IMPERIAL | Trusted AI ALLIANCE

Considerations regarding Sovereign AI and National AI Policy

By David L. Shrier[1,3], Ayisha Piotti[2], Alex Pentland[3,4] & A. Aldo Faisal[1,5,6]

[1] Imperial College London, [2] ETH Zurich, [3] Massachusetts Institute of Technology, [4] Stanford University, [5] Universität Bayreuth, [6] Alan Turing Institute

This paper was made possible through a collaboration with Payments Network Malaysia Sdn Bhd (Paynet). The authors would like to gratefully acknowledge the assistance of Farhan Ahmad and Endry Lim in crafting and revising this document, as well as research assistance from Adele Jashari and Gemma Bagchi and technical contributions from Yves-Alexandre de Montjoye.

V1.1 DRAFT November 2024

Contact:
david.shrier@imperial.ac.uk
aldo.faisal@imperial.ac.uk

Executive Summary

The concept of sovereign AI is gaining traction globally as nations seek to develop AI capabilities independent of a few US-based or PRC-based Big Tech companies (FANGMs plus BATs). Sovereign AI initiatives are under way in domiciles (or regional authorities within countries) including Singapore, Japan, Germany, France, the UK, KSA, UAE, India and others. Countries must balance the significant financial and technological investments that may be required with other national priorities. Options for nation-states, supranational federals (e.g., EU) and international alliances (e.g., ASEAN) include building sovereign AI, collaborating regionally, and leveraging open-source initiatives. Sovereign AI will participate in the broader structural shifts that AI is introducing globally, not only to economic and business structures, but in fundamental ways that people interact with each other, with companies, and with governments.

Part I - Context

Motivation for Sovereign AI:

Sovereign nations are initiating AI projects over which they have greater control and influence due to a number of factors, spanning political, economic and cultural spheres.
These include:

- Stragetic Autonomy
- Cultural and Linguistic Relevance
- Economic Benefits
- Protecting Privacy
- Mitigating Biases and Discrimination
- Addressing Large-Scale Misinformation and Deep Fakes
- Guarding Against Mass Surveillance and Economic Disruptions
- Navigating Jobs Disruption
- Protecting Intellectual Property Rights
- Building Resilience Vs. Trade Actions
- Protecting Against Existential Risk

Trusted AI ALLIANCE
aicollab.org

Overview of Technology:

- **What is a GPT?:** GPT (Generative Pretrained Transformer) is a powerful AI model that can perform various tasks like writing, answering questions, and translating languages by understanding and generating human-like text.
- **History and Evolution:** AI has evolved over 80 years, with significant milestones like machine learning and deep learning. The open-sourcing of TensorFlow and transformer technology in 2015 and 2018, respectively, revolutionized AI, particularly in natural language processing.

Use Cases:

- **Financial Services:** AI enhances fraud detection, personalized banking, algorithmic trading, and financial advice.

- **Healthcare:** AI aids in diagnosis, preventative care, drug discovery, and medical transcription.
- **Education:** AI supports personalized learning, administrative efficiency, language learning, and writing assistance.

Strategic Building Blocks:

Several critical elements form the strategic building blocks of Sovereign AI.

Governance and Regulatory Framework

Pro-Innovation AI Policy

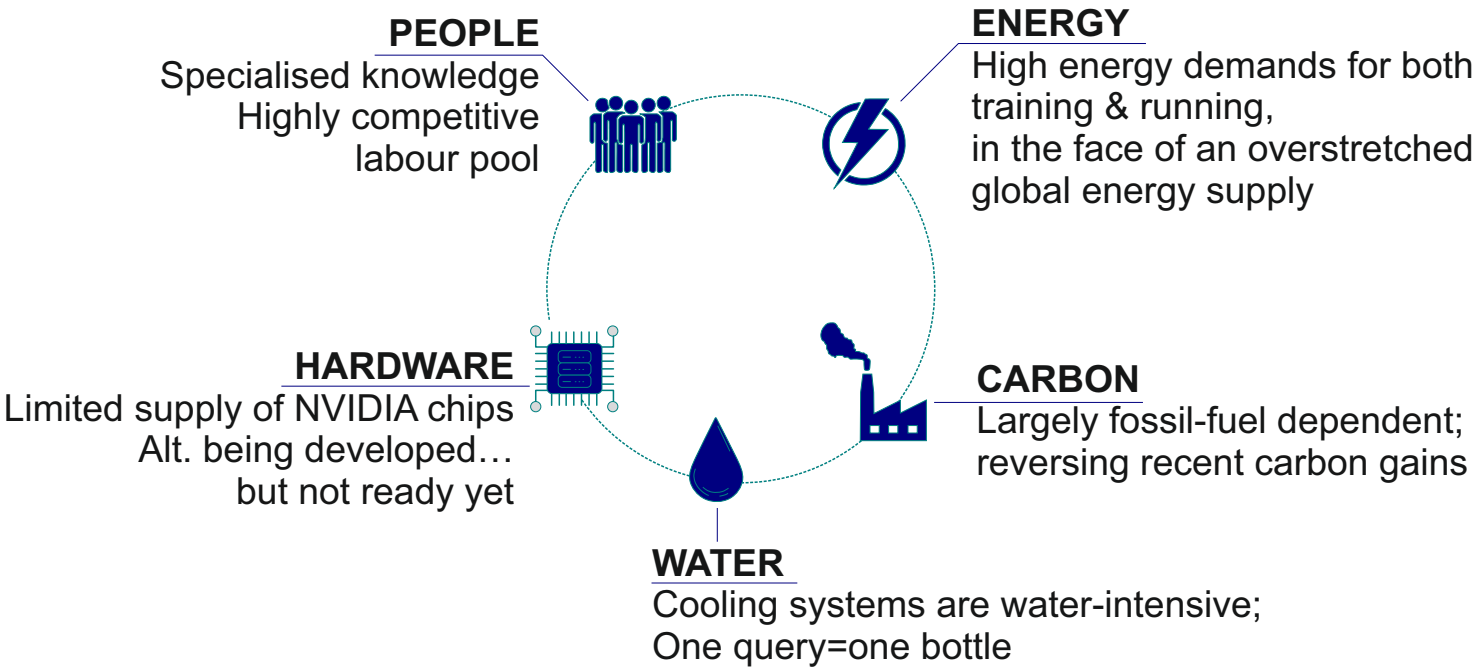
National AI Security & Safety

Data Sovereignty, Authenticity, Consent, and Provenance

Digital Infrastructure Requirements

Part II – Socio-technic Systems Considerations

There are 5 key building blocks in order to implement Sovereign AI capabilities, which have a number of embedded challenges



Determining Sovereign Value Proposition: Policymakers need to assess their domicile's infrastructure and capabilities against critical factors needed for creation and maintenance of sovereign AI, namely access to power, water, hardware, talent, and carbon.

Part III — Policy Frameworks

• **Principles and Regulations**: Core global AI principles such as transparency, safety, inclusivity, accountability, and fairness are emphasized, with examples like the EU AI Act and the Council of Europe Binding Convention demonstrating how these principles evolve into binding regulations. Adaptation to local contexts and the transition from principles to enforceable laws are highlighted.

• **Creating an Enabling Environment:** A domicile can foster an AI-friendly ecosystem through identifying policy gaps, adopting flexible regulations, building public sector capacity, supporting innovators, enhancing regulatory coherence, and raising public awareness. Sector-specific regulatory approaches and international collaboration are also crucial.

• **Horizon Scanning:** Brief overview of AI policies in various countries (e.g., EU, Japan, USA) and multilateral organizations (e.g., OECD, WEF, IEEE, UN), highlighting their focus areas and frameworks for responsible AI development. These insights aim to inform national AI strategy by drawing on global best practices and standards.

Part IV – Strategic Considerations

Policymakers have several options to take into account with respect to Sovereign AI:

●**Industry-Specific Sovereign AI:** Focusing on specific sectors like financial services.

● **Creating Sovereign AI:** Developing national AI capabilities with significant investments.

● **Partnering with Big Tech:** Collaborating with companies that support local capabilities.

● **Adapting Open Source Code:** Leveraging open-source AI projects for non-aligned development.

● **Harvesting Benefits of Other Initiatives:** Integrating national projects like digital identity into the AI strategy.

● **Exploring Alternatives to Sovereign AI:** investigate other choices versus a full-blown sovereign, AI, such as decentralized inference systems.

● **Wait and See:** Monitoring the technology landscape while assessing risks and opportunities.

Moving Forward:

● Form a high-level working group to assess options.
● Participate in multilateral dialogues for insights.
● Develop refined sovereign AI policy.

Table of contents

1 Introduction 8 8

2 Part I - Context 10 10

A. Motivation for a Sovereign GPT 10 10

B. Brief Explanation of Technology for Non-Technologists 14 14

C. Strategic Building Blocks for National AI Strategy 16 16

3 Part II - Socio-technic Systems Considerations 27 27

A. Capabilities and Motivation of Model Size 27 27

B. Small is Beautiful 30 30

C. Cybersecurity Requirements of Sovereign AI 30 30

D. Talent and AI Workforce Requirements 33 33

E. Innovation Ecosystem Support 36 36

F. Energy Requirements of AI 38 38

G. Net Carbon Impact of AI 38 38

H. Carbon and Energy Mitigation Strategies 39 39

I. Determining Sovereign Value Proposition 40 40

4 Part III - Policy Frameworks 41 41

i. Principles and Regulations 41 41

A. Core Principles forming the basis of Global policy initiatives 41 41

B. Adapting to Local Contexts 42 42

C. Evolution from Principles to Binding Regulations 42 42

ii. Creating an Enabling Environment 44 44

iii. Horizon Scanning 47 47

5 Part IV - Sovereign AI Strategic Considerations 53 53

1. Strategic Options 53 53

2. Creating sector-specific GPTs and Expert Systems 54 54

3. Counterpoint: Alternatives to a Sovereign AI 55 55

Next Steps 56 56

6 About the Authors 57 57

7 Appendix A: Select Survey of Open Source AI Systems 59 59

8 Appendix B: Summary of Relevant AI Policy Initiatives by Country 63

10 Appendix C: A Word About Quantum 117

11 Project History 118

Introduction

A topic that is gaining currency in policy circles globally is the idea that nation-states and supernational groupings (e.g., EU, African Union, GCC, ASEAN, etc.) should develop AI capability that is outside of the exclusive control of a handful of mostly US-based, Big Tech companies. This is happening at different levels of political organisation: nation-states like Japan, India, Switzerland and Singapore, subnational governments like Bavaria (Germany), Abu Dhabi (UAE), and supranationals and/or borderless initiatives like JAIS.

Artificial intelligence, and related technologies like quantum computing, are anticipated to deliver profound impacts to every manner in which we interact with the world around us. According to venture studio Visionary Future and investment bank Evercore ISI, artificial intelligence could add as much as $12 trillion to global GDP by 2032, perhaps doubling GDP growth over the next decade. AI will become increasingly integrated into everyday devices, enabling more natural interfaces, highly personalized experiences, and seamless automation of complex tasks.

As AI systems gain greater sophistication, we may see the emergence of bilateral and multilateral agent activity: an individual's AI agent will interact with the AI agents of governments and corporations, creating a new dynamic for the primary way in which people communicate with institutions (particularly so-called utility services). It will radically simplify lives, reduce errors and improve services.

In concert with these technologies, cybersecurity will undergo major shifts as new generation AI systems challenge existing encryption and security protocols. New approaches will be needed to safeguard data and protect consumers and organizations.
Technology disruption from AI will attenuate the ability of any one nation to protect domestic interests; instead, it is critical to have a global lens when considering the technologies, their impacts, and their potential applications.

Lessons for how sovereigns should address AI can be drawn from other strategic technologies, spanning an array as diverse as semiconductors, nuclear, biotechnology and space launch capabilities. In each instance, considerations that transcend purely cost- or market access-driven decisions have impacted national investments into long-cycle 'deep tech' ecosystems. Political leaders determined that access to and control over these technologies were sufficiently important and urgent as to dedicate financial resources, focus, and regulatory changes to accommodate pursuing capabilities in these areas.

However, it is worth noting that, unlike (for example) space launch, artificial intelligence permeates the everyday life of a citizen across privacy, financial, social and other dimensions, creating impact that is far more pervasive. Accordingly, this creates an imperative for political leaders to act, and do so in a considered manner that takes into account a number of nuances inherent to AI.

Regional and Global Considerations

Countries ranging across the Americas, Europe and Asia are all investing meaningfully into AI initiatives (see Appendix B for summary horizon scanning). Regional and global leadership around AI for the next several decades is being defined now and in the next 12 to 24 months, and this creates an imperative for policymakers to act sooner rather than later. Within nations, certain regions have decided to make significant investments, for example Bavaria in Germany and Flanders in Belgium.

Not all nations are equipped, financially or technologically, to pursue alone the considerable investments (billions of dollars US) associated with sovereign A. Even growing economic powers have numerous demands on government coffers, from health care to education, and will need to carefully consider decisions with respect to AI. The domain remains important to the national interest, and there are a number of options a sovereign policy body can consider alongside, or in lieu of, building its own sovereign AI, including collaborating with regional partners and leveraging open-source initiatives.

It's notable that centers of excellence around generative AI emerged in only a handful of countries. A concentration of factors made this possible, including: an open research landscape, access to specialised high-performance compute, a stable energy grid with suitable capacity, highly skilled talent with specific expertise around AI and a supportive innovation environment with appropriate capital availability.

In this paper, we will outline key considerations in creating sovereign AI, the risks and resource requirements associated with it, some alternatives to wholesale pursuit of independent GPTs, and potential steps forward for policymakers to consider.

Part I – Context

A. Motivation for Sovereign GPT

1. Strategic Autonomy

Ensuring national control over critical AI infrastructure. Developing sovereign AI capabilities is crucial for ensuring a nation's strategic autonomy. Relying on foreign AI technologies can create dependencies that may compromise a country's ability to make independent decisions in critical areas such as national security, economic policy, and technological innovation. By fostering domestic AI development, countries can retain control over their technological infrastructure, protect sensitive data, and reduce vulnerabilities to external pressures or coercion. This autonomy allows nations to pursue their strategic interests without undue influence from foreign entities, ensuring that their AI policies and applications align with national priorities and values.

2. Cultural and Linguistic Relevance:

AI technologies developed externally may not adequately capture or respect the cultural and linguistic nuances of a given country. This can lead to the deployment of AI systems that are ineffective or even harmful, as they may fail to understand or appropriately respond to local customs, languages, and societal norms. Developing sovereign AI capabilities allows nations to create technologies that are tailored to their unique cultural and linguistic contexts. This ensures that AI applications are more accurate, relevant, and respectful of local traditions and values, thereby enhancing user acceptance and trust in AI systems. In the MENA region, JAIS has emerged as an open-source Arabic GPT. Singapore has launched SEA-LION, an effort to provide a regional LLM to southeast Asia. Other efforts are emerging around culturally and linguisticallydistinct GPTs.

3. Economic Benefits

Enhancing local AI capabilities can drive innovation and economic growth. AI (including GPTs) is expected to generate as much as US$ 12 trillion of GDP growth by 2032 (Evercore ISI; Visionary Future LLC). Investing in sovereign AI capabilities can generate significant economic benefits for a country. By nurturing a domestic AI industry, nations can stimulate job creation, foster innovation, and attract investments in technology sectors. This can lead to the development of new industries and the revitalization of existing ones, driving economic growth and increasing global competitiveness. Additionally, retaining AI development within national borders helps to ensure that the economic value generated by these technologies, including profits, intellectual

property, and expertise, remains within the country. This not only strengthens the national economy but also reduces the risk of economic exploitation by foreign entities.

4. Protecting Privacy

The development and deployment of AI technologies by foreign entities pose significant risks to the privacy of citizens in any given country. When AI systems are created and managed by external parties, there is a higher likelihood of data breaches and unauthorized access to sensitive personal information. This can lead to various privacy violations, such as the misuse of data for surveillance or commercial purposes without individuals' consent. By establishing sovereign AI capabilities, countries can ensure that the handling and processing of personal data adhere to their specific legal frameworks and cultural norms, thereby protecting citizens' privacy more effectively.

5. Mitigating Biases and Discrimination

AI systems trained on data from different cultural and societal contexts may inadvertently embed biases and perpetuate discrimination when applied in a foreign setting. These biases can lead to unfair treatment and reinforce existing inequalities, as the AI may not accurately reflect the values or demographic realities of the country using the technology. Developing sovereign AI capabilities allows nations to tailor their systems to address and mitigate local biases, ensuring fairer and more equitable outcomes. This can foster trust in AI applications and promote social justice by making sure that the technology aligns with national ethical standards and diversity considerations.

6. Addressing Large-Scale Misinformation

The proliferation of AI-driven content creation tools has made it easier to generate and disseminate misinformation on a massive scale. When these tools are controlled by external entities, it becomes challenging to regulate and monitor the spread of false information that can undermine public trust and destabilize societies. Sovereign AI capabilities enable countries to develop and test robust mechanisms for detecting and countering misinformation tailored to their unique media landscapes and information ecosystems. This can help maintain the integrity of public discourse, protect democratic processes, and ensure that citizens receive accurate and reliable information.

7. Revealing Deep Fakes

The advent of deep fake technology, which uses AI to create highly realistic but fake audio and video content, poses a significant threat to national security and public trust. Deep fakes can be used to manipulate public opinion, discredit individuals, and incite social unrest. When the technology behind deep fakes is controlled by foreign entities, it becomes difficult for countries to develop effective countermeasures and regulatory frameworks. By investing in sovereign AI capabilities, nations can enhance their ability to detect and combat deep fakes, protecting their citizens from deception and ensuring the authenticity of digital media.

8. Defending Against Cyber Attack

More widely-available, more powerful AI systems have already increased risks of fraud and cyber crime across numerous nations. With resilience hindered by technology debt and often-limited budgets, nation-states and supranationals need to reconsider cybersecurity posture as generative AI creates a growing array of new threats. For several years, there has been a cat-and-mouse game of cyber infiltration and sabotage that includes state-sponsored activities by networks of cyber criminals. New generation AI systems have increased the sophistication, severity and scale of these attacks, creating new urgency for a better and better-coordinated defense.

9. Protecting Against Manipulation of Behavior

AI systems can be designed to influence and manipulate human behavior, often without individuals' awareness. This can be particularly concerning when such technologies are controlled by external actors with different agendas and interests. These AI-driven manipulations can affect consumer choices, voting behaviors, and even social dynamics, posing a threat to national sovereignty. Developing sovereign AI capabilities allows countries to implement ethical guidelines and oversight mechanisms to prevent undue manipulation and ensure that AI technologies are used to enhance, rather than exploit, human decision-making.

10. Guarding Against Mass Surveillance

The use of AI for mass surveillance by foreign entities raises significant concerns about national security and individual freedoms. When surveillance technologies are imported and controlled externally, there is a risk of sensitive data being accessed and exploited by foreign governments or corporations. Sovereign AI capabilities enable countries to develop and manage their surveillance systems in line with national security needs and privacy laws. This ensures that surveillance practices are transparent, accountable, and respectful of citizens' rights, while also protecting the nation from external threats.

11. Offsetting Economic Disruptions

The integration of foreign AI technologies into national economies can lead to economic dependencies and vulnerabilities. These technologies can disrupt local industries, alter labor markets, and create imbalances in economic power. By developing sovereign AI capabilities, countries can foster innovation, support local businesses, and create jobs within their own borders. This can lead to more resilient and self-sufficient economies that are better equipped to adapt to the rapid changes brought about by AI advancements.

12. Navigating Jobs Disruption

The adoption of AI technologies can significantly impact the job market, potentially leading to job displacement and changing the nature of work. When AI systems are developed and controlled by foreign entities, there is a risk that the economic benefits and job opportunities generated by these technologies will not be equitably distributed. By investing in sovereign AI capabilities, countries can ensure that the development and deployment of AI technologies are aligned with

national labor market strategies. This can help to create new job opportunities, facilitate workforce retraining, and promote inclusive economic growth.

13. Protecting Intellectual Property Rights

The control of AI technologies by foreign entities can lead to challenges in protecting intellectual property (IP) rights. When AI systems are developed externally, there is a risk of IP theft, patent infringement, and loss of competitive advantage. Sovereign AI capabilities enable countries to develop and protect their own AI innovations, ensuring that the economic and strategic benefits of these technologies are retained within national borders. This can foster a robust innovation ecosystem and protect the interests of domestic researchers, entrepreneurs, and businesses.

14. Building Resilience Against Trade Actions

The potential for a renewed US-Global trade war, particularly under a second Trump administration, could result in severe restrictions on the export of AI technologies. Such restrictions could limit access to critical AI tools and resources, hindering technological advancement and economic growth in affected countries. Developing sovereign AI capabilities allows nations to reduce their dependence on foreign AI technologies and mitigate the impact of trade restrictions. This can enhance national resilience and ensure continued progress in AI research and development, regardless of geopolitical tensions.

15. Protecting Against Existential Risk (X-Risk)

The rapid advancement of AI technologies has sparked discussions about their potential existential threat to humanity. When AI development is concentrated in a few powerful countries or corporations, there is a risk that these entities may not prioritize global safety and ethical considerations. Sovereign AI capabilities enable countries to participate actively in shaping the global AI landscape, ensuring that AI development is guided by diverse perspectives and aligned with human values. This collaborative approach can help to mitigate the risks associated with AI and ensure that its benefits are shared equitably across all of humanity.

Given how rapidly artificial intelligence is advancing, **we encourage governments to shape the next technological revolution in** the public's interest. What form that takes is being developed individually by each individual nation-state, and collaboratively in forums such as the World Economic Forum, the OECD and elsewhere. As part of this, the Trusted AI Alliance has initiated a series of dialogues around how AI should be oriented towards the benefit of humanity.

It should be noted that not all government motivations around sovereign AI are benign. Some governments are using these initiatives as a mechanism to exercise greater control over information, suppress dissidents, and limit freedom of movement and expression. When understanding cross-border and intercountry collaborations, the political context of a given instance of sovereign AI needs to be evaluated and understood, to frame how and where information, data and algorithms are permitted to interact with each other across different sovereign AI systems. The interoperability question has nontrivial ethical, moral and geopolitical dimensions.

B. Brief Explanation of Technology for Non-Technologists

• **What is a GPT?** GPT stands for “generative pretrained transformer”. It is a particularly flexible and powerful form of AI. Imagine having a computer program that can write essays, answer questions, create stories, and even translate languages just by understanding and generating human-like text, images and sounds. GPT’s don’t need specific instructions for each task; instead, they learn from vast amounts of text data to understand and generate language in a way that gives the appearance of human communication. While a national AI policy approach should not be limited to GPTs, they will form an integral part of a cohesive strategy.

• **History and Evolution:** AI has been developed over the past 80 years, after being invented during World War II. In the 1980s, machine learning was invented, a kind of AI that can find patterns within data. Earlier kinds of AI had to be specifically programmed for tasks, but machine learning systems can improve themselves without specifically being told what to do. Machine learning systems then gave rise to deep learning systems, where layer upon layer of neural networks were interconnected in a model that resembled how the human brain is structured. A proliferation of AI systems began in 2015 after Google open-sourced a major library of AI technology known as TensorFlow. They followed this in 2018 with the publication of research on transformer technology. Transformers are a type of model architecture used in artificial intelligence (AI), particularly in natural language processing (NLP). They have revolutionized the field by enabling machines to understand and generate human language with unprecedented accuracy and efficiency.

• **Use Cases:** AI can be used to improve processes and deliver new capabilities in a variety of industry sectors. Practical applications can be found in various sectors like financial services, healthcare, and education:

1. Financial Services:

• **Fraud Detection:** AI systems can analyze transactions in real-time to detect unusual patterns that might indicate fraud, helping to protect consumers and financial institutions.

• **Personalized Banking:** Banks use AI to offer personalized financial advice, helping customers manage their money more effectively by analyzing their spending habits and financial goals.

• **Efficiency Improvements:** Numerous middle and back-office tasks, which today require human beings, can be more effectively delivered by automated systems.

• **Algorithmic Trading:** More than 90% of trading volume on most major securities exchanges today are performed by computers, not people. AI trading continues to be a significant focus area for the financial services industry.

• **Finance Advice:** Increasingly, advice on investing and savings is being offered by computer systems rather than human financial advisors, making better guidance available to more of the population.

2. Healthcare:

• **Diagnosis and Treatment:** AI can assist doctors by analyzing medical images or patient data to help diagnose diseases more quickly and accurately. For instance, AI can help detect early signs of cancer from radiology scans.

• **Preventative Care:** Overall costs of healthcare in a country can be significantly reduced by wider adoption of preventative care, versus curative care. AI systems can help with tracking, reinforcement and recommendations for personalised care.

• **Drug Discovery:** AI speeds up the process of discovering new drugs by predicting which chemical compounds might be effective in treating certain diseases, saving time and resources in medical research.

• **Transcription:** Medical records automation is being assisted by AI transcription system, enabling healthcare providers to more rapidly convert patient notes into electronic form.

3. Education:

• **Personalized Learning:** AI can create customized learning experiences for students by adapting educational content to fit their individual needs, helping them to learn at their own pace.

• **Generative Instruction:** One-on-one instruction can be delivered to students at high quality with the introduction of properly-trained generative AI systems, that can create new content and interactively deliver existing content while offering a more individualised experience.

• **Administrative Efficiency:** Schools and universities use AI to automate administrative tasks, such as grading and scheduling, freeing up educators to focus more on teaching.

• **Language Learning:** AI can facilitate more rapid assimilation of new language skills, helping adapt a population to a globally-capable workforce.

• **Writing and Editing:** Students can develop strong writing and editing skills with the assistance of AI guidance.

C. Strategic Building Blocks for National AI Strategy

The strategic building blocks of national AI strategy include:

- Governance and Regulatory Framework
- Pro-Innovation AI Policy
- National AI Security & Safety
- Data Sovereignty, Authenticity, Consent, and Provenance
- Digital Infrastructure Requirements

i. Governance and Regulatory Framework on AI

Best practices exhibited in other national AI policy interventions provide guidelines for governance and for a regulatory framework regarding AI.

Governance Structure for AI

A robust governance structure is essential for the effective oversight and management of AI technologies. Establishing a central government unit dedicated to AI can provide a cohesive and coordinated approach to AI governance. This unit would serve multiple functions, including policy development, regulation, and coordination among various stakeholders. Key aspects of such a governance structure include:

1. **Centralized Leadership:** A central AI authority, such as a National AI Commission or Agency, can provide clear leadership and accountability. This entity would be responsible for setting national AI strategies, policies, and priorities, ensuring alignment with broader national interests and goals.

2. **Interagency Coordination:** The central AI unit would facilitate coordination between different government departments and agencies involved in AI-related activities. This ensures a unified approach to AI governance, avoiding duplication of efforts and ensuring consistency in policy implementation.

3. **Stakeholder Engagement:** A central AI unit would actively engage with a broad range of stakeholders, including industry leaders, academic researchers, civil society organizations, and the public. Regular consultations and collaborations with these stakeholders can ensure that AI policies are informed by diverse perspectives and address real-world challenges.

4. **Advisory Committees:** Establishing advisory committees composed of experts from various fields, such as technology, ethics, law, and social sciences, can provide valuable insights and guidance to the central AI unit. These committees can help in evaluating emerging trends, assessing risks, and recommending policy actions.

5. **International Collaboration:** The central AI unit would also play a crucial role in international collaboration, representing the country's interests in global AI forums, participating in international standard-setting, and fostering cross-border partnerships and cooperation.

Regulatory Framework

An effective regulatory framework for AI is crucial to ensure that AI technologies are developed and deployed responsibly, ethically, and safely. This framework should be comprehensive yet flexible enough to adapt to the rapidly evolving nature of AI. Key components of a robust AI regulatory framework include:

1. **Ethical Guidelines and Principles:** The regulatory framework should be grounded in clear ethical guidelines and principles, such as fairness, transparency, accountability, and respect for human rights. These principles should guide all aspects of AI development, deployment, and use.

2. **Risk-Based Regulation:** Adopting a risk-based approach to regulation, such as the approach taken by the EU, ensures that the level of regulatory oversight is proportional to the potential risks associated with different AI applications. High-risk applications, such as those in healthcare, finance, and public safety, would be subject to more stringent regulations and scrutiny, while lower-risk applications would face lighter regulatory requirements.

3. **Standards and Certification:** Developing and enforcing technical standards and certification processes for AI systems can ensure that these technologies meet specific safety, reliability, and ethical criteria. Certification programs can help build trust in AI technologies and provide assurance to users and consumers.

