IPS Working Papers

REGULATING ARTIFICIAL INTELLIGENCE: MAXIMISING BENEFITS AND MINIMISING HARMS

CAROL SOON
and
BEVERLY TAN

August 2023
IPS Working Papers No. 52

Institute of Policy Studies

About Institute of Policy Studies (IPS)

The Institute of Policy Studies (IPS) was established in 1988 to promote a greater awareness of policy issues and good governance. Today, IPS is a think-tank within the Lee Kuan Yew School of Public Policy (LKYSPP) at the National University of Singapore. It seeks to cultivate clarity of thought, forward thinking and a big-picture perspective on issues of critical national interest through strategic deliberation and research. It adopts a multi-disciplinary approach in its analysis and takes the long-term view. It studies the attitudes and aspirations of Singaporeans which have an impact on policy development and the relevant areas of diplomacy and international affairs. The Institute bridges and engages the diverse stakeholders through its conferences and seminars, closed-door discussions, publications, and surveys on public perceptions of policy.

IPS Working Papers No. 52

REGULATING ARTIFICIAL INTELLIGENCE: MAXIMISING BENEFITS AND MINIMISING HARMS

CAROL SOON

Principal Research Fellow

Head, Society & Culture

Institute of Policy Studies

carol.soon@nus.edu.sg

and

BEVERLY TAN

Research Assistant

Society & Culture

Institute of Policy Studies

August 2023

IPS Working Papers No. 52 (August 2023): Regulating Artificial Intelligence: Maximising Benefits and Minimising Harms by C. Soon and B. Tan.

CONTENTS

REGULATION OF ARTIFICIAL INTELLIGENCE: MAXIMISING BENEFITS AND MINIMISING HARMS

Executive Summary

Artificial Intelligence (AI) has made a resurgence in the last years, garnering the attention of the media, policymakers and the masses alike. The launch of free generative AI tools such as OpenAI's ChatGPT and Stability AI's Stable Diffusion have given rise to myriad concerns in different sectors like politics, business, education, arts and entertainment, and human rights. While predictive AI harnesses its training data to make complex calculations and predictions, generative AI takes what it learns from training data and examples or prompts to create new content.

Generative AI can be useful in increasing productivity by providing hyper-personalised support in various contexts. However, the technology can also easily be misused by malicious actors. These developments have raised alarm bells over issues concerning cybersecurity and hiring processes. Generative AI also allows for disinformation, such as deepfakes in the form of misleading images, video and audio or fake news articles, to be created at scale.

Countries are racing to regulate the use of AI to keep their people safe, while reaping the economic benefits the technology brings. In Singapore, the government has been regulating the development and application of AI systems since 2018 through voluntary frameworks and guidelines. The Ministry of Communications and Information has engaged various stakeholders on the responsible use of AI from 2018. It published a discussion paper through the Personal Data Protection Commission (PDPC), funded a five-year long AI research programme that started in 2018, and set up the Advisory Council on the Ethical Use of AI and Data. In terms of governance frameworks, the PDPC published the Model AI Governance Framework in 2019 that guides organisations on ethical and governance issues relating to the use of AI, while the Infocomm Media Development Authority (IMDA) published the AI Verify Framework and Toolkit in 2022 for companies to assess their AI systems according to international ethics standards. Singapore also has two sector-specific

guidelines that are publicly available, the Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector published by the Monetary Authority of Singapore (MAS), and the Public Sector AI Playbook published by the Smart Nation Digital Government Group (SNDGG) to guide public officers on how to use AI in their daily work.

Policymakers in other jurisdictions have responded to the benefits and harms brought about by AI in various ways. The European Union is the first group of countries to pass a legislation on AI, the Artificial Intelligence Act (AIA), in June 2023. The AIA takes a risk-based approach in governing the use of AI, setting out stringent requirements for AI applications deemed as high-risk, such as systems than scan job applications and rank applicants. The AIA also bans certain AI applications, such as government-run social-scoring tools, that are classified as having an unacceptable level of risk. More recently, China released its Interim Administrative Measures for Generative AI (Interim Measures) on 13 July 2023. The Interim Measures detail requirements for providers of generative AI products that are accessible to the Chinese public.

Other countries have published voluntary guidelines and frameworks that organisations can choose to adopt while developing or using AI. In the UK, Australia and the US, there is an added emphasis on the ability for individuals to contest decisions made by AI and seek remedies in the event of being wrongfully impacted. Japan and China are looking into AI education and literacy.

Currently, the adoption of Singapore's existing AI governance frameworks and toolkits are purely voluntary. As the country steps up its development and use of AI in various sectors, guardrails need to be put in place to ensure the fair and responsible use of the technology in various settings. Singapore can consider imposing data privacy and security requirements on companies developing or using AI systems, similar to what is proposed in the EU's AIA. Singapore can also consider imposing additional transparency requirements on labelling AI-generated content and on AI-products that impersonate humans, such as AI chatbots.

Singapore would also benefit from the added consideration of smaller companies in its efforts to regulate. While a handful of smaller or medium-sized companies have been consulted in the process of coming up with the existing frameworks and toolkits, these small and medium enterprises (SMEs) are typically in the finance and healthcare sectors. To better understand and address the needs of SMEs, a wider variety of SMEs in more sectors should be considered. In particular, the government can start with consulting SMEs from the high-value sectors detailed in the National AI Strategy, such as the education or transportation sector.

Finally, Singapore's existing AI regulation efforts revolve around organisations instead of individual users. Countries that have emphasised the importance of being able to challenge the decisions made by AI in their regulations are the EU, the UK, Australia and the US. Singapore can more strongly emphasise the importance of feedback mechanisms through which individuals can challenge decisions made by AI systems and seek remedies for any wrongful impact. To do this, Singapore can consider mandating companies using AI to provide a feedback mechanism for individuals to challenge AI-driven decisions and seek redress. AI literacy and education efforts in Singapore can be expanded to include more topics on the responsible and ethical use of AI. Existing efforts focus on explaining what AI is and building AI models while attempts at delving into the ethical issues of AI are limited to data governance. AI literacy programmes should cover the benefits and risks of using AI, as well as the implications of using AI. AI education should be available to all and can also be introduced in primary and secondary education to prepare young Singaporeans for AI. AI literacy programmes should also be available to non-profit organisations.

Recent developments in AI have cast a spotlight on the benefits as well as potential harms of integrating such technology into our daily lives, highlighting new set of challenges that policymakers need to urgently address. It is important for policymakers to look ahead and consider what needs to be done to expeditiously regulate the use of AI, while remaining adaptable and not stunting economic growth.

REGULATION OF ARTIFICIAL INTELLIGENCE: MAXIMISING BENEFITS AND MINIMISING HARMS

1. BACKGROUND

The recent launch of generative Artificial Intelligence (AI) tools such as OpenAI's ChatGPT, Microsoft's AI-powered Bing and Google's Bard have generated much media buzz and have given rise to myriad concerns in different sectors like education, arts and entertainment, business, politics and human rights. The term "Artificial Intelligence" was first coined in 1995 by Professor John McCarthy, as "the science and engineering of intelligent machines" (Manning, 2022, p. 1). The Stanford Institute for Human-Centered Artificial Intelligence defines AI as a machine that is intelligent, or in other words, able to "learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world" (Manning, 2022, p. 1). The key developments in AI in the last six to seven decades have been on machine learning and predictive AI — algorithms that use past data to perform complex calculations or make predictions, to observe and classify patterns in a wide array of settings (McKinsey & Company, 2023). Some examples include facial recognition, digital voice assistants like Siri and self-driving cars.

Unlike its predecessors, Generative AI takes what it learns from training data and examples or prompts to create new content. The new content that is created comes in the form of text, images, audio and videos (Google, 11 April 2023; McKinsey & Company, 2023; Stahl, 2023). For example, image generation tools like Dall-E, Stable Diffusion and Midjourney generate images based on text prompts, and chatbots like ChatGPT, Bard and Bing Chat help users compose emails, summarise content and generate social media captions based on the instructions they receive. While it seems like a brand-new innovation due to recent buzz, research has been conducted on generative AI models since the 1960s, as seen from ELIZA, an AI chatbot powered by natural language processing (White, 2023). However, the training models for modern generative AI models are based on much larger datasets given data availability and have vastly improved computational capabilities (White,

2023). In addition, given their usability and low to no cost, they are now accessible to the mass population, as opposed to only to niche groups like researchers and well-resourced organisations. Analogous to the transition from web 1.0 technologies to web 2.0 technologies, these tools have made AI much more accessible to the public.

The rapid developments in AI have cast a spotlight on the benefits and risks of integrating such technology into our daily lives, bringing about a new set of challenges that policymakers urgently need to address. To maximise the benefits of AI and minimise its risk, this working paper identifies what needs to be done in Singapore to govern the technology and its rapid developments effectively. We do so by reviewing existing regulatory developments in jurisdictions that have studied this issue and examine current measures adopted in Singapore. In so doing, we identify the gaps and propose recommendations for governance of AI use and innovation. We will continue to update this working paper to keep pace with developments in both the technology and policy responses around the world.

