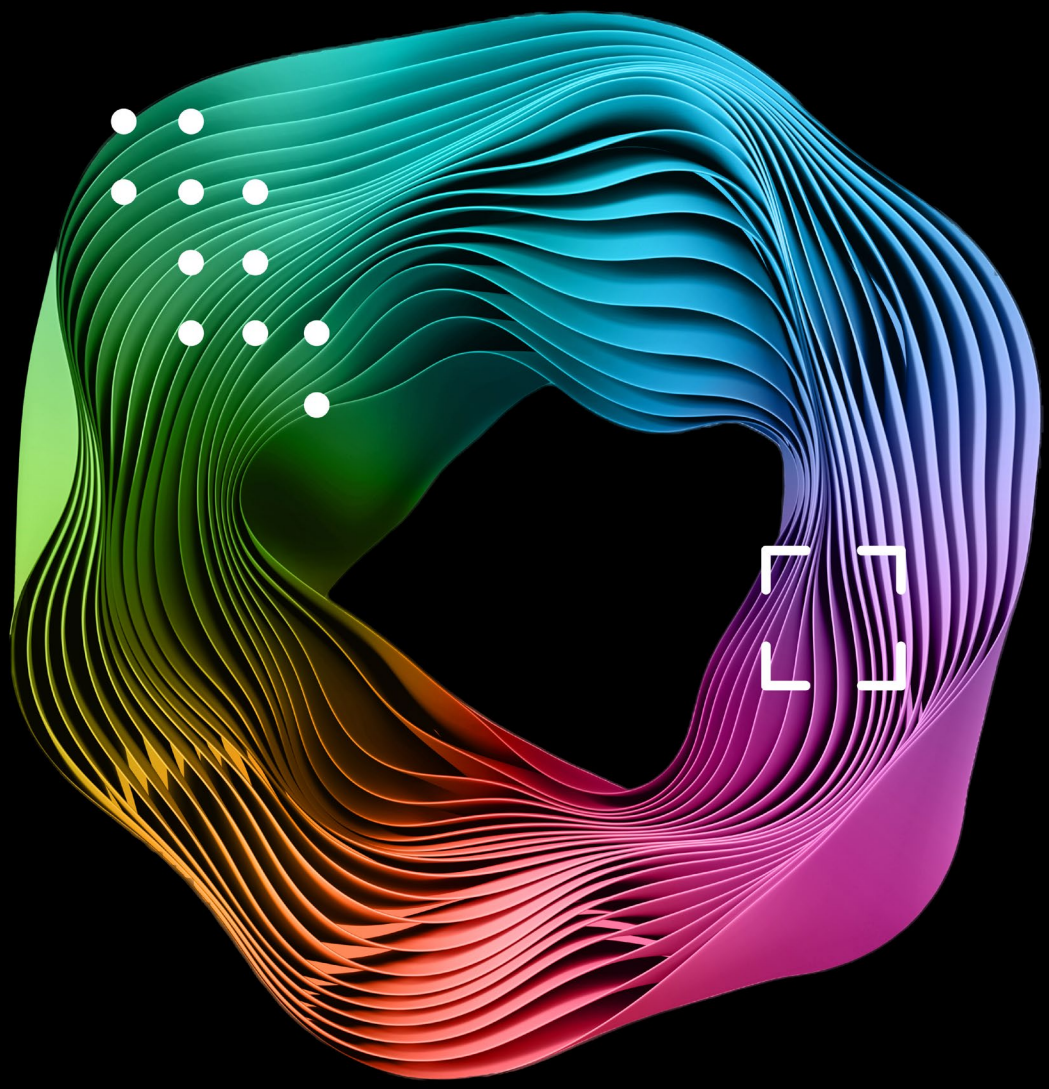
Deloitte.



