AI Governance: The path to responsible adoption of artificial intelligence.

BY BASIS AI

BASIS

Preamble.

AI is transforming how we design, build, work and live. In the next few decades, the effects of AI will be widely felt across all aspects of life.

As a prediction tool, AI is neutral. But how humans design and use AI, and manage potential biases have raised pertinent questions around trust, fairness and transparency.

At the heart of AI governance is the visibility of machine learning models - how to make these models more interpretable and explainable. It is automation through machine learning operations (MLOps) that enables this governance-by-design.

In this white paper, we unpack what AI governance is and is not, and review global developments in implementing AI governance. We also share our insights on how governance can be incorporated into MLOps, and provide practical tips to guide you in your AI adoption journey.

Introduction.

Artificial Intelligence (AI) now powers many real-world applications. AI is an umbrella term for different machine learning and deep learning capabilities to imitate intelligent human behaviour, such as analysing features and relationships between objects, reasoning and problem-solving. From voice recognition to forecasting demand, sectors that have successfully adopted AI stand to gain from the improvements in quality, efficiency and speed. These sectors include financial markets and services, healthcare, education, public sector, retail and e-commerce. Companies are forecast to spend $98 billion on AI worldwide in 2023, compared to $38 billion in 2019[1]. Management consulting group McKinsey forecasts that AI will add $13 trillion to the global economy by 2030[2].

The combination of man and machine makes AI a fundamental change agent since the Industrial Revolution. The development of AI may eventually usher in the singularity - the point at which AI surpasses human intelligence. Until then, AI can reduce the probability of human error, but it cannot fully eliminate error for as long as the human is involved in the process. AI can augment the scale and efficiencies of tedious tasks traditionally carried out by man, but it cannot replace the role of the human.

It is this interaction between man and machine that has also raised fears about the trade-offs between the benefits of AI and its risks. Particularly in governance dimensions where there are no clear answers yet. In Edelman's 2019 Artificial Intelligence Survey on the general public and tech executives, both survey groups acknowledge the benefits of AI, but there were also real concerns over its impact, such as smart toys that could invade children's privacy, and a loss of human intellectual capabilities[3]. 8 out of 10 survey respondents believe that advances in AI will likely cause a reactionary response from a society that feels threatened. A new report by the Centre for the Governance of AI likewise highlighted that more than 8 in 10 Americans believe AI and robots should be carefully managed. The American public also regards short and long-term AI governance issues, such as privacy, fairness to unemployment, as important for governments and technology companies to manage carefully[4]. The lack of societal consensus over the use of AI, and consumer distrust of AI, can impede AI adoption.

Global dialogue on AI governance.

There are a range of global frameworks on principles related to AI governance and ethics. Harvard University has produced a useful visualisation map of 32 sets of AI principles. Globally, there is convergence around a core set of AI principles, such as fairness, transparency and privacy. These principles[5] are functionally algorithm-agnostic, technology-agnostic and sector-agnostic.

However there is a lack of clarity over the definitions of these principles, making it hard to implement them. Existing generic guidelines do not provide a framework for technical implementation. No-go zones have yet to be defined. We expect to see more details emerge as governments and corporations push for a multi-jurisdictional approach to develop a common understanding. There is also a shared emphasis on engaging a diversity of stakeholders, such as policy makers, academia, private sector and industry experts, to share more industry-specific use cases and discuss cross-cutting implications on legal, risk and compliance domains.

A notable framework is Singapore's Model AI Governance Framework (Model Framework), the first to be developed in Asia. The strength of this framework is in the translation of these principles into a practical framework for action, thereby lowering the entry barriers to AI adoption. The Model Framework is founded on two high-level guiding principles -

that AI solutions should be human-centric, and that decisions made or assisted by AI should be explainable, transparent and fair. The four focal areas for action relate to key business concerns - internal governance structures and measures, determining the level of human involvement in AI-augmented decision-making, operations management, and stakeholder interaction and communication. To further augment accountability, Singapore also partnered the *World Economic Forum Centre for the Fourth Industrial Revolution to develop an Implementation and Self-Assessment Guide for Organisations and a Compendium of Use Cases* to help organisations assess and calibrate their progress in AI adoption.