2. A BOON OR BANE?

There are many use cases for how generative AI benefit businesses, governments and the public. Commercially, the use of AI in the sales and marketing industry has been touted to revolutionise customer experiences as businesses use generative AI to optimise audience segmentation and SEO marketing strategies more effectively, compared with AI models built using web-scraping and simple prioritisation (Deveau et al., 2023). On the labour and employment front, generative AI chatbots increase productivity by providing hyper-personalised support for a range of tasks, like drafting emails, solving coding problems and translating foreign languages (Marr, 2023). In politics, AI technologies are used to reduce campaigning costs during elections because they can be deployed to produce campaign emails and create instant responses to debate questions and attack advertisements (Hsu & Myers, 2023).

However, these benefits do not come without risks. Concerns have grown in recent years, particularly, over the threat AI poses to jobs and cybersecurity. The use of AI-powered recruitment tools has become more commonplace, as seen from how 99

per cent of Fortune 500 companies use AI-powered applicant tracking systems in their hiring process (Fuller et al., 2021). Even as these systems help businesses weed out unsuitable candidates quickly while supposedly eliminating bias and promoting diversity in the hiring process, they can exacerbate the problems they claim to solve (Halpert, 2022). On the security front, while AI has resulted in the creation of better threat detection and response tools, AI and machine learning can similarly benefit hackers by enabling them to guess passwords faster, send more convincing phishing emails, create malware and manipulate the training data of other machine learning models (Korolov, 2022; Bordoloi, 2023).

Another area of harms posed by generative AI is the risks the technology poses to individuals, societies and nations. ChatGPT has been shown to "hallucinate" or produce factual mistakes, thereby exacerbating the problem of "fake news" (Baum & Villasenor, 2023). There are growing concerns over how free applications like Wombo, Reface and MyHeritage are used to create deepfakes or synthetic media that are digitally manipulated to resemble someone (Barney & Wigmore, n.d.). This is because these new applications make it much easier for anyone, particularly bad actors, to produce disinformation at scale (Sha, 2019). The use of AI to create political campaigning material by politicians in Canada, New Zealand and the US has also raised the alarms of more widespread disinformation during election time (Hsu & Myers, 2023).

Researchers and activists have also highlighted the broader social and ethical issues arising from the use of these technologies, like the ethics of training AI models using copyrighted data and infringement of data privacy. As the publicly available data that is used to train generative AI models is collected without the meaningful consent of intellectual property owners, critical questions surrounding media and data ownership remain unresolved (Pandey, 2023). There are also potential harms of biases in AI models posed to communities and societies, as seen in how some generative AI models produce content that reinforces racial and gender stereotypes (Baum & Villasenor, 2023; Nicoletti & Bass, 2023).

3. SINGAPORE'S APPROACH TOWARDS RESPONSIBLE AI

The myriad of benefits and risks of AI have elicited a variety of reactions from policymakers in different jurisdictions. Initial responses to the launch of generative AI include the ban of ChatGPT in Italy (which has since been revoked), and in schools in the UK, the US and Australia. More recent policy responses suggest a more nuanced approach that acknowledges both the threats and opportunities that AI brings and the need to take on board different stakeholders' needs and perspectives. In June 2023, the Australian government released a discussion paper and launched an accompanying public consultation on how AI should be regulated. The purpose of the public consultation is to solicit views from Australians on how to mitigate the potential risks of AI while supporting safe and responsible AI practices (Department of Industry, Science and Resources, 2023b). The scope of the discussion paper, titled "Safe and Responsible AI in Australia", focuses on the governance mechanisms Australia should put in place to ensure the "safe and responsible use of AI" (Department of Industry, Science and Resources, 2023b, p. 4). The paper does not claim to cover all AI-related issues, nor does it cover the military-specific use of AI or implications of AI on labour markets, national security and intellectual property (Department of Industry, Science and Resources, 2023a). The EU has just passed the Artificial Intelligence Act (AIA) in June 2023 (see Section 4 for more details).

In Singapore, the government has been regulating the development and application of AI systems since 2018 through voluntary frameworks and guidelines. Developments in the applications of the technology like those mentioned above have increased the urgency at which the country is pushing out AI governance tools and guidelines. Examples include the launch of the AI Verify Foundation in June 2023 and the impending Advisory Guidelines on the Use of Personal Data in AI Systems which will be released by the Ministry of Communications and Information (MCI) by the end of 2023.

Singapore's AI regulation efforts are mostly led by MCI, the Personal Data Protection Commission (PDPC) and the Infocomm Media Development Authority (IMDA). There are also sector-specific efforts from Monetary Authority of Singapore (MAS) and Smart Nation Digital Government Group (SNDGG). This section examines the

various AI governance initiatives the Singapore government has introduced to support the responsible development and deployment of AI in Singapore (see Table 1 for a summary).

Table 1: Summary of AI governance tools in Singapore

AI Governance Tool	**Objectives**	**Scope**	**Stakeholders involved**
PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI (2018)	Provides a preliminary analysis of pertinent issues in the commercial development and adoption of AI solutions	The discussion paper goes over the common principles that surfaced during its research on the responsible use of AI and proposes an accountability-based framework for how issues related to the deployment of AI and the ethics, governance and consumer protection can be discussed	- PDPC
Funding for a Research Programme on the Governance of AI and Data Use (2018)	Funds a research programme at an institute of higher learning to research and inform AI and data policies and regulations, and to establish Singapore as a thought leader in AI and data policies	Research to develop cutting edge thinking and practices for AI and data policies for governance and regulation	- National Research Foundation (NRF) - IMDA - Institutes of higher learning
Advisory Council on the Ethical Use of AI and Data (2018)	Canvasses for views, from businesses, trade associations and chambers, and consumers, and makes recommendations on how organisations can develop and deploy AI solutions in a responsible and trusted manner	Makes recommendations and advises government on AI ethics, policy and governance issues	- Advisory council members - Businesses, trade associations and chambers - Consumers

National AI Strategy (2019)	Announces five national AI projects and sets the national agenda for AI, detailing the national AI vision	The National AI Strategy consolidates national efforts by identifying areas to focus attention and resources on; outlines how the government, industry and researchers can work together to reap the benefits of AI; addresses and manages change or new forms of AI-related risks	- Government - Businesses - Researchers
Model AI Governance Framework (2020)	Aims to provide guidance to organisations that are deploying AI solutions at scale on the key issues to be considered and measures that can be implemented	The voluntary model framework is a general, ready-to-use tool for organisations internal governance structures and measures, level of human involvement, operations management and stakeholder communication.	- PDPC - IMDA - Organisations that use AI solutions
AI Verify (2022)	Promotes transparency between companies and their stakeholders through a combination of technical tests and process checks; facilitates interoperability of AI governance frameworks; contributes to development of international standards on AI	Validates the performance of AI systems against a set of internationally recognised principles through standardised tests, and is consistent with international AI governance frameworks, such as those from EU, the Organization for Economic Co-operation and Development (OECD) and Singapore's Model AI Governance Framework	- IMDA - Organisations that use AI - AI solution developers
Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (2019)	Lays out foundational principles for the use of AI and data analytics in financial products and services to consider when using AI in decision-making; assists firms in contextualising and operationalising governance of use of AI in organisational	Use of AI and data analytics in decision-making in the provision of financial products and services	- MAS - PDPC - IMDA - Organisations in the financial industry that use AI in decision-making

	business models and structures		
Public Sector AI Playbook (n.d.)	Provides public officers with guidance on how AI can be used in their work	The playbook includes information on the basic concepts of AI, identifies opportunities to adopt AI in a government agency, how to start AI projects and get support for AI adoption	- SNDGG - Public officers

3.1 MCI Engagement Initiatives

In 2018, MCI announced that it had been engaging stakeholders through various initiatives related to the ethical use of AI (IMDA, 2018). The three initiatives announced were a discussion paper, written by the PDPC on the responsible development of AI, funding for an AI research programme and the formation of an advisory council.

3.1.1 PDPC Discussion Paper

The first initiative was the release of a discussion paper by the PDPC on the responsible development of AI in June 2018, titled “Discussion Paper on Artificial Intelligence (AI) and Personal Data — Fostering Responsible Development and Adoption of AI” (IMDA, 2018). The paper presents a preliminary analysis of the issues that might plague the commercial development and deployment of AI, and outlines PDPC’s recommendations for an AI governance framework.

The paper highlights that the principles of being explainable, transparent and fair should be incorporated to Singapore’s AI governance framework. It also points out the importance of AI being human-centric and proposes a system that “puts the individual customer or consumer front and centre of the design of the AI deployment” (PDPC, 2018, p. 5). Focusing on these principles, the paper details the objectives of having an AI governance framework and describes the importance of good organisational governance measures for companies developing and deploying AI.

