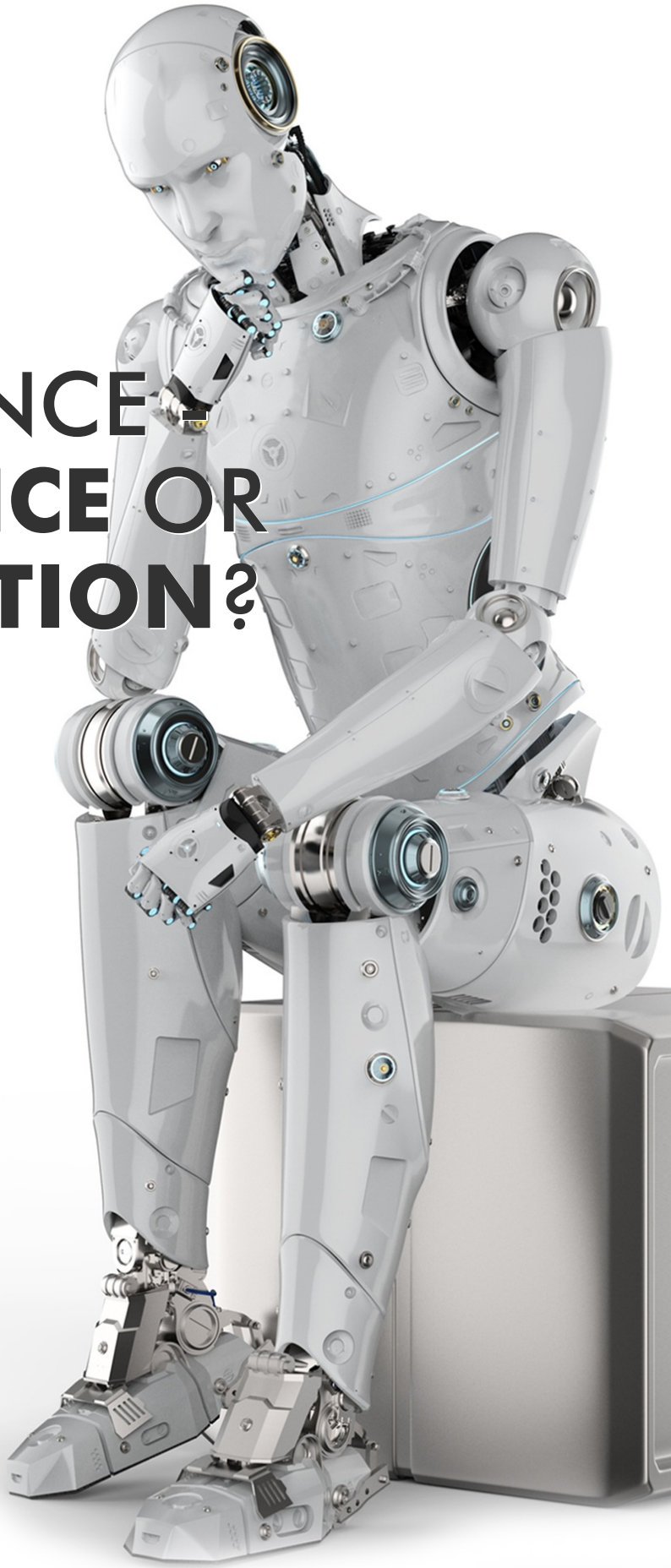
FUTUREMATTERS™
CENTRE OF EXCELLENCE

GLOBAL AI GOVERNANCE: **CONVERGENCE** OR **FRAGMENTATION**?