4. **Transparency and Explainability:** Requiring AI developers to ensure that their systems are transparent and explainable can help mitigate risks and enhance accountability. This includes mandates for clear documentation of AI algorithms, decision-making processes, and data usage.

5. **Data Protection and Privacy:** Strong data protection and privacy regulations are essential to safeguard individuals' personal information and prevent misuse of data. The regulatory framework should include provisions for data minimization, consent, anonymization, and secure data storage and sharing practices.

6. **Accountability Mechanisms:** Establishing clear accountability mechanisms, such as liability rules and enforcement procedures, ensures that AI developers and users are held

responsible for the outcomes of their AI systems. This can include requirements for regular audits, impact assessments, and reporting obligations.

7. Regulatory Sandboxes: Implementing regulatory sandboxes can allow for experimentation and innovation in a controlled environment. These sandboxes enable AI developers to test new technologies and business models under regulatory oversight, ensuring that they comply with ethical and safety standards before wider deployment. These instruments are also a way for regulatory authorities to learn from industry players, identify potential risks, and create practical regulatory solutions.

8. Public Awareness and Education: Promoting public awareness and education about AI technologies and their implications is crucial for fostering informed and engaged citizens. The regulatory framework should include initiatives to raise awareness, provide education and training, and facilitate public participation in AI governance.

9. Periodic Reviews and Updates: The regulatory framework should be dynamic and subject to regular reviews and updates to keep pace with technological advancements and emerging challenges. This ensures that regulations remain relevant, effective, and responsive to the evolving AI landscape.

By establishing a central governance structure and a comprehensive regulatory framework, nations can effectively oversee and manage the development and deployment of AI technologies. This approach ensures that AI advancements contribute positively to society while minimizing risks and addressing ethical, legal, and societal concerns.

ii. Pro-innovation AI Policy

To foster innovation in AI, nations need to move away from rigid legislative requirements that can stifle creativity and slow down technological advancement. Instead, a flexible and adaptive regulatory framework should be established. This framework should be characterized by:

1. **Agile Regulation:** Implementing a dynamic regulatory approach that can be quickly updated in response to technological advancements. This includes periodic reviews and updates of regulations to keep pace with the rapidly evolving AI landscape.

2. **Sandbox Environments:** Creating regulatory sandboxes where AI developers can test new technologies in a controlled environment without being subject to the full spectrum of regulations. This allows for experimentation and innovation while ensuring safety and ethical considerations.

3. **Outcome-Based Regulation:** Focusing on the outcomes rather than the processes. Regulations should set clear goals and desired outcomes, such as safety, transparency, and fairness, but allow flexibility in how these goals are achieved.

4. **Stakeholder Engagement:** Involving a broad range of stakeholders, including AI developers, industry leaders, academic researchers, and civil society, in the legislative process. This ensures that regulations are informed by diverse perspectives and address the practical realities of AI development and deployment.

Intellectual Property Rights and Data Sharing Frameworks

A pro-innovation AI policy should also address the critical areas of intellectual property (IP) rights and data sharing frameworks. These components are vital for fostering innovation and collaboration while protecting the rights of creators and users.

1. Intellectual Property Rights:

- **Balanced IP Laws:** Crafting IP laws that protect the interests of AI developers and innovators without hindering collaboration and the free flow of ideas. This might involve shorter patent periods for AI technologies, given the rapid pace ofadvancement in the field.

- **Open Source and Licensing Models:** Encouraging the use of open-source licenses and collaborative development models to promote innovation and the dissemination of AI technologies. Clear guidelines on licensing and usage rights can help mitigate legal uncertainties.

2. Data Sharing Frameworks:

- **Privacy and Security Standards:** Establishing robust standards for data privacy and security to protect individuals' data while enabling data sharing for AI development. This includes PII removal, anonymization and encryption techniques to safeguard personal information.

- **Data Trusts and Cooperatives:** Creating data trusts or cooperatives where data is shared and managed collectively by multiple stakeholders under agreed-upon rules. These entities can help facilitate data sharing while ensuring that data usage is transparent and aligned with ethical standards.

- **Interoperability Standards:** Promoting interoperability standards that enable seamless data exchange between different AI systems and platforms. This can accelerate innovation by allowing developers to build on existing data and technologies.

Principles-Based Policy Interventions and Capacity Building

Principles-based policy interventions offer a flexible and adaptive approach to AI regulation. These interventions are guided by overarching principles that can be adapted to specific contexts and evolving technologies. Coupling these with capacity building for government officials ensures effective implementation and oversight.

1. Principles-Based Policy Interventions:

- **Ethical AI Principles:** Establishing core ethical principles for AI development and deployment, such as fairness, transparency, accountability, and inclusivity. These principles serve as the foundation for more specific regulations and guidelines.

- **Risk-Based Approach:** Adopting a risk-based approach to regulation, where the level of regulatory scrutiny is proportional to the potential risks associated with the AI application. Low-risk applications may require minimal oversight, while high-risk applications undergo more stringent review.

- **Human-Centered AI:** Ensuring that AI technologies are designed and deployed with the well-being of humans in mind. This includes considerations of user rights, accessibility, and the impact on employment and social structures.

2. Capacity Building for Government Officials:

- **Training and Education:** Providing ongoing training and education for government officials to keep them updated on the latest developments in AI technologies, ethical considerations, and regulatory best practices.

- **Interdisciplinary Collaboration:** Encouraging collaboration between government officials and experts from various fields, including technology, ethics, law, and social sciences, to develop well-rounded and informed policies.

- **Resource Allocation:** Allocating sufficient resources for the development and enforcement of AI policies, including funding for regulatory agencies, research initiatives, and public awareness campaigns.

- **Pilot Programs and Case Studies:** Implementing pilot programs and analyzing case studies to learn from real-world applications of AI policies. These insights can inform future policy adjustments and improvements.

By rethinking legislative requirements, establishing balanced IP and data sharing frameworks, and implementing principles-based interventions coupled with capacity building, nations can create a pro-innovation AI policy that fosters technological advancement while safeguarding ethical and societal values.

iii. National AI Security & Safety

Developing Protocols to Ensure National AI Systems are Safe and Reliable

To safeguard national security and ensure the reliability of AI systems, it is essential to develop comprehensive safety protocols. These protocols should encompass a range of measures to prevent malicious use, ensure robust system performance, and protect against potential threats.
Key elements include:

1. **Risk Assessment and Management:** Implementing rigorous risk assessment frameworks to identify and mitigate potential threats posed by AI systems. This includes evaluating the risks associated with AI applications in critical infrastructure, defense, and public safety, and developing contingency plans for potential failures or attacks.

2. **Secure Development Practices:** Establishing secure development practices to prevent vulnerabilities in AI systems. This involves incorporating security measures at every stage of the AI lifecycle, from design and development to deployment and maintenance. Techniques such as secure coding practices, regular security audits, and penetration testing should be mandated.

3. **Access Control and Authentication:** Ensuring that AI systems are protected against unauthorized access through robust access control and authentication mechanisms. This includes implementing multi-factor authentication, role-based access control, and encryption to safeguard sensitive data and system functionalities.

4. **Incident Response and Recovery:** Developing comprehensive incident response and recovery protocols to quickly and effectively address security breaches or system failures. This involves establishing response teams, conducting regular drills, and maintaining backup and recovery systems to ensure continuity of operations.

5. **Monitoring and Surveillance:** Implementing continuous monitoring and surveillance of AI systems to detect and respond to anomalies, suspicious activities, and potential threats in real-time. Advanced monitoring tools and techniques, such as anomaly detection algorithms and threat intelligence feeds, should be employed.

6. **Collaboration with Security Agencies:** Coordinating with national security agencies, law enforcement, and international partners to share intelligence, best practices, and resources for safeguarding AI systems. This collaborative approach enhances the overall security posture and enables a unified response to emerging threats.

Ethical Guidelines

Implementing Ethical Guidelines to Ensure AI Aligns with Societal Values

Ethical guidelines are crucial to ensure that AI systems align with societal values and address issues such as bias, fairness, and transparency. These guidelines provide a framework for the responsible development and deployment of AI technologies. Key components include:

1. **Fairness and Non-Discrimination:** Ensuring that AI systems are designed and trained to avoid bias and discrimination. This involves using diverse and representative datasets, conducting regular bias audits, and implementing fairness-enhancing algorithms to promote equitable outcomes.

2. **Transparency and Explainability:** Mandating transparency in AI decision-making processes to build trust and accountability. This includes requiring AI developers to provide clear documentation of how algorithms work, the data used, and the rationale

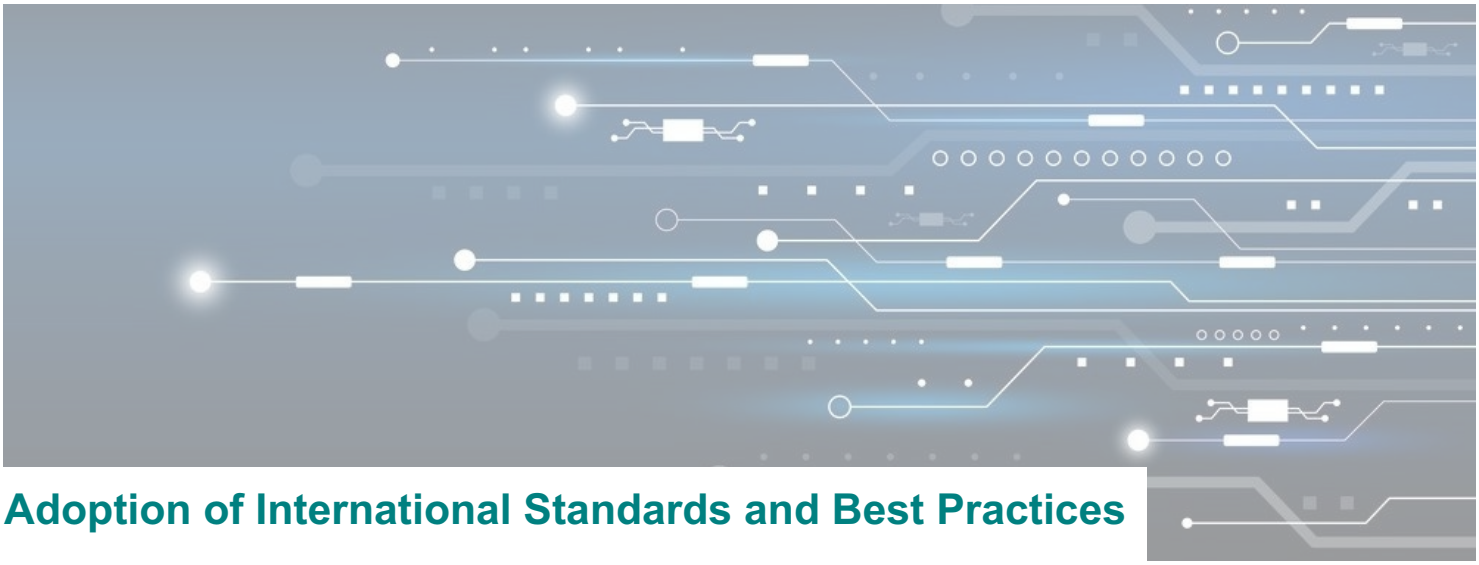
behind AI decisions. Explainable AI techniques should be employed to make AI systems more interpretable to users and stakeholders.

3. **Accountability and Responsibility:** Establishing clear accountability mechanisms to hold AI developers, operators, and users responsible for the outcomes of AI systems. This includes defining liability frameworks, enforcing compliance with ethical guidelines, and providing avenues for redress in case of harm or misuse.

4. **Human-Centric Design:** Prioritizing human well-being and autonomy in the design and deployment of AI systems. This involves ensuring that AI applications enhance, rather than undermine, human capabilities and decision-making. User-centric design principles should be applied to create AI systems that are intuitive, accessible, and aligned with human values.

5. **Privacy and Data Protection:** Upholding the highest standards of privacy and data protection in AI systems. This includes implementing data minimization, anonymization, and consent mechanisms to protect individuals' personal information. Ethical guidelines should also address the ethical implications of data usage and sharing.

6. **Ethical Review Boards:** Establishing ethical review boards to oversee the development and deployment of AI systems. These boards, composed of experts from various fields, can provide guidance, conduct ethical assessments, and ensure compliance with ethical standards.



Adoption of International Standards and Best Practices

Adopting international standards and best practices is essential for ensuring the interoperability, safety, and reliability of AI systems. These standards provide a common framework that facilitates global collaboration and ensures that AI technologies meet high-quality benchmarks. Key aspects include:

1. **Alignment with International Standards:** Ensuring that national AI policies and practices align with established international standards, such as those developed by the International Organization for Standardization (ISO) and the Institute of Electrical and Electronics Engineers (IEEE). This alignment promotes consistency and facilitates crossborder cooperation.

2. **Participation in Standard-Setting Bodies:** Actively participating in international standard-setting bodies to contribute to the development of new AI standards and best practices. This involvement ensures that national interests and perspectives are represented in global discussions and that emerging standards are relevant to the country's needs.

3. **Adoption of Best Practices:** Implementing best practices from leading AI research and industry organizations to enhance the safety, reliability, and effectiveness of AI systems. This includes adopting proven methodologies for AI development, testing, and deployment, as well as staying abreast of the latest advancements and innovations in the field.

4. **Interoperability and Compatibility:** Promoting interoperability and compatibility of AI systems with international standards to facilitate seamless integration and collaboration. This involves ensuring that AI technologies can operate across different platforms, devices, and environments, and that they can exchange data and information effectively.

5. **Certification and Accreditation:** Establishing certification and accreditation programs to verify that AI systems comply with international standards and best practices. These programs provide assurance to users and stakeholders that AI technologies meet rigorous quality and safety criteria.

6. **Capacity Building and Training:** Providing training and capacity-building initiatives to ensure that AI developers, operators, and regulators are familiar with international standards and best practices. This includes offering workshops, courses, and resources to enhance the skills and knowledge required to implement and enforce these standards effectively.

By developing robust national security and safety protocols, implementing comprehensive ethical guidelines, and adopting international standards and best practices, nations can ensure the responsible and secure development and deployment of AI technologies. These measures foster innovation while safeguarding societal values, protecting individual rights, and enhancing global

iv. Data Sovereignty, Authenticity, Consent, and Provenance for AI

Data Sovereignty – Ensuring Data Sovereignty

Data sovereignty refers to the principle that data is subject to the laws and governance structures within the nation where it is collected. Ensuring data sovereignty is critical for maintaining national control over data resources and protecting the privacy and rights of citizens. Key aspects include:

1. **National Data Storage Requirements:** Mandating that data collected within a country must be stored on servers physically located within national borders. This helps ensure that the data is subject to local laws and can be protected from foreign jurisdictions.
2. **Regulatory Compliance:** Implementing stringent regulations to govern data collection, storage, and processing within the country. These regulations should ensure that data handling practices comply with national privacy laws, security standards, and ethical guidelines.
3. **Data Access Controls:** Establishing robust access controls to prevent unauthorized access to sensitive data. This includes implementing strong authentication mechanisms, encryption protocols, and role-based access controls to safeguard data sovereignty.
4. **Collaboration with Local Entities:** Encouraging collaboration between government, private sector, and academic institutions to develop and manage data infrastructure. This can foster innovation and ensure that data governance practices are aligned with national interests.
5. **International Agreements and Standards:** Participating in international agreements and standards-setting bodies to advocate for data sovereignty principles and ensure that global data governance frameworks respect national sovereignty. This can help harmonize cross-border data flows while protecting local interests.

Data Authenticity – Ensuring the Reliability and Accuracy of Data Used

Data authenticity is crucial for ensuring that AI systems are built on reliable and accurate data. This enhances the trustworthiness of AI outputs and minimizes the risk of errors and biases. Key measures to ensure data authenticity include:

1. **Data Validation and Verification:** Implementing rigorous data validation and verification processes to ensure that data is accurate, complete, and reliable. This includes crossreferencing data sources, using checksums and hash functions, and employing data quality assessment tools.

2. **Provenance Tracking:** Maintaining detailed records of data provenance, including the origin, history, and transformations applied to the data. This helps trace the data's journey from collection to usage, ensuring its integrity and authenticity.

3. **Tamper-Evident Technologies:** Utilizing tamper-evident technologies, such as blockchain, to protect data integrity. These technologies can provide a secure and immutable record of data transactions, making it difficult for unauthorized modifications to go undetected.

4. **Automated Data Cleaning:** Deploying automated data cleaning tools to identify and correct errors, inconsistencies, and anomalies in datasets. This helps maintain high data quality and reduces the risk of inaccuracies affecting AI outcomes.

5. **Periodic Audits and Reviews:** Conducting regular audits and reviews of data management practices to ensure ongoing compliance with data authenticity standards. These audits can identify potential vulnerabilities and areas for improvement in data handling processes.

Consent and Provenance – Legal Frameworks Governing Data Consent and Strategies in Maintaining and Auditing Data Lineage

Ensuring proper consent and maintaining data provenance are critical for protecting individuals' rights and ensuring transparency in data usage. Legal frameworks and strategies to govern data consent and provenance include:

1. **Informed Consent Mechanisms:** Implementing robust informed consent mechanisms to ensure that individuals are fully aware of how their data will be collected, used, and shared. This includes providing clear and accessible information about data practices and obtaining explicit consent from individuals.

2. **Data Usage Transparency:** Mandating transparency in data usage practices to build trust and accountability. Organizations should publicly disclose their data handling policies, the purposes for which data is collected, and how it will be used and shared.

3. **Legal Safeguards and Compliance:** Establishing legal safeguards to protect individuals' rights and ensure compliance with consent requirements. This includes enforcing data protection laws, such as GDPR, which provide comprehensive guidelines on obtaining and managing data consent.

4. **Provenance Management Systems:** Developing and implementing provenance management systems to track and document the lineage of data. These systems should capture metadata about data sources, collection methods, transformations, and usage, providing a comprehensive audit trail.

5. **Periodic Consent Reconfirmation:** Requiring periodic reconfirmation of consent, particularly for long-term data usage or when data is repurposed for new applications. This ensures that individuals' consent remains valid and relevant over time.

6. **Auditing and Monitoring:** Conducting regular audits and monitoring of data consent and provenance practices to ensure compliance with legal and ethical standards. Independent oversight bodies can play a crucial role in verifying that data usage aligns with individuals' consent and that provenance records are accurate and complete.

7. **Education and Awareness Campaigns:** Implementing education and awareness campaigns to inform individuals about their rights regarding data consent and the importance of data provenance. Empowering individuals with knowledge can enhance their ability to make informed decisions about data sharing.

By ensuring data sovereignty, authenticity, consent, and provenance, nations can build robust frameworks that protect individual rights, enhance data reliability, and promote trust in AI systems. These measures are essential for creating an ethical and secure AI ecosystem that aligns with societal values and fosters innovation.

· Digital Infrastructure Requirements

- Infrastructure Requirements

- High-performance computing infrastructure (e.g., hyperscale data centres, 5G, cloud etc.)
- Energy
- Water
- Carbon offsets

- Algorithmic Suite

- Selection of algorithms and frameworks to optimise outputs of large-scale AI systems

- Data Requirements

- Volume and diversity of data needed to train the model (i.e., public datasets and proprietary data)
- Open source data utilisation (e.g., using repositories like Pentland's open-source data)

Counterpoints. Not all proponents of data sovereignty advocate for data localisation (national cloud). Data can instead have certain protections attached even if it crosses borders, such as contemplated by the US-EU Data Privacy Framework and Data Free Flow with Trust. While governments may appreciate the simplicity of requiring locally domiciled data, requiring servers to be housed with in a country may impede security and access to compute, particularly in developing nations that may lack the ability to fund large-scale AI infrastructure. Providers such as AWS and Microsoft offer solutions in some jurisdictions, but not every country has locally-sited infrastructure, reflective of a global shortage of high-intensity compute capabilities.

Part II - Socio-technic Systems Considerations

A. Capabilities and Motivation of Model Size

The Sovereign AI model should be able to offer complex problem solving abilities akin to ChatGPT 4o to be useful and crucially to act as reference, validate and respond to other foundational AIs.

Below is a table summarizing the minimum parameter requirements for various capabilities based on existing models and research:

Basic Language Understanding	1.5 billion (GPT-2)
Translation	10 billion
Coding	50 billion
Common Sense Reasoning	100 billion
Zero-shot Learning	175 billion (GPT-3)
Advanced Question Answering	500 billion
Complex Problem Solving	1+ trillion (GPT-4)

Source: Trusted AI Alliance, Imperial College London

Thus, the required minimum size is a trillion-parameter model so to handle a wide array of tasks with high proficiency.

To achieve the desired performance for a model with 1 trillion parameters, a sovereign AI will require substantial GPU resources. There are few published data points to estimate this, e.g. the training of GPT-3 with 175 billion parameters already demanded extensive computational power.
Based on the computational scaling trends being systematically linear, a 1 trillion-parameter model would require approximately 10,000 NVIDIA H200 GPUs to train within feasible time scales.
This estimate considers the increased need for processing power due to the larger model size and the complexity of training tasks and effectively having to develop lessons learned from the 4 Big Foundational commercial models.

The proposed hardware requirements for the sovereign AI system, a GPT-based model, are driven by the need to handle a large-scale language model with approximately 1 trillion parameters. Such a model is essential to achieve advanced intellectual capabilities, including natural language understanding, reasoning, coding, and more. This justification outlines why the specified hardware and infrastructure are necessary, citing relevant research and established benchmarks in AI development. This model size requirement thus yields a pathway to specify the hardware and infrastructure requirements and costs both in terms of one-off investment and operational costs.

1. Storage Capacity and Backup

To accommodate the extensive data requirements for our sovereign AI system, a storage capacity of 5 petabytes is essential. This capacity will allow for efficient data handling, processing, and storage, which is critical for the performance of an AI system of this scale. To ensure data integrity and availability, a mirrored backup system will be implemented, effectively doubling the storage needs to 10 petabytes. The estimated cost for this storage solution is approximately US$ 20 million. This investment includes the cost of high-capacity, high-reliability storage hardware, and the necessary backup systems to prevent data loss.

2. GPU Compute Hardware and Interconnect

The core computational power for the AI system will be provided by 10,000 units of NVIDIA H200 GPUs, organized within GH200-type servers. These GPUs are selected for their high performance and efficiency in handling the complex calculations required by AI workloads. The cost for this GPU compute hardware is estimated at around US$ 560 million. This significant investment ensures that the system will have the necessary processing power to handle largescale AI tasks efficiently.

Space Requirements and Infrastructure

The physical space required to house the server racks and associated hardware is approximately 850 square meters. This space calculation includes not only the server racks but also the necessary passages for workforce movement, ensuring efficient operation and maintenance. The server environment will need to incorporate underfloor cooling ducts, assumed to be watercooled, to maintain optimal operating temperatures for the hardware. Overfloor power distribution is required, given potential limitations in ceiling carry capacity for power ducts.

Existing buildings may be suitable for this setup, but they must be evaluated for their structural capacity, especially concerning cooling systems and housing. Cooling solutions should ideally be located on the roof to leverage natural evaporation and air movement, minimizing ground-level space usage. Modern buildings and warehouses often lack the necessary structural integrity to support such loads, making them unsuitable for this purpose.

A potential solution to the infrastructure challenge is the use of modular container units. These units can house servers, cooling systems, and power distribution, offering a quick and effective deployment solution. The additional cost for using modular containers is estimated at 30% (TBC)

over the cost of individually purchasing and installing infrastructure components, but this method offers significant benefits in terms of deployment speed and flexibility.

3. Power Requirements

The site will require a substantial power delivery capacity of approximately 18 megawatts to support the compute hardware and cooling systems. It is critical to ensure that the power supply is stable and meets G7 standards for electricity quality, including the stability of the sine wave. Additionally, a reliable power backup system is necessary to mitigate any potential power outages. To support ecologically sustainable operations, the system should incorporate heat energy recovery/reuse and explore renewable energy sources such as solar and wind power. These measures will not only reduce the environmental impact but also potentially lower operating costs in the long term.

4. Operating Costs

The operating costs for the AI system will be primarily driven by energy consumption for both computation and cooling, each accounting for roughly half of the total energy cost. These costs will vary based on local energy prices. Additionally, maintaining the hardware will require a dedicated team of approximately 30 personnel. This estimate assumes an issue resolution rate of one per 1,000-1,500 GPUs per week, ensuring prompt maintenance and minimal downtime.

Summary of Investment and Operating Costs

Initial Investments:

Storage Capacity and Backup: US$ 20 million
GPU Compute Hardware: US$ 560 million
Infrastructure : Dependent on site and location

Operating Costs:

Energy Costs (Compute and Cooling): Dependent on local energy prices
Hardware Maintenance Staff: Approximately 30 personnel
Total One-off Investments:
Estimated at US$ 560 million plus 30% cooling etc (approx. US$ 168 million), bringing the total to approximately US$ 720 million.

Ongoing Operating Costs:

Highly variable, influenced by local energy prices and the chosen sustainability measures.
This comprehensive outline addresses the key hardware requirements for deploying a sovereign AI system, ensuring both operational efficiency and scalability.

Source: Trusted AI Alliance, Imperial College London

B. Small Is Beautiful

The Emerging Advances in Small Language Models and Small LLMs

Given the challenges associated with scale of 1 trillion parameter ("1T") models, a movement has arisen to explore much smaller, multi-layer, multi-agent systems that keep proprietary data safe, are much less expensive to build and maintain, and much less prone to hallucinations. Researchers from firms ranging from Meta to LinkedIn have recently published on significant advances in such 'small language models'. Capabilities are delivered not only in a more efficient package, but also through the coordination of a combination of small expert models. New 'swarm' systems are being developed that integrate multiple small models in a coordinated fashion.

Likewise, in the hardware domain, NVIDIA's GPU dominance is propelled by a highly developed software stack. As more and better software systems that make use of CPUs and edge computing emerge, we will see more choices around the hardware bottleneck.

Sovereign Exploration of Small AI

Several nation-states are engaging the sovereign AI enterprise through the lens of "small". The UAE, Singapore, India and other countries are building sovereign tools on top of open source AI (e.g., LLama) to build specialized models. They are augmenting this approach by concentrating expertise within the country (bringing talent to bear on the problem, including talent acquired from elsewhere - see "Solving the Talent Shortage: Policy Options" below).

They are also addressing the talent shortage by constructing a network of experts to collaborate with them. Academia remains a repository of expert talent and knowledge that has not entirely migrated into proprietary, closed-source code bases and companies. Sovereigns can bridge the talent shortage through robust partnerships with academic organizations, tapping into cuttingedge research and expertise. In addition, academic experts are typically accustomed to capacity building and knowledge sharing in ways that corporate groups are not, and a properly-constructed academic partnership can facilitate the creation of more expertise locally in different domiciles.

C. Cybersecurity Requirements of Sovereign AI

Measures to Secure Data Pipelines and Model Integrity

Ensuring the security of data pipelines and the integrity of AI models is paramount for the reliable and safe operation of sovereign AI systems. Effective measures to achieve this include:

1. Data Encryption:

- At Rest: Encrypt data stored in databases, data lakes, and other storage mediums to prevent unauthorized access.

- In Transit: Use secure communication protocols such as TLS/SSL to encrypt data as it moves between different components of the data pipeline, ensuring that it remains confidential and unaltered during transmission.

2. Access Controls:

- Role-Based Access Control (RBAC): Implement RBAC to ensure that only authorized personnel have access to data and AI models based on their roles and responsibilities.
- Multi-Factor Authentication (MFA): Use MFA to add an extra layer of security for accessing critical systems and data, reducing the risk of unauthorized access.

3. Data Provenance and Lineage Tracking:

- Metadata Management: Maintain detailed metadata for all data sources to track the origin, transformations, and usage of data. This helps in auditing and ensuring the authenticity of the data used in AI models.
- Blockchain Technology: Utilize blockchain to create tamper-proof records of data transactions and transformations, enhancing data integrity and traceability.

4. Regular Audits and Monitoring:

- Automated Audits: Use automated tools to regularly audit data pipelines for compliance with security policies and detect any anomalies or unauthorized changes.
- Continuous Monitoring: Implement continuous monitoring of data flows and model performance to quickly identify and respond to potential security threats or performance issues.

5. Model Protection Techniques:

- Model Encryption: Encrypt AI models to protect their intellectual property and prevent reverse engineering.
- Adversarial Robustness: Incorporate techniques to enhance the robustness of models against adversarial attacks, such as adversarial training and defensive distillation.