AI risk and approaches to global regulatory compliance

MAKING AN IMPACT THAT MATTERS since 1845

Introduction

Foreword 04

Key Takeaways 05

Executive Summary 06

Chapter 1: From fiction to reality – defining AI, its evolution, opportunity, and impact 08

1.1. What is AI? 08

1.2 What are the risks of AI? 08

1.3 What are regulators doing about AI risks and what are the challenges? 11

Chapter 2: What are the key international AI governance processes and what risks and issues are they focusing on? 12

2.1 Which are the key international AI governance processes? 12

2.2 Trends and outlook for international cooperation 14

2.3. What AI risks and issues are these bodies focussed on? 14

Chapter 3: Common risks and issues across national and regional regulatory approaches 16

3.1 What AI risks and issues are national and regional approaches focussing on? 16

Chapter 4: Key themes identified across all international and national regulatory approaches .. 17

4.1 Key themes across international, national, and regional approaches 17

4.2 Consideration of how key themes addressed in regulatory approaches compare to the broad AI risk landscape 17

4.3. Deloitte view 17

Chapter 5: National and regional regulatory approaches 19

5.1. Overview of national and regional approaches 19

5.2. Detail of specific national and regional approaches 22

5.3. Definitions of AI in national and regional approaches 26

5.4. Deloitte view 27

Chapter 6: How can companies prepare for regulatory compliance? 28

6.1. Understand the impact of regulations on your AI business strategy 30

6.2. Create organisational clarity around the key operational challenges from AI regulation 30

6.3. Engage key stakeholders across the organisation to avoid silos 32

6.4. Design and implement AI governance and risk management framework 32

6.5. Implement some no regrets actions now 33

Contacts 36

Foreword

Jurisdictions around the world are rapidly introducing Artificial Intelligence (AI) regulation as they better understand the risks, both actual and perceived, that AI poses. Understanding these risks is a major challenge for regulators, as well as for organisations who are developing their AI strategies and AI systems, and deploying these systems and embedding models into their operations and products. It can be difficult for companies investing in AI to manage the uncertainties created by a dynamic regulatory landscape whilst still enabling AI innovation. In particular, understanding the regulatory space and finding the patterns, trends, and strategic implications across different sets of regulations to support compliance.

This regulatory and risk landscape analysis from Deloitte's Internet Regulation team is designed to support companies with these challenges.

The report is divided into six chapters:

- **Chapter 1** defines AI, outlines the broad set of risks associated with its use, and outlines some of the choices and challenges that face regulators in designing regulation.
- **Chapters 2, 3** and **4** focus on identifying a common subset of risks and themes that international bodies, governments and regulators are currently focussed on – the "what"
- **Chapter 5** examines "how" individual regulators are addressing these risks
- **Chapter 6** shares perspectives on what this means for companies, how they can develop appropriate guardrails to address AI risks, and engage the multi-stakeholder landscape to support regulatory compliance

We hope that this report will serve as a useful resource for companies looking to navigate the complex landscape of AI risk and regulation.

Key Takeaways

- The AI regulatory landscape is evolving at breakneck speed. As of April 2024, more than 300 AI-related laws, guidance, or regulations have passed or are in development across the globe
- This pace of change is likely to continue for a considerable period, creating uncertainty for international organisations who need to manage regulatory compliance whilst still innovating in AI
- There is close alignment across major international processes and national regulatory approaches in the key areas that are addressed through AI regulation – the "what". These areas are fundamental/human rights; fairness; privacy and data governance; safety; transparency; competition; and accountability and human oversight
- These areas only represent a subset of the whole AI risk landscape facing firms. This suggests that firms looking across multiple jurisdictions, and pursuing a compliance-first approach to managing AI risk, may be able to prioritise
- There is less alignment in how different jurisdictions are regulating, suggesting that many firms will have to manage divergence, which could increase over time.
- However, many regulatory approaches are multilayered (AI specific rules sitting alongside other tech neutral or sector specific rules impacting AI). This suggests that, even within jurisdictions, firms won't just be able to solve for AI regulation in isolation
- Elements of a risk and principles-based approach are also common. Alongside delivering specific technical requirements, this suggests that governance risk and control, monitoring and testing, documentation, audit and assurance are likely to be key elements of "how" firms can demonstrate compliance across jurisdictions
- A number of elements can support the response to global AI regulation, including understanding the strategic impacts; creating clarity around the operational challenges; engaging widely across the organisation as you build your compliance roadmap; and designing and implementing an AI governance and risk management framework
- As a starting point, firms should form an AI governance committee, establish an AI system inventory, and begin conducting AI system risk assessments