These recommendations would later be translated into the Model AI Governance Framework (Model Framework) by the PDPC.

3.1.2 Funding for AI Research

The second initiative announced was the funding of a five-year research programme on the governance of AI and data use from the NRF and the IMDA. A grant call was put out to institutes of higher learning to set up a research programme with the following objectives:

> (1) Promote cutting edge thinking and practices in AI and data policies and regulations; (2) Inform AI and data policy and regulation formulation in Singapore through research publications and stakeholder engagement events and activities; and (3) Establish Singapore as a global thought leader in AI and data policies and regulations. (NRF & IMDA, 2018, p. 1)

Following this grant call, the Centre for AI and Data Governance (CAIDG) was set up at the Singapore Management University (SMU) (IMDA, 2018). SMU's CAIDG has since contributed to discussions and research on AI, publishing over 90 papers on AI ethics and AI's impact on society, business and in specific industries (CAIDG, n.d.).

3.1.3 Advisory Council on the Ethical Use of AI and Data

The Advisory Council on the Ethical Use of AI and Data (Advisory Council) was set up to advise the government on the ethical, policy and governance issues that could arise from using data-driven technologies in the private sector, and to support the government in providing guidance to businesses to minimise ethical, governance and sustainability risks to the business and their customers (IMDA, 2018; PDPC, n.d.). It was also involved in advising the PDPC and IMDA on how to update the Model AI Governance Framework (Model Framework) that was first published in 2019 (MCI, 2020).

3.2 National AI Strategy

The government launched the National AI Strategy in 2019 to coordinate efforts on a national level to encourage the use of AI in various sectors to transform the economy (Smart Nation Singapore, n.d.). *The document outlines goals for Singapore to be a leader in the development and deployment of AI solutions in nine high-value sectors by 2030.* These sectors are transport and logistics, manufacturing, finance, safety and security, cybersecurity, smart cities and estates, healthcare, education and government. The document also highlighted the importance of establishing trust from citizens on the responsible use of AI.

3.3 PDPC's Model AI Governance Framework

The first edition of the Model Framework was published in 2019 by the PDPC (PDPC, 2020). *The Model Framework is a voluntary framework that aims to provide guidance to organisations on ethical and governance issues relating to the use of AI, regardless of the technology they use or the sector in which they operate* (PDPC, n.d.). In the Model Framework, a variety of industries ranging from technology (e.g., Grab, Suade Labs, Facebook), finance (e.g., Mastercard), human resources (e.g., Harver, formerly known as Pymetrics) and pharmaceutical companies (e.g., MSD) are featured as case studies and best practices. The companies that are included as case studies are mostly large corporations with the exception of two medium-sized technology companies based overseas.

According to the PDPC (2020, p. 64), the stated AI ethics principles are for organisations using AI to "consider incorporating these principles into their own corporate principles, where relevant and desired." Based on two high-level guiding principles, the Model Framework aims to promote explainable, transparent and fair use of AI-decision making in organisations, as well as promote the use of AI to enhance human well-being and safety.

The second edition of the Model Framework published in 2020 was refined to include more practical considerations such as interactions with more stakeholders (PDPC, 2020). The updates were made with inputs from the Advisory Council on the Ethical

Use of AI and Data (see Section 3.3.3). The updated Model Framework comes with industry examples to illustrate how organisations can integrate AI governance practices. It also clarifies the recommendations for the various levels of human involvement in AI-powered decision-making. Additionally, the updated Model Framework describes how organisations can take a risk-based approach by considering the impact of the AI-powered tools on the stakeholders involved and introduced the concepts of robustness, reproducibility and auditability with the relevant industry examples. Another notable change in the second edition of the Model Framework is the section on stakeholder interaction and communication, which provides guidance on what to consider when communicating with internal and external stakeholders. The updated framework also clearly states that it "uses existing and common AI ethical principles" (PDPC, 2020, p. 11), and includes a list of these principles in its Annex (see Section 3.6). According to MCI, the Model Framework is adopted by companies like DBS, HSBC, Visa and Microsoft (MCI, 2023a).

The Model Framework is also accompanied by the Implementation and Self-Assessment Guide for Organisations (ISAGO) that helps organisations determine whether their AI practices align with the Model Framework. To help companies manage the impact that AI might have on their employees, the PDPC has also released an industry-agnostic *Guide to Job Redesign in the Age of AI* in 2020 (PDPC, n.d.). The guide and ISAGO provide case studies that help organisations contextualise the Model Framework further.

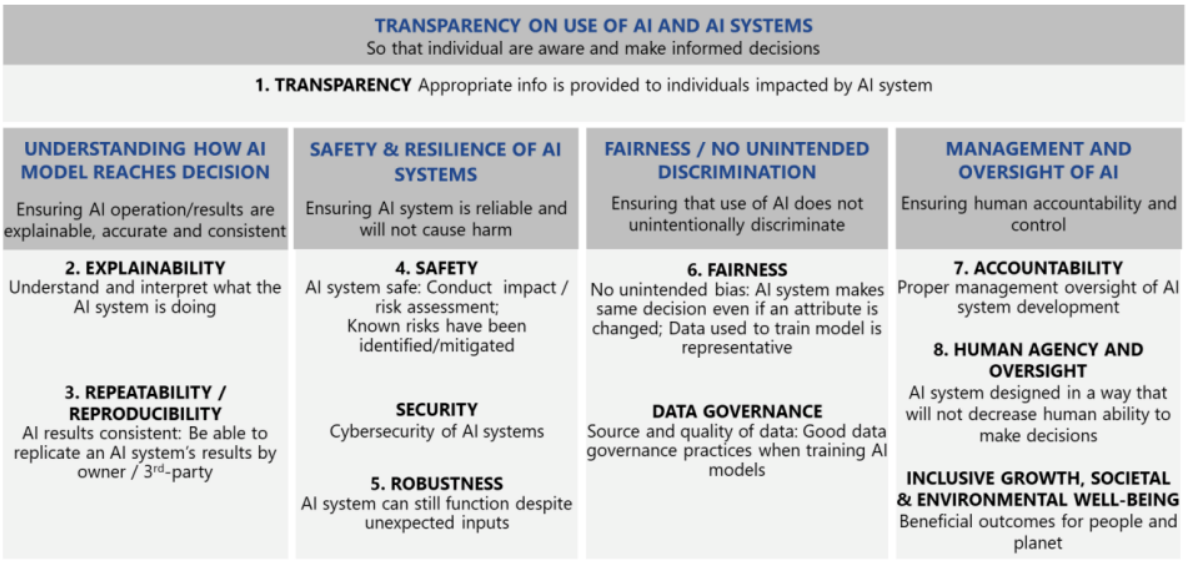
3.4 IMDA's AI Verify

In May 2022, the IMDA launched AI Verify as a minimum viable product (MVP)[1] for pilot testing among a small group of industry partners (IMDA, 2022a; IMDA, 2022b). *AI Verify is an AI governance testing framework and toolkit that companies can use as a self-assessment tool for their AI systems based on 11 international AI ethics principles, which are categorised into five pillars* (see Figure 1). According to an

[1] According to the IMDA's press release on 25 May 2022, a minimum viable product is a product with just enough features for early adopters to test and provide feedback for product development.

article co-written by the Deputy Commissioner of the PDPC, a director and a manager from IMDA, AI Verify is not a tool that sets the ethical standards by which companies should operate; instead, it is a tool for validating companies' claims about their AI systems based on "internationally accepted AI governance principles that countries coalesce around and on which Singapore's AI governance initiatives also stand" (Yeong et al., 2023, para. 7).

Figure 1: AI Verify ethics principles



Source: "Invitation to Pilot AI Verify AI Governance Testing Framework & Toolkit" (IMDA, 2022a)

3.5 Sector-Specific Guidelines

The government has also launched sector-specific guidelines for industries that are expected to see an increased use of AI-powered tools, in some of the sectors mentioned in the National AI Strategy (see Section 3.2). The two publicly accessible guidelines are the *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector* by the MAS and the *Public Sector AI Playbook* by the SNDGG.

3.5.1 MAS FEAT Principles

MAS's FEAT outlines the key principles for firms providing financial products and services to consider when using AI and data analytics in decision-making and to provide the context for AI governance and the responsible use of AI in these firms (MAS, 2018). FEAT principles are intended to be complementary to the voluntary Model Framework.

In 2022, an MAS-led consortium of industry players released the first version of the Veritas Toolkit, which conducts a fairness assessment on the IT system being used by the financial institution (MAS, 2022). The toolkit has since been updated in 2023 to include an improved fairness assessment and assessments for ethics, accountability and transparency (MAS, 2023).