6. Secure Development Practices:

- Code Reviews: Conduct regular code reviews and vulnerability assessments to identify and fix security issues in AI software.
- DevSecOps: Integrate security practices into the DevOps pipeline (DevSecOps) to ensure that security is considered throughout the software development lifecycle.

Addressing Cybersecurity Threats

To effectively address cybersecurity threats, sovereign AI systems must implement comprehensive strategies and practices that encompass prevention, detection, response, and recovery.
Key aspects include:

1. Threat Intelligence and Analysis:

- Cyber Threat Intelligence (CTI): Gather and analyze threat intelligence from various sources to stay informed about the latest threats and vulnerabilities that could impact AI systems.
- Behavioral Analytics: Use AI and machine learning to analyze patterns of behavior and detect anomalies that may indicate cyber threats.

2. Intrusion Detection and Prevention Systems (IDPS):

- Network-Based IDPS: Deploy network-based IDPS to monitor network traffic for signs of malicious activity and prevent potential intrusions.
- Host-Based IDPS: Use host-based IDPS to monitor critical system activities, including file access and process execution, to detect and respond to threats.

3. Incident Response Planning:

- Incident Response Teams: Establish dedicated incident response teams to quickly address cybersecurity incidents and mitigate their impact.
- Incident Response Plans: Develop and regularly update incident response plans that outline the procedures for detecting, responding to, and recovering from cyber attacks.

4. Regular Security Assessments:

- Penetration Testing: Conduct regular penetration testing to identify and address security weaknesses in AI systems.
- Vulnerability Scanning: Use automated vulnerability scanning tools to continuously assess the security posture of AI infrastructure and applications.

5. User and Employee Training:

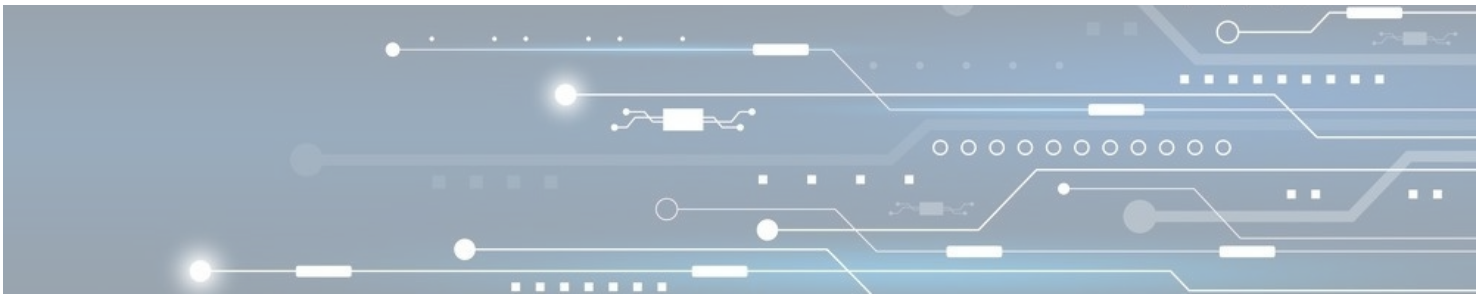
- Security Awareness Training: Provide ongoing training for users and employees to recognize and respond to cybersecurity threats such as phishing, social engineering, and insider threats.
- Phishing Simulations: Conduct regular phishing simulations to test employees' awareness and preparedness to handle phishing attacks.

6. Data and Model Resilience:

- Data Backups: Implement regular data backup procedures to ensure that data can be restored in case of a cyber attack or data loss incident.
- Model Redundancy: Develop redundant AI models and systems to ensure continuity of operations in the event of a cyber attack or system failure.

7. Collaboration and Information Sharing:

- Public-Private Partnerships: Foster collaboration between government agencies, private sector organizations, and academia to share information on threats, vulnerabilities, and best practices.
- International Cooperation: Engage in international cooperation to address crossborder cybersecurity threats and share threat intelligence globally



D. Talent and AI Workforce Requirements

Key Strategies to ensure the nation has the right talent and workforce for the AI-era include

1. Educational Reform and Curriculum Development:

- STEM Education: Emphasize science, technology, engineering, and mathematics (STEM) from early education through higher education.
- AI and Data Science Courses: Integrate AI, machine learning, and data science courses into school and university curricula.
- Interdisciplinary Learning: Encourage interdisciplinary programs that combine AI with other fields such as humanities, social sciences, and business.

2. Skill Development and Lifelong Learning:

- Technical Skills Training: Provide training in AI-specific skills like programming, data analysis, and algorithm development.
- Soft Skills Development: Emphasize soft skills such as critical thinking, problemsolving, and adaptability.
- Online Learning Platforms: Promote the use of online courses and certifications to facilitate continuous learning and upskilling.

3. Government Policies and Incentives:

- Funding and Grants: Provide funding for research and development in AI and related fields.
- Tax Incentives: Offer tax incentives to companies investing in AI training and education for their employees.
- Immigration Policies: Implement policies to attract and retain global AI talent.

4. Public-Private Partnerships:

- Industry Collaboration: Foster collaboration between academia, industry, and government to align education and training with industry needs.
- Internships and Apprenticeships: Encourage companies to offer internships and apprenticeship programs focused on AI skills.

5. Research and Development:

- Innovation Hubs: Establish AI research centers and innovation hubs to drive research and provide hands-on experience.
- Collaboration with Leading Institutions: Partner with leading global institutions for research and knowledge exchange.

6. Awareness and Inclusivity:

- Public Awareness Campaigns: Run campaigns to raise awareness about the importance of AI and the opportunities it presents.
- Inclusivity Programs: Ensure diversity and inclusion in AI education and careers to leverage a broad range of perspectives and talents.

7. Ethics and Governance:

- Ethics Education: Incorporate ethics and responsible AI practices into the curriculum to ensure that the workforce is equipped to handle AI responsibly.
- Regulatory Frameworks: Develop regulatory frameworks that support innovation while addressing ethical and societal implications of AI.

8. International Collaboration:

- Global Cooperation: Engage in international collaborations to share knowledge, resources, and best practices in AI education and workforce development.
- Participation in Global Initiatives: Participate in global AI initiatives and forums to stay abreast of the latest developments and trends.

The Female Quotient

While women comprise half the world's population, they make up perhaps one fifth to one quarter of AI professionals (Stanford HAI 2024). Policy that advances the integration of female talent into the broader AI conversation can enable a sovereign grouping to gain competitive advantage through greater diversity of inputs supporting entrepreneurial and societal outputs, a larger talent pool to address critical needs for AI competitiveness, and alignment with global goals on gender equity.

Solving the Talent Shortage: Policy Options

Many nations (outside perhaps the US and the PRC) will not be able to rely solely on homegrown talent, but instead needs a comprehensive and competitive strategy to attract and retain foreign talents through well thought out policies, processes, and incentives. A potential solutionset could include education, immigration policy, public-private partnerships, infrastructure investments and a communications plan.

- **Education:** Collaborate with local and global universities, and corporate employers to develop curricula and devise vocational training. Establish public-private partnerships to facilitate integration of real-world needs with formal education. Leverage new tools and technologies to align global expertise and knowledge with local delivery.

- **Immigration Policy:** Expand and accelerate high-talent visa programs streamlined for critical skills workers to attract foreign talent (perhaps temporary 1- or 2-year visas, modeled on the digital nomad programmes, to draw in expertise and facilitate knowledge transfer); research grants and scholarships for existing foreign university

students to conduct a semester, summer, or year in coiuntry; provide incentives such as tax breaks, grants, or accelerated permanent residency to retain foreign-trained AI professionals.

- **Public-Private Partnerships:** Increase public investment in AI R&D through national grants, research centers, and collaborations between universities and the private sector (building capabilities in AI while also generating intellectual property); Develop AI innovation hubs that bring together talent, investors, and researchers to foster collaboration and innovation (perhaps around specific industry sectors such as financial services, health, and manufacturing); Launch government-sponsored AI competitions and hackathons to stimulate creativity and develop solutions to real-world problems.

- **Infrastructure Investments:** Addressing AI infrastructure requirements can not only assist with national competitiveness and economic development, but also serve as a talent attractor and a foundation on which local talent can be grown and developed. Improving access to computing power ensure the availability of high-performance computing infrastructure necessary for AI research and development; refining National AI Strategy (to the extent it already exists) can align vision, investment plans, and regulatory frameworks to promote AI growth; integrating data sharing frameworks with AI policy can accelerate AI success through clear guidelines for data collection, sharing, and privacy (which are essential for AI development, especially for training machine learning models).

- **Communications Plans:** Raise awareness of AI's importance through media campaigns, workshops, and seminars to encourage more students and professionals to enter AI-related fields; encourage underrepresented groups (such as women or ethnic minorities) to pursue careers in AI through targeted programs and scholarships.

E. Innovation Ecosystem Support

Building a robust ecosystem of startups, research institutions, and industry players can be facilitated by the following strategies:

1. Funding and Financial Incentives:

- **Research Grants and Funding:** Provide grants and funding for AI research projects at universities, research institutions, and startups.
- **Venture Capital Support:** Establish government-backed venture capital funds to invest in AI startups, particularly in early stages.
- **Tax Incentives:** Offer tax credits and deductions for companies investing in AI research and development.

2. Regulatory Frameworks:

- **AI-friendly Regulations:** Create a regulatory environment that encourages innovation while addressing ethical and privacy concerns.
- **Intellectual Property Protection:** Strengthen intellectual property laws to protect AI innovations and encourage investment.
- **Data Governance Policies:** Develop clear policies on data sharing, data privacy, and data protection to build trust and facilitate data-driven AI research.

3. Infrastructure Development:

- **High-Performance Computing:** Invest in high-performance computing infrastructure and cloud services accessible to startups and researchers.
- **Innovation Hubs and Incubators:** Establish AI innovation hubs, tech parks, and incubators to provide shared resources and collaboration spaces.
- **Digital Infrastructure:** Enhance digital infrastructure, including high-speed internet and robust cybersecurity measures.

4. Education and Workforce Development:

- **AI Education Programs:** Integrate AI and machine learning courses into school and university curricula to build a skilled workforce.
- **Vocational Training:** Offer vocational training programs focused on AI and related technologies.
- **Lifelong Learning:** Promote continuous learning initiatives to help workers update their skills in line with AI advancements.

5. Public-Private Partnerships:

- **Collaborative Projects:** Facilitate partnerships between universities, research institutions, startups, and established companies for joint AI projects.
- **Shared Resources:** Support the creation of shared research facilities, labs, and datasets for collaborative AI research.
- **Industry Networks:** Encourage the formation of AI industry networks and associations for knowledge sharing and collaboration.

6. Market Access and Export Support:

- **Export Incentives:** Provide incentives for AI startups to export their technologies and services.
- **Trade Agreements:** Negotiate trade agreements that promote the export of AI innovations.
- **Market Development Programs:** Assist AI startups in understanding and entering international markets.

7. Ethics and Governance:

- **Ethical AI Guidelines:** Develop and promote ethical guidelines for AI development and deployment.
- **Transparency and Accountability:** Encourage transparency in AI algorithms and decision-making processes.
- **Responsible Innovation:** Support responsible AI innovation practices that consider societal impacts and ethical implications.

8. International Collaboration:

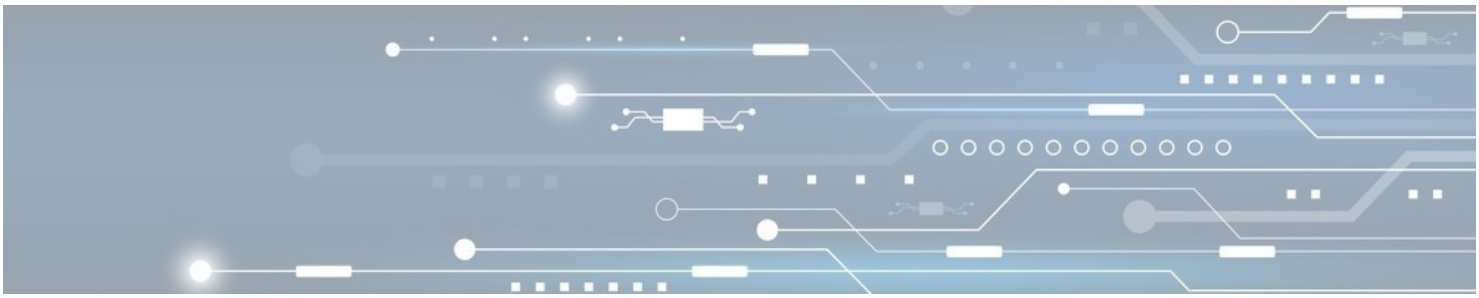
- **Global Partnerships:** Foster international collaborations with other countries, multinational corporations, and global AI research institutions.
- **Cross-Border Initiatives:** Participate in cross-border AI initiatives and consortia focused on addressing global challenges.
- **Knowledge Exchange:** Facilitate the exchange of knowledge, expertise, and best practices in AI on a global scale.

9. Networking and Community Building:

- **AI Conferences and Workshops:** Host AI conferences, workshops, and hackathons to bring together stakeholders from different sectors.
- **Mentorship Programs:** Develop mentorship programs that connect AI startups with experienced industry leaders and researchers.
- **Online Platforms:** Create online platforms for AI professionals to network, collaborate, and share insights.

10. Inclusion and Diversity:

- **Diverse Talent Pool:** Implement policies to ensure diversity and inclusion within the AI ecosystem.
- **Support for Underrepresented Groups:** Provide support and resources for underrepresented groups in AI, including women and minorities.
- **Inclusive Policies:** Ensure AI policies and programs are inclusive and consider the needs of all segments of society.



F. Energy Requirements of AI

AI usage is fueling a significant increase in global energy demand. CNBC reports that electricity demand is forecast to grow as much as 20% by 2030, with AI data centers alone expected to add about 323 terawatt hours of electricity demand in the U.S. and more in other markets. Closer in time, the annual demand for electricity by artificial intelligence worldwide could increase from 85.4 to 134 terawatt hours by 2027. We will discuss energy and AI more extensively in a forthcoming white paper.

1. Training AI Models:

- **High Computational Power:** Training AI models, especially deep learning models, requires substantial computational power. This involves using GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units) that consume a lot of energy. GPT-4 took an estimated 49 GWh to train, and larger models will take more.
- **Data Centers:** AI training typically takes place in data centers, which house numerous servers and require significant amounts of electricity for both computation and cooling.

2. Inference and Deployment:

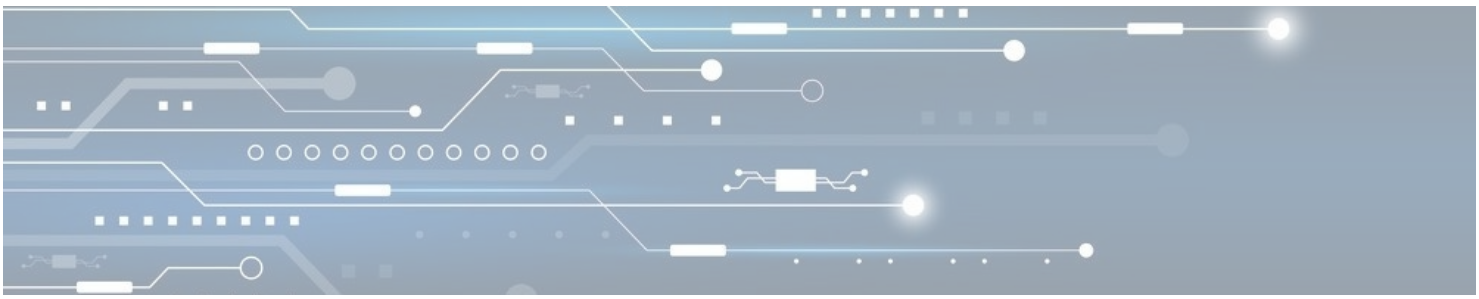
- **Edge Devices:** Running AI models on edge devices (like smartphones, IoT devices, etc.) consumes less power compared to training but still contributes to overall energy usage.
- **Cloud Services:** AI services provided over the cloud require ongoing computational power, contributing to continuous energy consumption.

G. Net Carbon Impact of AI

AI systems are primarily powered by fossil fuels. GPT-4's carbon footprint is estimated at as much as 800 tons of CO2. Bloom, another generative AI model, was estimated at 120 tons of carbon for training. For comparison, the average carbon footprint of an individual human is about 6 tons globally (and as much as 21 tons in the USA).

1. Carbon Footprint of Data Centers:

- **Electricity Consumption:** Data centers are major energy consumers. The source of electricity (renewable vs. non-renewable) significantly affects their carbon footprint.
- **Cooling Systems:** Cooling systems in data centers are essential to prevent overheating, but they also add to the energy consumption and carbon emissions.



A single query to a generative AI model consumes as much water as contained in an entire bottle.
- Google's carbon emissions, for example, have increased by 50% since 2019 due primarily to AI, according to its 2024 environmental impact report.

2. Lifecycle Emissions:
- **Hardware Production:** The production of GPUs, TPUs, and other hardware components involves significant carbon emissions due to manufacturing processes and raw material extraction.
- **Operational Emissions:** The operational phase, including the energy consumed during training and inference, is the primary contributor to the carbon footprint of AI.

H. Carbon and Energy Mitigation Strategies

1. Improving Energy Efficiency:
- **Efficient Algorithms:** Developing more efficient algorithms that require less computational power can reduce energy consumption.
- **Hardware Innovations:** Innovations in hardware design to improve energy efficiency (e.g., specialized AI chips) can also help.

2. Renewable Energy Sources:
- **Green Data Centers:** Transitioning to renewable energy sources for powering data centers can significantly reduce the carbon footprint.
- **Carbon Offsetting:** Companies can invest in carbon offsetting initiatives to counterbalance their emissions.

3. Optimized Data Management:
- **Data Pruning:** Reducing the amount of data processed by pruning irrelevant or redundant data can lower energy consumption.
- **Efficient Storage:** Using energy-efficient storage solutions can also contribute to reducing the overall energy requirements.

4. Policy and Regulation:
- **Sustainability Standards:** Governments and regulatory bodies can set sustainability standards and guidelines for AI development and data center operations.
- **Incentives for Green Practices:** Providing incentives for companies that adopt green practices in AI development can encourage more sustainable approaches.

5. Awareness and Collaboration:
- **Industry Collaboration:** Collaborating across the AI industry to share best practices and technologies for reducing energy consumption and carbon emissions.

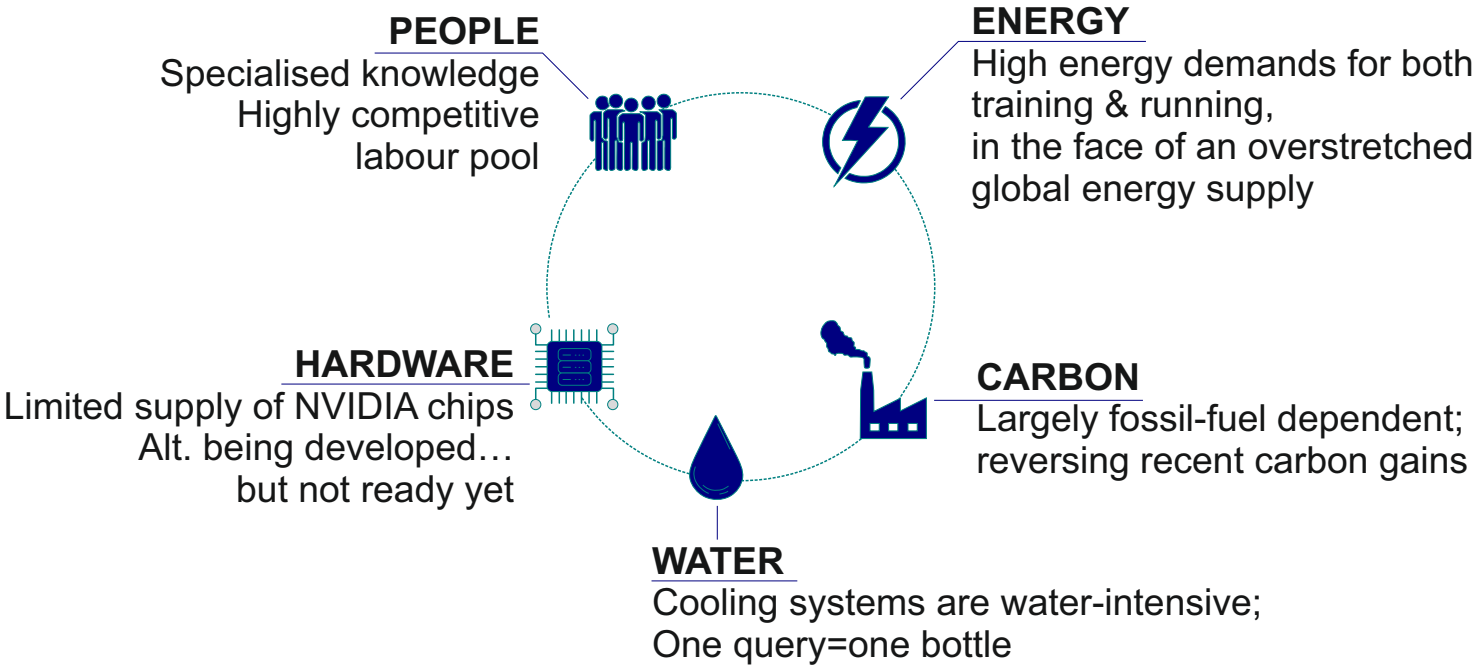
- **Public Awareness:** Raising awareness about the environmental impact of AI and encouraging responsible usage and development.

I. Determining Sovereign Value Proposition

When considering a sovereign AI initiative, a policy body benefits from considering how its resources and capabilities map to the critical inputs necessary for sovereign AI.

- Digital penetration, infrastructure and readiness
- Resource availability (energy, water and land)
- Human capabilities (talent skilled in advanced AI systems)

The major rate-limiting factors to AI growth can be summarised as follows:



By evaluating its capabilities against international benchmarks (see Exhibit B), a sovereign policy body can determine both where it has critical gaps that need to be addressed, and which solutionpaths are optimal to competing in the AI future.

Part III: Policy Frameworks

As a political body develops its national AI strategy, there are a number of broad policy considerations to take into account (not only with respect to sovereign AI, but across the board). These concepts, in turn, will inform decision-making with respect to sovereign AI.

i. PRINCIPLES AND REGULATIONS

A. Core Principles forming the basis of Global policy initiatives

While specific prioritization might vary, several core principles form the foundation of global policy initiatives on AI, as evidenced by the work of the OECD (2018) and UNESCO (2021). These principles are:

1. Transparency & Explainability:

- Users should understand how AI systems work and the rationale behind their decisions.
- This fosters trust, allows for human oversight, and helps identify potential biases.
- The level of transparency may differ based on the risk associated with the AI system.

2. Safety & Security:

- AI systems should be designed and operated to minimize risks of harm, both physical and psychological.
- This includes ensuring data security, robustness against attacks, and mitigating unintended consequences.
- Safety becomes especially crucial for applications in critical sectors like healthcare and transportation.

3. Inclusive Growth & Sustainability:

- AI development and deployment should promote broad societal benefits and economic opportunities for all.
- This includes addressing potential job displacement and ensuring equitable access to AI technologies.
- AI should be used responsibly to address global challenges like climate change and poverty.

4. Accountability:

- There should be clear lines of responsibility for the development, deployment, and use of AI systems.
- This includes identifying who is accountable for decisions made by AI and potential harms caused.
- Legal frameworks and mechanisms for redress need to be established.

5. Fairness & Non-discrimination:

- AI systems should be designed and trained to avoid bias and discrimination based on factors like race, gender, or religion.
- This necessitates diverse datasets and ongoing monitoring to mitigate bias creep.
- Fairness considerations are particularly important for AI used in areas like law enforcement and recruitment.

B. Adapting to Local Contexts:

As highlighted by the Digital Policy Alert (https://digitalpolicyalert.org/), the prioritization of these principles can be adapted to local cultural norms and values.

- For example, a country with a strong emphasis on individual privacy might prioritize transparency more highly.
- Another country with a focus on social harmony might prioritize fairness and nondiscrimination to a greater degree.

C. Evolution from Principles to Binding Regulations

The journey from broad principles to binding regulations in AI governance involves several steps, reflecting increasing specificity, enforceability, and scope. This progression ensures that ethical guidelines and best practices evolve into legally enforceable frameworks, promoting responsible AI development and deployment. Two significant examples of this evolution are the EU AI Act 2024 and the Council of Europe Binding Convention on AI.

EU AI Act 2024

Principles to Regulations:

1. Ethical Principles:

- Initially, the EU, like many other regions, developed ethical guidelines for AI based on principles such as transparency, safety, fairness, and accountability. These principles provided a foundation for trustworthy AI and were outlined in documents like the EU's Ethical Guidelines for Trustworthy AI (2019).

2. Guidance and Recommendations:

- These ethical principles were further detailed into more specific recommendations and frameworks by expert groups and advisory bodies. These frameworks guided AI development and implementation within the EU, encouraging best practices and voluntary compliance.

3. Drafting of the AI Act:

- Recognizing the need for enforceable regulations, the European Commission drafted the AI Act. This draft was developed through extensive consultations with stakeholders, including industry experts, civil society, and member states. The draft aimed to translate ethical principles into specific legal requirements.

4. Legislative Process:

● The draft AI Act underwent rigorous scrutiny and amendment by the European Parliament and the Council of the European Union. This process included debates, expert testimonies, and impact assessments to ensure the regulation was comprehensive and balanced.

5. Binding Regulation:

● The EU AI Act 2024 became the first comprehensive regulation on AI, incorporating guidelines on Generative AI and other advanced AI technologies. It classifies AI systems based on their risk levels (unacceptable, high, and limited) and sets strict requirements for high-risk AI applications, ensuring safety, transparency, and accountability.

Key Features of the EU AI Act:

● **Risk-Based Classification:** AI systems are categorized based on their potential risks, with high-risk systems subject to stringent requirements.
● **Generative AI Guidance:** Specific provisions address generative AI, ensuring transparency, traceability, and robustness.
● Compliance Obligations: Mandatory compliance assessments, documentation, and reporting requirements for AI developers and deployers.
● Enforcement and Penalties: Establishes enforcement mechanisms and significant penalties for non-compliance, ensuring adherence to the regulations.

Council of Europe Binding Convention on AI

Principles to Regulations:

1. Ethical Guidelines:

● The Council of Europe (CoE) initially developed ethical guidelines for AI, emphasizing human rights, democracy, and the rule of law. These guidelines aimed to ensure that AI technologies respect fundamental rights and freedoms.

2. Framework for AI Governance:

● Building on these ethical guidelines, the CoE created a governance framework that outlined best practices for AI development and deployment. This framework served as a reference for member states to develop their own AI policies and regulations.

3. Drafting the Convention:

● The CoE recognized the need for a binding legal instrument to ensure consistent and enforceable AI governance across its member states. A draft convention was prepared, detailing specific legal obligations for AI developers, users, and regulators.

4. Consultation and Revision:

● The draft convention underwent extensive consultation with various stakeholders, including member states, international organizations, academia, and civil society. This collaborative process ensured that the convention addressed diverse perspectives and concerns.

5. Adoption of the Binding Convention:

● The CoE Binding Convention on AI was adopted, creating a legally enforceable framework that member states are obligated to implement. This convention

harmonizes AI regulations across Europe, ensuring a unified approach to AI governance.
Key Features of the CoE Binding Convention on AI:
- **Human Rights Protection:** Ensures that AI systems respect and promote human rights, with specific provisions to prevent discrimination and bias.
- **Democratic Oversight:** Establishes mechanisms for democratic oversight of AI technologies, involving public participation and transparency.
- **Legal Obligations:** Defines clear legal obligations for AI developers and users, including requirements for accountability, transparency, and safety.
- **International Cooperation:** Promotes international cooperation in AI governance, facilitating knowledge sharing and collaborative enforcement of regulations.