Proactive governance and accountability practices are increasingly seen as a differentiating factor for businesses to brand themselves as trustworthy. The EU's GDPR audit process has a specific set of provisions that relate to a consumer's right to explanation when companies use algorithms to make automated decisions. The EU is also likely to be the first to introduce AI regulation laws. The EU Guidelines contain the highest number of requirements. Arguably, it sets the highest bar by requiring that trustworthy AI include assessments on wider social and environmental implications and potential dual-use of technology. In the US, the Algorithmic Accountability Act, requires major companies with access to large amounts of information, to audit machine-learning powered systems for accuracy, fairness, privacy and security risks.

Trust in technology.

In recent years, there has been a trust deficit in the digital economy, or “techlash” - defined as a strong reaction against the major technology companies, as a result of concerns over their power, users’ privacy, the possibility of political manipulation etc. Trust is easy to lose and difficult to win back.

“If the lifeblood of the digital economy is data, its heart is digital trust,” cited in a PWC report[6], that claims companies that will set themselves apart are the ones that prioritize security, reliability, privacy, and data ethics. What’s common across these components is the centrality of data in these discussions, and how they all relate to enabling consumer trust in the use of technology.

Threat scenario

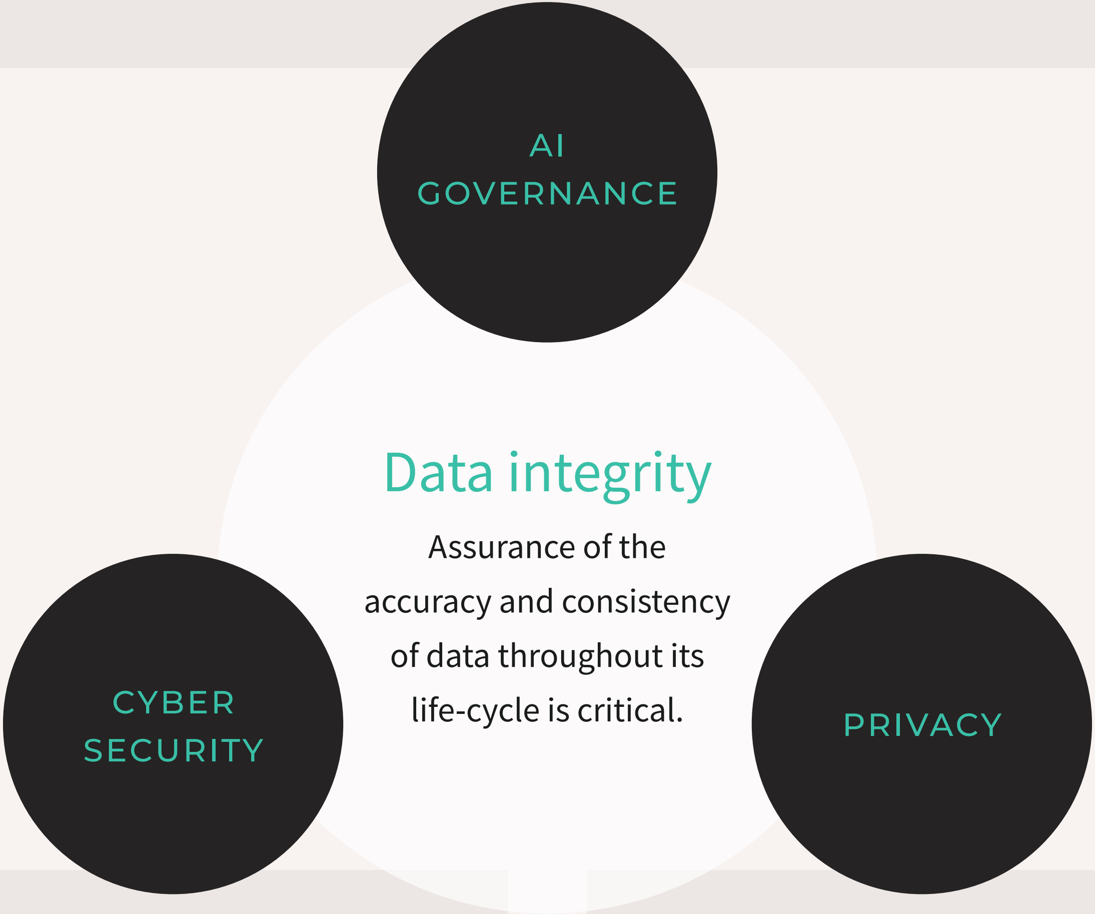
Data provenance - Unintended and undetected biases.