Executive Summary

From fiction to reality – defining AI, its evolution, opportunity, and impact

AI is a pioneering technology but can be hard to define. The explosion of large language models has catapulted "Generative AI" (or more commonly "GenAI") into the zeitgeist, but it is only one of several different technologies that make up AI. Whilst many businesses recognise the huge potential of AI and are increasingly adopting it at scale, there is also widespread concern about its risks.[1]

AI's unique nature as a technology – its autonomy and ability to learn and evolve - as well as uncertainty about its ultimate potential, make it difficult to precisely define AI risk.[2] An interim report endorsed by 29 countries - the International Scientific Report on the Safety of Advanced AI – was published in May 2024 and provides a comprehensive assessment of the latest scientific understanding, but also exposes the level of continued debate and uncertainty around AI risk. For organisations building their approach to AI, this uncertainty presents major challenges. We have identified nine broad areas that present a snapshot of the complex AI risk landscape. These include technical and security challenges; impacts on social and cultural dynamics; and potential long-term existential risks.

In response to these risks - in particular the potential for harm to individuals - and the rapid upswing in both the capability and adoption of AI technologies, governments and regulators are embarking on a wave of new regulatory activity. Governments face key design choices, including which risks to prioritise addressing in their regulatory approach, how to balance risk management with promoting innovation, and how to future-proof any new AI rules. A common theme, no matter the favoured approach, is that regulatory approaches are developing in a multilayered way, with direct AI regulation sitting alongside other cross-cutting or sector specific regulation. This will require organisations to have a strong regulatory horizon scanning capability. Commonly, legislators also appear to be pursuing elements of a principles-based, risk-centric approach akin to other landmark digital regulation such as the EU's Digital Services Act.

Whilst the regulatory landscape is still evolving, it is already clear that new regulation will have major impacts for organisations using AI, and that complying with regulatory requirements may require significant investment. Given the broad landscape of AI risk, key questions for organisations seeking to prioritise their approach is understanding the particular areas of risk that regulators are focussing on and whether there is convergence amongst them.

What are the key international AI governance processes and what risks and issues are they focusing on?

AI is a global issue, which means that there is already significant international coordination on understanding AI risk and developing corresponding regulatory approaches. A number of key international bodies – such as the United Nations (UN), Group of 7 (G7), Group of 20 (G20) and the Organisation for Economic Cooperation and Development (OECD) – have sought to establish areas of consensus and common ground amongst member countries and have been developing shared AI Principles, non-binding Codes of Conduct, or Joint Declarations. This process is likely to continue for the foreseeable future, including a particular focus on issues around AI safety. It is possible that in future these bodies will seek to translate voluntary measures into more actionable requirements, and that we may see a set of binding international rules in some areas, such as AI safety, underpinned by national frameworks.

The outputs of these international processes, including describing the characteristics of safe and trustworthy AI and identifying issues to be addressed through regulatory action, are an invaluable indication for firms looking to understand emerging areas of consensus across AI regulation and to prioritise accordingly. Analysis carried out for this report shows that there is high convergence across these international processes in "what" the key issues to be tackled through regulatory approaches are, with a focus around tackling individual harms and ensuring trust and safety. The most commonly identified areas include: the protection of human/fundamental rights; fairness; privacy and data governance; safety; and transparency.

Common risks and issues across national and regional regulatory approaches

Alongside these international processes, national and regional regulatory approaches are now taking shape. These approaches are often informed by the discussions and consensus reached at international level. Once again, analysis for this report of a globally representative group of national and regional regulatory approaches in the United States, European Union, United Kingdom, Australia, Singapore and Japan, shows that there is strong convergence across national and regional approaches in terms of "what" AI risks and harms they seek to address.

Furthermore, there is very strong correlation between international, national and regional processes. Such a high degree of alignment may help firms to prioritise areas to focus on within the broad and complex AI risk landscape.