In light of the differences in legal systems around the world, the convention offers parties (countries) some flexibility in compliance mechanism for the private sector. For the private sector, parties may opt to be directly obliged by the relevant convention provisions or, as an alternative, take other measures to comply with the treaty's provisions while fully respecting their international obligations regarding human rights, democracy and the rule of law.

ii. CREATING AN ENABLING ENVIRONMENT

An enabling environment for an AI ecosystem can be fostered through a set of actions that include:

- Identification of Policy Gaps
- Exploration of Flexible Regulatory Approaches
- Evaluating specific risks of open source systems (eg Cybersecurity) and exploring ways to mitigate them
- Capacity building of the public sector (investments capabilities, tools, and institutions)
- Providing support for innovators to navigate the regulatory landscape
- Improved regulatory coherence and international cooperation
- Awareness building of the broader society regarding risks and opportunities
- Allocating resources for platforms promoting Multi-Stakeholder Dialogue

Screening of Current Legislation and Identifying Gaps to be Addressed

- **Sector and Use Case Specific Approach:** The rapid advancement of AI technologies necessitates a thorough screening of current legislation to identify gaps that must be addressed to ensure robust governance. This process involves evaluating existing laws and regulations to determine their adequacy in addressing the unique challenges posed by AI in various sectors. By adopting a sector and use case-specific approach, policymakers can tailor regulatory frameworks to the distinct needs and risks associated with different industries, such as healthcare, finance, and transportation. This ensures that regulations are not overly broad or restrictive, allowing for innovation while safeguarding public interests and ethical standards. For instance, AI applications in healthcare might require stringent data privacy measures, while AI in transportation might focus more on safety and reliability standards.

Exploring Flexible Regulatory Approaches

● **Co-Regulation Models:** Flexible regulatory approaches can facilitate innovation by providing a balance between oversight and freedom. Co-regulation models are particularly effective, where legislation outlines desired outcomes but leaves room for industries to innovate on how to achieve compliance. This approach allows industries to develop best practices and standards that are more adaptable to technological changes, fostering an environment of continuous improvement and innovation.

● **Creating Space for Experimentation:** Regulatory sandboxes provide a controlled environment for testing new AI technologies without the full burden of regulatory constraints. These sandboxes enable developers to experiment with innovative solutions while regulators observe and learn, ensuring that any potential risks are identified and mitigated before broader deployment. This approach encourages innovation by reducing the risk and uncertainty associated with regulatory compliance.

● **Building on Current Best Practices:** Leveraging existing best practices and frameworks from leading organizations can streamline the development of effective AI regulations. Collaborating with standard-setting organizations such as ISO, IEEE, and IEC, and supportive organisations like NIST, ensures that new regulations are aligned with global standards and reflect the latest advancements and consensus in AI governance. This cooperation helps harmonize regulations across jurisdictions, reducing barriers to innovation and facilitating international collaboration.

Evaluating Specific Risks of Open Source Systems

● **Using Red Teaming Techniques:** Open-source AI systems, while fostering innovation and collaboration, also pose unique cybersecurity risks. Red teaming techniques involve simulating cyberattacks to identify vulnerabilities and test the robustness of AI systems. By proactively assessing potential threats, organizations can strengthen their security measures and develop strategies to mitigate risks.

● I**ntellectual Property Assessment:** New techniques such as ‘copyright traps’, a form of digital watermarking, can be employed to assess incorporation of protected material in large language models, potentially in violation of local law.

● **Collaboration with Other Countries and Academic Institutes:** Tackling the cybersecurity risks associated with open-source systems requires global collaboration. Working with other countries and academic institutions that are researching this domain can lead to shared insights, resources, and solutions. Such partnerships can enhance the collective understanding of cybersecurity challenges and drive the development of more secure open-source AI systems.

● **Soliciting Specific Research:** Encouraging targeted research into the cybersecurity aspects of open-source AI systems can provide valuable insights and innovations. Governments and institutions can fund specific research projects aimed at identifying and mitigating risks, fostering a deeper understanding of the challenges and opportunities in this area.

Capacity Building of the Public Sector

● **Establishment of a Robust Skills Training Framework:** Building capacity in the public sector is crucial for effective AI governance. A comprehensive skills training framework should be established, encompassing primary, tertiary, and on-the-job training. This ensures that public sector employees are well-equipped with the knowledge and skills required to manage and regulate AI technologies effectively.

● **Investment in Continuous Education:** Continuous education and policies that incentivize lifelong learning are essential to keep pace with rapid technological advancements. Governments should invest in programs that provide ongoing training and professional development opportunities for public sector employees, ensuring that their skills remain relevant and up-to-date.

● **Creation of a Fertile Ground for Attracting the Right Talent:** To foster a thriving AI ecosystem, it is important to attract and retain skilled professionals. Creating an environment that supports innovation, offers competitive compensation, and provides opportunities for career growth can help attract the right talent to the public sector. Providing Support for Innovators to Navigate the Regulatory Landscape

● **Online Platforms for Guidance:** Innovators often face challenges in understanding and complying with complex regulatory requirements. Online platforms that provide clear guidance, resources, and direct connections to regulators can help innovative tech companies navigate the regulatory landscape. These platforms can offer tailored advice, streamline compliance processes, and facilitate communication between innovators and regulatory bodies.

Improved Regulatory Coherence and International Cooperation

● **Building the Bridge with Like-Minded Countries:** International cooperation and regulatory coherence are vital for effective AI governance. Building partnerships with likeminded countries can lead to the development of harmonized regulations and standards, reducing barriers to innovation and facilitating cross-border collaboration. Such cooperation can also enhance the ability to address global challenges and leverage collective expertise.

Awareness Building of the Broader Society Regarding Risks and Opportunities

● **Fostering Trust and Adoption:** Raising awareness about the risks and opportunities of AI among the broader society is essential for fostering trust and encouraging adoption. Public education campaigns, transparent communication about AI technologies, and engagement with communities can help demystify AI and address public concerns. Building a knowledgeable and informed society ensures that AI advancements are embraced and used responsibly.

Allocating Resources for Platforms Promoting Multi-Stakeholder Dialogue

● **Broader Participation in the Policy Debate:** Promoting multi-stakeholder dialogue is crucial for developing inclusive and effective AI policies. Allocating resources to platforms that facilitate broad participation in policy discussions ensures that diverse perspectives

are considered. This includes involving smaller companies, underrepresented stakeholders, and civil society in the debate, leading to more balanced and comprehensive AI governance frameworks.

By implementing these measures, countries can develop robust, flexible, and inclusive AI policies that promote innovation while ensuring safety, fairness, and accountability. These steps are essential for harnessing the full potential of AI technologies and addressing the complex challenges they present.

iii.HORIZON SCANNING

Below is a high-level summary of AI policy from domiciles across the globe. More detail is provided in Appendix B.

Country and Regional

1. European Union (EU):

- Focuses on a human-centric approach to AI, emphasizing ethical principles and risk mitigation.
- The proposed AI Act aims to regulate high-risk AI applications and establish clear requirements for transparency, fairness, and accountability.
- The EU is also investing heavily in research and development of trustworthy AI.

2. Japan:

- Emphasizes collaboration between industry and government to promote responsible AI development.
- Adopted "Guidelines for Ethical Considerations on AI Development and Use" in 2017.
- Focuses on AI for social good, such as healthcare and disaster management.

3. United Kingdom (UK):

- Aims to be a global leader in ethical AI development.
- Published a white paper on AI outlining principles for responsible AI development.
- Investing in research on AI safety and security.

4. Singapore:

- Aims to be a hub for AI innovation, focusing on economic benefits.
- Established an AI Ethics & Governance framework with principles for responsible AI development.

- Promotes collaboration between government, industry, and academia.

5. United States of America (USA):

- Primarily relies on industry-led innovation in AI.
- Focuses on technological advancement and economic competitiveness.
- However, growing concerns exist regarding ethical implications of AI, leading to calls for increased regulation.

6. Germany:

- Strong emphasis on data privacy and human rights in AI development.
- Developed the Ethics Commission for Autonomous Driving to guide development of selfdriving cars.
- Promotes research on explainable and trustworthy AI.

7. France:

- Similar to Germany, prioritizes data privacy and human rights in AI development.
- Introduced a national AI strategy focusing on ethics, education, and investment.
- Supports research on responsible AI development.

8. Switzerland:

- Focuses on AI for the benefit of society and the economy.
- Established an interdepartmental working group on AI to develop guidelines for responsible development.
- Promotes international cooperation on AI governance.

9. Canada:

- Prioritizes an inclusive and human-centred approach to AI development.
- Developed the "Algorithmic Transparency and Accountability Framework" focusing on transparency and fairness.
- Invests in research on responsible AI development and public education.

10. People's Republic of China:

- China is focused on becoming a global leader in AI by 2030.
- The country's approach to AI regulation is characterized by a strong emphasis on national security, public interest, and alignment with socialist values.
- China's policies are designed to promote technological innovation while ensuring that AI development is ethical, safe, and beneficial to society.

- The government has introduced various regulations, such as the Deep Synthesis Provisions and the draft Artificial Intelligence Law, to manage AI's development and mitigate associated risks.

11. India:

- India is committed to becoming a global AI powerhouse by leveraging AI to stimulate economic growth, enhance public services, and improve the quality of life for its citizens. The government aims to integrate AI across critical sectors such as agriculture, healthcare.
- India has undertaken several policy actions including National Strategy for Artificial Intelligence (2018), Responsible AI for All: Part 1 and Part 2 (2021), Digital Personal Data Protection Act (2023) and the National Data Governance Framework Policy (Draft, 2022)

Multilateral Organizations

Several international organisations have developed AI policy recommendations. Below are a selection of notable projects:

	OECD	WEF	IEEE	UN
Type of framework	Non-binding Principles	Collaboration & Best Practices	Technical Standards & Ethical Guidelines	Binding Recommendation (UNESCO)
Key Areas of Focus	Trustworthy AI, Human-centricity	Responsible AI Governance, Best Practices	Technical Standards, Ethical Considerations	Ethics, Human Rights, Sustainability
Examples	OECD AI Principles, Classification Framework	AI Governance Alliance reports & initiatives	EAD, Initiative, Software Engineering Standards	UNESCO Recommendation on the Ethics of AI

OECD (Organisation for Economic Co-operation and Development):

- **Focus:** Non-binding principles for trustworthy AI.
- **Key Features:**
  - The OECD AI Principles (2019, updated 2024) advocate for human-centric AI development that respects human rights and democratic values.
  - They cover fairness, transparency, accountability, human oversight, and safety & security.
  - The OECD Framework for the Classification of AI Systems helps assess risks and opportunities of different AI types.

World Economic Forum (WEF):

- **Focus:** Multi-stakeholder collaboration and best practices for responsible AI governance.
- **Key Features:**

○ WEF fosters collaboration to develop best practices, not a single overarching framework.
○ The AI Governance Alliance tackles responsible AI integration across industries, promoting regulatory frameworks and technical standards for safe and advanced AI.
○ Reports and initiatives address specific areas like responsible AI in healthcare or finance.

IEEE (Institute of Electrical and Electronics Engineers):

- **Focus:** Technical standards and ethical guidelines for responsible AI development.

- **Key Features:**
  - IEEE develops technical standards for AI development, like software engineering practices for building trustworthy AI systems.
  - IEEE's Ethically Aligned Design (EAD) initiative provides guidance for ethical considerations throughout the AI development lifecycle.
  - These standards are not mandatory but offer a valuable reference for developers and companies.

United Nations (UN):

- **Focus:** Global ethical considerations and responsible AI development for social good.

- **Key Features:**
  - UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) sets the first global standard for ethical AI development and use.
  - It emphasizes human rights, fairness, sustainability, and accountability.
  - Other UN bodies like the International Labour Organization (ILO) address the impact of AI on work and the future of jobs.

Aside from IEEE, other standards bodies active in this area include NIST (USA), CEN CENELEC (EU), ISO and IEC.

There are also confederations such as the **Global Partnership on Artificial Intelligence (GPAI)**. GPAI is an initiative for global multi stakeholder cooperation around the topic of AI, and recently adopted (wholesale) the OECD Recommendation on Artificial Intelligence. Among other initiatives, GPAI is

- Seeking to achieve sustainable development using AI, while mitigating risks (notably around advanced AI systems);
- Fostering multistakeholder collaboration (across private sector, government, research & academia and other participants) to promote open source solutions as well as standards; and

● Cultivating a multi-stakeholder expert network, to better enable access to key insights and capabilities as well as foster dissemination of principles and ideas.

Part IV – Sovereign AI strategic considerations

1. Strategic Options

As policymakers evaluate options with respect to sovereign AI, there are several considerations which will inform its decision-making:

a) Wait and see: In fast evolving technology space, sometimes "do nothing and monitor" is an acceptable, or even recommended, policy response. For example, we recommended this path to several governments in 2013-2014 with respect to digital currencies such as bitcoin, since early intervention would have suppressed innovation without delivering clear public benefit. Prior to that, a wait-and-see attitude in the mid 1990s towards eCommerce allowed development of the field, so that subsequent regulations emerging in several countries in the late 1990s and early 00's could intelligently determine how issues such as tax should apply to the then-nascent field. However, given the speed with which artificial intelligence is being adopted at scale, and the emerging risks it is creating across different sectors, we do not recommend a completely passive response to AI disruption. This is not necessarily the same as recommending a proprietary sovereign AI, a federated sovereign AI, or other path, merely that 'do nothing' will likely not produce the optimal outcome for a given nation or region.

b) Considerations of scale: a viable path forward could be to focus on an industry-specific sovereign AI, or set of sovereign AI's. For example, a financial services AI, oriented around a discrete set of problems (fraud, security, credit, stability/risk management, customer service for banking etc.) would require less resourcing and less training time, on the one hand, and could have cross-border applicability facilitating regional or supra-regional collaboration.

c) Creating Sovereign AI: A nation or supranational could consider the advantages of creating its own AI system entirely within local borders. It would require considerable resources and represent a major national initiative. Advantages include demonstrating innovation and technological sophistication (a 'lighthouse project' that seeks to attract enterprise and foreign direct investment through signaling). Disadvantages include the fact that relative to overall GDP and government budget, the revenue requirements could be considerable for many countries, allocating resources away from other government priorities.

d) Partnering within Big Tech: not all 'big tech' are created equal, and some of the large platform companies are more amenable to creating local capability, segregating data to mitigate data leakage concerns, and taking other steps to make a viable path forward for nation-states. These options should be investigated in more depth as part of a structured process.

e) Adapting Open Source Code and Data: a number of open source code projects are currently in development, and could represent a 'non-aligned' base from which to develop sovereign AI (whether general purpose or industry-specific). See Appendix A for a review of a selection of open source AI projects. The Data Provenance Explorer project provides an assessment of a large number of open source repositories and investigates their constituent components.

f) Harvest Benefits of Other Government Initiatives: National AI strategy, digital strategies, and national digital identity project can be a core component contributing to shaping of its approach to sovereign AI.

2. Creating sector-specific GPTs and Expert Systems

One path forward could be the creation of a sector-specific GPT (such as a financial GPT or a health GPT). While investing in creating a ground-up, proprietary, multi-modal GPT would be both cost-prohibitive and duplicative to several existing systems from large, well-funded private companies (and some established open-source projects), a political body could carve out a distinct niche not only within its borders, but trans-nationally, by focusing development on asubject-matter-intensive AI system in a specific vertical market.

Benefits of such a specialized system include:

- A model trained on a discrete data set would deeply understand the domain-specific terminology, regulations, products, and services. This would make it more accurate in interpreting questions, generating responses, and performing analysis that fits the industry's unique needs.
- Some domains have strict regulatory and compliance requirements. A specialized GPT would be well-versed in these areas, helping users navigate regulatory landscapes, ensure compliance, and provide guidance on risk management.
- It could provide personalized advice to users, with a better understanding of risk tolerance, goals, and market conditions, thanks to the specialized data and models it is trained on.
- It could serve as an effective tool for client-facing operations, answering complex queries from customers, explaining products or services and resolving account or service issues.
- Trained on patterns of fraud or digital crime, the model could provide better identification of potentially fraudulent behavior or anomalous transactions.
- The GPT could provide more accurate risk assessments by analyzing profiles and historical data, enhancing risk decisions.
- A model specifically trained around a particular domain would have knowledge of the industry's data privacy requirements, which may require special handling of data (such as health data or financial data), enabling better compliance.

A verticalised sovereign GPT could not only enhance government services, reduce risk and fraud, and improve cyber security within a nation, but it could provide regional collaboration and benefits

across aligned countries, who could be contributing partners to a federated model. In such a system, there would be a ‘base layer’ of capabilities that would be jointly developed and shared across countries, and then each country would create its own private derivation.

3. Counterpoint: Alternatives to a Sovereign AI

New developments are creating additional choices in lieu of committing to a full-scale sovereign AI.

For example, the proposed strategy of ‘centralized training, decentralized inference’ suggests that the expensive base model training is still conducted in central clusters, perhaps still on GPUs, but that edge devices (e.g., even mobile phone processors) and/or lower-cost and more widelyavailable CPUs are used for inference (answering queries). While this does not ameliorate a number of the concerns such as those regarding export restrictions, values and ethics, it can help address questions around population surveillance, data privacy and aspects of behavior manipulation and security.

Other developments in terms of training algorithms that require less power and less compute in order to achieve comparable results to extant large-scale systems.

It is essential to understand that current technology and engineering limit the ability to create very large scale AI systems to a handful of the wealthiest and most technologically sophisticated nations.

Nation-states also need to consider the technology change question, when thinking about allocating substantial resources to sovereign AI. Generative AI, in particular, is evolving rapidly, as is the hardware that underpins it – with new generations appearing annually or even every few months. Constructing large-scale local data centers, a decision that has a multi-year lead time, exposes nations to the risk that they have made an incorrect bet on a particular technology stack.

An alternative is to leverage the investments of others, and continue to ‘rent vs buy’ on the question of AI technology. This approach fails to address a number of the considerations that drive one to pursue sovereign AI in the first place, but offsets the risk of picking the wrong technology foundation.

Next Steps

To investigate its options, if a sovereign grouping does not already have an effort under way, it immediately should create a high-level working group that incorporates industry, government and academic experts to help assess its options.
In parallel, and as part of this process, participation in multi-lateral dialogues on the topic will help educate officials about the nuances associated with each of its options, and the timing considerations associated with the domain area.
Cabinet-level ministries should be closely integrated into the process.
We recommend that the high-level working group provide a report back in 2025, to enable key government officials to consider policy formulation as to a given course of action to be undertaken.

About the Authors

David Shrier

David Shrier is a Professor of Practice, AI and Innovation, with Imperial College London, where he leads the Trusted AI Alliance (a multi-university collaborative focused on building responsible & trustworthy AI). He has helped over 100 governments to develop technology policy including advising the European Parliament on the EU AI Act, and leading the team that created the Commonwealth Fintech Toolkit. Other relevant work includes creating the Leadership and Diversity for Regulators programme when he was with University of Oxford, helping more than 50 nations shape financial inclusion policy. His latest book, Basic AI: A Human Guide to Artificial Intelligence was published in 2024.

.

Ayisha Piotti

Ayisha Piotti is the head of the AI Policy Summit, an annual event jointly organized by the ETH Zurich Center for Law & Economics and RegHorizon. She has over 20 years of experience in the private & public sector, including with the United Nations.

About the Authors

Alex Pentland

Alex Pentland has helped create and direct the MIT Media Lab and the Media Lab Asia in India, and is a HAI Fellow at Stanford. He is one of the most-cited computational scientists in the world, and Forbes declared him one of the "7 most powerful data scientists in the world" along with Google founders and the Chief Technical Officer of the United States. He co-led the World Economic Forum discussion in Davos that led to the EU privacy regulation GDPR, and was one of the UN Secretary General's "Data Revolutionaries" helping to forge the transparency and accountability mechanisms in the UN's Sustainable Development Goals.

Aldo Faisa

Aldo Faisal holds a prestigious UKRI Turing AI Fellowship, is a Professor of AI at Imperial College London and holds the Chair in Digital Health at the Universtät Bayreuth (Germany). He is since 2019 the founding director of the £50 million UKRI Centers in AI for Healthcare and AI for Digital Health in London. Since 2024 he is Director of Science and Innovation at the Alan Turing Institute responsible for the Grand Challenge in Health, the national AI research institute of the United Kingdom. His research focussing on Generative AI for health and Human-AI interfacing has won numerous international prizes and in 2024 the Federal German Government and Parliament appointed him as member of the German Ethics Council. He is an author of the Amazon Global Top 10 textbook Mathematics for Machine Learning.

The authors would also like to gratefully acknowledge the feedback and input we have received from Cameron Kerry, Lisette Cipriano, Alex Kotyck, Sabah Carter and Michael Huth. As we evolve our thinking around Sovereign AI we expect their input to be further reflected in subsequent drafts of this document.

Appendix A

Select Survey of Open Source AI Systems

A country aiming for "non-aligned" sovereign AI development can leverage open-source projects as a foundation. Here's how some major projects contribute:

NEAR (Neural Enhanced Re-ranking): Strengthens search and recommendation algorithms by refining results based on user interaction. A country can use NEAR to tailor search engines or recommendation systems for its citizens, prioritizing local content or addressing specific cultural nuances.

- **Merits:** NEAR is a blockchain-based open-source platform known for its focus on scalability, speed, and user-friendly development environments. It offers tools for building decentralized applications (dApps).
- **Adaptation for Sovereign AI:** Countries can use NEAR to develop secure and scalable AI-powered applications, especially those requiring decentralization. This is useful for enhancing transparency and trust in public services and financial systems.

Alpaca: Focuses on Natural Language Processing (NLP) tasks, particularly machine translation. A country with diverse languages can leverage Alpaca to build custom translation tools, promoting communication and information access within its borders.

- **Merits:** Alpaca is an open-source project focused on creating high-quality, performant, and easy-to-use machine learning tools. It emphasizes accessibility and efficiency in AI development.
- **Adaptation for Sovereign AI:** Alpaca can be used by countries to democratize AI development, making it accessible for smaller enterprises and educational institutions. This promotes widespread AI literacy and innovation.

Vicuna: Another NLP project specializing in question answering. By adapting Vicuna, a country can develop its own virtual assistants or chatbots that understand local languages and cultural references.

- **Merits:** Vicuna is an open-source initiative aimed at improving data analysis and machine learning integration in scientific research. It offers robust tools for managing and analyzing large datasets.
- **Adaptation for Sovereign AI:** Governments can leverage Vicuna to enhance research capabilities in healthcare, environmental science, and other critical sectors. This supports data-driven decision-making and scientific advancement.

Bloom: A large language model (LLM) trained on a massive dataset of text and code. A country can fine-tune Bloom on its own data to create a powerful AI for various tasks, from generating educational content to summarizing scientific research.

● **Merits:** Bloom is a large-scale open-source initiative focusing on AI for social good. It aims to leverage AI to tackle global challenges, from healthcare to climate change.
● **Adaptation for Sovereign AI:** By adopting Bloom, countries can develop AI solutions aimed at social impact, addressing local and global issues. This aligns AI development with national and humanitarian goals.

Falcon: An open-source framework for building and deploying computer vision models. A country can use Falcon to develop AI for tasks like traffic monitoring, security applications, or analyzing medical imagery specific to its population's health concerns.
● **Merits:** Falcon is an open-source framework designed for high-performance machine learning and deep learning tasks. It offers efficient tools for building and deploying AI models.
● **Adaptation for Sovereign AI:** Falcon's high performance makes it suitable for developing advanced AI applications in sectors such as defense, energy, and transport. This supports national strategic initiatives with powerful AI capabilities.

Mistral: Concentrates on Explainable AI (XAI), helping understand how AI models arrive at decisions. This transparency is crucial for building trust in sovereign AI. By using Mistral, a country can ensure its AI systems are fair, unbiased, and accountable to its citizens.
● **Merits:** Mistral is an open-source project focused on providing robust, scalable, and efficient tools for AI and machine learning. It is particularly known for its applicability in large-scale data environments.
● **Adaptation for Sovereign AI:** Mistral can be employed by countries to manage and analyze vast amounts of data in sectors like finance, telecommunications, and urban planning. This facilitates informed policy-making and efficient resource management.
While not strictly "open-source AI projects" in the same way as NEAR or Bloom, the following frameworks and libraries play a crucial role in building and deploying AI models.

TensorFlow & PyTorch: These are dominant open-source frameworks for building and training machine learning models. A country can leverage them to develop its own AI applications without vendor lock-in. TensorFlow offers flexibility for diverse tasks, while PyTorch is known for its ease of use in research.

Tensorflow

● **Merits:** Developed by Google Brain, TensorFlow is a highly flexible and comprehensive open-source library for machine learning and neural network research. It supports a wide range of applications from training models to deploying them across different platforms.
● **Adaptation for Sovereign AI:** Countries can leverage TensorFlow's flexibility to build customized AI models tailored to national needs. Its extensive documentation and community support make it easier to train local talent and adapt the technology for various industry-specific applications.

PyTorch

● **Merits:** Backed by Facebook's AI Research lab, PyTorch is known for its dynamic computational graph, ease of use, and strong support for research and development. It has become a favorite among researchers due to its intuitive design and robust community.
● **Adaptation for Sovereign AI:** PyTorch's dynamic nature allows for quick prototyping and iterative development, which is beneficial for experimenting with new AI techniques. Governments can use it to foster innovation in AI research and to build adaptable AI solutions that meet local requirements.

Hugging Face Transformers: This library provides pre-trained models for various NLP tasks and integrates seamlessly with TensorFlow and PyTorch. A country can use it to accelerate NLP development in its sovereign AI projects, like building chatbots or sentiment analysis tools.
● **Merits:** Hugging Face Transformers provides state-of-the-art natural language processing (NLP) capabilities through pre-trained models. It supports a wide range of tasks like text classification, translation, and summarization.
● **Adaptation for Sovereign AI:** Utilizing Hugging Face's models, countries can develop powerful language models tailored to local languages and dialects. This is crucial for applications in government, healthcare, and education where understanding local languages is vital.

Apache MXNet: Another open-source deep learning framework, known for its scalability and efficiency. A country with limited computational resources might choose MXNet for specific AI tasks due to its ability to handle large datasets on modest hardware.
● **Merits:** Supported by Amazon Web Services, Apache MXNet is designed for high efficiency, flexibility, and scalability. It supports multiple languages and provides deep learning capabilities.
● **Adaptation for Sovereign AI:** MXNet's scalability makes it ideal for large-scale AI deployments. Governments can use it to develop AI infrastructure that supports national projects in smart cities, transportation, and public safety.

Open Neural Network Exchange (ONNX): An open standard for representing deep learning models. By adopting ONNX, a country can ensure its AI models are interoperable across different frameworks, fostering collaboration and easier deployment of sovereign AI solutions.
● **Merits:** ONNX is an open standard for representing machine learning models, allowing models to be transferred between different frameworks. This interoperability is key for ensuring flexibility in AI development.
● **Adaptation for Sovereign AI:** By adopting ONNX, countries can ensure their AI systems are interoperable and can be integrated with various AI frameworks. This promotes flexibility and reduces dependency on any single technology provider.

Keras: A high-level API built on top of frameworks like TensorFlow. It simplifies the process of building and training deep learning models, making AI development more accessible for a wider range of programmers within a country

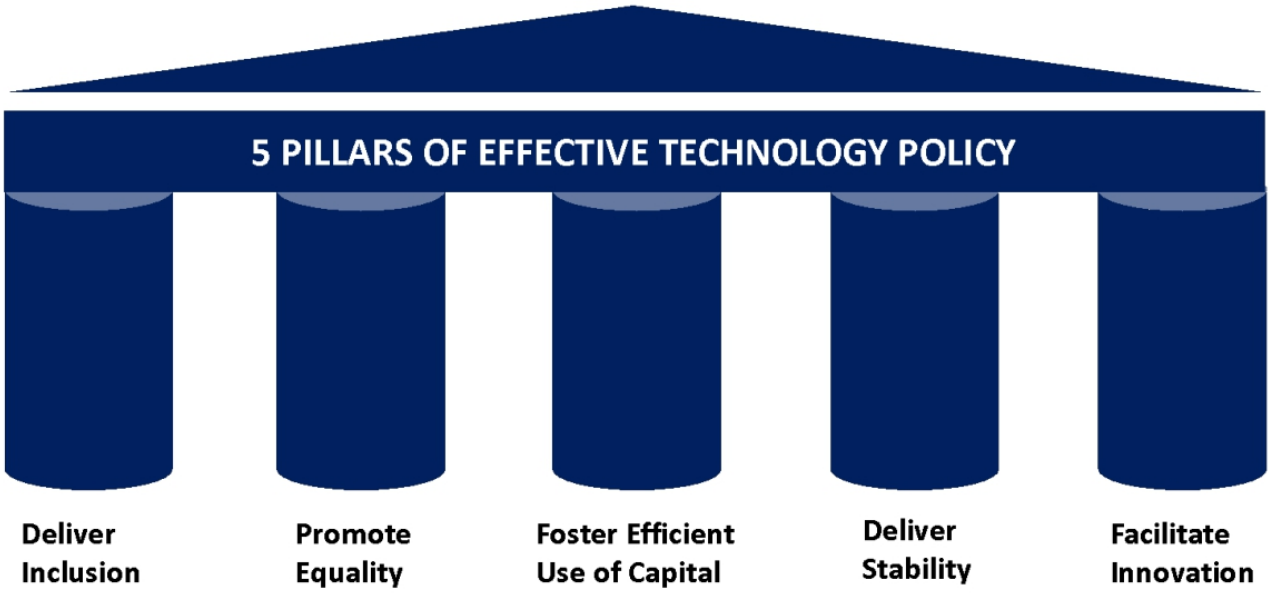
● **Merits:** Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It is user-friendly and modular, making it suitable for rapid prototyping.
● **Adaptation for Sovereign AI:** Keras's simplicity and ease of use make it an excellent tool for training local developers and researchers. It can be used to quickly develop and test AI models for various applications, promoting local innovation and rapid deployment.