March 2024

CONTRIBUTOR

Shameek Kundu

Chief Strategy Officer

TruEra

JAPAN FINTECH FESTIVAL™

INTRODUCTION

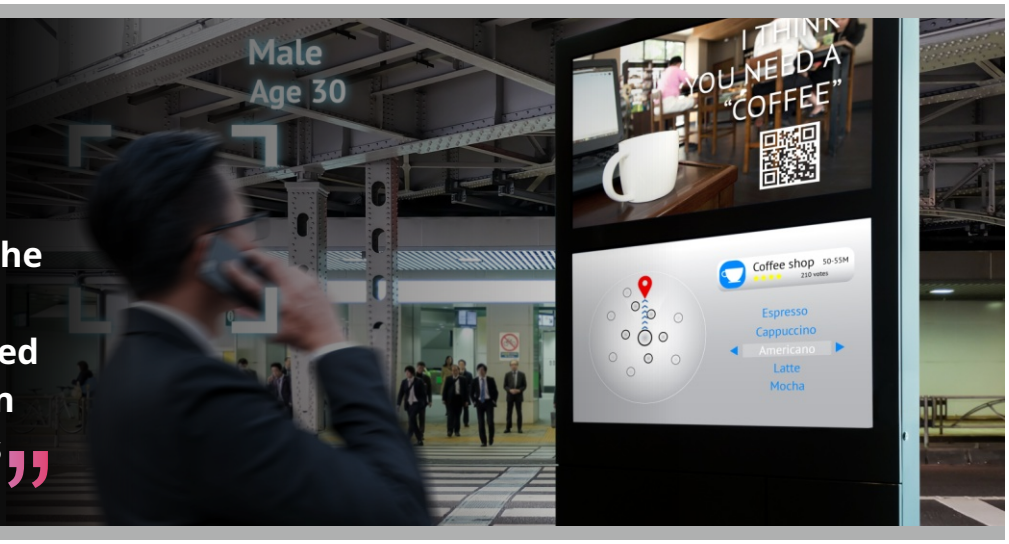
At least 70 countries have put forward guidelines, consultations and (occasionally) regulations to enable the safe, responsible and beneficial use of AI. In theory, that can make life very difficult for financial institutions operating in multiple jurisdictions, as they try to manage compliance against a diverse set of regulatory expectations.

This article starts with a representative overview of some of the major initiatives worldwide, with a particular focus on financial services. It highlights the dimensions on which countries differ in their approach, but also argues that they have a lot in common. Most countries have converged on the spirit, if not the letter, of what they are trying to achieve in AI governance. The article concludes by providing practical suggestions on how Financial Institutions can safely and responsibly ramp up AI adoption in a scalable manner.

From A Trickle To A Flood

Arguably, the first ever regulation impacting the use of predictive models (not limited to AI) goes back to 2011, when the US **Office of the Comptroller of the Currency** (OCC) and **Federal Reserve** published the SR 11-7 guidelines on Model Risk Management. 7 years later, the **Monetary Authority of Singapore's** Fairness, Ethics, Accountability and Transparency (FEAT) guidelines were the first to be specifically focused on the use of AI in financial services.

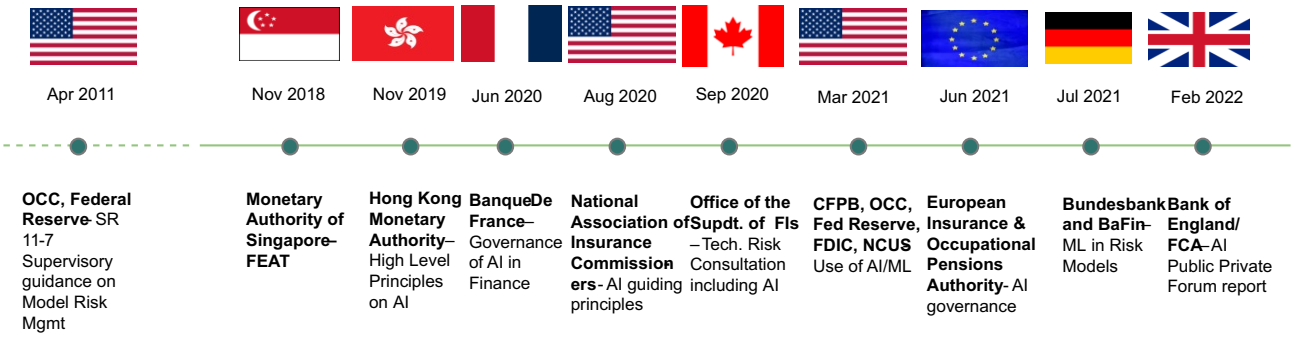
> **"The Monetary Authority of Singapore's FEAT guidelines were the first to be specifically focused on the use of AI in financial services"**



In the years since FEAT, a plethora of regulations, guidelines and consultation papers have emerged from regulators all over the world. Some have been specific to financial services, such as those from the **HKMA** in 2019, the **Banque De France,** Canadian **OSFI** and the **US National**

Association of Insurance Commissioners in 2020, the **Bundesbank/ BaFin**, **ECB** and the **European Insurance and Occupational Pensions Authority** in 2021, the **Bank of England/ FCA** in 2022 and the 2023 requirements on AI in insurance in the US state of Colorado.