3.5.2 SNDGG Public Sector AI Playbook

The playbook by SNDGG seeks to guide public officers on how to use AI in their daily work, providing examples of the different types of AI applications and possible use cases (SNDGG, n.d.). The playbook focuses on three recommendations from an internal *AI Strategy for the Government* document, which includes identifying common AI applications that can be more widely used, identifying signature AI use cases and expanding on AI capabilities by increasing technical capabilities. The playbook includes chapters on understanding AI, how AI can be applied in the public sector, case studies of how various government entities have used AI, how to start an AI project and the AI-related courses public officers can consider based on the type of work they do.

3.6 Benchmarking Singapore's Efforts Against International Ethics Guidelines

The government has stated multiple times that the existing governance frameworks and guidelines do not seek to establish a new set of ethical principles for AI. These guidelines are largely based "existing and common international AI ethical principles" (PDPC, 2020, p. 11) and are "aligned… with internationally accepted AI ethics principles, guidelines and frameworks" (IMDA, 2022a, p. 3). For example, PDPC's

Model Framework was developed based on existing AI ethics principles from organisations such as the Institute of Electrical and Electronics Engineers (IEEE), the Software and Information Industry Association (SIIA), the European Commission, the OECD, and Fairness, Accountability and Transparency in Machine Learning. The Model Framework also includes principles raised through consultation with the industry; however, there is no distinction between the principles derived from the consultation and the principles adapted from the organisations mentioned earlier. The IMDA's AI Verify framework does not clearly state which organisations or countries it takes reference from, apart from briefly mentioning the EU and the OECD as examples of sources referred to.

Table 2 presents a comparison of the principles used in each governance tool. The principles in Table 2 include those from Singapore's AI governance tools and frameworks as well as the other international guidelines and frameworks discussed in Section 4. As the Public Sector AI Playbook does not specifically articulate the AI ethics principles on which it is based, it is excluded from the table. Similarly, MCI's engagement initiatives, apart from the PDPC discussion paper, do not articulate AI ethics principles and are excluded from the table.

Table 2: Comparison of focus areas in Singapore's AI frameworks

AI governance framework/ AI ethics principles covered	**AI Verify Framework and Toolkit**	**PDPC Discussion Paper**	**Model AI Governance Framework**	**Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT)**
Fairness	✓	✓	✓	✓
Explainability	✓	✓	✓	
Accountability	✓			✓
Transparency	✓	✓	✓	✓
Safety	✓		✓	
Security	✓			
Robustness	✓			
Repeatability/reproducibility	✓			

Data governance	✓			
Human agency and oversight	✓		✓	
Inclusive growth, societal and environmental well-being	✓	✓	✓	
Ethics				✓
Contestability				
Literacy/Education				

The framework that covers the most principles is the AI Verify Framework and Toolkit which, in its full form, covers 11 "internationally accepted" AI ethics principles (IMDA, 2022a, p. 3). These are transparency, explainability, repeatability/reproducibility, safety, security, robustness, fairness, data governance, accountability, human agency and oversight, and inclusive growth, societal and environmental well-being. In the initial pilot, the principles of data governance, security and inclusive growth, societal and environmental well-being were omitted for testing due to practical considerations, such as the availability of existing testing regimes and open-source tools or established methodologies that can be used for testing (IMDA, 2022a). The Model Framework focuses on the principles of explainability, transparency and fairness in AI-decision making, and inclusive growth, societal and environmental well-being (PDPC, 2020). While the invitation to pilot for AI Verify includes the brief definitions for each principle, the Model Framework includes broader explanations for internationally accepted AI ethics principles, instead of definitions for the principles it adopts (see Table 3).

Table 3: Model Artificial Intelligence Governance Framework's compilation of existing AI ethical principles

AI Ethical Principle	**Description**
Accountability	Ensure that AI actors are responsible and accountable for the proper functioning of AI systems and for the respect of AI ethics and principles, based on their roles, the context and consistency with the state of art.
Accuracy	Identify, log and articulate sources of error and uncertainty throughout.
Auditability	Enable interested third parties to probe, understand and review the behaviour of the algorithm through disclosure of information that enables monitoring, checking or criticism.

Explainability	Ensure that automated and algorithmic decisions and any associated data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.
Fairness	a. Ensure that algorithmic decisions do not create discriminatory or unjust impacts across different demographic lines (e.g., race, sex, etc.). b. Develop and include monitoring and accounting mechanisms to avoid unintentional discrimination when implementing decision-making systems. c. Consult a diversity of voices and demographics when developing systems, applications and algorithms.
Human centricity and well-being	a. Aim for an equitable distribution of the benefits of data practices and avoid data practices that disproportionately disadvantage vulnerable groups. b. Aim to create the greatest possible benefit from the use of data and advanced modelling techniques. c. Engage in data practices that encourage the practice of virtues that contribute to human flourishing, human dignity and human autonomy. d. Give weight to the considered judgements of people or communities affected by data practices and to be aligned with the values and ethical principles of the people or communities affected. e. Make decisions that should cause no foreseeable harm to the individual or should at least minimise such harm (in necessary circumstances, when weighed against the greater good). f. Allow users to maintain control over the data being used, the context such data is being used in and the ability to modify that use and context. g. Ensure that the overall well-being of the user should be central to the AI system's functionality.
Human rights alignment	Ensure that the design, development and implementation of technologies do not infringe internationally recognised human rights.
Inclusivity	Ensure that AI is accessible to all.
Progressiveness	Favour implementations where the value created is materially better than not engaging in that project.
Responsibility, accountability and transparency	a. Build trust by ensuring that designers and operators are responsible and accountable for their systems, applications and algorithms, and to ensure that such systems, applications and algorithms operate in a transparent and fair manner. b. Make available externally visible and impartial avenues of redress for adverse individual or societal effects of an algorithmic decision system,and designate a role to a person or office who is responsible for the timely remedy of such issues. c. Incorporate downstream measures and processes for users or stakeholders to verify how and when AI technology is being applied. d. Keep detailed records of design processes and decision-making.
Robustness and Security	AI systems should be safe and secure, not vulnerable to tampering or compromising the data they are trained on.
Sustainability:	Favour implementations that effectively predict future behaviour and generate beneficial insights over a reasonable period of time.

Source: Model Artificial Intelligence Governance Framework Annex A (PDPC, 2020)

FEAT from MAS focuses on the principles of fairness, accountability and transparency (MAS, 2019). There is also an added principle of “ethics” in FEAT that considers the values of the organisation using AI-driven decisions and whether the decisions made by AI are held to at least the same standards as those made by humans. FEAT provides detailed definitions and applications of each principle for financial organisations using AI-driven decision-making to consider. For example, for the principle of fairness, FEAT includes specific illustrations of how AI decision-making could impact the organisation or its customers — such as when AI is used to approve requests for an increase in credit card limits or for revising parameters of AI decision-making tools due to recent policy or legal developments in the industry.

The Public Sector AI Playbook does not explicitly adhere to any AI principles. Instead, it is underpinned by a more practical approach to describing how AI can be used in the daily work of public officers. This playbook provides individual level guidance to AI use, whereas the previous frameworks and guidelines provide higher-level, organisational level guidance. While there is a brief mention of the data quality principles of completeness, accuracy and veracity, the document also points public officers to an internal Data Strategy Playbook to help them identify the data requirements of a given use case (SNDGG, n.d.).

Apart from publishing guidelines and frameworks, Singapore is re-solidifying its commitment to enabling the responsible use of AI, evident from its recent engagements at the global stage. In June 2020, Singapore became a founding member of the Global Partnership on AI (GPAI) (OECD, 2020). GPAI is a multi-stakeholder initiative that aims to enable and facilitate research how AI can be used responsibly, in a manner that respects human rights and democratic values (GPAI, n.d.). The Association of Southeast Asian Nations (ASEAN), of which Singapore is a member, also announced the development of an ASEAN Guide on AI Governance and Ethics that is slated for completion at the end of 2023 (Potkin & Wongcha-um, 2023).

In May 2023, MCI announced that the PDPC will issue Advisory Guidelines on the Use of Personal Data in AI Systems under the Personal Data Protection Act (PDPA) within the year, to support the responsible deployment and development of AI (MCI,

2023b). In June 2023, Minister for Communications and Information Josephine Teo also announced that Singapore will be launching the AI Verify foundation “to harness the collective power and contributions of the global open-source community to develop AI Verify testing tool for the responsible use of AI” (PDPC, 2023, para. 1).

As reviewed in this section, current efforts by the Singapore government to govern the use of AI mainly targets AI developers and the organisations that deploy AI. The initiatives are either broad-based or sector-specific, and aimed at businesses and organisations from different industries that are not limited to high-value ones, and the public sector. In the next section, we examine key initiatives that are underway in other jurisdictions.

4. INTERNATIONAL AI POLICY LANDSCAPE

Governments across the world are racing to regulate AI, drafting legislations and guidelines that address the potential threats of AI while attempting to maximise its benefits. Many of these regulations build on the OECD AI Principles (see Table 4). The OECD provides some of the first intergovernmental standards for AI policies and 42 countries have formally agreed to adopt these principles (OECD, 2019). They include the US, the UK, Japan, South Korea, Australia and European countries that are part of the EU including Germany, France and Italy (OECD, n.d.).