Scikit-learn: While not specifically focused on deep learning, Scikit-learn provides a comprehensive library for traditional machine learning tasks like classification, regression, and clustering. This allows a country to incorporate various AI techniques into its sovereign AI strategy, even for projects that don't require deep learning models.
● **Merits:** Scikit-learn is a robust library for machine learning in Python, offering simple and efficient tools for data mining and data analysis. It is built on NumPy, SciPy, and matplotlib.
● **Adaptation for Sovereign AI:** Scikit-learn's versatility in handling basic to intermediate machine learning tasks makes it ideal for educational purposes and small to medium-sized AI projects. Governments can use it to build foundational AI capabilities and train the workforce.

Appendix B

Summary of Relevant AI Policy Initiatives by Country

Our preliminary horizon scanning reveals a number of commonalities across jurisdictions, with a few select differences emerging depending on alignment across key dimensions such as protection of personal rights versus enabling private sector innovations.
In assessing different policy interventions across jurisdictions, it is useful to review the 5 principles of effective regulation:



There is a tension between private sector imperatives around growth and market access (such as the goals of having as much data as possible to train generative AI systems) with the objective of protecting personal data privacy (putting more value in the hands of the individual). Different jurisdictions place different emphasis. For example, the US has relatively weak personal data protection standards (excepting California, whose CCPA regulation consciously imitates the strong protections of the EU's GDPR) and has robust commercial activity in AI. The EU has a stronger emphasis on personal data privacy, and has seen recent actions such as Apple's announcement that it will withhold its latest generation technology, Apple Intelligence, because of restrictions placed by the Digital Markets Act. California recently institutes new restrictions around synthetic media and other AI innovations, which may create both limits on private sector activity in that state, and open opportunity in others.

As governments consider design of AI policy, they will face critical decisions regarding the balance between creating attractive economic opportunity for private sector actors and protecting citizens from exploitation.

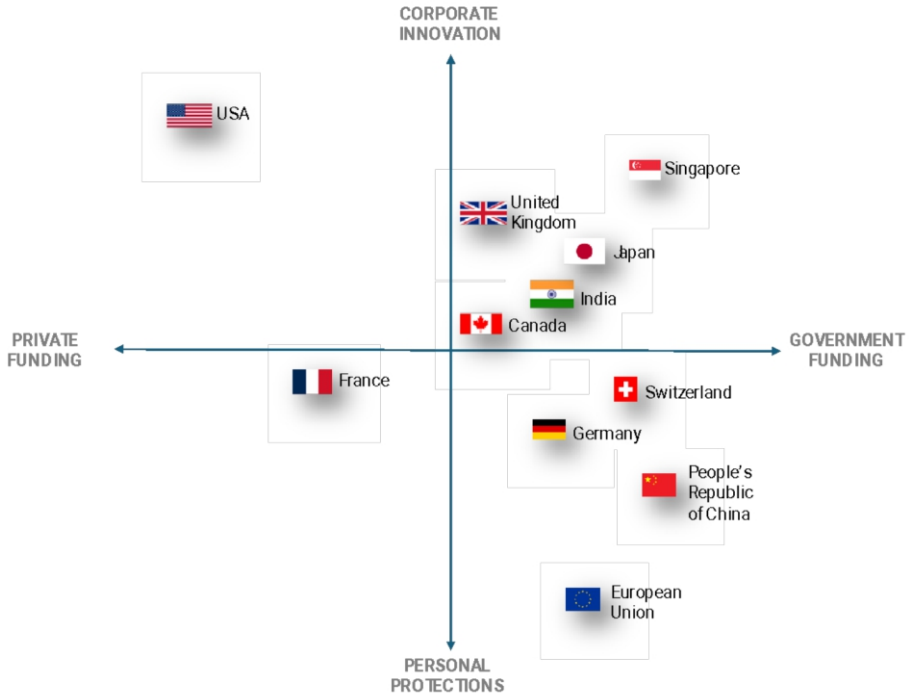
Major AI policy interventions in other jurisdictions can be found on the following pages, prefaced by a summary table, which is proceeded by domicile-by-domicile analysis. We summarised activity in the following domiciles:

1. EU
2. Japan
3. Singapore
4. France
5. USA
6. Germany
7. Switzerland
8. UK
9. Canada
10. People's Republic of China
11. India

Each jurisdiction is presented at the axis of the following information:

A. Policy Summary
B. Laws & Regulations Implemented
C. Policy Initiatives
D. Consultations in Process
E. Proposed Legislation
F. Commentary

When we compare across countries and regions, we can provide a simplified analysis on the dimensions of innovation vs. personal protection, and reliance on private funding versus government support:



We note that many governments are investing a significant amount of capital, relative to the sizes of their economies and AI productivity bases. The US is a notable exception, with a relatively small federal government commitment to AI – a fraction of the size of its defense budget – and a heavier reliance on private sector capital and actors. The US also hosts some of the world's largest AI companies – Alphabet (Google), Microsoft, Amazon, Meta, OpenAI, Anthropic and NVIDIA among others, with several trillion dollars of market capitalization, so perhaps feels less urgency to address the issue with direct government funds,

and instead relies on using US government moral suasion to encourage private sector investments.

When we look at research output, the top countries publishing AI research (aside from the US) also see government supporting significant private sector activity in AI:

Country	AI Publications (in 000s) 2016-2020	AI Index Rank
China	76.3	2
United States	44.4	1
India	27.0	14
United Kingdom	16.0	4
Japan	13.0	12
Germany	12.9	8
France	10.9	13
Spain	9.7	21
Italy	9.1	23
Australia	8.5	15
Canada	8.2	5
South Korea	6.5	6
Brazil	6.0	35
Poland	5.8	27
Iran	5.7	*Not ranked*
Turkey	5.6	39
Russia	4.8	30
Taiwan	4.8	26
Singapore	4.1	3
Hong Kong	4.1	32
Malaysia	3.8	*Not ranked*
Netherlands	3.6	11
Switzerland	3.3	9
Portugal	3.2	29
Saudi Arabia	2.9	31

Sources: Statistica 2023; Tortoise Media 2023

Assessing AI Competitiveness

Sheer volume of publication does not automatically confer overall AI competitiveness, when taking into account other factors like talent availability, infrastructure, ease of doing business and other dimensions of the operating environment, and development of research into commercial activity.

Policymakers would be well served to understand why they might have high research productivity but relatively low AI competitiveness. What are the rate limiting factors?

Typically, dimensions that we see when evaluating countries, that either show high or low overall AI competitiveness, include:

- **Ease of doing business** (incorporation, licensing, hiring)
- **Access to talent** (how well or poorly has the country enabled high skill workers to migrate and be employed? What is the strength of the immigrant-worker visa programme?)
- **Access to capital** (how robust is the venture capital market? How large, and how easy is it for new entrants to access capital?)
- **Translation of research to innovation** (how incentivised are the universities to commercialise their work? How easy is it to spin out research into a company?)
- **Protection of intellectual property** (how strong or weak are the IP protections?)
- **Business environment** (how willing are commercial actors to adopt AI technology?)
- **Rule of law** (how strong are the courts and the jurisprudence foundations? How easily are protections applied against monopolistic, unethical or illegal acts?)

Comparison of AI Policy Initiatives by Domicile

	EU	Japan	Singapore	France	USA	Germany	Switzerland	UK	Canada	China	India
Values/ Ethics	Human-centric and trustworthy AI	Human-centric AI and fairness	Fairness, ethics, accountability and Transparency (FEAT)	Transparency and Fairness	Democratic values and human rights	Responsible and public welfare-oriented AI	Human-centric approach	Transparency and explainability	Alignment with international human rights and responsible governance	Alignment with socialist values and national security, ensuring ethical AI development.	Ethical AI development, emphasizing fairness, transparency, and accountability
Economic	Support for AI start-ups and SMEs	Support for Startups and R&D, Innovation	Economic and Business support (SMEs), National AI R&D plans	Innovation and Economic support of SMEs, R&D	Innovation and Competition, investments in R&D	Research and Innovation, Technology Leadership, support for startups	Encouraging innovation	Investment in R&D and economic growth	Economic growth and commercialization-support for business	Aiming to lead the global AI market by 2030 with significant investments in AI infrastructure.	AI for national development, enhancing sectors like agriculture, healthcare, and education
Safety	Ethical guidelines for AI development and regulation	Privacy Protection, ensuring security	Trusted development and deployment, protection of data	Data protection and Privacy	AI Bill of rights, Safe, secure and trustworthy development of AI, privacy and civil liberties	Data security, explainability	Transparency, Traceability and Explainability. Safe, robust systems	Public Trust and Responsible Innovation	Transparency, accountability and trust	Introduced measures to regulate synthetic content and ensure AI safety	Compliance with stringent data privacy requirements
Risk management	Risk-based approach for legislation	Guidelines to address AI risks, transparency and accountability	Accountability, Responsibility-testing and assurance	Transparency of algorithms, ethical code for programmers	Evaluation and policies to test and mitigate AI risks, consumer protection	Setting national AI standards, Protecting personal rights	Accountability and liability. Putting people first	Clear, adaptable and trustworthy regulatory regime. Responsible innovation	Creation and adoption of standards related to AI	Focused on national security and strict AI regulations	Ethical guidelines ensuring that AI is deployed transparently and safely

	EU	Japan	Singapore	France	USA	Germany	Switzerland	UK	Canada	China	India
International Collaboration / Standards	Ethical Framework and Standards across the EU	International Collaboration and Standards	Global Outlook and international collaboration	European and International Cooperation	International Cooperation	International cooperation and social dialogue	Engagement in Global AI governance	Global Leadership and cooperation, promote cross-border interoperability	International collaboration and standards	Setting national AI standards that align with international norms	Active in global AI partnership
Infrastructure	Infrastructure Investment - AI factories, data infrastructure	Data management and infrastructure	Infrastructure Investment and Data management, sectoral focus	Infrastructure and Data management	Infrastructure investment and data management	Infrastructure investment - computational resources, AI competence centers	Digital Infrastructure	Investment in digital Infrastructure-smart homes, electro mobility	Research institutes, AI compute infrastructure	Investing heavily in AI infrastructure and core technologies	AI across sectors to drive inclusive growth and modernization
Workforce	Education, skills and Talent Development	Education and skills development	National Competency, Education and Talent Development	Education and Talent Development	Investment in Education and AI training, focus on workers rights	National training strategy	Investment in education and building new skills	Create new AI-related jobs	Development research Talent, education and training	Promoting AI-driven productivity in key industries	Upskilling and integrating AI in public services
Legal Framework	Liability and Legal Framework	Fair competition and market regulation, Soft law	Soft Law and flexible governance	Legal Framework via EU AI Act	(Proposed) Federal and state regulations	Legal Framework via EU AI Act	Soft law guidelines,AI regulation, accountability and proposed legislation	Soft law approach	(Proposed)federal regulation and ethical framework	Active approach – Implemented laws like the draft AI Law and generative AI rules	Existing laws govern AI, with new legislation in progress

	EU	Japan	Singapore	France	USA	Germany	Switzerland	UK	Canada	China	India
Commentary	Lack of emphasis on changes to workforce/ work environment, job insecurity	Ethics Ineffectiveness, Corporate Agenda	Good Adaptability	Implementation Challenges, Investment needed	Favoring Big Tech, regulatory Gaps	Data Protection, Talent shortage	Skills training, public skepticism	Economic focus, alignment versus enforcement on ethical standards	Challenges concerning regulatory clarity and public trust	Stringent security standards posing challenged for AI service providers	AI generated content under the government strict supervision

Source: Trusted AI Alliance analysis, Imperial College London 2024

Detail on Government Policy Initiatives in Select Domiciles

1. European Union

A. Policy Summary

Political Goals :

● **Build a resilient Europe for the Digital Decade:** people and business should be able to enjoy the benefits of AI while feeling safe and protected.
● Make the EU a **world-class hub** and ensuring that AI is **human-centric and trustworthy**. Such an objective translates into the European approach to excellence and trust through concrete rules and actions.
● **Strengthen Europe's potential to compete globally.**

Laws regulating AI : Artificial Intelligence Act, discussions on AI Liability Directive (proposal)

Policy Initiatives : support for AI start-ups and SMEs, financial and talent support, development of AI factories, AI office, EDICs.

Public Consultations : AI in Financial Sector

Commentary :

- AI regulations in the EU needs clearer employer responsibilities for AI-related Safety
- Psychosocial Risks: Addressing stress and job insecurity due to AI.
- Trustworthiness and Rights: Lack of mechanisms to ensure AI protects individual rights.
- Supply Chain Control: Challenges in managing complex digital supply chains.
- Fragmented Regulation: Inconsistent risk assessments and terms with technology providers.
- Interoperability Issues: Limited promotion of cross-border interoperability.
- Digital Skills Shortage: Dependency on tech providers due to public sector's limited skills.

B. Laws & Regulations Implemented

The EU AI ACT

- First-ever legal framework on AI Liability Directive
- Future-proof legislation - rules can adapt to technological change
- The EU AI Act aims to set clear requirements for AI developers and deployers, reducing burdens for SMEs, and promoting trustworthy AI. It ensures AI systems respect fundamental rights, safety, and ethical principles, addressing risks from powerful AI models.
- Enforcement and implementation -managed by the European AI office

The EU AI Act operates on a risk-based approach, categorizing AI systems into four levels of risk:

1. Unacceptable Risk: AI systems that pose a clear threat to safety, livelihoods, and rights (e.g., social scoring by governments, dangerous behavior-inducing toys) are banned.

2. High Risk: AI used in critical areas (e.g., infrastructure, education, safety, employment, essential services, law enforcement, migration, justice) must meet strict requirements like risk assessments, high-quality datasets, traceability, documentation, human oversight, and robustness. Remote biometric identification has strict regulations with limited exceptions

3. Limited Risk: AI systems with transparency issues (e.g., chatbots) must ensure users are informed when interacting with AI. AI-generated content must be identifiable, especially in matters of public interest and deep fakes.
Minimal or No Risk: Low-risk AI applications (e.g., video games, spam filters) are freely usable. Most EU AI systems fall into this category.

C. Policy Initiatives

In January 2024, the Commission launched measures to support startups and SMEs in developing trustworthy AI. This package includes several initiatives:

1. Supercomputing Access: proposal to provide privileged access to supercomputers to AI startups and the broader innovation community.

2. AI Factories: The **EuroHPC Regulation** will be amended, a new initiative within the EU's supercomputers Joint Undertaking. These factories will acquire, upgrade, and operate AIdedicated supercomputers for fast machine learning and training of large General Purpose AI (GPAI) models. They will provide access to these supercomputers for public and private users, including startups and SMEs, and serve as a one-stop shop for AI development, testing, and validation, offering supercomputer-friendly programming facilities and other enabling services.

3. AI Office: An AI Office within the Commission will coordinate AI policy, ensure implementation of the AI Act, and promote EU AI governance internationally.

4. Financial and Talent Support: Through Horizon Europe and the Digital Europe programme, the EU will provide around €4 billion in public and private investments (until 2027), along with initiatives to enhance AI talent through education and training.
- The Commission will also mobilise additional investments from the private sector and the Member States in order to reach an **annual investment volume of €20 billion over the course of the digital decade**.

5. GenAI4EU: The GenAI4EU initiative aims to support the development of novel use cases and emerging applications in Europe's 14 industrial ecosystems, as well as the public sector. Application areas include robotics, health, biotech, manufacturing, mobility, climate and virtual worlds.

The Commission is also establishing two European Digital Infrastructure Consortiums (EDICs): The ALT-EDIC will develop language technologies to support linguistic diversity, while the CitiVERSE EDIC will use AI to enhance smart city operations.

Other Fields of Action:
Transfer to the market :
- Networking of European AI excellence research centres
- Construction of test facilities (e.g. for connected and autonomous driving)
- Developing platforms and large-scale pilot projects with AI elements in areas such as energy, healthcare, manufacturing, geospatial information and agriculture

- Promoting the integration of AI and data analytics in lighthouse initiatives in manufacturing, mobility, personalized medicine
- Transition from individual test projects to concrete, far-reaching value creation activities
Infrastructure

Education, research and skills :
- Promoting AI-relevant skills
- Retaining and attracting the best scientists
- Pooling of competence centres
- Supporting changes in the labour market through specific training programmes financed by the European Social Fund

Infrastructure :
- Investments in high-performance computers, quantum computers, and AI and data infrastructure
- Further development of the European Cloud for Open Science
- Development of a Europe-wide high-performance computing infrastructure

Ethical Framework :
- Development of standards that specify the contents of the AI Act for practical application (European Committee for Standardization CEN, European Committee for Electrotechnical Standardization CENELEC)
- Development of concepts for monitoring and certification of risky AI systems
- Establishment of ethical guidelines for the development of AI (based on the EU Charter of n Fundamental Rights); close cooperation in the “European AI Alliance”
- Orientation towards the principles of European data protection and European product liability directives

Strategic measures and coordination :
- Further development of Member States’ national AI strategies
- Development of sector-specific joint actions by Member States

Vision: Establish AI policies and make investments

D. Consultations in Process

Targeted consultation on AI in the financial sector :

- The Directorate-General for Financial Stability, Financial Services and Capital Markets Union is conducting a consultation to collect feedback from financial services stakeholders on AI systems. Open from 18 June to September 2024, the consultation aims to understand AI’s application and impact in financial services. This will help the Commission assess market developments, risks, and support the implementation of the AI Act in the financial sector.

- The questionnaire is divided in three parts: general questions on AI development, questions on specific use cases in finance, and questions related to the AI Act and the financial sector.

https://finance.ec.europa.eu/regulation-and-supervision/consultations0/targeted-consultation-artificial-intelligence-financial-sector_en

E. Proposed Legislation

AI Liability Directive Proposal

(https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2022/0303(COD)&l=en)

The European Commission published a proposal for a directive on adapting non-contractual civil liability rules ro AI in Sept 2022. Awaiting Committee decision.

- Current fault-based liability rules are not suited to handling liability claims for damage caused by AI-enabled products and services. The existing EU liability framework consists of the Product Liability Directive 85/374/EEC(the 'PLD') and of national liability rules that apply in parallel.

- The Commission proposes to complement and modernise the EU liability framework to introduce new rules specific to damages caused by AI systems. The new rules intend to ensure that persons harmed by AI systems enjoy the same level of protection as persons harmed by other technologies in the EU. The AI liability directive would create a rebuttable 'presumption of causality', to ease the burden of proof for victims to establish damage caused by an AI system. It would furthermore give national courts the power to order disclosure of evidence about high-risk AI systems suspected of having caused damage.

Objective :

- Improve the functioning of the internal market by establishing uniform non-contractual civil liability rules for AI-related damages.
- Promote trustworthy AI and ensure victims receive the same protection as for other products.
- The proposal also aims to reduce legal uncertainty for businesses developing or using AI regarding their possible exposure to liability and prevent fragmented AI-specific adaptations of national civil liability rules.

F. Commentary

Maciej Jarota, 'Artificial intelligence in the work process: A reflection on the proposed European Union regulations on artificial intelligence from an occupational health and safety perspective' (2023) 49 Computer Law & Security Review 105825.

● The European Commission's proposed AI regulation inadequately addresses the employer-employee relationship, focusing solely on supplier liability and requiring employers to follow AI usage instructions. This approach fails to assign clear responsibility for workplace safety and health conditions involving AI, potentially leading to blurred accountability. The regulation should explicitly identify employers as responsible for these conditions to ensure comprehensive safety measures.

● Similar to the Machinery Directive (Directive 2006/42/EC), the proposed regulation does not address employer obligations towards employees, despite AI's unpredictability and potential to introduce new risks. Autonomous AI operations can impose unforeseen challenges, including physical risks and psychosocial stress for employees. Therefore, the development of clear rules for human-AI collaboration is essential to safeguard working conditions.

● EU law mandates general occupational health and safety (OHS) obligations under Directive 89/391/EEC (article 6(1) and (2)).

While these provisions establish a

framework for worker health protection, including with AI, they are general and may not adequately address the specific challenges posed by AI's evolving role in the workplace. As AI systems can autonomously allocate tasks, employers must retain control over these systems to ensure fair work distribution and safety.
● EU does not adequately address psychosocial risks, such as stress from AI management, job insecurity, and potential burnout, which can significantly impact employee well-being. The legislator should explicitly state in EU law the main risks associated with using artificial intelligence by employers and the recommended responses to them.
● The proposed regulations lack specific obligations for employers to communicate AIrelated OHS risks to workers. While Article 10 of Directive 89/391/EEC mandates employers to inform workers about safety and health risks, it does not explicitly address AI-specific risks or methods for effective communication within large organizations.
This information gap, exacerbated by asymmetry between employees and managers, can lead to inadequate health protection measures. The EU should implement mechanisms to ensure employers effectively communicate all AI-related risks provided by suppliers, fostering a well-informed workforce capable of addressing these challenges.
● Despite extensive obligations under Directive 89/391/EEC, implementation shows that employers are not adequately monitoring OHS measures. There is a lack of a coordinated OHS approach among European employers, with varying practices across EU countries. Some focus on psychosocial risks while others do not. The inconsistency in addressing new and psychosocial threats highlights the need for dedicated EU
regulations on AI-related occupational safety.

Albert Sanchez-Graells, 'Public Procurement of Artificial Intelligence: Recent Developments and Remaining Challenges in EU Law' (2024) 2 LTZ (Legal Tech Journal) 122, 25 January 2024.

● The author criticizes the current EU public procurement legislation for lacking binding mechanisms to ensure that procured technologies are "trustworthy" and align with the protection of individual rights and freedoms or broader digital regulation goals such as those outlined in the European Declaration on Digital Rights and Principles.
● The existing procurement mechanisms are also criticized for their limitations in controlling increasingly complex digital supply chains, posing significant technical and commercial challenges, particularly concerning cybersecurity.
● Under the current, fragmented regulation, Member States and contracting authorities at all government levels can procure any digital technology and implement it without necessarily conducting a proper risk assessment of its implications. They are also free to agree on any terms with technology providers, irrespective of potential effects on the digitalization of other services or future procurements
● Efforts to promote cross-border interoperability through the proposed Interoperable Europe Act are limited to cases of "high impacts on cross-border interoperability." This limitation may lead to insufficient promotion of future interoperability for public services not yet integrated across borders.
● There is a general shortage of digital skills within the public sector in most Member States. Contracting authorities often face significant risks of being overly dependent on technology providers due to their limited digital skills.

2. Japan

A. Policy Summary

Political Goals :

- Increasing the productivity of society and the creativity of the population
- Leading the way in healthcare and welfare technology by leveraging big data
- Improved travel for citizens, increasing environmental friendliness and preventing all accidents by 2030
- Creating a robust economic development of artificial intelligence, accompanied by appropriate evaluation criteria and prices

Japan currently has no specific laws governing AI. No active public consultations currently.
Proposed legislation : Basic Act on the Advancement of Responsible AI.
Policy Initiatives : Social Principles of Human-Centric AI(2019); Hiroshima AI Process Comprehensive Policy Framework (2023) ; AI Guidelines for Business Version 1.0

Fields of action:

1. Research and development
2. Data management
3. Productivity
4. Human resources
5. Mobility
6. Health, medical care and welfare

B. Laws & Regulations Implemented

At this time, there are no laws or regulations in Japan specifically enacted to govern the development, use, or provision of AI. Instead, Japan is extending existing laws and regulations, on the one hand, and providing AI-focused guidelines and principles, on the other.

- The current approach focuses on supporting innovation while minimizing harms.
- Specific legislation is being developed, but has not yet been adopted.

Other laws **indirectly** affecting AI:

- **Digital Platform Transparency Act:** Ensures transparency and fairness in transactions for large online malls, app stores, and digital advertising businesses.
- **Financial Instruments and Exchange Act:** Requires businesses involved in algorithmic high-speed trading to register, establish risk management systems, and maintain transaction records.
- **Civil Code:** Allows tort claims against individuals instructing AI to produce and publish defamatory content.
- **Copyright Act and Act on the Protection of Personal Information:** Apply to inappropriate uses of AI, protecting intellectual property and personal information.

C. Policy Initiatives

In **2019,** the Japanese government published the Social Principles of Human-Centric AI as principles for implementing AI in society.
The Social Principles set forth three basic philosophies: **human dignity, diversity and inclusion, and sustainability**.
- AIM: Not to restrict the use of AI in order to protect these principles but rather to realize them *through* AI.

To achieve these goals, **the Social Principles set forth seven principles surrounding AI**:
1. Human-centric;
2. Education/literacy;
3. Privacy protection;
4. Ensuring security;
5. Fair competition;
6. Fairness, accountability, and transparency;
7. Innovation.

Hiroshima AI Process Comprehensive Policy Framework:

Launched by the G7 under Japan's presidency in May 2023 to address the global challenges and opportunities posed by advanced AI systems, including generative AI. **The framework aims to promote safe, secure, and trustworthy AI development and use.**
- Includes the **"Hiroshima Process International Guiding Principles for All AI Actors"** and the **"Hiroshima Process International Code of Conduct for OrganizationsmDeveloping Advanced AI Systems."**

Hiroshima Process International Guiding Principles for All AI Actors :

- **Transparency and Accountability:** AI developers are required to publicly report the capabilities of advanced AI systems and domains of inappropriate use.
- **Protection of Intellectual Property:** Ensuring that AI innovations are protected and used responsibly.
- **Digital Literacy:** Encouraging users to improve their understanding of AI technologies to mitigate risks like disinformation.

Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems
- **Risk Management:** Detailed guidelines for developers on identifying and addressing risks associated with AI.
- **Ethical Use:** Encouraging the development of technologies that allow users to identify AIgenerated conten

AI Guidelines for Business Version 1.0

Published on **April 19, 2024, the guidelines are not legally binding** but aim to encourage voluntary compliance with recognized AI principles by developers, providers, and business users.
AIM - Promote **"agile governance"** involving continuous and rapid cycles of environment and risk analysis, goal setting, system design, operation, and evaluation.

The Guidelines provide certain general principles which **AI businesses actors are expected to incorporate into the training and deployment of their products and services**; however, it is up to each business actor, taking into consideration the likely risks, to determine how to give effect to the principles. Principles :

1. **Human centric** - AI must respect fundamental rights
2. **Safety** - Avoid harm to lives, minds and properties
3. **Fairness** - Eliminate bias and discrimination
4. **Privacy Protection** - Respect and protect privacy
5. **Transparency** - Provide necessary information to stakeholders and ensure AI systems verifiability
6. **Accountability** - Ensure traceability and conform to guiding principles based on roe and risk
7. **Education/ literacy** - Educate employees and stakeholders on AI use, misinformation and potential misuse.
8. **Ensuring fair competition** - Maintain a fair competitive environment for new AI businesses and services.
9. **Innovation** - Promote innovation, considering interconnectivity and interoperability.

Fields of action :

Research and development :

- Tripling corporate investment in universities and R&D facilities by 2027
- Promotion of skilled workers, especially in the top-level area of industry-science cooperation
- Creating a working and scientific environment that incentivizes AI experts
- Improving energy efficiency and reducing space requirements of supercomputers

Data management :

- Collecting necessary information to create an effective data environment and embedding input/output devices (e.g. sensors)
- Creating incentives for companies to provide their data
- Increasing data permeability using standard profiles and standardized data formats

Human resources :

- Reforms of the education system, from primary school to university
- Programming classes starting in primary school and a focus on AI-related degrees at universities

Mobility :

- Increasing freedom of movement, environmental friendliness and safety in travel
- Establishment of AI-based sharing services for people and goods

Productivity :

- Customized mass production through automation and optimization of production systems
- Improvements in the service sector
- Supporting start-ups through “open innovation”

Health, medical care and welfare :

- Preventing diseases and increasing life expectancy through AI-based preventive medicine
- Reduction in social spending
- Tackling the problems of a shrinking workforce

D. Consultations in Process

Currently, there are no active public consultations on AI policy in Japan.

