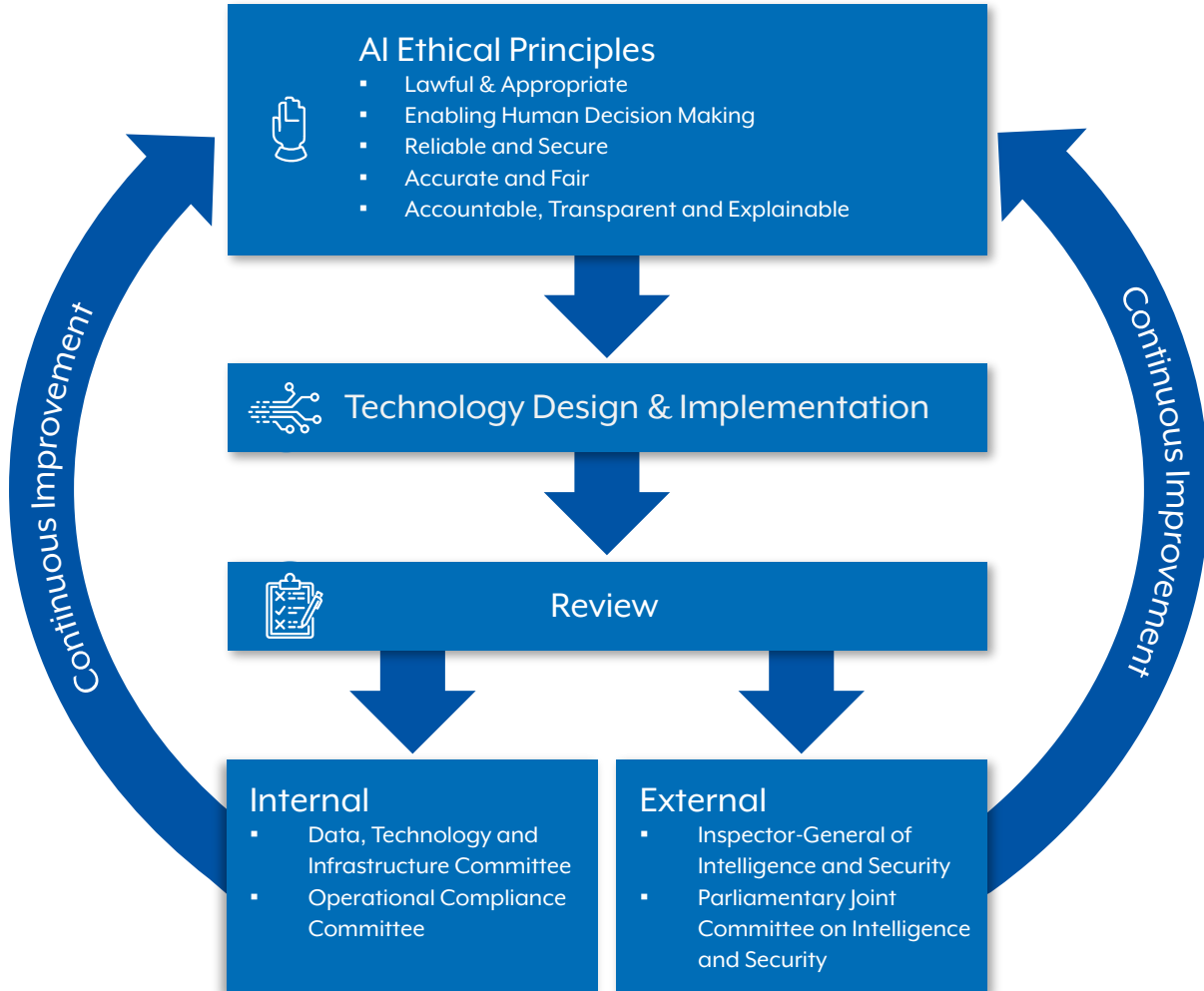


FIGURE 1: ASD Ethical AI Framework

AI Ethical Principles

We recognise that the promise of AI raises questions around how ASD manages complex systems that can have a direct impact on the privacy and security of Australians. The AI Ethical Principles below help answer these questions. They also provide appropriate guidance for ASD staff involved in the development, deployment and maintenance of AI systems.

The principles

- **Lawful and appropriate:** ASD's development, management, and deployment of AI capabilities are consistent with the legislation, policies, processes and frameworks that govern ASD's functions and protect the privacy of Australian citizens.
- **Enabling human decision-making:** Human assessment and judgement will remain central. ASD will use AI to support and enable its highly-skilled workforce to best fulfil ASD's functions in protecting Australia and Australians from threats.
- **Reliable and secure:** AI capabilities developed by ASD will be managed to ensure that they are reliable, continue to meet their intended purpose, and remain protected from external interference.
- **Accurate and fair:** ASD will endeavour to remove unintended bias from its AI systems so that they produce results that are balanced, accurate and fair.
- **Accountable, transparent and explainable:** AI-based capabilities will have human oversight and control, with clear accountabilities in place for all stages of the development life-cycle. ASD is committed to meeting the ethical need for its decisions and actions to be accountable, transparent and explainable to ensure it operates appropriately and proportionately. This will be balanced with the need to protect sensitive equities to ensure ASD is able to continue to perform its critical intelligence and security functions.

The AI Ethical Principles are an important step in a program of work within ASD to build world-class AI capabilities that support our organisation to defend Australia from global threats and advance Australia's national interests. ASD will continue to develop the technical, policy and governance frameworks required to help us manage our AI capabilities in line with these principles and we will continue to invest in our most valuable asset, our workforce, to fully realise the benefits that AI offers us as an organisation.