Table 4: Organization for Economic Co-operation and Development’s (OECD) AI principles

OECD Principle	Definition
Inclusive growth, sustainable development and well-being	- Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.
Human-centred values and fairness	- AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights. - To this end, AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.

Transparency and explainability	- AI actors should commit to transparency and responsible disclosure regarding AI systems. - To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: o to foster a general understanding of AI systems o to make stakeholders aware of their interactions with AI systems, including in the workplace o to enable those affected by an AI system to understand the outcome o to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.
Robustness, security and safety	- AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. - To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art. - AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.
Accountability	- AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.

Source: Recommendation of the Council on Artificial Intelligence (n.d.), OECD

Table 5 shows the comparison of focus areas of each supra-national or national regulation in relation to the principles covered in Singapore's AI Verify framework.

Table 5: Comparison of focus areas in international frameworks and legislation

AI governance framework or legislation/ AI ethics principles covered	**AI Verify Framework and Toolkit**	**EU Artificial Intelligence Act (AIA)**	**China Interim Measures for Generative Artificial intelligence Services[2]**	**UK Pro-Innovation Approach to AI Regulation**	**Australia's Artificial Intelligence Ethics Framework**	**US Blueprint for an AI Bill of Rights**	**China Ministry of Science and Technology's Ethical Norms for New Generation Artificial Intelligence**	**Japan's Social Principles of Human-Centric AI**
Fairness	✓	✓	✓	✓	✓	✓	✓	✓
Explainability	✓	✓	✓	✓	✓	✓		✓
Accountability	✓	✓	✓	✓	✓	✓	✓	✓
Transparency	✓	✓	✓	✓	✓	✓		✓
Safety	✓	✓	✓	✓	✓	✓	✓	✓
Security	✓	✓	✓	✓	✓		✓	✓
Robustness	✓	✓		✓	✓			
Repeatability/ Reproducibility	✓				✓			
Data governance	✓	✓	✓	✓	✓	✓	✓	✓
Human agency and oversight	✓	✓			✓	✓	✓	✓
Inclusive growth, societal and environmental well-being	✓	✓	✓		✓	✓	✓	✓
Contestability		✓	✓	✓	✓	✓		
Literacy/Education							✓	✓

[2] China does not explicitly state if the legislation was crafted based on any AI ethics principles. However, many provisions are similar to the regulations from overseas which are based on AI ethics principles. The ticks in this column indicate that there are provisions in China's law that overlap with the provisions based on AI ethics principles from the other countries.

4.1 AI Legislation

4.1.1 EU: Artificial Intelligence Act

While various countries like Canada, Australia and the US are in the process of drafting AI laws and conducting public consultations, the EU Parliament had recently passed the world's first AI law — the Artificial Intelligence Act (AIA) — in June 2023 (European Parliament, 2023).

The AIA was drafted based on the recommendations from the High-Level Expert Group on Artificial Intelligence (AI HLEG) set up by the European Commission (The AI Act, 2021). The ethical principles laid out by the AI HLEG are: (1) respect for human autonomy, (2) prevention of harm, (3) fairness and (4) explicability (AI HLEG, 2019).

The AIA will govern the development and use of AI systems in the EU through a tiered risk-based approach that imposes more requirements on AI systems deemed as posing higher risks (The AI Act, 2021). The AIA classifies AI applications and systems into three categories namely, (1) unacceptable risk, (2) high-risk, and (3) low-risk. AI tools that create an unacceptable level of risk (e.g., government-run social-scoring tools) will be completely banned under the AIA. Specific legal requirements are imposed on high-risk AI applications such as CV-scanning tools that rank job applicants. Applications that do not fall into the categories previously mentioned are low-risk and are left mostly unregulated and able to voluntarily apply the codes of conduct that relevant to them. High-risk systems need to conform to legal requirements on the following fronts:

1. Risk management system
2. Data and data governance
3. Technical documentation
4. Record-keeping
5. Transparency and provision of information to users
6. Human oversight
7. Accuracy, robustness and cybersecurity

In addition to categorising and regulating AI applications according to their risk levels, the AIA also imposes transparency obligations for certain AI applications that impersonate humans, such as chatbots. The developers of these AI applications will have to notify users that they are interacting with an AI system, that emotional recognition or biometric categorisation systems are being used, and to label deepfakes.

4.1.2 China: Interim Administrative Measures for Generative Artificial Intelligence Services

China unveiled its Draft Measures for Generative Artificial Intelligence Services (Draft Measures) earlier in April 2023 for public consultation via the Cyberspace Administration of China (CAC). The Draft Measures were met with criticism for being too onerous on generative AI service providers, whom under the Draft Measures, would need to ensure the accuracy and veracity of the training data and output of the model (Sheehan, 2023). The finalised version was released in July 2023, with the government making it clear that the regulation is put in place to support the growth of the technology as well as ensure the safety of the public (Ye, 2023). The Interim Administrative Measures for Generative Artificial Intelligence Services (Interim Measures) have been in effect since 15 August 2023.

It is currently unclear if the legislation takes a principles-based approach to AI legislation. The law, which has been translated by China Law Translate (2023), does not state that it is based on any AI ethics principles. The summaries of the legislation's key provisions and changes also make no mention of China adhering to common AI ethics principles, or that the country is taking a principles-based approach (Hurcombe et al., 2023; Wu, 2023). However, as shown in Table 5, many provisions overlap with other countries' regulations that are based on AI ethics principles. The Interim Measures only apply to generative AI products that the general public in China can access and exclude the use of AI in enterprises, research and academic institutions, and the government.

The newly enacted legislation requires generative AI service providers to carefully consider how their products will impact individual users (Hurcombe et al., 2023).

Service providers are required to monitor the content generated and report to the authorities if it is found that illegal content that has been generated by their model. The service providers are required to swiftly take action by removing the illegal content, halting the transmission of the illegal content and refining the model so it stops generating more illegal content. Service providers are also required to test, train and optimise their models so that they "increase the truth, accuracy, objectivity, and diversity of training data" (China Law Translate, 2023, para. 20). AI-generated content has to be labelled under the provisions. The Interim Measures also seeks to strengthen user protection by mandating that service providers should not collect or store personal data unnecessarily, and must abide by user requests to access, delete or correct their personal data. Another provision that seeks to protect users is the requirement for service providers to set up a feedback mechanism to collect and address complaints from users. Lastly, "security assessment and algorithm filing" requirements are imposed on generative AI products and services "associated with public opinion or social mobilisation" (Wu, 2023, para. 18). Generative AI services under this category would have to file their algorithms with the CAC's algorithm registry. The registry collects information from the service providers on the source of the training data sets, the type of input information required by the algorithm, as well as descriptions of how the algorithm works (Sheehan & Du, 2022).

4.2 Non-Legislative Guidelines and Norms

The EU's AIA is the world's first comprehensive AI law that has been passed, with other countries focusing more on regulating AI using voluntary guidelines and frameworks. The countries that have published their own national guidelines or frameworks include the UK, the US, Australia, China and Japan. The frameworks from the various countries selected in this paper use principles-based approaches and cover a variety of domains, ranging from organisations developing AI to individual users or consumers who interface with AI. The selected frameworks also span across the world, providing insights on what countries in Europe, Asia and North America are doing when it comes to AI regulation.

4.2.1 UK: Pro-Innovation Approach to AI Regulation

The UK's Pro-Innovation Approach to AI Regulation is a framework that focuses on five principles: (1) safety, security and robustness, (2) appropriate transparency and explainability, (3) fairness, (4) accountability and governance, and (5) contestability and redress (Gov.uk, 2023). The framework aims to help regulate the design, development and use of AI, based on the five principles to promote the responsible use of AI. *The framework is largely based on the OECD principles, but has highlighted contestability and redress, by making it one principle instead of embedding it into the principle of transparency and explainability.* This highlights the importance of allowing users of AI to "contest an AI decision or outcome that is harmful or creates material risk of harm" to them (Gov.uk, 2023). Under this principle, users of an AI system or those impacted by an AI system can seek some form of remedy or compensation if they have been harmed or wronged. Additionally, this principle imposes the need for AI developers or owners to implement proper feedback mechanisms through which this contest and redress can occur.

4.2.2 Australia: Artificial Intelligence Ethics Framework

Another country that has highlighted the importance of contestability is Australia. Australia's Artificial Intelligence Ethics Framework is developed based on eight key principles: (1) human, societal and environmental wellbeing, (2) human-centred values, (3) fairness, (4) privacy protection and security, (5) reliability and safety, (6) transparency and explainability, (7) contestability and (8) accountability (Department of Industry, Science and Resources, 2022). *Australia's principle of contestability highlights the importance of a timely process that people can use to challenge the use or outcomes of an AI system and recommends the "appropriate use of human judgement" in feedback process* (Department of Industry, Science and Resources, 2022, para. 23).