The most recent consultation, which shought public comments on the Draft AI Guidelines for Business to Open, opened on the 20th of January and ended the 19th February 2024.
- This consultation, organized by the Ministry of Economy, Trade and Industry (METI) and the Ministry of Internal Affairs and Communications (MIC), aimed to update and integrate existing AI guidelines in response to rapid technological advancements, including generative AI.

E. Proposed Legislation

On **May 22, 2024, the Council submitted draft discussion points** (Draft Discussion Points) concerning the advisability and potential scope of any future regulation.
- The Draft Discussion Points identify **several risks that Japan should prioritize**, including **safety, privacy and fairness, national security and crime, property protection, and intellectual property**.
- Some government officials advocate for a risk-based approach, **suggesting hard law regulations for high-risk AI systems while maintaining soft law measures for other areas**.
- The Draft Discussion Points propose **a classification system dividing AI developers, providers, and users into "large impact and high risk"** and "little impact and low risk" groups, tailoring the regulatory approach accordingly.

Basic Act on the Advancement of Responsible AI :
A working group has proposed **implementing hard law regulations** for certain generative AI foundation models.
- The government would designate specific AI systems and developers for regulation.
- Imposes obligations on vetting, operation, and output of these AI systems.
- Requires **periodic reports** concerning regulated AI systems.
- Obligations **likely similar to voluntary commitments by major US AI companies to the Biden Administration in July 2023.**
- Government would monitor AI developers and enforce compliance through **fines and penalties**.
- Marks a shift from soft law to hard law approach in Japan.

F. Commentary

James Wright, 'The Development of AI Ethics in Japan: Ethics-washing Society 5.0?' (2023)
- **Munn argues that AI ethical principles are largely ineffective** due to their contested nature, isolation from practical applications, and alignment with corporate agendas, making them difficult to apply and lacking in meaningful consequences.
- **The creation of the Social Principles was seen as a balancing act to ensure acceptability by both the US and People's Republic of China**.

The inclusion of a "uniquely Japanese" principle, specifically one prohibiting military use of AI, was debated but ultimately dropped to achieve international consensus.
- Kozuka notes that Japan's approach to AI ethics reflects its **broader strategy of stateinduced self-regulation in science and technology governance**. This approach may evolve to address regulatory gaps, potentially aligning with international standards such as the EU's AI Act.
- The Japanese government has invested significantly in developing ethical AI principles to position itself as a leader on the global stage, particularly ahead of regional competitors like China and South Korea.
- **The content of the principles appears secondary to Japan's role in contributing to an international alignment of states committed to ethical AI. Japan aims to lead in setting ethical principles, governance rules, and international technological standards**.
- **Japan's AI ethics are influenced by science fiction**, with references to Asimov, Astro Boy, Doraemon, and the idea that future AI agents should follow ethical guidelines.The government's vision includes a futuristic Society 5.0, imagined as a super-smart global utopia enabled by AI and other advanced technologies, where all societal barriers are removed.
- Current research and development are not impeded by stringent guidelines since these are expected to be applied only to more advanced future technologies.
- **This future-oriented approach can lead to the overlooking of present-day issues, such as gender discrimination in the AI sector..**
- **Japan retains the Ethical, Legal, and Social Implications (ELSI) approach**, but there is a significant gap between the techno-utopian promises of government and industry and the reality of research and development outcomes.
- Japan's approach is not unique but reflects global trends like the commercialization of science, strategic government framing of AI, and ethics washing

3. Singapore

A. Policy Summary

Political Goals :
- Developing national competencies in artificial intelligence for social and economic progress
- Promoting AI experts and building an AI-supported economic system
- Promoting business-related research into applications in the digital city (including life, transport, health)

No specific laws, statutory rules, or regulations in Singapore that directly regulate AI.

Policy Initiatives: National AI Strategy 2.0 (NAIS 2.0) (2023); Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems; Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT); Artificial Intelligence in Healthcare Guidelines; Model AI Governance Framework for Generative AI (30 May 2024)

Currently, there is no active public consultation on the topic of AI.

Commentary: Guidelines are adaptable and flexible; creation of AI adapted to South-East Asian context may be an example for other nations to follow, Talent shortage and Dependency on Foreign Talent.

B. Laws & Regulations Implemented

Currently, there are no specific laws, statutory rules, or regulations in Singapore that directly regulate AI.
Other laws **indirectly** affecting AI:
- The Road Traffic Act 1961, which was amended in 2017 to allow for the testing and use of autonomous motor vehicles
- The Health Products Act 2007, which requires medical devices that incorporate AI technology to be registered before they are used

C. Policy Initiatives

National AI Strategy 2.0 (NAIS 2.0) (2023) (update of the 2019 National AI Strategy)

Key shifts :

1. **From Opportunity to Necessity**: AI is now essential for national prosperity and relevance.
2. **From Local to Global:** Emphasis on a global outlook, connecting with international networks and pooling resources.
3. **From Projects to Systems:** Focus on enhancing infrastructure and resources, and accelerating idea exchanges to impact various sectors broadly.

Twin Goals :

1. Excellence: Develop peaks of excellence in AI, addressing global challenges like population health and climate change.
2. Empowerment: Enable individuals, businesses, and communities to use AI confidently and effectively, promoting AI as an equalizer.
To achieve the vision and goals of NAIS 2.0, Singapore will focus on **three systems** supported by **ten enablers**, implementing **fifteen actions** (available here) over the **next 3-5 years**.
System 1 - activity drivers (enablers: industry, government, research)
System 2 - people & communities (enablers : Talent, capabilities, placemaking)
System 3 - Infrastructure & environment (enablers : Compute, Data, Trusted Environment, Leader in Thought and Action)

Soft law strategy

Regulatory agencies issue **non binding guidelines and recommendations** rather than enforceable regulations.
1. The Personal Data Protection Commission of Singapore issued the **Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems**(here) in 2024 to provide organizations with certainty on when they can use personal data to develop and deploy systems that embed machine-learning models
2. The Monetary Authority of Singapore issued the **Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT)** in the **Use of Artificial Intelligence and Data**

Analytics in Singapore's Financial Sector (here)in 2018 (updated in 2019) to provide a set of foundational principles for firms to consider when using AI in decision-making in the provision of financial products and services
3. The Ministry of Health, Health Sciences Authority and Integrated Health Information Systems jointly issued the **Artificial Intelligence in Healthcare Guidelines**(here) in 2021 to improve the understanding, codify good practice and support the safe growth of AI in healthcare Model AI Governance Framework for Generative AI (30 May 2024) (GenAI Framework)
The GenAI Framework is based on insights from key jurisdictions, international organizations, and leading AI entities. It outlines nine dimensions to foster a trusted AI ecosystem:

1. Accountability: Establishing incentives for AI system developers to be responsible to end-users.
2. Data: Ensuring data quality and pragmatically addressing contentious training data.
3. Trusted Development and Deployment: Enhancing transparency around safety and industry best practices in AI development, evaluation, and disclosure.
4. Incident Reporting: Implementing a system for timely notification, remediation, and continuous improvement.
5. Testing and Assurance: Providing external validation through third-party testing and developing common AI testing standards.
6. Security: Addressing new threat vectors arising from generative AI models.
7. Content Provenance: Ensuring transparency about the origins of AI-generated content.
8. Safety and Alignment R&D: Accelerating research and development through global cooperation to align AI models with human intentions and values.
9. Evolution from Traditional AI Framework: Promoting responsible AI use to benefit the public, including democratizing access, improving public sector adoption, upskilling workers, and developing sustainable AI systems.

Fields of action :
Grand challenges :
- Supporting interdisciplinary research teams in solving AI-specific challenges in the areas of health, digital city and finance

100 Experiments :
- Financial, personnel and organizational support for industry in AI-specific problems as well as in the assembly of expert teams

AI Apprenticeship Program
- Promoting local talent through courses, training in real industrial applications and AIrelated scholarships

AI for industry :
- Teaching practical AI fundamentals in the Python programming language through face-toface workshops and online training for engineers, software developers and managers

AI for everyone :
- Training and teaching of basic AI methods (e.g. machine learning, deep learning) and their use in companies

Focus on further initiatives :
- Support for small and medium-sized enterprises in the use of digital technologies
- Facilitating partnerships between local companies and AI application manufacturers

D. Consultations in Process

Currently, no public consultation on the government's artificial AI strategy is active.

The most recent public consultation on AI in Singapore took place from January 16 to March 15, 2024.
- Conducted by the Infocomm Media Development Authority (IMDA) and the AI Verify Foundation, it focused on a draft Model AI Governance Framework for Generative AI.
- This framework addresses challenges such as misinformation and ethical concerns while promoting innovation.
- It includes nine key dimensions: accountability, data quality, transparency, security, and public good, aiming to create a comprehensive and trusted AI ecosystem.

E. Proposed Legislation

At the moment, Singapore intends to rely on existing laws, such as data protection, copyright and other sectoral legislation, to regulate AI at this time.
The soft law guidelines introduced intend to complement existing laws and to lay the groundwork for the possible enactment of future general AI regulations.

F. Commentary

- **The strength of Singapore's approach to AI lies in its adaptability. Guidelines can be amended (or new guidelines issued) quickly to adapt to any changes.** Guidelines are developed in close consultation with industry stakeholders, incorporating feedback to ensure relevance and effectiveness. Technology and its use cases are studied thoroughly before making permanent legislative changes, allowing for a more informed and adaptable approach.
- There are concerns about the quality and quantity of AI talent in Singapore. Despite initiatives to train 25,000 professionals in basic AI skills and broader AI literacy efforts, critics argue that putting individuals through training does not guarantee deeper capabilities. More emphasis is needed on solving common business problems and closing the gap between available AI solutions and their adoption (Raffles Institution Singapore).
- Singapore's strategy involves significant collaboration with international AI experts and reliance on foreign AI platforms. **There is a push for more self-reliance by developing local AI talent and infrastructure, but the current dependency poses a challenge to achieving true autonomy in AI capabilities** (Raffles Institution Singapore).
- **AI Singapore launched SEA-LION, a family of LLMs trained on data sets in 11 regional languages to better cater to Southeast Asian contexts. The launch of SEALION thus heralds a potentially significant shift in the way AI is developed and deployed in the Asian context.** Similar developments, for instance in Sweden, strongly suggest that LLMs should be trained for and tuned to local contexts in order to be most effective.

4. France

A. Policy Summary

Political Goals :
- Open data policy for the implementation of AI applications
- Focus on four sectors: health, environment, mobility, security/defense
- European cooperation in the field of AI

No specific laws directly regulating AI in France; the EU AI Act will apply. Legislative proposal (September 2023) to amend copyright provisions in the French Intellectual Property Code (IPC) to address AI.

National AI Strategy - divided into 2 phases (2018-2022 and 2021-2025).
CNIL AI Action Plan (2024)

Fields of action :
- AI ecosystem for France and Europe
- Legal and ethical issues
- Open data

CNIL public consultation (July 2 - September 1, 2024) on AI system development.

B. Laws & Regulations Implemented

Currently, **there are no specific laws, statutory rules, or regulations in France that directly regulate AI.** France is not expected to enact its own comprehensive AI regulation. The EU AI Act is superceding legislation in France.

Other laws **indirectly** affecting AI: :
- The Law of a Digital Republic
- French Antitrust and competition law, especially the Commercial Code
- The Public Health Code
- Laws relating to data protection and the GDPR
- Civil and product liability laws
- Security and cybersecurity legislation
- Intellectual property laws

C. Policy Initiatives

The National AI Strategy (available here), launched in 2018, is divided into 2 main phases :
1. First phase (2018-2022) :
- Objective : To provide France with competitive research capacities
- Budget : 1.5 billions euros.
- 45% of the dedicated budget is devoted to the National AI Research Program (PNRIA)
2. Second phase (2021-2025):

- Objective : To integrate AI within the economy and support development in priority areas
- **Trusted Artificial Intelligence Demonstrators** (DIAC) (operated by Bpifrance). Aims to support the development of hardware, software, and system innovations aimed at maturing and demonstrating critical functional systems integrating trusted artificial intelligence (in fields such as safety and security, robustness…)
- **Demonstrators of Artificial Intelligence in Territories (DIAT)** (operated by the Banque des Territoires). Aims to support technology demonstrator projects based on data science and artificial intelligence responding to territorial challenges
- **Relaunch of the Technological Maturation and Demonstration of Embedded Artificial Intelligence Technologies** (operated by Bpifrance). Aims to develop and test embedded AI solutions under real-life conditions. It also supports the development of advanced hardware architectures for deploying algorithms on embedded systems
- **Digital Commons for Generative AI**. Aims to build and make Digital Commons available for Generative AI.
- **IA Booster France 2030**(available here). Aims to support French SMEs and intermediate sized enterprises in all sectors. Priority is given to companies with between ten and 2,000 employees and sales in excess of 250,000 euros
- **IA-cluster** (available here) - Aims to establish 5-10 universities/schools as global AI leaders, aiming to position at least three French institutions in the world's Top-50 AI institutions.

In May 2024, **the CNIL published its AI action plan** (available here). It is structured around 4 objectives :

1. Understanding the functioning of AI systems and their impacts for people
- Addressing issues of data processing transparency and fairness.
- Protecting publicly available data from scraping and ensuring secure handling of user data.
- Ensuring individuals' data rights are maintained, including data collected for model learning and data produced by AI systems.
- Protecting against biases and discrimination inherent in AI systems.
- Addressing unprecedented security challenges posed by AI tools.

2. Enabling and guiding the development of AI that respects personal data
- Providing guidance on GDPR application to AI, including training of generative AI.
- Publishing fact sheets and guides to support compliance with data protection laws.
- Engaging in consultations and publishing guidelines on data sharing, re-use, and compliance with data protection principles.

3. Federating and supporting innovative players in the AI ecosystem in France and Europe
- Launching sandboxes to support innovative AI projects, especially in health, education, and public sector applications.
- Offering enhanced support programs to assist AI companies in complying with GDPR.
- Encouraging dialogue with research teams, R&D centers, and AI developers to ensure compliance with personal data protection rules.

4. Audit and control AI systems and protect people
- Developing tools to audit AI systems, ensuring they respect individual rights and freedoms.
- Monitoring the use of enhanced video surveillance, AI in fraud detection, and investigating complaints related to AI systems.

- Ensuring that actors have conducted DIAs to document risks and have measures in place to mitigate them.

Other fields of action :

AI ecosystem for France and Europe

- Establishment of a national coordination office led by INRIA to network French AI expertise
- Establishment of four to five AI alliances with partners from science and industry (e.g. PRAIRIE)
- Expanding AI research and recruiting international researchers

Legal and ethical issues :

- Adaptation of the legal framework to AI development (e.g. for autonomous driving by 2022; regional experiments with exemptions in advance)
- Dialogue on ethical issues at European and international level (especially Canada) with the aim of creating an organization modeled on the IPCC (Intergovernmental Panel on Climate Change) for AI
- Transparency of algorithms
- Ethical Code for Programmers

Open Data :

- Provision of public data for publicly funded projects
- Establishment of common data platforms for the public and private sectors
- Opening the database at European level
- Public discourse on data handling

D. Consultations in Process

CNIL opened a new public consultation on the development of AI systems.

- **From the 2 July 2024- 1st September 2024**
-

What is the consultation about?

1. A new series of how- to sheets subjected to public consultation :

- Legal basis for legitimate interest and development of AI systems
- Legitimate interest: focus on open-sourcing models
- Legitimate interest: focus on webscraping
- Informing data subjects
- Respecting and facilitating the exercise of data subjects' rights
- Annotating data
- Ensuring the safe development of an AI system

2. A questionnaire on the application of the GDPR to AI models

- The CNIL invites providers and users of AI systems, as well as all relevant actors, to shed light on the conditions under which AI models can be considered anonymous or must be regulated by the GDPR and on the consequences of such a qualification.
- With this questionnaire, the CNIL is consulting all stakeholders in the AI sector in order to adapt its future recommendations to the real risks for the people concerned and the sector's capacity to reduce them.

E. Proposed Legislation

Legislative proposal in Sept 2023 to amend the copyright provisions of the French Intellectual Property Code (IPC) to properly account for AI.
- The IPC's legislative proposal is available here.
- Intended to encourage AI systems to respect copyright by establishing a framework to protect the rights of artists and authors.
- The IPC amendment applies to the exploitation of AI-generated works.

Under the proposed IPC amendment, developers, deployers and users of AI systems will have to ensure that they:
- Obtain authorization to use the right-holders work to develop the AI-generated content
- Assign ownership of any fully AI-generated work to the authors/right holders
- Comply with any transparency obligations requiring AI-generated work. In accordance with the proposed IPC amendment, the exploitation of AI-generated works should be strictly controlled
The IPC's legislative proposal is available here

F. Commentary

- Noah Greene from the AI Safety and Stability Project at Washington-based think tank Center for a New American Security (CNAS) says **the French government "flipped a switch" when it decided to become an AI champion. But making that ambition a reality might be an uphill struggle, he told DW**. The US leads the AI market, followed by China and the UK, with France and Germany lagging due to non-technological factors like **complex labor laws**. Large US tech firms have struggled with France's labor code.
- Christine Dugoin highlights the **need for more and bigger supercomputers in Europe to compete in AI**. She advocates for a Europe-wide AI approach to tackle global competition and AI-based disinformation, particularly from Russia.
- Philippe Aghion suggests that AI could boost French GDP by 0.8% annually over the next decade. **Calls for substantial state investment in AI, proposing at least €25 billion to realize its potential**.

5. United States of America (USA)

A. Policy Summary

Political Goals :
- Use of AI based on democratic values and respecting the rights and security of individuals
- Developing and using AI systems to promote public welfare
- Focus on international cooperation

Legislation: There is currently no comprehensive federal legislation or regulation in the United States that governs the development or use of AI. However, numerous state legislatures have introduced significant bills aimed at regulating AI. Several federal proposed laws related to AI include: SAFE Innovation AI Framework; REAL Political Advertisements Act; Stop Spying Bosses Act; Draft No FAKES Act; AI Research Innovation and Accountability Act

Policy Initiatives: The White House Executive Order on AI : 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' (2023); The White House Blueprint for an AI Bill of Rights and other specific fields of action.

Several active public consultations launched by : The U.S. Department of the Treasury, U.S. Department of Transportation's ARPA-I, U.S. Patent and Trademark Office (USPTO) and Department of Commerce (DOC)

Criticisms include: Lack of robust AI regulations compared to AI, regulatory capture by Big Tech, absence of a comprehensive federal data privacy framework

B. Laws & Regulations Implemented

Currently, there is no comprehensive federal legislation or regulations in the US that regulate the development of AI or specifically prohibit or restrict their use.

However, there are existing federal laws that concern AI, albeit with limited application. For example :
- Federal Aviation Administration Reauthorization Act, which includes language requiring review of AI in aviation
- National Defense Authorization Act for Fiscal Year 2019, which directed the Department of Defense to undertake various AI-related activities, including appointing a coordinator to oversee AI activities.
- National AI Initiative Act of 2020, which focused on expanding AI research and development and created the National Artificial Intelligence Initiative Office that is responsible for "overseeing and implementing the US national AI strategy."

There is also a material 'soft law' included in the NIST Risk Management Framework for AI. It "can help organizations identify unique risks posed by generative AI and proposes actions for generative AI risk management".

States legislatures have introduced a substantial number of bills aimed at regulating AI, notably:

- On May 17, 2024, Colorado enacted the first comprehensive US AI legislation, the Colorado AI Act. This legislation imposes duties on AI developers and deployers, particularly focusing on automated decision-making systems. It defines a high-risk AI system as one that significantly contributes to making consequential decisions. The Act emphasizes preventing bias and discrimination, requiring developers and deployers to exercise reasonable care to avoid such issues in AI systems involved in critical decisionmaking processes. The Colorado AI Act will come into effect in 2026.

- The California Consumer Privacy Act (here),which contains provisions on the use of automated decision-making tools.

Additionally, the California Privacy Protection Agency has released draft rules on these provisions (here), addressing consumer notice, access, and opt-out rights concerning these technologies.

These draft rules define automated decision-making technology broadly and require significant disclosures about businesses' implementation and use of such tools. These rules are expected to be formalized sometime in 2024.

More than 40 state AI bills were introduced in 2023, with Connecticut(here) and Texas (here) adopting statutes. Both of those enacted statutes establish state working groups to assess state agencies' use of AI systems to ensure they do not result in unlawful discrimination
Other States legislation enacted and proposed here -> AVAILABLE HERE

C. Policy Initiatives

The White House Executive Order on AI : 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence' (2023) (available here)

By requiring federal contractors to comply with agency requirements, the US Government is inducing broader-based compliance with a set of AI policies.

It is premised on the understanding that "[h]arnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks." The executive order focuses on federal agencies and developers of foundation models, mandates the development of federal standards, and requires developers of the most powerful AI systems to share safety tests results and other critical information with the U.S. government. The Executive Order also calls on the Department of Commerce to issue guidance for content authentication and watermarking to label AI-generated content.

8 key principles and priorities :

1. AI Safety and Security
- Develop standardized evaluations and policies to test and mitigate AI risks.
- Address security risks in critical areas like biotechnology and cybersecurity.
- Implement effective labeling for AI-generated content.

2. Promoting Innovation and competition
- Invest in AI education, training, and research.
- Encourage a competitive AI ecosystem and protect small developers.

3. Supporting American Workers
- Adapt job training for new opportunities created by AI.
- Ensure AI does not undermine worker rights or job quality.

4. Advancing Equity and Civil rights
- Prevent AI from exacerbating discrimination and bias.
- Ensure AI complies with federal laws promoting equity and justice.

5. Consumer Protection: Uphold consumer protections in AI applications, especially in critical fields like healthcare and finance.

6. Privacy and Civil Liberties: Protect personal data privacy and prevent AI misuse of sensitive information.

7. Federal Government Use
- increase federal capacity to regulate and use AI responsibly.
- Attract and retain AI professionals in the public sector.

8. Global Leadership: Promote responsible AI principles globally and collaborate with international partners.

The White House Blueprint for an AI Bill of Rights (available here).

The AI Bill of Rights provides five principles and associated practices to help guide the design, use and deployment of "automated systems". The associated "From Principles to Practice" handbook provides detailed steps for incorporating these protections into technological processes. Aim: To harness AI's benefits while mitigating its risks, ensuring that advancements in technology do not come at the expense of civil rights or democratic values.

5 core principles :

1. Safe and effective systems

- Automated systems must be developed with input from diverse stakeholders to identify and mitigate risks.
- Systems should undergo rigorous pre-deployment testing, continuous monitoring, and independent evaluations to ensure they are safe and effective.
- Proactive measures should be in place to protect against foreseeable harms, including from unintended uses.

2. Algorithmic Discrimination Protections

- AI systems must be designed to avoid discrimination and ensure equitable treatment across all protected categories (e.g., race, gender, religion).
- Continuous assessments and audits should be conducted to identify and mitigate biases.
- Systems should be transparent, with public reporting on measures taken to prevent algorithmic discrimination.

3. Data Privacy

- Individuals should have control over how their data is collected, used, and shared.
- Data practices should include built-in privacy protections, with the minimum necessary data collected.
- Consent should be meaningful, understandable, and respect user agency.
- Enhanced protections should be in place for sensitive data, and surveillance technologies should be subject to strict oversight.

4. Notice and Explanation

- Users should be informed when automated systems are in use and understand how these systems impact them.
- Clear, accessible explanations should be provided about the functioning of AI systems and the rationale behind their decisions.
- Continuous updates and public reporting should be maintained to ensure transparency.

5. Human alternatives, consideration and fallback :

- Users should have the option to opt-out of automated systems in favor of human alternatives where appropriate.
- There should be accessible and effective mechanisms for human review and intervention in case of system failures or disputes.

Particularly in sensitive areas like criminal justice, employment, and healthcare, human oversight should be integrated to ensure fairness and accountability.

FIELDS OF ACTION

Long-term investments in fundamental and responsible AI research
- Promoting data-oriented methods and federated machine learning approaches
- Development of AI systems and simulations in real and virtual environments and improvement of the perception capabilities of AI systems
- Understanding the theoretical capabilities and limitations of AI

Measuring and evaluating AI systems
- Increasing the availability of AI test environments
- Involvement of the AI community in standards and benchmarks

Method development for collaboration between humans and AI
- Search for improved models and metrics
- Building trust and greater understanding in human-AI interaction

Better understanding of national AI skills needs
- Evaluating the need for AI professionals
- Training/retraining the workforce and developing AI expertise, taking into account ethical, legal and societal implications
- Exploring the impact of diverse and multidisciplinary expertise

Considering the ethical, legal and social impacts of AI
- Investing in basic research to promote fundamental values
- Understanding and mitigating social and ethical risks of AI and using AI to solve ethical, legal and societal problems

Public datasets and environments for AI training and testing
- Development of shared advanced computing and hardware resources and open source software libraries and toolkits

Expansion of public-private partnerships
- Involvement of more diverse interest groups
- Improving, expanding and creating mechanisms for R&D partnerships

Creating a principled and coordinated approach to international cooperation in AI research
- International cooperation for global challenges (e.g. sustainability, healthcare, manufacturing)
- Development of international standards and guidelines for trustworthy AI

Ensuring the security of AI systems
- Developing secure AI and securing AII

D. Consultations in Process

1. The U.S. Department of the Treasury is seeking comment on the use of artificial intelligence in the financial services sector. The purpose of this request for information is to understand the uses and opportunities of artificial intelligence in financial services, as well as risk management practices and impacts on consumers and other end users.
- This consultation seeks public input on how AI can be leveraged to enhance financial services while addressing potential risks such as bias, security concerns, and regulatory compliance.
6th June 2024 - 11th August 2024

2. The U.S. Department of Transportation's Advanced Research Projects Agency - Infrastructure (ARPA-I) is seeking input on the potential applications of AI in transportation,

as well as emerging challenges and opportunities in creating and deploying AI technologies across all modes of transportation.
3. The U.S. Patent and Trademark Office (USPTO) is seeking public feedback on how the proliferation of AI could affect USPTO evaluations on patentability, including what qualifies as prior art and the assessment of the level of ordinary skills in the art. Comments are due by July 29, 2024.
4. The Department of Commerce (DOC) is seeking input on AI and Open Government Data Assets. DOC is interested in how generative AI systems can utilize the department's public data assets, with the goal of democratizing access to its public data. This RFI seeks valuable insights on the development of open data assets for AI and modernized data dissemination standards. Comments are due by July 16, 2024.

E. Proposed Legislation

On September 12, 2023, the US Senate held public hearings regarding AI (see here),which laid out **potential forthcoming AI regulations**. Possible legislation could include requiring licensing and creating a new federal regulatory agency. Additionally, US lawmakers held closed-door listening sessions with AI developers, technology leaders and civil society groups on September 13, 2023 in a continued push to understand and address AI(see here).
There are several proposed federal laws related to AI. A non-exhaustive list of key examples includes:
- The SAFE Innovation AI Framework (here) which is a bipartisan set of guidelines for AI developers, companies and policymakers. This is not a law, but rather a set of principles to encourage federal law-making on AI
- The REAL Political Advertisements Act (here)which aims to regulate generative AI in political advertisements
- The Stop Spying Bosses Act (here)which aims to regulate employers surveilling employees with machine learning and AI techniques
- The Draft No FAKES Act (here and here) which would protect voice and visual likenesses of individuals from unauthorized recreations from Generative AI
- The AI Research Innovation and Accountability Act (here)which calls for greater transparency, accountability and security in AI, while establishing a framework for AI innovation. It would create an enforceable testing and evaluation standard for high-risk AI systems and require companies that use high-risk AI systems to produce transparency reports. It also empowers the National Institute of Standards and Technology to issue sector-specific recommendations to regulate them

F. Commentary

- Critics argue that the current AI framework, especially the one proposed by Senate Majority Leader Chuck Schumer, seems to favor large tech companies. Evan Greer, director of Fight for the Future, commented that the framework "reads like it was written by Sam Altman and Big Tech lobbyists." She highlighted that the roadmap emphasizes funding AI research and development but lacks substantial measures to address urgent

issues such as AI's impact on policing, immigration, and workers' rights (Enterprise Technology News and Analysis).
- There is a significant concern that the U.S. is lagging in establishing robust AI regulations compared to other regions like the EU. The EU's AI Act, which demands transparency and accountability from companies developing high-risk AI systems, is seen as a model that the U.S. could follow. The U.S., however, is perceived to be taking a more fragmented and less stringent approach, potentially stifling its ability to lead in AI governance globally
(MIT Technology Review).
- The absence of a comprehensive federal data privacy framework in the U.S. is another point of criticism. Despite ongoing discussions, the U.S. has not made significant progress in establishing laws that would protect citizens' data from misuse, which is crucial for responsible AI deployment. This gap in legislation could undermine efforts to build public trust in AI technologies (Enterprise Technology News and Analysis).
- The Department of Defense's AI strategy, focusing on enhancing decision-making capabilities for military operations, emphasizes speed and adaptability. While this aims to maintain U.S. military superiority, there are concerns about the ethical implications and potential misuse of AI in warfare. Critics argue for more stringent oversight and international cooperation to ensure responsible use of AI in military contexts (Defense).