Implementing AI ethically

There are challenges associated with the fair and transparent use of AI in any industry. Meeting these challenges in a national security context is particularly demanding because ASD cannot always be completely transparent about its operations. Our AI Ethical Principles are the first step in addressing these challenges. Some specific challenges that ASD seeks to address, through the application of these principles, are:

- **Minimising unintended bias:** AI models can reflect biases made by system creators in their training data and design decisions. This can lead to biased results such as disadvantaging certain demographics. Appropriate guidelines and governance arrangements will be developed to ensure that ASD minimises unintended bias and that the true and intended outcome of an AI system aligns with ASD's licence to operate.
- **Appropriate use of AI models:** Using a model for a purpose for which it was not originally designed can have unintended consequences that impact on individuals. Even when a model is used for its intended purpose, model drift over time may introduce biases or inaccuracy, meaning that it is no longer suitable for its intended purpose. Governance related to the ongoing applicability of AI models and standards for accuracy and reliability will be developed to ensure the continued appropriate use of AI systems in ASD.
- **Enabling human decision-making:** Effective human-machine teaming is important for users to understand how to make informed decisions based on AI system outputs, and to maintain trust in AI systems. ASD's workforce will have a solid foundational knowledge of how AI works, enabling them to use the right model on the right data and the confidence to judge the level of certainty of the results.
- **Securing AI systems:** Attacks by adversaries could affect the behaviour of AI systems and the decisions made as a result of their outputs. Standards for system monitoring, reliability and life-cycle management will be developed to safeguard the health and security of AI systems.

Governance and oversight of AI

As an intelligence agency, ASD has been entrusted with sensitive powers and must be accountable to the public, through the Government, for everything it does. This accountability is exercised through multiple external oversight bodies. In particular:

- the Inspector-General of Intelligence and Security (IGIS), who has the powers of a standing Royal Commission – provides independent assurance that ASD acts legally, with propriety and consistent with human rights
- the Parliamentary Joint Committee on Intelligence and Security (PJCIS), which provides oversight of ASD's administration, expenditure and enabling legislation.

Additionally:

- the Independent National Security Legislation Monitor (INSLM) independently reviews the operation, effectiveness and implications of national security and counter-terrorism laws in which ASD has equities
- the Auditor-General and the Australian National Audit Office (ANAO) provides assurance that ASD is operating and accounting for its performance in accordance with the Parliament's intent through independent audit reporting.

Internally, the Data, Technology and Infrastructure Committee (DTIC) is ASD's senior data and technology decision-making committee. The DTIC ensures that ASD makes evidence-based decisions on data management and meets the highest standards of portfolio management practice and technology delivery.

Additionally, ASD's Operational Compliance Committee helps ensure ASD remains compliant in its intelligence activities, cyber security activities and offensive cyber operations, as well as ensuring ASD establishes and maintains best practice in its operational policies and procedures. This includes compliance with ASD's Ethical AI Framework and AI Ethical Principles.

ASD has also implemented recommendations from the [Comprehensive Review of the Legal Framework of the National Intelligence Community](#) (Comprehensive Review) to strengthen governance and oversight of AI. This includes providing the PJCIS with an annual submission on the development of its AI-based intelligence capabilities and maintaining human involvement in significant or difficult AI-based decisions.

ASD will continue to refine and adapt its enterprise processes and governance structures as AI technologies evolve. Initiatives, such as a catalogued and discoverable database of AI tools and models that have been proven to meet our AI Ethical Principles, will ensure that these technologies are not only harnessed across all of ASD's missions, but are also implemented in a consistent, legal and ethical manner. Such initiatives may also enhance the sharing of AI technologies and management practices with ASD's Five-Eyes Partners.

Partnerships

ASD has a long and proud tradition of adopting new technologies and using them to drive better intelligence and security outcomes for Australia. We have operated in the slim area between the difficult and the impossible since our inception, and the Five-Eyes Partnership, forged during the Second World War, has brought immeasurable benefit, including in the field of AI and machine learning. Without this partnership, ASD would never have been able to independently achieve for Australia the advances we have in shared technology, innovation, capability and 'reach'; we are stronger together.

Through continued collaboration with our Five-Eyes Partners, and with industry and academic partners, ASD's AI capabilities can advance in step with emerging technologies. This will drive improvements in quality, efficiency, and timeliness across all aspects of our business including the triage of large volumes of data to identify high value intelligence, the automation of routine tasks and the early detection of anomalous cyber activity. ASD's approach is to build a team that operates seamlessly with these new technologies. It is the interaction between subject matter experts and AI capabilities that will unlock the true potential of partnership.

In this spirit, the ASD-ANU Co-Lab partnership brings together problem-solvers from a variety of disciplines to conduct complex research on Australia's toughest national security problems. The CoLab brings the next generation of talented scientists and mathematicians into ASD's partnerships, and together we are tackling projects on topics such as machine learning and AI ethics, cryptography, cyber security and vulnerability analysis.

ASD will further extend its partnership with academia through the establishment of an AI Hub in 2023 as part of REDSPICE. The AI Hub will focus on acquiring and integrating AI technologies into production systems, and will be a driving force in partnering industry with academia to bring cutting-edge research to new product development.

Conclusion

With the exponential growth in information globally, ASD must adapt and innovate to meet its strategic objectives. Harnessing the full potential of AI will deliver strategic advantage for Australia, freeing up ASD's analysts to focus on the greatest threats. ASD's international, industry and academic partnerships will help drive the development and implementation of cutting-edge AI applications. ASD will apply these capabilities in a way that is secure, trusted and ethical, and it will be as transparent as possible about their use.