Underlying issues

Complex algorithms that are not easily understood.

Solve

Better transparency and auditability of AI systems.



AI GOVERNANCE

Data integrity

Assurance of the accuracy and consistency of data throughout its life-cycle is critical.

CYBER SECURITY

PRIVACY

Threat scenario

Data breach - Malicious external party tries to tamper with data or compromise availability of data.

Underlying issues

Security vulnerabilities, outsider and insider theft, human error.

Solve

Invest in technologies and risk management approaches to protect networks and systems.

Threat scenario

Data misuse - An organisation misuses personal data entrusted by an individual.

Underlying issues

Mismatch between regulatory and behavioural approaches when handling data.

Solve

Internal processes for handling personal data and customer consent.

What is AI governance?

> *A strong AI governance framework is the means by which AI adoption can be accelerated and used responsibly. It is the foundation for sustainable AI growth.*

AI is increasingly being used as an automated decision-making system for businesses. Some of these are straightforward predictions that do not require human oversight. Others are critical decisions that are made with little human oversight. In these situations, the risk and impact of an inaccurate prediction is high. As organisations calibrate the extent of human involvement in AI, it is also crucial to define what AI governance means for them.

AI governance is a framework and process for organisations to ensure that their AI systems work as intended, in accordance to customer expectations, organisational goals and societal laws and norms. When integrated with other parts of the organisation, decision trade-offs can be made in view of overall compliance and risk management perspectives. AI governance comprises two broad components:

Guidelines for action

This is a set of consistent principles that guide the design, development and deployment of AI in a way that is explainable, transparent and ethical. This leads to responsible use of AI, which ultimately builds trust in AI.

Systems and processes

This is a set of quantifiable metrics, clear roles, actionable steps and auditable processes such as thresholds and controls on implementation.

A strong AI governance framework is the means by which AI adoption can be accelerated and used responsibly. It is the foundation for sustainable AI growth.

Organisations that have the foresight to invest in developing an implementable AI governance framework stand to harness the full benefits of AI in the long-term, such as:

- Engenders a strong brand reputation for the organisation.
- Accelerates AI adoption and the use of AI responsibly.
- Creates an ecosystem of trust amongst stakeholders such as customers, investors and regulators, leading to a virtuous circle for AI adoption.
- Educated customers can lead to greater participation rates, and in turn generates useful data on which organisations can build and fine-tune new technologies.
- Enables organisations to stay ahead of regulatory and governance standards, and retain the space for exploration of innovative AI tools and methods.

#1

How does AI ethics relate to AI governance?

AI governance and AI ethics are often discussed in tandem. AI ethics refers to a set of moral principles that help define the boundaries and responsibilities for action, especially in highly ambiguous situations.

The use of AI will invariably bring about ethical dilemmas. From the boundaries of privacy, to the impact of social media algorithms on our mental and emotional wellbeing, ethical decisions need to be made by leaders, not machines. This entails a moral judgement on what is good, and what is harmful. For example:

- How much should we trade off profit and transparency?
- How do we define a fair decision-making process?
- How far should we trust AI systems to make important decisions without humans-in-the-loop?
- What features (e.g. gender, age, income) are justifiable in the delivery of differential treatment or services?

AI governance can help leaders make the implications of these ethical decisions more transparent, by designing metrics to evaluate ethical trade-offs. For example, we can compare a model designed to have maximum effectiveness but no fairness constraints, versus another model that is designed to be fair, to see the relative drop in effectiveness.

#2

How does MLOps relate to AI governance?

"In order to build systems that robustly behave well, we of course need to decide what good behavior means in each application domain. This ethical dimension is tied intimately to questions of what engineering techniques are available, how reliable these techniques are, and what trade-offs are made - all areas where computer science, machine learning, and broader AI expertise is valuable" - An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence

Discussions on trust in AI often do not lead to tangible action. There is a gulf between the amorphous governance concepts that policy makers describe, and the technical processes required to build these systems. MLOps (a compound of "machine learning" and "operations") has proved a useful foundation for the implementation of sound AI governance.