6. Germany

A. Policy Summary

Germany currently lacks specific AI regulations, with only a minor reference to AI in labor law. Instead, the country relies on the EU AI Act. The German government is actively assessing the necessity for additional AI-specific legislation to address emerging challenges and ensure robust governance, and has made a significant commitment to AI as a national priority.

Policy Initiatives : National AI Strategy (updated in 2020), 2023 AI Action Plan and other specific fields of action.

Individual states within Germany have also taken action with respect to AI. For example, Bavaria has established the “first AI university in the world” in Nuremberg, which (together with a proposed German LLM project) represent as much as a €4 billion commitment, separate and aside from the federal commitment of €5 billion).

No new public consultations on AI as of 2024.

Commentary and criticism :
- Strict regulations like GDPR challenge data acquisition for AI. Solutions include promoting data partnerships and synthetic data production
- Germany struggles to attract and retain AI talent due to stringent labor laws and insufficient incentives. More investment in education and training is needed
- The complex regulatory framework may hinder innovation. Agile and flexible regulations are necessary to support SMEs and adapt to advancements
- Germany’s AI investment is lower compared to the US and China. Increased funding and robust infrastructure are crucial for competitiveness

B. Laws & Regulations Implemented

With a legislative amendment in mid-2021, a reference to AI was included in **three provisions of the German Works Constitution Act** (Betriebsverfassungsgesetz; see Section 80(3), Section 90(1) No. 3 and Section 95(2a) Works Constitution Act).
- The provisions include a right to information for works councils if AI is to be used in the workplace and facilitate the consultation of an expert by works councils related to the introduction or use of AI.
- These amendments are narrow in scope and do not significantly alter German labor law.

Albeit this minor reference to AI in labor law, **Germany has no specific laws regulating AI. Germany will rely on the upcoming EU AI Act.**

Other laws indirectly affecting AI:

- The German civil code (BGB)
- The Product Liability Act (ProdHaftG)
- The Copyright Act (UrhG), in particular Sections 44b and 60d concerning text and data mining
Intellectual property laws that may affect several aspects of AI development and use

C. Policy Initiatives

National AI Strategy (launched in 2018 and updated in 2020).
- https://www.ki-strategie-deutschland.de/
- updated AI strategy here

Political Goals:

1. Technology leadership and quality seal “AI Made in Germany”
2. Responsible and public welfare-oriented development and use of AI
3. Development of AI solutions as a contribution to the environment and climate protection
4. Broad social dialogue
5. Building a European AI ecosystem that increases the competitiveness of industry and research, promotes diverse AI applications in the interest of society and is based on European values

The Federal Government will provide around **€5 billion for this purpose by 2025**

National AI Strategy :

12 Fields of Action :

1. Strengthen research in Germany and Europe to drive innovation
2. Innovation competition and european innovation clusters : Foster disruptive ideas, new solutions, business models, start-ups, talent, trends, and multidisciplinary insights.
3. Transfer to the economy, strengthen small and medium-sized business (Better access to AI technologies for companies, especially for medium-sized companies via the Mittelstand 4.0 competence centers)

4. Awaken start-up dynamics and lead to success : through funding programs (e.g. EXIST) and venture capital
5. Working world and labour market: shaping structural shape. Support employees with a national training strategy during AI-driven changes.
6. Strengthen training and recruit skilled workers/ experts. Broaden AI understanding, especially among youth, and improve conditions to attract and retain scientists and establish new professorships.
7. Use of AI for sovereign tasks and adapt the powers of the administration
8. Make data available and facilitate use - Implement measures to significantly increase the availability of high-quality data, establishing Germany as a leading AI hub while protecting personal rights and informational self-determination.
9. Adapt the regulatory framework - Adjust legal frameworks, if necessary, to protect against AI-related abuses (distortion, discrimination, manipulation).
10. Set standards : The federal government will collaborate with business representatives to promote AI standards and norms at national, European, and international levels through DIN/DKE.
11. National and international networking - Expand international, bilateral, and multilateral cooperation in the field of AI.
12. Conduct dialogues in society and further develop the political framework for action. The federal government will intensify social dialogue and education on AI, involving participatory processes to create an informed society. Transparent discussions on AI's opportunities, risks, and challenges will be promoted to elevate research, development, and application of AI in Germany to a world-leading level.

2023 AI ACTION PLAN : available here

Key initiatives and projects :
- **AI Competence Centers:** Establishing leading research institutions.
- **High-Performance Computing:** Developing infrastructure for top-tier computational resources.
- **AI Talent Development:** Training and retaining top AI talent.
- **Health AI Projects:** Advancing AI applications in healthcare for better patient outcomes.
- **Sustainability and AI:** Using AI to support environmental sustainability and resilience.
- **Educational AI Projects:** Enhancing AI integration in education systems.

Fields of action :

Expansion of AI research :
- Establishment of a national network of at least twelve centres and application hubs
- Establishment of a world-leading European AI network under the umbrella brand “AI – Made in Europe”
- Establishment of at least 100 additional AI professorships and strengthening of teaching and promotion of young talent in the field of AI
- Retaining and attracting the best minds through attractive working conditions and remuneration
- Establishment of a German-French research and innovation network (“virtual centre”)
- Strengthening interdisciplinary research on AI

- Implementation of AI challenges and establishment of a German award for “AI Made in Germany”

Knowledge transfer, application and entrepreneurship

- Faster transfer of research into concrete AI applications through test fields, real-world laboratories, model experiments, regional clusters and innovative funding formats
- Better access to AI technologies for companies, especially for medium-sized companies via the Mittelstand 4.0 competence centers
- Promoting the start-up dynamics for AI start-ups through funding programs (e.g. EXIST) and venture capital
- Creating an AI map of applications and actors; AI monitoring and networking of companies and institutions
- Establishment of an agency for breakthrough innovations with AI as one of the main focuses
- Creation of a European innovation cluster on AI and implementation of innovation competitions
- Public presentation of best practices (especially with the help of the Learning Systems platform)

Change in work :

- Holistic and humane approach based on the self-determined development of skills and talents, social security and the health of employees
- Establishment of regional competence centers for work research and design
- Skilled worker monitoring and national training strategy to promote the skills of employed persons, particularly with regard to digital change and AI
- Investigation of the impact of AI in the workplace in company experimentation spaces and early involvement of works councils in the introduction of AI applications
- Qualification of human resources managers, personnel and works councils (e.g. in future centres)
- Development of an AI-supported online entry portal for professional training

Data use, data security, law and ethics

- Promoting research into the control and traceability of algorithmic forecasting and decision-making systems as well as consumer protection and privacy
- Making data available, e.g. through potential data partnerships between companies and research institutions and by setting up incentives and framework conditions for the voluntary and data protection-compliant sharing of data (including from publicly funded research projects)
- Adaptation of competition and copyright law to increase the amount of usable data without disclosing personal data or operational know-how (Commission Competition Law 4.0)
- Adaptation of labour law and employee data protection law
- Creating a legally secure regulatory framework for AI actors

International and social dialogues

- European and transatlantic dialogue on the human-centered use of AI in the world of work
- Broad social dialogue on the ethical, legal, cultural and institutional design of AI – the Platform for Artificial Intelligence will play a key role here

- Development of an ecosystem for common good-oriented AI under the label " Civic Coding – Innovation Network AI for the Common Good", including implementation of the projects *Civic Innovation Platform , Civic Data Lab and AI Ideas Workshop for Environmental Protection*

D. Consultations in Process

Germany is currently not conducting any new public consultations specifically focused on artificial intelligence as of 2024.

E. Proposed Legislation

According to several official statements, **the German government continues to evaluate the need for additional AI-specific national legislation**

- For example, a statement on the National AI Strategy website: "The Federal Government will review the legal framework for algorithm- and AI-based decisions, services and products and **possibly adapt it to ensure that effective protection against bias, discrimination, manipulation or other misuse is possible."** (available here)

- See also the National AI Strategy (2018) itself, page 9, II, j.: ""We want to [...] **examine whether the regulatory framework needs to be further developed to ensure a high degree of legal certainty [...]"** (available here)

F. Commentary

- Germany's strict data protection laws, including GDPR, pose a challenge in acquiring the large datasets needed for AI development. The government is working on solutions like promoting data partnerships and synthetic data production to mitigate this issue. However, the balance between protecting privacy and facilitating data access remains a delicate one (IIoT World).

- A notable critique is the shortage of AI talent in Germany. While the country is known for its strong research sector, it struggles to retain and attract top AI professionals due to stringent labor laws and lack of sufficient incentives. The German government has been urged to invest more in education and training programs to build a larger AI-skilled workforce (FAU Erlangen-Nürnberg) (POLITICO).

- There is concern over Germany's complex regulatory environment, which may stifle innovation. The need for more agile and flexible regulations is highlighted, especially to allow for faster adaptation to technological advancements and to support SMEs, which make up a large portion of the German economy (reframe[Tech]) (IIoT World).

- Critics point out that Germany's investment in AI, though significant, lags behind the US and China. Germany plans to spend €3 billion by 2025, which is relatively small compared to the billions spent annually by these leading nations. Experts emphasize that for Germany to remain competitive, it must significantly increase funding and create more robust infrastructure for AI research and development (POLITICO) (IIoT World)

7. Switzerland

A. Policy Summary

There are no specific laws, statutory rules, or regulations directly governing AI in Switzerland. Regulatory Exploration (2023): **DETEC to explore AI regulatory approaches by end of 2024**, aligning with EU AI Act and Council of Europe's AI Convention. Switzerland has made a significant commitment to AI relative to its GDP, seeking to leverage its research institutions and noting the impact of AI on financial services, a major sector of the Swiss economy.

- **Policy Initiatives :** Guidelines on Artificial Intelligence for the Confederation (2020); Digital Switzerland Strategy 2023; Recommendations from Swiss Federal Data Protection and Information Commissioner; FINMA Risk Monitor 2023
- **Consultation in process** - Action Plan Proposals: Inviting proposals to implement the Digital Switzerland Strategy, focusing on AI.

Commentary, recommendations:

- Investment in skills training and continuing education is critical as GenAI shifts the workplace.
- Professionals are uninformed about GenAI policies; immediate action is needed to inform and protect citizens and support innovation.
- Authorities should reduce disinformation to bridge the "trust gap" and promote AI benefits.
- Address data protection standards to support GenAI learning while ensuring privacy.
- Enhance data sharing incentives and align with EU AI Act to mitigate reliance on foreign platforms.

B. Laws & Regulations Implemented

Currently, **there are no specific laws, statutory rules or regulations in Switzerland that directly regulate AI.**

Indirect laws affecting AI:

- The revised Federal Act on Data Protection (FADP) entered into force on September 1, 2023 and includes provisions on automated decision-making in relation to personal data
- Intellectual property laws may affect several aspects of AI development and use (particularly the Copyright Act and Patents Act)
- Civil law (such as the Swiss Civil Code, Code of Obligations or the Product Liability Act)
- Product safety laws, both general (such as the Product Safety Act) and sectoral (such as the Therapeutic Products Act10 and Medical Devices Ordinance)
- Non-discrimination laws in the areas of gender equality and protection of disabled people (Gender Equality Act and Disability Discrimination Act)
- Swiss Criminal Code
- General human rights legislation (such as the Federal Constitution and Convention for the Protection of Human Rights and Fundamental Freedoms)

C. Policy Initiatives

Guidelines on Artificial Intelligence for the Confederation (2020)

The guidelines must be adhered to: When developing sectoral AI strategies; When introducing or adapting specific, sectoral regulations; When developing and using AI systems within the Federal Administration; When helping to shape the international regulatory framework on AI.

1. Putting People First
- Core Principle: Prioritize human dignity, well-being, and the common good in AI development and use.
- Preserve self-determination and ensure AI improves quality of life.
- Ensure AI supports equal opportunities and access to education, goods, services, and technology.
- Protect fundamental rights through "ethics by design" and impact assessments for AI applications.
- Ensure AI used by the Confederation respects privacy and complies with data protection laws.

2. Regulatory conditions for the development and application of AI
- Core Principle: Create favorable regulatory conditions for AI to exploit opportunities.
- Maintain Switzerland's leadership in AI research and development.
- Balance regulation to support innovation and legal certainty.
- Encourage education, research, and innovation to strengthen AI competencies.
- Promote economic growth, security, and sustainability through AI.

3. Transparency, traceability and explainability
- Core Principle: Ensure AI is transparent, traceable, and explainable to foster trust.
- Clearly identify AI-based decision-making processes.
- Disclose AI functioning and purposes to comply with legal standards.
- Ensure data quality and documentation for AI training.
- Balance privacy protection with data usage.

4. Accountability
- Core Principle: Clearly define liability and responsibility in AI usage.
- Ensure responsibilities are clarified to address damages or violations.
- Prevent the delegation of responsibility to machines.

5. Safety
- Core Principle: Design AI systems to be safe, robust, and resilient.
- Prevent misuse and misapplication of AI.
- Implement safeguards against serious misuse.
- Monitor AI impacts on individuals, society, economy, and environment.

6. Actively shape AI governance
- Core Principle: Engage in global AI governance to influence standards and norms.
- Participate in international organizations and processes (e.g., UN, OECD).
- Align AI governance with human rights and responsible corporate governance.
- Monitor EU and NATO developments in AI.

7. Involve all relevant national and international stakeholders
- Core Principle: Include diverse stakeholders in AI governance.
- Engage national governments, private sector, civil society, and technical experts.
- Ensure accountability for AI usage.
- Promote networking and cooperation, strengthening Geneva as a digital AI governance hub.

Digital Switzerland Strategy 2023

- Swiss's approach to regulating AI is a focus theme, along with cybersecurity and Application programming interfaces (APIs)
AIM- The Digital Switzerland Strategy sets the guidelines for Switzerland's digital transformation. It is binding for the Federal Administration and serves as an orientation for all other actors involved in digitisation. The aim is for the population as a whole to benefit from a sustainable and responsible digital transformation.

The Digital Switzerland Strategy 2023 is structured around five long term domains
1. Education and skills - People, businesses and public authorities have sufficient skills to make the most of new technologies and are able to question them.
2. Security and trust - People in Switzerland can move around safely in the digital environment; privacy is protected.
3. Framework -Businesses and society can rely on a reliable and advantageous framework for the digital environment.
4. Infrastructure - Public authorities promote and operate reliable and resilient physical as well as digital infrastructure.
5. Digital Public services - Public authorities offer their services digitally as standard (digital first)

Other soft law guidelines :
- Recommendations from the Swiss Federal Data Protection and Information Commissioner about data processing in relation to AI (2023)(See here)
- Expectations from the Swiss Financial Market Supervisory Authority (FINMA) for the use of AI by regulated institutions (FINMA Risk Monitor 2023) (See here)

D. Consultations in Process

The Federal Chancellery is currently inviting organizations, companies, communes, and cantons to submit proposals for the action **plan to implement the Digital Switzerland strategy.**

- The Digital Switzerland section of the Federal Chancellery will accept proposals based on set criteria regarding proposal relevance, broad sponsorship, readiness for implementation, time frame and quality

- Artificial intelligence is a core theme of the Digital Switzerland Strategy Form : here

E. Proposed Legislation

On November 22, 2023, the Swiss Federal Council directed the Federal Department of the Environment, Transport, Energy and Communications (DETEC) to explore potential regulatory approaches to artificial intelligence (AI) by the end of 2024.
- This initiative aims to leverage AI's benefits while mitigating societal risks. The overview will consider current Swiss laws, align with the upcoming EU AI Act and the Council of Europe's AI Convention, and focus on ensuring compliance with fundamental rights.
- The analysis will address **legal, economic, and European policy aspects**, involving interdisciplinary cooperation across federal departments.

- The goal is to establish a foundation for an AI regulatory proposal by end of 2024 and clarify responsibilities.
- See here

F. Commentary

Charting the Future: Switzerland's path to Generative AI leadership in 2024 and Beyond, IMD (see here)

Demand for skills training:
- As generative AI (GenAI) technologies drive significant shifts in the workplace, there is a clear need for increased investment in skills training. This encompasses not only new training programs but also transforming continuing education to meet the evolving needs.
- Government, organizations, and providers of continuing education must invest now in the resources to successfully meet this swelling demand.

Awareness and Regulation :
- Swiss professionals are broadly uninformed about official policies governing the use of GenAI. While GenAI adoption is growing in Swiss organizations across functions, most still do not have clear company guidelines on its use.
- Swiss regulators and authorities must take immediate and decisive actions to inform and safeguard citizens and organizations, while supporting the growth and innovation opportunities that GenAI delivers.

'Trust Gap' - Public skepticism :
- There is a general growing skepticism among the Swiss public about AI. While there is cautious optimism about its potential benefits, there is also a significant "trust gap" driven by imbalanced media coverage. To bridge this gap, it is recommended that Swiss authorities increase efforts to reduce disinformation.

Privacy and Data Protection :
- The Swiss Federal Act of Data Protection lacks broad and coherent standards, particularly in addressing the deletion of data which can impede the learning processes of GenAI.There are concerns about the unchecked adoption of AI, especially among younger, digitally native generations, which is outpacing current regulations.

Innovation and Competitiveness :
- Swiss innovation and competitiveness in GenAI are at risk of stalling due to a lack of data. Switzerland's small population and data production limit its capacity to innovate in GenAI, leading to heavy reliance on foreign GenAI platforms.
- Regulation is needed, say our experts, along with incentive systems for data sharing, so that global viewpoints and culture differences like those of Switzerland are fully and appropriately represented in AI. Experts suggest creating a Swiss Data Act or aligning Swiss regulations with the EU AI Act to enhance data sharing incentives and ensure Switzerland's representation in global AI development.

Decision makers must :
- Act now to enact a robust and flexible educational and political framework to ensure that Switzerland disrupts – and is not disrupted by – GenAI. Besides country-level regulations, it is crucial for organizations to develop internal rules and guidelines for managing these technologies.

- Invest in the transformation of continuing education and the training of educators at all levels to meet growing and changing needs, leveraging GenAI technologies to enhance training capabilities.
- Prioritize the critical, judicious, and vigilant use of GenAI to harness opportunities while minimizing risk.

8. United Kingdom

A. Policy Summary

Political Goals :
- Establishment as an attractive technology location for founders and entrepreneurs
- Expansion of the digital infrastructure
- Promoting AI-related cooperation

No AI-specific regulations or laws enacted in the UK. Potential statutory duty on regulators to respect cross-sectoral principle to be introduced, but still uncertain.

Policy Initiatives: White Paper 2023, Government response ' a pro-innovation approach to AI regulation' and other fields of action such as investments in digital infrastructure, training and further education for AI

Consultation in Process for Cyber Security of AI, plans to launch more in 2024.

Commentary : UK is very focused on economic plans with a stated "pro-innovation" approach and lack of emphasis on ethical and transparent AI integration; No clear methodology or stakeholder involvement in developing AI regulations; UK prioritizes technology performance and innovation, possibly compromising transparency and data protection. Several initiatives have been initiated to address deficiencies, such as Responsible AI UK and the AI Safety Institute.

B. Laws & Regulations Implemented

The UK government's AI Regulation White Paper (3/08/2023) indicates that the **UK does not intend to enact horizontal AI regulation in the near future.**
- "New rigid and onerous legislative requirements on businesses could hold back AI innovation and reduce our ability to respond quickly and in a proportionate way to future technological advances." - White paper 2023
- The UK considers that a **non-statutory approach** to the application of the framework offers **"critical adaptability"** that keeps pace with rapid and uncertain advances in AI technology.
With the new Labour government taking power in 2024, this position may potentially be revisted.

Other Indirect laws affecting AI :
- Data protection laws

- Intellectual Property laws
- Human rights laws (particularly the Equality Act 2010 and the Human Rights Act 1998)
- Consumer and competition laws

C. Policy Initiatives

White Paper 2023

Aims of the Regulatory framework :

1. Drive growth and prosperity
- Facilitate responsible innovation and reduce regulatory uncertainty
- Encourage AI investment and adoption, creating jobs and improving efficiency.
- Remove barriers to innovation swiftly, enabling AI companies to leverage early successes and gain long-term market advantages.

2. Increase Public Trust
- Address AI risks and protect fundamental values to build trust.
- Demonstrate that the regulatory framework effectively mitigates risks, encouraging AI adoption and innovation.

3. Strengthen Global leadership
- Shape international AI governance and regulation to promote responsible innovation.
- Address global challenges like climate change and pandemics.
- Lead in the global AI assurance industry, including auditing and safety.
- Promote interoperability with other regulatory systems and minimize cross-border frictions.
- Engage with global partners to influence and adapt to evolving AI regulations.

Essential characteristic of UK's regulatory regime :

Pro-innovation; Proportionate: Avoids unnecessary burdens on businesses and regulators; **Trustworthy:** Avoids unnecessary burdens on businesses and regulators; **Adaptable:**Quickly responds to emerging AI opportunities and risks; **Clear:** Ensures businesses understand the rules, applicability, enforcement, and compliance; **Collaborative:** Encourages cooperation among government, regulators, industry, and the public.

4 key elements of the framework :
1. Defining AI: Establishes a definition based on AI's unique characteristics to facilitate regulator coordination.
2. Context-Specific Approach: Adopts a tailored approach to AI regulation based on specific contexts and sectors.
3. Cross-Sectoral Principles : 1. Safety,security and robustness, 2. appropriate transparency and explainability, 3. fairness, 4. accountability and governance, and 5. contestability and redress. => principles to guide regulators in addressing AI risks and opportunities.Initially, application of principles will be **discretionary for regulators**, with a possible future statutory duty for compliance.
4. Central Support Functions: Introduces new central functions to aid regulators in implementing the AI framework. Ensures coherence and maximizes the benefits of an iterative regulatory approach.

Fields of action :

Investment-friendly environment
- Increase government spending on research and development
- Tax incentives for research and development in the field of AI

Training and further education for AI
- Investments in STEM teaching and digital education
- Further training for all population groups in the field of AI, especially in structurally weak regions

Investments in digital infrastructure
- Expansion of digital mobility concepts and promotion of electromobility and digitally supported charging
- Promoting the construction of smart homes
- Investments in modern broadband expansion

Economic development
- Promoting partnerships between government and industry, particularly in the areas of health, mechanical engineering and automotive
- Establishment of an investment fund for the British economy, supported by the British Business Bank
- Evaluation of all policy measures to improve economic innovation capacity

D. Consultations in Process

Cyber Security of AI : Call for views

The UK Department for Science, Innovation and Technology seeks input on enhancing AI's cyber security to ensure AI's safe integration into daily life. This initiative outlines specific measures to secure AI, maximizing its benefits while addressing potential risks. Stakeholders are invited to provide their views on these interventions to help shape robust security frameworks for AI technologies. The consultation closed on August 9, 2024.

E. Proposed Legislation

The prior administration announced in its response to White Paper consultation (here) that they anticipate introducing a **statutory duty on regulators requiring them to have due regard to the cross-sectoral principles** after reviewing an initial period of non-statutory implementation. Additionally, beyond data protection laws, **the government is exploring measures to ensure responsible AI development, including transparency in training data, robust risk management, and corporate governance standards**. Bese efforts aim to ensure safe and principled AI development.

The new administration is still formulating its positions with respect to AI policy and legislation.

F. Commentary

Montasari, R., 'National Artificial Intelligence Strategies: A Comparison of the UK, EU and US Approaches with those Adopted by State Adversaries' in Countering

Cyberterrorism

- Brexit has allowed the UK to take a much more US-centric approach to its framework, focusing on the economic outcomes of its strategies. However, this has raised questions as to whether this is a step in the right direction and whether following in the footsteps of the US is a sensible approach for data protection rights in the UK.
- The main criticism of the UK's AI national strategy is that, despite being presented as a comprehensive plan for artificial intelligence, it predominantly focuses on future economic plans, geopolitical considerations, and how AI can support these areas. In contrast, other strategies like the EU's emphasize developing innovative, ethical, and well-governed AI systems and technologies, integrating them transparently into modern environments. The UK's strategy is seen as lacking in this regard. Achieving ethical innovation requires multidisciplinary expertise and global diplomatic efforts.
- Another criticism of the UK's AI strategy is the lack of clarity regarding responsibility and liability for AI regulations. There is no defined methodology for developing these regulations, nor any statements on how stakeholders will be involved in the process (Cooper et al., 2021). This lack of clear legal and ethical guidelines needs to be addressed for the UK to match the ethical standards of AI integration seen in the EU.
- The UK's AI strategy seems to prioritize technology performance and innovation, potentially at the expense of full transparency and data protection. This raises concerns about whether the strategy is genuinely pro-innovation or a regression in data protection rights. The EU views enabling standards and regulations as innovative, a perspective that appears to clash with the UK's approach. Addressing this discrepancy requires aligning the definition of 'innovation' internationally, as the EU's focus on standards and regulations as part of innovation does not currently align with the UK strategy.

9. Canada

A. Policy Summary

The Artificial Intelligence and Data Act is expected to regulate AI at the federal level (post2025).

Policy Initiatives :

- Pan-Canadian strategy : commercialisation, standards, talent and research
- Canadian AI Sovereign Compute Strategy
- AI Compute Access Fund
- Aligning Canadian AI practices with international human rights and responsible corporate governance.
- Ensuring AI systems are transparent, explainable, and accountable

Active public consultation : Consultation on Artificial Intelligence (AI) Compute.

Commentary :

- Criticism of the AIDA proposal : potential conflict with existing law, difficult to understand, exclusion of government use

B. Laws & Regulations Implemented

The Artificial Intelligence and Data Act is expected to regulate AI at the federal level (see E. Proposed legislation).

There are existing laws that may affect the development or use of AI in Canada :
- The Privacy Act and the Personal Information Protection and Electronic Documents Act (PIPEDA)
- At a provincial level, the Personal Information Protection Act, SA 2003 (in Alberta); the Personal Information Protection Act (in British Columbia); and the Act Representing the Protection of Personal Information in the Private Sector (in Quebec).
- Intellectual property laws may affect several aspects of AI development and use.
- Competition law may have influence over the structure of the AI market in Canada.
- The Canada Consumer Product Safety Act
- The Food and Drugs Act
- The motor Vehicle Safety Act
- The Bank Act
- The Criminal Code
- The Canadian Human Rights Act and provincial human rights laws

C. Policy Initiatives

Pan-Canadian AI Strategy :
- Launched in 2017, the Pan-Canadian Artificial Intelligence Strategy (PCAIS) was the first fully-funded national AI strategy in the world.
- Key Strategic Priorities : Advancing AI Science, AI for Health, AI for Energy and the Environment, AI commercialisation

3 Pillars of the Strategy :

1. Commercialisation
National Artificial Intelligence Institutes :
- The Mila-Quebec Artificial Intelligence Institute in Montréal, the Alberta Machine Intelligence Institute (Amii) in Edmonton, and the Vector Institute in Toronto.
- These institutes translate AI research into commercial applications and support businesses in adopting AI technologies.
- Supported by US$ 60 million from Budget 2021, each institute is eligible for up to US$ 20 million over five years (2021-2026).
Canada’s Global Innovation Clusters :
- Canada's Global Innovation Clusters – Digital Technology, Protein Industries Canada, Next Generation Manufacturing Canada, Scale AI, and Canada's Ocean Supercluster – are strengthening Canada's innovation landscape by promoting the adoption of made-in Canada artificial intelligence technologies by businesses in key industries, and by public and not-for-profit entities.
- The government is supporting this initiative with US$ 125 million in funding provided in Budget 2021, over five years, from 2021-2022 to 2025-2026.