4.2.3 US: Blueprint for an AI Bill of Rights

The Blueprint for an AI Bill of Rights published in October 2022 focuses on the individual rights of AI users, unlike the previously discussed regulations which have

focused on AI owners, deployers and developers. The blueprint is based on the principles of (1) safe and effective systems, (2) algorithmic discrimination protections, (3) data privacy, (4) notice and explanation and (5) human alternatives, consideration and fallback (The White House, 2022). The blueprint describes each principle, its importance as well as how the principles can be applied in a real-life setting. *The blueprint is also unique in its inclusion of the fifth principle of human alternatives, consideration and fallback, where users are entitled to "opt out… and have access to a person who can quickly consider and remedy problems" encountered when using AI.* While this is similar to the earlier point of being able to contest decisions made by AI in the UK's and Australia's approach to AI regulation, there is an added requirement for users to be able to access a human, as well as an automated feedback system.

4.2.4 China: Ethical Norms for New Generation Artificial Intelligence

China has various legislations to govern the use of various AI and machine learning applications, such as the Interim Measures (see Section 4.1.2). The country also has released several guidelines that regulate the use of AI, including the Internet Information Service Algorithmic Recommendation Management Provisions and city-level guidelines like the Regulations on Promoting Artificial Intelligence Industry in Shenzhen Special Economic Zone. These regulations focus on setting out various technical requirements of service providers who develop or deploy AI in their businesses and do not lay out ethical principles, instead mandating the formation of an ethics council that will develop the ethical standards for the use of AI by businesses (Wu, 2022).

China's Ethical Norms for New Generation Artificial Intelligence (Ethical Norms), introduced by the country's Ministry of Science and Technology, focuses on articulating AI ethics principles that should be adhered to on a national level and was thus selected for the comparison in this paper. *The Ethical Norms are far-reaching and applies to any person and institution engaged in any AI-related activity including organisations involved in the research and development of AI, organisations supplying AI solutions, organisations deploying AI solutions and individuals who use AI for any purpose.* The norms are based on six principles: (1) advancement of

human welfare, (2) promotion of fairness and justice, (3) protection of privacy and security, (4) assurance of controllability and trustworthiness, (5) strengthening accountability and (6) improvements to the cultivation of ethics (Center for Security and Emerging Technology, 2021). Under the point of improving the cultivation of ethics, the Ethical Norms state that those developing or deploying AI should understand, consider and discuss AI ethics issues, as well as promote good AI governance.

4.2.5 Japan: Social Principles of Human-Centric AI

Japan's Social Principles of Human-Centric AI (Social Principles) are based on seven key principles: (1) human-centric, (2) education or literacy, (3) privacy protection, (4) ensuring security, (5) fair competition, (6) fairness, accountability and transparency, and (7) innovation (Cabinet Secretariat, 2019). The principles are for developers and business operators of AI to consider in their research, development and implementation of AI systems (Habuka, 2023). The principles of fair competition and innovation are clearly targeted at businesses and AI developers to encourage the use of AI for economic growth.

Japan's approach emphasises the importance of understanding AI, or AI literacy and education. The Social Principles recommend policymakers and those in the management of businesses to properly understand AI and its ethical issues. They also recommend that users of AI should have a general understanding of AI and have the necessary knowledge about AI to use it. AI developers are expected to master the fundamentals of AI technology and are also encouraged to consider how AI can be used in different business models. The Social Principles suggest that there is a need for an educational system that enables its citizens to acquire the basics of AI, mathematics and data science as well as ethical issues surrounding the use of data (Cabinet Secretariat, 2019).

As seen from the regulations highlighted in this section, there is significant convergence in the principles used by different jurisdictions to guide their regulation of AI. While worded differently, many countries tend to look to the OECD's AI ethics principles to guide their AI regulations.

5. RECOMMENDATIONS FOR SINGAPORE

At the national level, while the main approach is to regulate AI through voluntary frameworks, guidelines and codes of conduct, there is a move towards using legislation. With the AIA, the EU is the first to pass legislation that imposes strict requirements on high-risk AI systems, and even banning certain uses of AI completely. China's Interim Measures also require an AI provider to label certain types of AI-generated content and disallow the production of certain types of AI-generated content. AI providers who fail to comply will be punished according to the relevant existing laws; they are also required to cooperate with the inspection by providing the necessary information and support to the regulators (Luo et al., 2023).

In Singapore, the existing initiatives cover a broad range of AI ethics principles that are relevant and pertinent (see Table 2). However, they are voluntary and there are no disincentives or costs to organisations which do not adopt the guidelines. Additionally, the most comprehensive framework in Singapore, AI Verify, is meant to serve as a toolkit for organisations to voluntarily assess their own AI systems. It does not impose requirements on the AI ethics principles that should be followed. While the existing frameworks are important and a necessary first step towards promoting responsible use of AI, their efficacy depends mainly on voluntary compliance. Moving forward, the government might consider imposing legal requirements to ensure that AI developers perform due diligence when developing their AI systems, and that deployers of AI properly assess the risks of the AI systems they use. For example, the developers of AI could be legally required to ensure the security and privacy of the data they collect and use to train their models, similar to the EU's AIA and China's Interim Measures.

Additionally, Singapore can take a leaf from the EU AIA's tiered, risk-based approach in regulating AI. While Singapore's Model Framework references a risk-based approach, there is no clear categorisation or definition of what different levels of risk constitute. Organisations and developers are left to decide if the risk is high or low by themselves. Currently, Singapore's guidelines are industry- and technology-agnostic, with the exceptions of the finance industry and public sector that have their own guidelines. However, as suggested by the EU, the use of AI in an increasing number

of settings — like biometric identification, law enforcement, employment management, and administration of legal processes (e.g., using AI for legal research or interpreting facts) — highlights the pressing need to regulate high-risk AI systems that can exert a serious impact on people's lives (The AI Act, 2021). To facilitate the responsible use of AI, developers and users of high-risk AI systems should abide by a set of baseline requirements to ensure quality of output, which can be mandated by the law.

Furthermore, additional requirements can be placed on AI systems that are set up to act like a human being, for example chatbots, like the additional requirement imposed by the EU's AIA. China's Interim Measures also sets out a requirement for AI service providers to label content that is generated by AI. According to a 2023 survey conducted on almost 2,000 Americans by Tooltester, participants were only able to identify ChatGPT's AI-written content less than half of the time (Brandl & Ellis, 2023). While Singapore's frameworks include the principle of transparency, the developments in AI have made it increasingly hard to distinguish between content generated by people and that of an AI. As such, the developers or deployers of these AI systems should be required to label an AI system or the content it generates, so that users who interact with the AI system are aware that they are not interacting with a human but with an AI model.

In addition to legislating certain requirements, ongoing efforts to improve the governance of AI need to include a wider range of actors and sectors. Existing regulations seem to focus on large finance and technology organisations (e.g., HSBC, DBS, Microsoft and Google), as seen from the companies that participated in the pilot for AI Verify and the companies featured in the Compendium of Use Cases that accompanies the Model Framework (PDPC, n.d.). However, at the time of writing, there are close to 600 AI-related start-ups registered in Singapore's Startup SG ecosystem, and possibly many small-medium enterprises (SMEs) that deploy AI in their work processes (Startup SG, n.d.). SMEs function very differently from bigger tech or finance corporations. Often working on limited resources, these smaller businesses may be unwilling or simply unable to comply with stringent regulations on AI (Cheng, 2023). Thus, it is important to consider the needs and constraints of SMEs, like their level of technical expertise and available resources to implement

any changes necessitated by AI regulations. To do this, more SMEs from various sectors could be involved in engagement or consultations to better understand their capabilities, needs and constraints.

While the government has committed to becoming a leader in the development and deployment of AI in nine high-value sectors[3], the current focus is the healthcare and finance sector when it comes to testing frameworks or providing examples. Most recently, the AI Trailblazers initiative launched by MCI includes the Propertyguru Group, a property technology company, as well as the Ministry of Manpower and the Government Technology Agency. While this initiative seems to include a larger variety of sectors, it is unclear how many of the companies participating in the initiative are smaller enterprises and if there will be enough representation across the various sectors. It is also important to involve companies in various engagements from the other sectors, such as transport, education and beyond, to better understand the general and specific needs of small- and medium-sized companies in various industries. This would allow for a better transfer of regulatory principles across sectors and develop more focused, sector-specific considerations.