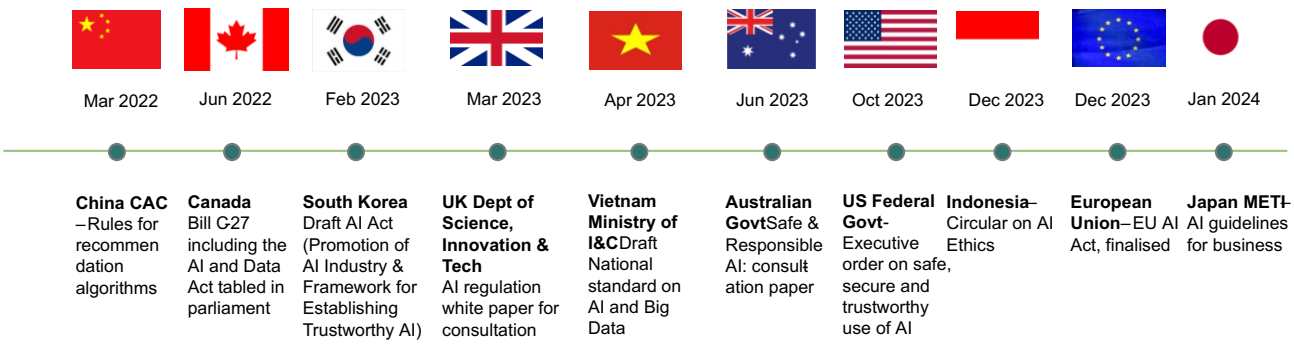
Regulatory initiatives around AI - Financial services examples



There have also been multiple cross-industry initiatives that can impact financial services. Actual laws or regulations dedicated to AI are still relatively rare. Examples include the recently finalised **EU** AI Act, and multiple targeted laws in **China** such as those on Generative AI, deep fakes and recommendation algorithms. Countries like **Canada** and **South Korea** have AI-related laws making their way through the legislative process. The **United States** published a National AI risk framework, and backed it up with an Executive Order in 2023.

Regulatory guidelines - often drafts for consultation, but sometimes formalised - have been issued in many other countries. These include: **Japan**'s 2024 Draft AI Guidelines for Business as well as previous principles and guidelines since 2019; consultations by the governments of the **UK** and **Australia** on their respective approaches to regulating AI, **Vietnam**'s Draft AI standard and **Indonesia**'s draft AI Ethics Circular in 2023; **Thailand**'s draft royal decree on AI in 2022; **India**'s NITI Aayog principles for Responsible AI in 2021; Singapore's Model AI Governance Framework; and the EU's trustworthy AI principles in 2019. There have also been a few attempts at multilateral coordination, including the **Hiroshima Process** Guiding Principles for organisations developing advanced AI systems (2023) and the **OECD** AI principles (2019).

Regulatory initiatives around AI - Industry agnostic examples



More In Common Than What Divides Us

If you are sitting in a multinational organisation, you would be forgiven for feeling overwhelmed in navigating through these numerous documents. From the very definition of AI to the mechanisms to ensure compliance, there are multiple dimensions on which countries, and individual regulators within countries, diverge in their approach to AI risk.

Sr	Dimension	Illustrative differences in regulatory positions	
1	AI Definition	Narrow focus on Machine Learning/ Deep Learning	Broad ranging definition including all statistical models and even rule-based systems
2	Objective	Safety/ harm-prevention as well as AI innovation/ investment	Prevent harm / ensure safety
3	Scope	Specific to an industry (e.g., insurance) or use-case (e.g., recruitment)	Horizontal, cutting across industries and use-cases
4	Graded	Limited to high risk use cases or scenarios	Applicable broadly irrespective of perceived risk levels (though the provisions may be relaxed for lower risk instances)
5	Mandate	Advisory guidelines for voluntary adoption	Hard regulation, with significant cost of non-compliance
6	Level of detail/ prescription	Not prescriptive on specific metrics or thresholds to use (e.g., "prevent unfair bias"). Potentially accompanied by illustrative examples	Prescriptive on specific thresholds on metrics, - e.g., assessing fairness using "Disparate Income <20%"
7	Technology-specificity	Technology-agnostic wording	Specific wording to address particular technologies (e.g., large language models)
8	Accountability across AI value chain	Clear distinction in expectations from AI developers (e.g., for foundation models), implementers and users	Overlapping expectations - e.g., placing obligations related to upstream foundation models on those implementing AI systems
9	Awareness of adjacent regulation	Piggy backing on existing regulatory requirements such as privacy, security, competition, fair treatment of customers, data quality and model risk	Wording in AI regulation/ guidelines overlap with (rather than just refer to) adjacent requirements
10	Enforcement roadmap	Structured pathway from regulation/ guidelines to standards to enforcement bodies	Unclear pathway - e.g., high level principles followed by case-by-case enforcement
11	Testing/ certification	Self-assessment as primary mechanism	Insistence on 3rd party testing / audit/ certification

However, the good news is that despite the range of potential design choices that individual countries or regulators can make, the core principles around AI governance/ Responsible AI are increasingly converging on the following set:

1. **Human-centricity:** human agency and oversight; respect for fundamental rights and/ or nationally shared values
2. **Safety and security:** Physical safety; prevention of harm to customers, staff or broader society; protection against malign attacks
3. **Robustness/ resilience:** Model accuracy and robustness in real-life situations including stress scenarios; operational resilience of the broader system and process in which the AI model is embedded (e.g., fall back plans)
4. **Fairness:** Prevention of unjust bias against one or more groups
5. **Transparency and explainability:** Transparency into the model itself (e.g., purpose, scope, limitations) and how it is trained/ used; appropriate level of model explainability; making model information available to customers and stakeholders
6. **Data privacy and governance:** Respect for applicable data privacy/ protection regulation; attention to quality and representativeness of the data used to train and run the model
7. **Accountability:** Clear articulation of roles and responsibilities; auditability throughout model lifecycle; demonstration of appropriate consideration of tradeoffs at relevant point (e.g., not choosing a complex model when simpler alternative suffices)
8. **Environmental and social responsibility:** energy consumption; social impact
9. **Respect for Intellectual Property considerations:** respect for IP rights over the data used for model training; clarity on IP rights of the output from the AI model

> **"The core principles around AI governance / responsible AI are increasingly converging"**

10. **Ensuring fair competition:** competition within the firm's own industry; avoiding excessive dependency on specific providers of AI models or associated infrastructure

Generative AI has shaken up some aspects of AI guidelines/ regulation. This is reflected in

- Greater focus on safety (both in the near term and at a more existential level), robustness (accuracy), data governance and environmental obligations
- A change in the approach to explainability (recognising that 'traditional' AI explainability concepts may be meaningless with large foundation models), and
- New(ish) interest in the risks arising from IP considerations and potential concentration of models, compute and data in a few companies/ countries

However, reassuringly, anyone who might have read the EU Trustworthy AI guidelines in 2019 will readily recognise most of these principles. Fast forward 5 years, and you can see these reflected in the recently published "AI Guidelines for Business" from January 2024 too.

> **"Generative AI has changed the equation somewhat when compared to the previous wave of AI guidelines/ regulation"**

Great, What Does It All Mean For Me?

Understanding the differences, and similarities, in AI guidelines/ regulations is a useful starting point, particularly for Financial Institutions (FIs) that operate in multiple countries. The next step is to try to meet these emerging requirements in a way that is **efficient, effective**, **scalable** and **agile** enough to align to ongoing changes.

What does that mean in practice? For established FIs, with existing risk management frameworks, that translates into 4 actions.

1. **Create your own baseline set of "AI governance" requirements** based on the existing/ emerging AI regulations/ guidelines in your footprint countries. As noted earlier, the core principles are unlikely to vary too much. However, your geographic presence, use cases and maturity of AI adoption will impact the next level of detail underneath.

   For example, FIs operating in the US will typically need to aim for a higher bar in model risk management, due to the long heritage of SR 11-7. FIs that are adopting Generative AI aggressively may need to increase their focus on security, privacy, operational resilience and accuracy requirements. In many cases, an individual FI may choose to set its minimum standards at a level consistent with the requirements from its most "aggressive" major regulator.

2. **Enhance existing risk policies and standards to reflect the AI Risk baseline:** With well-established risk management frameworks that encompass multiple risk types, incumbent FIs are unlikely to need to add any new risk types into their risk taxonomy. Almost every requirement falling under the AI Governance baseline is likely to be either already present in existing risk policies/ standards, or can be accommodated with relatively minor enhancements.

   For example, requirements around data representativeness can be easily accommodated by updating data quality considerations. Specific additions around explainability and conceptual soundness can make Model Risk Management policies AI-ready. The figure below illustrates this mapping to existing risk types in banking.

Mapping existing Bank risk types to AI risk considerations

3. **Reposition from "compliance" to business need:** Almost everything required by the emerging set of AI regulations/ guidelines is first and foremost a business necessity. It is not in the interest of a bank to build a model that breaks at the first sign of interest rate changes, or a chatbot that antagonises its customers through foul language. It is not in a data scientist's interest to make it difficult for business stakeholders to understand how the model is making the predictions. Justifying AI governance as a compliance requirement is therefore a massive missed opportunity.

4. **Embed into the model lifecycle:** FIs should aim to make the work needed for compliance as "invisible" and automated as possible. This means, for example, that testing for model fairness or robustness starts right from the model development stage, rather than being a standalone step at the time of formal validation. Or that the ML Ops technology stack supports such testing and ongoing monitoring by default.

These 4 steps also apply to "disruptor" fintechs. However, the effort required can be somewhat different - they may find step 4 somewhat easier, and step 2 more difficult, than their established, incumbent peers.

I write about trust in Data and AI, and the adoption of AI in Financial Services

The world of AI governance can be overwhelming, particularly for those tasked with interpreting different requirements across countries and navigating their multinational employers through them. However, at least in the world of financial services, the problem is less intractable than it might appear. Done right, FIs' response to regulatory requirements can *be a strong enabler* for AI adoption at scale.