2. Standards
- Through the Standards Council of Canada, the Government of Canada is supporting efforts to advance the development and adoption of standards related to artificial intelligence.
- The government is supporting this initiative with US$ 8.6 million in funding provided in Budget 2021, over five years, from 2021-2022 to 2025-2026.

3. Talent and Research
CIFAR
● CIFAR is enhancing programs to attract, retain and develop academic research talent, and maintain centres of research and academic training at Amii, Mila, and the Vector Institute. In addition, CIFAR is renewing its advanced research, training, and knowledge mobilization programs.
● The government is supporting these initiatives with US$ 208 million in funding provided in Budget 2021, over ten years, from 2021-2022 to 2030-2031 Compute
● The Digital Research Alliance of Canada is providing dedicated computing capacity for artificial intelligence researchers across Canada to support the objectives of the strategy.
● The government is supporting this initiative with US$ 40 million in funding provided in Budget 2021, over five years, from 2022-2023 to 2026-2027.

As part of Budget 2024 investments the Government is launching a new Canadian AI Sovereign Compute Strategy and AI Compute Access Fund. This is a US$ 2.0 billion investment, to launch two new initiatives that will provide Canadian researchers and AI companies with the tools needed to be competitive in a rapidly advancing global AI landscape:

- **a Canadian AI Sovereign Compute Strategy** that will guide Canada's efforts to develop AI compute infrastructure over the long term. Investments made under this strategy would be directed at incentivizing and developing compute capacity; and

- **a new AI Compute Access Fund** that will provide direct support to Canadian AI researchers and developers to access the compute they need in the near term.

These initiatives will enable Canada to secure its globally competitive position by ensuring that both AI industry and researchers have access to affordable and cutting-edge infrastructure.

Other key facts and figures :
- The Government has taken significant steps to support both AI and high-performance computing (HPC) for researchers in Canada. Budget 2018 committed US$ 572.5 million to the Digital Research Infrastructure Strategy to enhance digital research infrastructure, supporting advanced computing data management and high-speed networking for the academic research community.
- Created three national AI institutes to be global centers of training and research excellence—the Mila-Quebec Artificial Intelligence Institute (Mila) in Montréal, the Alberta Machine Intelligence Institute (Amii) in Edmonton, and the Vector Institute in Toronto.
- Budget 2021 committed US$ 40M though the Digital Research Alliance of Canda in collaboration with the three AI institutes - Vector, Mila and Amii - to provide dedicated computing capacity for AI researchers across Canada to support the objectives of the PanCanadian AI Strategy (PCAIS).

D. Consultations in Process

On 26th June 2024, the Minister of innovation, Science and Industry announced the Consultation on Artificial Intelligence (AI) Compute.
This initiative aims to inform the development of a new AI Compute Access Fund and a Canadian AI Sovereign Compute Strategy, proposed in Budget 2024 with a US$ 2 billion investment. The consultation features the AI Blueprint, a discussion paper outlining Canada's AI opportunities and ambitions.
The consultation will involve Canadian researchers, innovators, businesses, civil society, Indigenous groups, and other stakeholders to identify optimal investment strategies for Canada's AI future. It will include various engagement methods, including online platforms.

E. Proposed Legislation

Canada's Proposed AI regulation through the **Artificial Intelligence and Data Act (AIDA)**
- Forms part of Bill C-27, which is an Act to enact the Consumer Privacy Protection Act (an update ot the current federal privacy law), the Personal Information and Data Protection Tribunal Act and AIDA
- As of 2024, Bill C-27 is under committee consideration in the House of Commons. Its enactment timeline will depend on the legislative processes progression. Provisions expected to come into force after 2025.

Purpose and Goals:
- AIDA aims to regulate AI at the federal level, ensuring the responsible development and deployment of AI technologies in Canada.
- Seeks to protect Canadians, foster innovation, and position Canadian firms and values prominently in global AI development.
- Aligns Canada's approach with international standards and establishes a risk-based regulatory framework focusing on high-impact AI systems.

Scope and Applicability :
- Applies to private sector entities involved in the design, development, or deployment of AI systems in international and interprovincial trade and commerce.
- Mandates rigorous assessment and mitigation of risks for high-impact AI systems, ensuring adherence to safety and ethical guidelines.

Key Obligations and Requirements:
- Entities under AIDA must conduct risk assessments, establish risk mitigation measures, ensure continuous monitoring, and publicly disclose information about the functioning, intended use, and risk management of high-impact AI systems.
- The Minister of Innovation, Science, and Industry is empowered to enforce compliance, conduct audits, and impose penalties for violations.

The Innovation Council of Quebec has recommended amending Quebec labor law to account for the impacts of AI.

F. Commentary

Teresa Scassa, 'Regulating AI in Canada: A Critical Look at the Proposed Artificial Intelligence and Data Act' (2023) No 1.
- Canada's early entry into AI governance in 2017–along with the country's large-scale and continuous advancements in AI governance across multiple sectors and levels of government (Brandusescu, 2021; Frost, 2020)–have made Canada into a uniquely information-rich national AI governance context.
- The bill (AIDA) is challenging to fully understand, as many of these obligations are left to be fleshed out in regulations, including even the definition of the 'high impact' AI, to which the AIDA will apply
- The AIDA may conflict or overlap with norms or rules in other legislation such as data protection laws, as well as emerging laws for issues such as platform governance
The federal government's proposed AI legislation misses the mark on protecting Canadians, 04/2024 (here)
- The first issue is that AIDA as presently drafted does not address government use. This is despite widespread use across the public sector.The Canadian Tracking Automated Governance (TAG) register lists 303 applications of AI within government agencies in Canada. The fact that AIDA as presently drafted will not apply to government use means this legislation is out of step with AI governance in other AI leading nations and the expressed interests of government employees.
- AIDA will not apply to public sector uses of AI, despite the widespread use of AI and automated systems. This runs counter to expressed concerns of public sector workers.
The Canadian Union for Public Employees, the Professional Institute for Public Employees and the Canadian Labour Congress have called for AIDA to apply to government departments, agencies and crown corporations.
- AIDA has been widely criticized for not providing the protections Canadians need.
- AIDA was rushed and there has been no meaningful consultation with the public.
- Canadians have one of the lowest levels of trust in AI, even though Canada has had one of the first national AI strategies.
- Given these limitations, AIDA is already out of step with the needs of Canadians. Canadian legislation also falls short of the regulatory approaches taken by other nations.

10. People's Republic of China

A. Policy Summary

China is focused on becoming a global leader in AI by 2030. The country's approach to AI regulation is characterized by a strong emphasis on national security, public interest, and alignment with socialist values. China's policies are designed to promote technological innovation while ensuring that AI development is ethical, safe, and beneficial to society. The government has introduced various regulations, such as the Deep Synthesis Provisions and the draft Artificial Intelligence Law, to manage AI's development and mitigate associated risks.

Policy Initiatives :
● **Next Generation Artificial Intelligence Development Plan** - Aims to position China as the global leader in AI by 2030 with significant investments in AI infrastructure and research. (CGTN)

● **Deep Synthesis Provisions** - Regulates the use of technologies that generate synthetic content (e.g., deepfakes) to ensure transparency and prevent misuse. (Reed Smith LLP).
● **Generative AI Measures (2023)** - Requires AI service providers to conduct security assessments and ensure transparency in AI-generated content.(Reed Smith LLP).
● **Artificial Intelligence Law (Draft, 2024)** - A comprehensive draft law governing AI activities, focusing on ethical development, national security, and public interest. (CSET).

B. Laws & Regulations Implemented

China's three most concrete and impactful regulations on algorithms and AI are its **2021 regulation on recommendation algorithms, the 2022 rules for deep synthesis** (synthetically generated content), and the **2023 draft rules on generative AI**. Information control is a central goal of all three measures, but they also contain many other notable provisions.

The rules for recommendation algorithms bar excessive price discrimination and protect the rights of workers subject to algorithmic scheduling. The deep synthesis regulation requires conspicuous labels be placed on synthetically generated content. And the draft generative AI regulation requires both the training data and model outputs to be "true and accurate," a potentially insurmountable hurdle for AI chatbots to clear. All three regulations require developers to make a filing to China's algorithm registry, a newly built government repository that gathers information on how algorithms are trained, as well as requiring them to pass a security self-assessment.

Deep Synthesis Provisions (2022) - These provisions regulate the creation and distribution of AI-generated synthetic content, such as deepfakes. The regulations require that AI-generated content be clearly labeled, and they impose strict measures to prevent the use of such technologies for illegal activities.

On May 23, 2024, the National Information Security Standardization Technical Committee (NISSTC) released new draft regulations titled Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence (AI) Service (Chinese) - Implemented to manage the rapid development of generative AI technologies, these measures include requirements for security assessments, data privacy protections, and transparency in AI services. The measures highlight the balance between encouraging innovation and ensuring safety.

Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions from Scholars , 2nd May 2024)- This Law is enacted in accordance with the Constitution in order to promote technological innovation in artificial intelligence (AI), facilitate the healthy development of the AI industry, regulate AI product and service development, provision, and use activities, as well as their supervision and management, safeguard national security and the public interest, and protect the legitimate rights and interests of individuals and organizations.

C. Policy Initiatives

● A Next Generation Artificial Intelligence Development Plan) - Launched in 2017, this strategic plan outlines China's vision to become a global leader in AI by 2030. The plan emphasizes the integration of AI into various sectors such as manufacturing, healthcare, and transportation. It also includes specific goals like developing core AI technologies, creating a favorable AI ecosystem, and establishing China as a premier AI innovation center.

● National AI Standardization Guidelines - (2023) These guidelines are part of China's broader strategy to create a standardised framework for AI development, ensuring that AI technologies are safe, reliable, and interoperable across different sectors. The guidelines focus on creating a unified approach to AI standards, covering areas such as data management, algorithm transparency, and system interoperability.

- **Standard Development:** Establishing national standards for AI that align with international norms, particularly in data security and privacy.
- **Implementation Framework:** Creating a roadmap for implementing these standards across industries, with regular updates to adapt to technological advancements.

● **AI Innovation Action Plan for Industries (2021-2023)** - This action plan was designed to accelerate the integration of AI into key industries, focusing on enhancing productivity and innovation in sectors such as manufacturing, smart cities, and agriculture. The plan outlines specific goals and initiatives to promote the widespread adoption of AI technologies, improve industry standards, and support AI-driven economic

- **Sector-Specific AI Applications:** Developing AI solutions tailored to specific industries, such as predictive maintenance in manufacturing and precision farming in agriculture.
- **Support for AI Startups:** Providing funding and resources for AI startups to develop and commercialize innovative AI technologies.

D. Consultations in Process

None at present. China recently held a consultation on new Draft Regulations on Generative AI ((NISSTC) and draft Requirements for Security of Generative AI Services in Cybersecurity Technology Tc260)

E. Proposed Legislation

China Releases New Draft Regulations on Generative AI (May 2024)

● On May 23, 2024, the National Information Security Standardization Technical Committee (NISSTC) released new draft regulations titled Cybersecurity Technology – Basic Security Requirements for Generative Artificial Intelligence (AI) Service.

● China is striving to regulate generative AI while promoting innovation and technological advancement. The NISSTC has issued draft regulations outlining security measures for generative AI service providers, underscoring China's commitment to responsible AI development.

The draft serves as a reference for both service providers and regulatory authorities. It offers guidance for conducting security assessments and establishing pertinent regulations.

TC260 requests public comments on draft Requirements for Security of Generative AI Services in Cybersecurity Technology (May 2024)
● On May 23, 2024, the National Information Security Standardization Technical Committee (TC260) requested public comments on the draft national standard titled: 'Basic Requirements for Security of Generative Artificial Intelligence Services in Cybersecurity Technology.'
● In particular, the draft national standard specifies the basic security requirements for generative artificial intelligence (AI) services, including training data security, model security, and security measures. It is set to apply to service providers conducting security assessments and may be used as a reference by relevant competent
authorities.
● Lastly, the draft national standard contains an annex outlining the main security risks of training data and generated content.

F. Commentary
● **Regulatory Scope and Intent:** The draft regulations, proposed by China's National Information Security Standardization Technical Committee (TC260), aim to set stringent security standards for generative AI services. These standards cover various aspects, including the security of training data, the safety of AI models, and the transparency of AIgenerated content. The TC260's approach is seen as an effort to bolster China's regulatory framework around AI, ensuring that these technologies are developed and deployed safely and in alignment with national security and social stability goals. The emphasis on content safety throughout the AI model's lifecycle and the requirement for real-time security checks are particularly notable, reflecting China's focus on preventing misuse of AI (DataGuidance and Mondaq)

● **Challenges for AI Service Providers:** One of the critical challenges highlighted by commentators is the complexity of compliance with these new regulations. The draft requirements mandate thorough security assessments before AI services can be filed for regulatory approval. Providers must either conduct these assessments internally or entrust them to third-party organizations. This process is expected to increase operational costs and may pose significant hurdles, particularly for smaller AI firms. Moreover, the requirement for service providers to demonstrate the legality of their data sources and to manage intellectual property rights carefully adds another layer of complexity Mondaq and China Law Vision).

● **Impact on Innovation and International Collaboration:** The draft regulations have sparked discussions about their potential impact on innovation within China's AI sector. While the removal of certain restrictive language—such as the requirement that only locally registered models could be used—suggests some flexibility, there remains a concern that the stringent regulatory environment might stifle innovation, particularly for companies relying on foreign models or collaborating internationally. This balancing act between regulation and innovation is a recurring theme in the commentary, with some experts cautioning that over-regulation could hinder China's ambitions to lead in AI technology (The National Bureau of Asian Research and China Law Vision).

11. India

A. Policy Summary

India is committed to becoming a global AI powerhouse by leveraging AI to stimulate economic growth, enhance public services, and improve the quality of life for its citizens. The government aims to integrate AI across critical sectors such as agriculture, healthcare.

These policies and frameworks represent India's strategic approach to harnessing AI for national development while addressing ethical, legal, and societal challenges. They are part of a broader effort to position India as a global leader in AI, with a focus on responsible innovation and inclusive growth.

● National Strategy for Artificial Intelligence (2018) - This strategy, developed by NITI Aayog, serves as India's foundational document for AI development. It outlines the vision of AI being a force for inclusive growth and highlights the potential of AI in sectors such as healthcare, agriculture, education, smart cities, and infrastructure. The strategy also emphasizes the need for ethical AI, data privacy, and the development of AI technologies that are aligned with Indian values.
NITI Aayog - National Strategy for AI

● Responsible AI for All: Part 1 and Part 2 (2021) - These documents, also from NITI Aayog, provide guidelines for the ethical deployment of AI in India. Part 1 outlines the principles of responsible AI, including fairness, transparency, and accountability, while Part 2 focuses on operationalizing these principles through specific actions for government and private sector stakeholders.
NITI Aayog - Responsible AI

● Part 2 - Operationalizing Principles for Responsible AI (2021), which focuses on operationalizing principles for responsible AI. The report breaks down the actions that need to be taken by both the Government and the private sector, in partnership with research institutes, to cover regulatory and policy interventions, capacity building, incentivizing ethics by design, and creating frameworks for compliance with relevant AI standards.

● Digital Personal Data Protection Act (2023) - This act was enacted to regulate the processing of digital personal data in India. While not solely focused on AI, it is a critical component of the AI regulatory framework, addressing concerns about data privacy, a key issue in AI development and deployment. The act provides a legal structure for protecting personal data, which is vital for AI systems that rely on large datasets.
(Meity.gov)
● National Data Governance Framework Policy (Draft, 2022) - Released by the Ministry of Electronics and Information Technology (MeitY), this draft policy aims to modernize data governance in India, creating a conducive environment for AI and data-driven research. The policy is expected to streamline data management practices across government departments and establish a comprehensive repository of datasets that can be used for AI development.
Access Partnership - Key Policy Frameworks

B. Laws & Regulations Implemented

Currently, there is no AI-specific regulator in India. As such, the Ministry of Electronics & Information Technology is the executive agency for AI-related strategies and has constituted committees to bring in a policy framework for AI. The Ministry of Commerce and Industry has also established an 'Artificial Intelligence Task Force,'12 with the aim of eventually establishing some form of AI regulatory authority
The list below provides a perspective of how India's AI regulatory landscape has evolved, with a mix of foundational laws, ethical guidelines, and emerging policies designed to support and regulate AI technologies.

Information Technology Act (2000) - The IT Act is the foundational law that governs cyber activities in India, including AI. While not originally designed for AI, it covers aspects such as data protection, cybersecurity, and digital governance, which are crucial for AI implementation. The act has been used to regulate certain AI applications, especially in cases involving data breaches, privacy violations, and cybercrimes. (eProcedure)

Responsible AI for All: Principles for Responsible AI (2021) - Released by NITI Aayog, this document outlines ethical guidelines for AI development in India, focusing on fairness, transparency, and accountability. Although not legally binding, it serves as a de facto regulatory framework for AI ethics. It provides guidance on the ethical development and deployment of AI systems, influencing both public and private sector AI projects. (NITI Aayog)

Digital Personal Data Protection Act (2023) - This policy, still in its draft stage, aims to modernize the way data is governed in India. It focuses on creating a robust data governance framework that supports AI development and innovation. The policy proposes the creation of a comprehensive repository of datasets that can be used for AI research and development.
Current Status: The policy is currently being refined based on public consultations and stakeholder feedback. (meity.gov)

C. Policy Initiatives

India is a member of the Global Partnership on Artificial Intelligence (GPAI). The 2023 GPAI Summit was recently held in New Delhi, where GPAI experts presented their work on responsible AI, data governance, and the future of work, innovation, and commercialization.
GPAI's experts produce deliverables that can be integrated into members' national strategies to ensure the inclusive and sustainable development of AI.

Under the 2023 themes of climate change, global health, and societal resilience, experts worked to ensure that AI is used responsibly to address current challenges around the world.
GPAI's members, on the other hand, adopted the 2023 Ministerial Declaration, reaffirming their commitment to the trustworthy stewardship of AI in line with the Organization for Economic

Cooperation and Development (OECD) AI Principles, as well as their dedication to implementing those principles through the development of regulations, policies, standards, and other initiatives. In doing so, they highlighted efforts to bridge the gap between theory and practice and advance AI that is responsible, sustainable, and inclusive for all.

These initiatives represent the ongoing efforts of the Indian government to integrate AI into various sectors and ensure that its development aligns with national priorities and ethical standards. Each initiative is at a different stage of implementation, reflecting the dynamic and evolving nature of AI policy in India.

IndiaAI Program (2023) - The IndiaAI program, launched by the Ministry of Electronics and Information Technology (MeitY), aims to consolidate all AI-related research, innovation, and policy initiatives under one umbrella. This initiative is designed to position India as a global hub for AI by focusing on cutting-edge research, fostering innovation, and ensuring the ethical use of AI across sectors.

Current Status: A task force has been established to create a roadmap for the development and functioning of the IndiaAI program. This roadmap will guide the structure and implementation of AI projects across the country. Access Partnership

Development of Indian Standards for AI (Ongoing) - The Bureau of Indian Standards (BIS) is actively working on creating specific standards for AI technologies in India. These standards will cover aspects such as AI safety, ethics, and interoperability, and are intended to ensure that AI technologies developed in India are reliable, safe, and aligned with international norms.

Current Status: Committees have been set up to draft these standards, with ongoing discussions and revisions.

AI for Social Empowerment Program - This initiative is focused on using AI to empower marginalized communities and enhance social welfare in India. It includes projects aimed at improving access to healthcare, education, and financial services for underprivileged sections of society.

Current Status: The program is being rolled out in phases, with pilot projects already underway in certain states. (India AI)

D. Consultations in Process

Framing Key Standards (2023 - Ongoing) - The Bureau of Indian Standards (BIS) has established committees to propose Indian standards for AI. These standards are intended to guide the development, safety, and ethical deployment of AI technologies across various sectors. These standards will serve as benchmarks for AI safety and ethics, influencing how AI technologies are developed and used in India.

Opened consultation for proposals on AI and Competition Market Study
The Competition Commission of India (CCI) initiated a consultation on 22 April 2024, inviting proposals for a Market Study on Artificial Intelligence (AI) and Competition in India, with a submission deadline of 3 June 2024. The study aims to explore the implications of the use of AI for competition, identify potential competition issues, and understand the implications of AI applications on market competition, efficiency, and innovation.

E. Proposed Legislation

India does not have a legislative framework that expressly regulates the development and use of AI and ML tools/technologies. It is expected that this sector will be governed by the Digital India Act. This law is expected to facilitate AI development by 'safeguarding' innovation in AI, ML and other emerging technologies. The Government of India has indicated that while it will support monetisation of AI/ML technology in India, this process should be regulated by specific compliances for high-risk use cases, including human intervention and oversight, and ethical use of AI/ML tools and technology.

The Ministry of Electronics and IT (MeitY) is drafting a new law focused on artificial intelligence (AI), which will notably avoid prescribing penal consequences for violations, recognising the technology's significant benefits, according to a report by The Indian Express. This legislation will be a standalone law, which will require social media platforms such as Facebook, Instagram, YouTube, and X to include watermarks and labels on AI-generated content. MeitY is also exploring legal frameworks to mandate companies developing large language models to train their systems on Indian languages and context-specific content.

On March 1, 2024, the Indian government issued an advisory instructing platforms to obtain explicit permission from the Ministry of Electronics and Information Technology (MeitY) before implementing any "unreliable Artificial Intelligence (AI) models /Large Language Models (LLM)/Generative AI, software or algorithms" for users accessing the Indian Internet.
Furthermore, intermediaries or platforms are required to ensure that their systems do not facilitate bias, discrimination, or compromise the integrity of the electoral process. Additionally, they must label all artificially generated media and text with unique identifiers or metadata to facilitate easy identification.

India is taking significant steps towards regulating artificial intelligence (AI) with the introduction of a new draft to enhance and extend protections provided by the Digital Personal Data Protection (DPDP) Act to AI systems. The Act outlines a comprehensive framework to govern high-risk AI systems, ensuring they are developed and deployed responsibly. In this blog we will explore the draft, how it augments the DPDP Act, and the evolving privacy landscape in India.

F. Commentary

Last November, Union IT Minister Ashwini Vaishnaw announced plans to regulate the spread of deepfakes on social media, identifying them as a 'threat to democracy'. Vaishnaw highlighted the government's strategy focusing on deepfake detection, prevention, reporting, and public awareness.

On March 1, The Ministry of Electronics and IT, MeitY also issued an advisory mandating the labelling of under-trial AI models and prohibiting unlawful content. This directive has been reinforced by a recent advisory from the ministry requiring all AI-generated content to be labelled uniformly.

In May, IT Secretary S Krishnan further reassured the industry that while the government seeks to regulate AI, it will not stifle innovation. Reflecting on the approach taken with the Digital Personal Data Protection (DPDP) Act, Krishnan stated, “We will ensure that both the interests of innovation and protection of vital interests will come in in the future.”

Appendix C

A Word About Quantum

We would be remiss if we did not acknowledge that potential evolution may emerge once quantum computing reaches a sufficient level of capability. Our current view is that 'Y2Q' is still anywhere from 4 to 8 years off, but this could change with technology breakthrough. We therefore suggest that governments continue to monitor quantum, and have contingency plans in place, but that they do not need to integrate quantum AI into immediate-term plans.

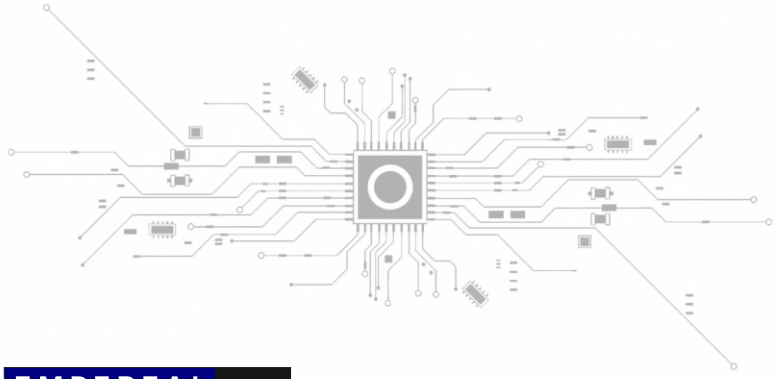
Quantum computing will solve problems that classical computers cannot, such as optimizing complex systems (logistics, energy distribution, drug discovery, etc.), accelerating breakthroughs in materials science, cryptography, and AI itself.

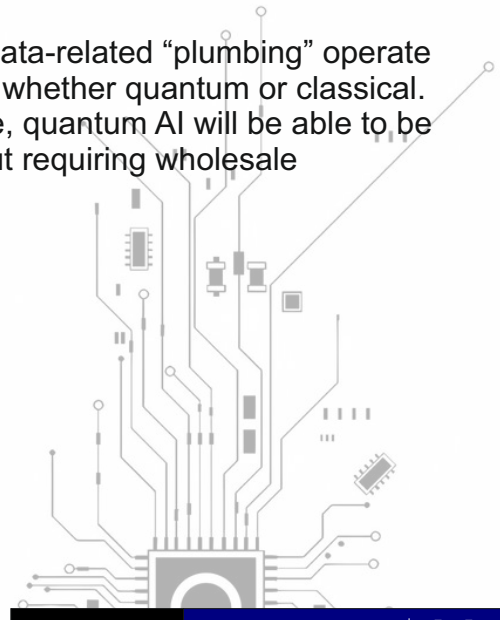
Quantum computers leverage the principles of quantum mechanics—like superposition and entanglement—to process information in fundamentally different ways than classical computers. By using quantum bits, or qubits, which can exist in multiple states simultaneously, they can evaluate many possibilities at once. This allows quantum computers to solve certain complex problems much faster than classical computers. Benefits include accelerating computations, and optimizing complex systems , all essential operations in AI. Overall, quantum computers have the potential to tackle problems that are currently infeasible, yet while the field has advanced significantly, it remains in its developmental stages.

Researchers have created quantum processors with increasing numbers of qubits, enhancing their stability and reducing error rates. Qubits are the key parameter measuring the complexity of problems that a quantum computer can solve and current achievements are on the order of 100 to a 1000 qubits. Therefore, building a fully functional, large-scale quantum computer is still challenging due to issues like qubit decoherence and the need for effective error correction.

Of relevance immediate to sovereign AI initiatives is the novel field of Quantum machine learning. Its aim is to combines quantum computing with machine learning techniques. Quantum machine learning explores how quantum computers can enhance the processing and analysis of data by using quantum phenomena like superposition and entanglement. These quantum properties have the potential to make certain computations significantly faster or more efficient than classical computers can achieve. In principle the operational stages of training an AI on a quantum computer are structurally the same to training an AI on a classical computer - data needs to be assimilated and setup for training in analogous ways.

This means that a sovereign AI initiative framework and a lot of the data-related "plumbing" operate independent from the nature of the involved computational principle, whether quantum or classical. This implies that aside from the very different hardware infrastructure, quantum AI will be able to be integrated into a properly constructed sovereign AI framework without requiring wholesale reformulation of the government's approach.





Origins of this Report

In the spring of 2023, Farhan Ahmad, the CEO of Payment Systems Malaysia (PayNet), participated in the World Innovation Network (TWIN) roundtable organised by Prof. Rob Wolcott of University of Chicago and Northwestern. At this roundtable, Mr. Ahmad met David Shrier, Professor of Practice in AI & Innovation with Imperial College London. Ahmad subsequently visited the Imperial campus and met with Shrier and his colleagues to discuss the importance of AI for Malaysia's financial system, and opportunities to leverage academic research and insights
to address fraud and cyber crime.

This led to a multi-year collaboration with Imperial's Trusted AI Alliance, a network of AI researchers spanning multiple institutions, seeking to deliver responsible and trustworthy AI for the world. As the conversations evolved, it emerged that Malaysia was revisiting its national AI strategy and was seeking to evaluate the broader context of AI policy and global interdependence, as well as address the needs for AI systems that are tailored to suit Malaysia's citizens.

In the course of this exploration the project team realised that the same questions that Malaysia faces, and the principles to develop answers to those questions, are also relevant to a broader set of policymakers and innovators in numerous other domiciles. PayNet authorised the creation of a report that not only could be useful in Malaysia, but that could help other ASEAN countries (as Malaysia assumes the Chair of the ASEAN group) and other political bodies globally, in navigating national AI policy and sovereign AI.