Lastly, at the individual level, Singapore's national governance framework should be expanded to include more values that focus on the user. Currently, the accent of Singapore's national AI governance frameworks is on empowering businesses in their development and deployment of AI. However, as shown in Table 2, Singapore places less importance on contestability. While the Model Framework does include a mention of allowing consumers impacted by decisions made by AI to contest these decisions, the Model Framework does not provide detail on what this would look like in practice. The voluntary frameworks from the UK and Australia articulate the importance of allowing consumers to report and challenge decisions made by AI that harms them, while the EU AIA takes into consideration if AI systems have such mechanisms in place during their risk assessment. The US takes it a step further by encouraging that the feedback mechanisms should also grant access to a human who can address or remedy the problems caused by or encountered when

[3] The high value sectors are: (1) Transport and logistics, (2) Manufacturing, (3) Finance, (4) Safety and security, (5) Cybersecurity, (6) Smart cities and estates, (7) Healthcare, (8) Education and (9) Government.

interacting with AI systems. Singapore could consider more strongly encouraging these practices as well, particularly in contexts where the decisions made by the AI can impact the life or quality of life of an individual, so that citizens can be adequately protected from the potential risks of AI. To ensure that consumers are adequately protected and that companies developing or deploying AI stay accountable, Singapore could also consider a legal requirement that allows individuals who have been wrongfully impacted by AI-powered decision-making to challenge the result and seek redress.

Another area where Singapore has opportunity for improvement is in AI literacy. As shown earlier in Tables 2 and 5, Japan and China highlight the importance of education and literacy of the population. While Singapore is certainly not ignoring the importance of AI literacy, as seen by the efforts of AI Singapore (AISG) in putting out free AI-related courses and workshops, more can be done. Existing AI literacy efforts in Singapore focus on helping people understand what AI and machine learning is, or how to build an AI system. There is less focus on the potential benefits and harms of AI and its ethical implications. Two free courses by AISG — AI for Everyone (AI4E) and AI for Industry (AI4I) — briefly cover some AI ethics principles and concepts. However, the content in these courses focuses on encouraging good data collection practices and the importance of high data quality to reduce biases in machine learning algorithms; without delving into the importance of using AI responsibly. While data governance is an important topic, the users of AI should also be aware of the limitations of AI and the importance of other AI ethics principles such as robustness, transparency or human agency.

Additionally, Singapore should seriously consider expanding existing AI literacy efforts, making such education available to more groups of people. For example, AI content can be integrated into existing school curriculum in primary and secondary schools, like what is being done in Zhejiang, China, and as per United Nations Educational, Scientific and Cultural Organization's (UNESCO) recommendations for platform-agnostic and brand-agnostic education for K1-K12 (Liu, 2023; UNESCO, 2023). Singapore can also step up its efforts on educating non-profit organisations about AI. To do this, we can take reference from NetHope's *Artificial Intelligence (AI) Suitability Toolkit for Nonprofits* (NetHope, n.d.).

Contemporary developments in AI have cast a spotlight on the benefits as well as risks of integrating such technology into our daily lives, as they bring about a new set of challenges that policymakers urgently need to address. As AI technology continues to advance, more benefits and risks of using AI will emerge. It is important for policymakers to adopt an anticipatory approach and consider what more needs to be done in terms of AI regulation, while remaining adaptable and not stunting economic growth.

6. REFERENCES

Barney, N., & Wigmore, I. (n.d.). *What is deepfake AI? A definition from TechTarget*. WhatIs.Com. Retrieved June 28, 2023, from https://www.techtarget.com/whatis/definition/deepfake

Baum, J., & Villasenor, J. (2023, May 8). *The politics of AI: ChatGPT and political bias*. Brookings. https://www.brookings.edu/articles/the-politics-of-ai-chatgpt-and-political-bias/

Bordoloi, P. (2023, March 21). AI may get scarier, govts must tame it in time. *Analytics India* Magazine. https://analyticsindiamag.com/ai-may-get-scarier-govt-regulations-must-tame-it-in-time/

Brandl, R., & Ellis, C. (2023, March 8). *Survey: ChatGPT and AI content — can people tell the difference?* Tooltester. https://www.tooltester.com/en/blog/chatgpt-survey-can-people-tell-the-difference/

Cabinet Secretariat. (2019). *Social principles of human-centric AI*. Cabinet Secretariat. https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf

Center for Security and Emerging Technology. (2021, October 21). *Translation — Ethical norms for new generation artificial intelligence released*. Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

Centre for AI & Data Governance. (n.d.). *Research*. Retrieved June 28, 2023, from https://caidg.smu.edu.sg/research

Cheng, M. (2023, June 14). Unlike big tech, some AI startups aren't at all ready to invite regulation. *Yahoo News*. https://sg.news.yahoo.com/unlike-big-tech-ai-startups-154400322.html

China Law Translate. (2023, July 13). Interim measures for the management of generative artificial intelligence services. *China Law Translate*. https://www.chinalawtranslate.com/generative-ai-interim/

Department of Industry, Science and Resources. (2023a). *Safe and responsible AI in Australia*. Australian Government. https://storage.googleapis.com/converlens-

au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf

Department of Industry, Science and Resources. (2022, October 5). *Australia's AI ethics principles: Australia's artificial intelligence ethics framework*. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

Department of Industry, Science and Resources. (2023b, June 1). *Supporting responsible AI: discussion paper*. https://consult.industry.gov.au/supporting-responsible-ai

Deveau, R., Griffin, S. J., & Reis, S. (2023, May 11). *Marketing and sales soar with generative AI*. McKinsey. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/ai-powered-marketing-and-sales-reach-new-heights-with-generative-ai

European Parliament. (2023, June 8). *EU AI Act: First regulation on artificial intelligence.* Updated June 16, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Fuller, J. B., Raman, M., Sage-Gavin, E., & Hines, K. (2021). *Hidden workers: Untapped talent*. Harvard Business School. https://www.hbs.edu/managing-the-future-of-work/Documents/research/hiddenworkers09032021.pdf

Gov.uk. (2023, June 22). *A pro-innovation approach to AI regulation*. GOV.UK. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Habuka, H. (2023). *Japan's approach to AI regulation and its impact on the 2023 G7 presidency*. Centre for Strategic and International Studies. https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency

Halpert, M. (2022, October 9). AI-powered job recruitment tools may not improve hiring diversity, experts argue. *Forbes*. https://www.forbes.com/sites/madelinehalpert/2022/10/09/ai-powered-job-recruitment-tools-may-not-improve-hiring-diversity-experts-argue/

High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission.

Hsu, T., & Myers, S. L. (2023, June 25). A.I.'s use in elections sets off a scramble for guardrails. *The New York Times*. https://www.nytimes.com/2023/06/25/technology/ai-elections-disinformation-guardrails.html

Hurcombe, D. P.-L., Neo, H. Y., & Wong, D. (2023, July 19). *China: New measures on generative artificial intelligence*. Lexology. https://www.lexology.com/library/detail.aspx?g=c7ca9dd4-7220-457c-b272-79c35f58ba91

Infocomm Media Development Authority. (2018). *Artificial intelligence governance and ethics initiatives*. Infocomm Media Development Authority. https://www.imda.gov.sg/-/media/imda/files/about/media-releases/2018/2018-06-05-fact-sheet-for-ai-govt.pdf

Infocomm Media Development Authority. (2022a). *Invitation to pilot A.I. Verify AI governance testing framework & toolkit*. Infocomm Media Development Authority. https://file.go.gov.sg/aiverify.pdf

Infocomm Media Development Authority. (2022b, May 25). Singapore launches world's first AI testing framework and toolkit to promote transparency; invites companies to pilot and contribute to international standards development [Press Release]. https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2022/singapore-launches-worlds-first-ai-testing-framework-and-toolkit-to-promote-transparency-invites-companies-to-pilot-and-contribute-to-international-standards-development

Korolov, M. (2022, June 13). *9 ways hackers will use machine learning to launch attacks*. CSO Online. https://www.csoonline.com/article/3250144/6-ways-hackers-will-use-machine-learning-to-launch-attacks.html

Liu, C. (2023, March 17). School students in East China's Zhejiang to study AI as compulsory course. *Global Times*. https://www.globaltimes.cn/page/202303/1287483.shtml

Luo, B. L.-Y., Dan, X., & Ponder, J. (2023, July 14). *Key takeaways from China's finalized generative artificial intelligence measures*. Lexology. https://www.lexology.com/library/detail.aspx?g=1bd75a42-358c-4984-a0df-31d12f37b9a8

Manning, C. (2022). *Artificial intelligence definitions.* Stanford university Human-Centred AI. https://hai.stanford.edu/sites/default/files/2023-03/AI-Key-Terms-Glossary-Definition.pdf

Marr, B. (2023, March 1). *The best examples of what you can do with ChatGPT*. *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/03/01/the-best-examples-of-what-you-can-do-with-chatgpt/

McKinsey & Company. (2023, January 19). *What is ChatGPT, DALL-E, and generative AI?.* McKinsey. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

Ministry of Communications and Information. (2020, February 3). *MCI's response to PQ on the work of the Advisory Council on the Ethical Use of AI and Data*. Question for Written Answer, Parliament Sitting on 3 February 2020. https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2020/2/mci-response-to-pq-on-the-work-of-the-advisory-council-on-the-ethical-use-of-ai-and-data