MLOps arose out of the recognition that it is challenging to build a performant machine learning engine and subsequently maintain them. MLOps is an emergent machine learning practice that draws from DevOps approaches to increase visibility, automation and availability in machine learning systems. MLOps encourages:

Visibility of metrics. Enabling the ability to monitor and understand the way your systems are functioning at every level.

Version control. Ensuring you have the ability to collaborate, iterate and roll-back where necessary.

Automation. Automating various tasks of data scientists and engineers so that they can focus on the creative aspects.

#3

Attributes of well-governed AI systems.

Attributes	Benefits
OVERSIGHT AND VISIBILITY	
Comprehensive picture of all the AI models operating in the organisation.	■ Allows for scalability of governance framework across different AI models.
Visibility into the performance of AI models, both before deployment and on an ongoing basis.	■ Ensures AI models are working as intended. ■ Allows for deviations to be addressed. ■ Prevents AI models from going stale as data and conditions evolve.
EXPLAINABILITY AND INTERPRETABILITY	
Identify key factors that lead to the AI's decisions being made, even with complex models	■ Allows for decisions to be easily justified and communicated to stakeholders including executives, customers and regulators.
Inspect model's treatment of individual cases as well as overall properties of the model	■ Enables the technical teams to intervene quickly to diagnose and debug the algorithms when they go wrong.

Attributes	Benefits
TRACEABILITY AND MAINTAINABILITY	
Trace the provenance of how models are built, deployed and updated, using version control and logging of parameters and data sources.	▪ Systematic record of vital data on the AI build process is crucial for scalability, troubleshooting and accountability.
Diagnose quickly when AI models are deviating from their intended effect or detect unintended effects.	▪ Data provenance and diagnosis through metrics equip engineers to quickly detect and fix issues that can otherwise go undetected.
Fix problems decisively upon detection with the ability to roll back to safe versions.	▪ Ability to course correct, before damage is done.
Audit by knowing what you are looking for.	▪ Ensures processes implemented were as planned and in compliance with legal and regulatory rules.
FAIRNESS	
Identify unfair treatment of different groups of people which have been deemed to be ethically unacceptable e.g. race or gender	▪ With complex AI algorithms, bias is not intentionally introduced, but powerful algorithms working on vast datasets can find correlations with prohibited variables and cause unintended bias. Active steps have to be taken to assess for unintended bias.
Quantify the trade-offs between fairness and effective algorithms	▪ Provides context and a basis for making ethical trade-offs.
Correct bias while retaining effectiveness of algorithms	▪ Allows organisations to address and fix issues of unintended bias so that there is a path to AI governance and effective use of AI.

Building an AI governance culture.

An integrated AI strategy that translates governance principles into actions and behaviours is one of the key success factors in unlocking the full potential of AI. Besides the realm of machine learning, this must also be done in an organisation’s culture, mindset, structure and processes. However each organisation needs to customise its approach depending on where they are at in the AI adoption journey. In this section, we have short-listed best practices that have enabled organisations to pivot to being an early adopter of AI:

Sponsorship, ownership of AI strategy and clear accountability by the leadership	■ Ensure appropriate board-level endorsement of AI strategy, and accountability for the outcomes of AI. ■ Set up appropriate management committees to oversee the implementation of the AI governance framework. ■ Promote a culture of discussions on ethical dilemmas and speaking up on perceived ethical implications.
Clear organisational structures, roles and responsibilities to support the ethical deployment of AI	■ Develop a common AI project management framework to enable cross-functional teams to work together to define the problem and design the AI solution. ■ Internalise and ensure AI governance outcomes are embedded into the product development process by business units.
Holistic mitigation of enterprise risks	■ Incorporate AI governance framework with existing risk management framework. ■ Adopt a risk-based approach to assessing materiality - not all applications of AI need to be scrutinised as closely.
Inspect model’s treatment of individual cases as well as overall properties of the model	■ Invest in technology that enables you to introduce AI governance-by-design rather than pit technology and governance against each other. ■ Hire a diverse range of tech talent, or upskilling existing employees to grow into new AI governance related roles.
Ongoing education, communication and engagement with key stakeholders	■ Regular monitoring of trust levels in AI in the market. ■ Anticipate and shape customer sentiment through clear principles and guidelines for communicating AI, such as when addressing customer queries and appeals. ■ Internally communicate and engage employees to understand the the latest AI guidelines and best practices.

Conclusion.

> “
>
> *AI can be very powerful in helping you and your organisations make better decisions with data. But you must be able to trust these systems and understand the predictions they make. AI governance is the key to supercharging businesses through AI-driven products that are trusted by end users.*

The introduction of new technology has often come with a combination of hype, excitement and fear. The path to mainstream adoption of new technologies always requires better ways of managing risk. AI is no different. What’s challenging about AI is how complex it is to mimic the human mind, and how opaque it can be when implemented. As AI becomes more pervasive in decision-making, trust in AI becomes more vital.

The concerns over trust in the use of AI cannot be addressed solely by technology. It must be accompanied with AI governance, which is an intersection of two practices: one originating from ethics; the other from engineering processes such as MLOps. Leaders who are able to hold the tension between both practices will be well poised to lead their organisations on the path to AI adoption.

AI Governance: 10 key concepts.

1.

AI Ethics refers to moral principles that articulate the moral obligations and duties of an individual or group in AI adoption. These principles serve as a guide for decision-making during potentially ambiguous, uncertain or context-dependent situations.

2.

AI Governance is a framework and process that guides the design, development and deployment of AI in a way that is explainable, transparent and ethical. It comprises guidelines for actions, as well as systems and processes.

3.

Algorithm is a formula given to a machine (e.g. a computer) in order for it to complete a task. It comprises step-by-step instructions that form the building blocks of artificial intelligence and machine learning.

4.

Artificial intelligence is a subset of computer science that aims to build machines capable of doing tasks similar, equal or superior to that of the human (e.g. decision-making, object classification and detection, speech recognition and translation).

5.

Machine learning is a subset of AI. It is a set of algorithms that enables machines to learn from data sets to perform, and subsequently autonomously improve on a specific task.

6.

Deep learning is a subset of machine learning that uses artificial neural networks to process complex structures and relationships among data. Artificial neural networks are algorithms that have layers of connected neurons sending information to each other, mimicking the human brain, where "deep" refers to the number of layers.

7.

Explainable AI (X.A.I) is AI that is capable of showing its human operators how it came to its conclusions.

8.

Fairness is a social construct, with more than 20 mathematical definitions. A violation of an aspect of fairness can lead to bias. This typically happens when an algorithm produces results that are systematically prejudiced, consciously or unconsciously. Choosing one definition of fairness will mean violating aspects of others. A constructive approach is to focus on fairness goals - the definition of fairness that is in line with the organisation's risk appetite, and will enable the AI system to meet the business objectives and customer expectations.

9.

AI model training is a process of training the AI model. AI models can learn from massive amounts of unstructured data. This involves feeding data into the system, validating the algorithm and testing the results against real-life use. Once trained, the AI model is ready for inference.

10.

Inference is the process of making predictions or recommendations from an AI model. Typically AI systems are able to make these predictions at scale and have a prediction that is personalised or contextualised to each new situation, event or case it is presented with.

References.

[1] International Data Corporation, “Worldwide Spending on Artificial Intelligence Systems will be nearly $98 billion in 2023, According to New IDC Spending Guide”

4 SEP 2019

https://www.idc.com/getdoc.jsp?containerId=prUS45481219

[2] McKinsey Global Institute, “Notes from the AI frontier: Modeling the impact of AI on the world economy”

SEP 2018

https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-AI-frontier-modeling-the-impact-of-ai-on-the-world-economy

[3] Edelman, “2019 Edelman AI Survey”

11 MAR 2019

https://www.edelman.com/sites/g/files/aatuss191/files/2019-03/2019_Edelman_AI_Survey_Whitepaper.pdf

[4] Centre for the Governance of AI, “Artificial Intelligence: American Attitudes and Trends”

JAN 2019

https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/executive-summary.html

[5] Berkman Klein Center, “Principled Artificial Intelligence”

4 SEP 2018

https://ai-hr.cyber.harvard.edu/primp-viz.html

[6] PWC, “The Journey to Digital Trust”

2019

https://www.pwc.com/sg/en/publications/assets/the-journey-to-digital-trust-2019.pdf

ABOUT BASIS AI

BasisAI builds responsible augmented intelligence software for data-driven enterprises.

For more information on AI governance and responsible AI, or to speak to an AI specialist, visit basis-ai.com or get in touch via contact@basis-ai.com.