Ministry of Communications and Information. (2023a, April 21). *MCI Response to PQ on Regulatory Framework for Artificial Intelligence Governance in Singapore. Ministry of Communications and Information*. Question for Oral Answer, Parliament Sitting on 21 April 2023. https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2023/4/mci-response-to-pq-on-regulatory-framework-for-artificial-intelligence-governance-in-singapore?category=Cyber+Security

Ministry of Communications and Information. (2023b, May 9). *MCI Response to PQ on Ensuring Development and Maintenance of Ethical Artificial Intelligence Standards*. Question for Oral Answer, Parliament Sitting on 9 May 2023. https://www.mci.gov.sg/pressroom/news-and-stories/pressroom/2023/5/mci-response-to-pq-on-ensuring-development-and-maintenance-of-ethical-artificial-intelligence-standards?category=Cyber+Security

Monetary Authority of Singapore. (2018a). *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*. https://www.mas.gov.sg/-/media/MAS/News-and-Publications/Monographs-and-Information-Papers/FEAT-Principles-Updated-7-Feb-19.pdf

Monetary Authority of Singapore. (2018b, November 12). *MAS introduces new FEAT Principles to promote responsible use of AI and data analytics* [Press

Release]. https://www.mas.gov.sg/news/media-releases/2018/mas-introduces-new-feat-principles-to-promote-responsible-use-of-ai-and-data-analytics

Monetary Authority of Singapore. (2022, February 4). *MAS-led industry consortium publishes assessment methodologies for responsible use of AI by financial institutions* [Press Release]. https://www.mas.gov.sg/news/media-releases/2022/mas-led-industry-consortium-publishes-assessment-methodologies-for-responsible-use-of-ai-by-financial-institutions

Monetary Authority of Singapore. (2023, June 26). *MAS-led industry consortium releases toolkit for responsible use of AI in the financial sector* [Press Release]. https://www.mas.gov.sg/news/media-releases/2023/toolkit-for-responsible-use-of-ai-in-the-financial-sector

National Research Foundation & Infocomm Media Development Authority. (2018). *CFP — research programme on the governance of AI and data use*. https://researchgrant.gov.sg/Pages/GrantCallDetailDocs.aspx

NetHope. (n.d.). *The artificial intelligence (AI) suitability toolkit for nonprofits*. Retrieved June 28, 2023, from https://nethope.org/toolkits/the-artificial-intelligence-ai-suitability-toolkit-for-nonprofits/

Nicoletti, L., & Bass, D. (2023, June 9). Humans are biased. generative AI is even worse. *Bloomberg*. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Pandey, K. (2023, June 12). *UNESCO discusses IP and data ownership in the era of generative AI.* MediaNama. https://www.medianama.com/2023/06/223-unesco-ip-generative-ai-data-ownership-copyright/

Personal Data Protection Commission Singapore. (n.d.). *Singapore's approach to AI governance*. Retrieved June 28, 2023, from https://www.pdpc.gov.sg/Help-and-Resources/2020/01/Model-AI-Governance-Framework

Personal Data Protection Commission Singapore. (2018). Discussion paper on AI and PD. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf

Personal Data Protection Commission Singapore. (2020). *Model Artificial Intelligence Governance Framework Second Edition*. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

Personal Data Protection Commission Singapore. (2023, June 7). *Launch of AI Verify Foundation to shape the future of AI standards through collaboration*. https://www.pdpc.gov.sg/News-and-Events/Announcements/2023/06/Launch-of-AI-Verify-Foundation-to-Shape-the-Future-of-AI-Standards-Through-Collaboration

Potkin, F., & Wongcha-um, P. (2023, June 16). Exclusive: Southeast Asia to set "guardrails" on AI with new governance code. *Reuters*. https://www.reuters.com/technology/southeast-asia-set-guardrails-ai-with-new-governance-code-sources-2023-06-16/

Sha, A. (2019, September 6). *10 best deepfake apps and websites you can try for fun*. Beebom. https://beebom.com/best-deepfake-apps-websites/

Sheehan, M. (2023, July 10). *China's AI regulations and how they get made*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

Sheehan, M., & Du, S. (2022, December 9). *What China's algorithm registry reveals about AI Governance*. Carnegie Endowment for International Peace. https://carnegieendowment.org/2022/12/09/what-china-s-algorithm-registry-reveals-about-ai-governance-pub-88606

Smart Nation Digital Government Group. (n.d.). *Public Sector AI Playbook*. Singapore Government Developer Portal.

Smart Nation Singapore. (n.d.). *National AI strategy*. Retrieved June 28, 2023, from https://www.smartnation.gov.sg//initiatives/artificial-intelligence/

Stahl, A. (2023, April 10). *The rise of generative AI: A new era of creativity and innovation*. LinkedIn. https://www.linkedin.com/pulse/rise-generative-ai-new-era-creativity-innovation-alexander-stahl/

Startup SG. (n.d.). *Startup SG — the Singapore startup ecosystem*. Retrieved July 3, 2023, from https://www.startupsg.gov.sg/directory/startups/?sector=638&sort=-changed&q

The AI Act. (2021, September 7). *The Artificial Intelligence Act*. https://artificialintelligenceact.eu/

The Global Partnership on Artificial Intelligence. (2020, June). *About — GPAI*. https://gpai.ai/about/

The Organization for Economic Co-operation and Development. (n.d.). *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments. Retrieved June 28, 2023, from https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

The Organization for Economic Co-operation and Development. (2019, May 22). *Forty-two countries adopt new OECD principles on artificial intelligence*. https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm

The Organization for Economic Co-operation and Development. (2020, June 15). *OECD to host secretariat of new global partnership on artificial intelligence*. https://www.oecd.org/newsroom/oecd-to-host-secretariat-of-new-global-partnership-on-artificial-intelligence.htm

The White House. (2022). *Blueprint for an AI Bill of Rights*. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

United Nations Educational, Scientific and Cultural Organization. (2023, April 20). *UNESCO releases report on the mapping of K-12 artificial intelligence curricula*. https://www.unesco.org/en/articles/unesco-releases-report-mapping-k-12-artificial-intelligence-curricula

White, M. (2023, January 8). A brief history of generative AI. *Medium*. https://matthewdwhite.medium.com/a-brief-history-of-generative-ai-cb1837e67106

Wu, Y. (2023, July 27). *China's interim measures to regulate generative AI services: Key points*. China Briefing News. https://www.china-briefing.com/news/how-to-interpret-chinas-first-effort-to-regulate-generative-ai-measures/

Wu, Y. (2022, October 14). *AI in China: Regulations, market opportunities, challenges for investors*. China Briefing News. https://www.china-briefing.com/news/ai-in-china-regulatory-updates-investment-opportunities-and-challenges/

Ye, J. (2023, July 14). China says generative AI rules to apply only to products for the public. *Reuters*. https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/

Yeong, Z. K., Lee, W. S., & Tan, W. R. (2023, January 9). *How Singapore is creating a global trustworthy AI solution*. World Economic Forum.

https://www.weforum.org/agenda/2023/01/how-singapore-is-demonstrating-trustworthy-ai-davos2023/

APPENDIX: ABOUT THE AUTHORS

Carol **SOON** is Principal Research Fellow at the Institute of Policy Studies (Lee Kuan Yew School of Public Policy, National University of Singapore) where she heads the Society and Culture department. Her research interests include false information, media regulation, digital inclusion and public engagement. She has published her research in books and peer-reviewed journals such as the *Journal of Computer-Mediated Communication*, *Asian Journal of Communication* and *Public Integrity*. She is also Associate Director of the Asia Journalism Fellowship and Vice Chair of Singapore's Media Literacy Council.

Beverly **TAN** is a Research Assistant at the Institute of Policy Studies (Lee Kuan Yew School of Public Policy, National University of Singapore). Her work focuses on the social and policy implications of digital media and the Internet, including online harms and safety, digital literacy and artificial intelligence. She holds a bachelor's degree in communication studies from the Wee Kim Wee School of Communication and Information at Nanyang Technological University.

About IPS Working Paper Series

The IPS Working Papers Series is published in-house for early dissemination of works-in-progress. This may be research carried out by IPS researchers, work commissioned by the Institute or work submitted to the Institute for publication.

The views expressed in the Working Papers are strictly those of the author(s) alone and do not necessarily reflect the views of the IPS.

Comments on the Working Papers are invited. Please direct your comments and queries to the author(s).

IPS Working Papers are available from the IPS at $7.00 each (before GST). Postage and handling charges will be added for mail orders.

For more information, please visit www.lkyspp.nus.edu.sg/ips or
contact email: ips@nus.edu.sg or tel: 6516-8388.

Institute of Policy Studies
Lee Kuan Yew School of Public Policy
National University of Singapore
1C Cluny Road House 5
Singapore 259599

Tel: (65) 6516 8388
Web: www.lkyspp.nus.edu.sg/ips
Registration Number: 200604346E