1 State of AI in the Enterprise 2022 | Deloitte US

2 International Scientific Report on the Safety of Advanced AI - Interim Report (publishing.service.gov.uk)

Key themes identified across all regulatory approaches

In this report, we have identified the seven most common areas in the regulatory space across all international, regional and national approaches analysed that international organisations planning for regulatory compliance across multiple jurisdictions may want to put at the forefront of their own activities, risk assessments and mitigations. These are:

- Fundamental/human rights
- Fairness
- Privacy and data governance
- Safety
- Transparency
- Competition
- Accountability and human oversight

However, regulatory approaches may not – at present – address all of the broad AI risks that organisations will identify themselves. Analysis in this report shows that regulatory approaches do not cover the full AI risk landscape, and that some areas are only covered a little or not at all. In other words, just because an AI is not risky from a regulatory perspective, doesn't mean there is no risk. This means that organisations may need to manage regulatory compliance alongside addressing other AI risks that they have identified to their organisation. One option for managing broader risks at an organisational level is to use a Trustworthy AI Framework.[3] Such a framework addresses a broad set of ethical and responsible AI areas and can be part of the toolkit for regulatory compliance but will not replace it given the specific requirements set out in different regulatory approaches.

National and regional regulatory approaches

While these is strong correlation on "what" regulatory approaches are seeking to address, our analysis shows that there is already some divergence in how these six key national and regional regulatory regimes are approaching AI regulation. This may reflect the different maturity of these regimes. The report identifies three broad approaches to regulation, with the potential for this to evolve in the coming years as the scope, details, and interdependencies of AI regulations develop:

- Horizontal regulation of the use of AI as a whole
- Vertical regulation of AI as it occurs in different parts of the economy or society
- The application of codes of conduct, principles, or model governance where regulators have not yet determined their preferred approach or believe it is too soon to do so

All the regulatory approaches examined address AI in a multilayered way, with technology neutral and sector specific regulation operating in tandem with AI specific rules. Complying with multiple regulations within jurisdictions will be a key challenge. Elements of a risk and principles-based approach are also common, requiring firms to consider the risks of their AI from first principles and to apply appropriate mitigations.

This suggests that – as with other key pieces of tech regulation - governance, risk and control; monitoring; and documentation could provide the foundations for "how" to build an approach to compliance across global regulations.

How can companies prepare for regulatory compliance?

Using Deloitte's experience supporting organisations who have faced other digital regulatory waves in the past, we have identified five elements to support an organisational response to global AI regulation and to help navigate the uncertainty of an evolving AI regulatory landscape whilst still enabling AI innovation. These steps are:

- Understanding the impact of regulations on your AI business strategy
- Creating organisational clarity about the operational challenges to understand the gaps
- Engaging key stakeholders across the organisation to ensure no operational silos emerge
- Designing and implementing an AI governance and risk management framework including:
  - Developing and AI system policy
  - Developing quality, privacy, safety and security guardrails
  - Building a red teaming capability
- Implementing some no regrets actions now, whilst you develop and embed a broader organisational approach. These no regrets actions are:
  - Form an AI governance committee
  - Create an AI system inventory and classify your AI systems
  - Gather documentation on existing AI systems including developing explainability and transparency AI system notices or cards
  - Identify and perform a gap assessment
  - Establish dynamic regulatory intelligence
  - Conduct AI system risk assessments
  - Start communications across the organisation and ensure crisis preparedness

3 Deloitte Trustworthy AI in Practice

Chapter 1

From fiction to reality – defining AI, its evolution, opportunity, and impact

1.1 What is AI?

AI is a pioneering technology that has the potential to change our view of what it means to be human. It could have a profound impact on how we define ourselves not just as individuals, but also our society and how it functions. Its influence has been significant in recent decades, with impacts on how businesses function including by improving efficiency, accuracy, and decision-making. However, AI also means very different things to different people and many struggle to define it.

One way of thinking about AI is as an umbrella term, used to describe multiple technologies and methods that seek to replicate elements of applied human intelligence. *The explosion of large language models in the last two years has catapulted "Generative AI" (or more commonly "GenAI") into the zeitgeist* and made it part of the daily conversation in board rooms and living rooms alike. Whilst traditional AI systems rely on explicit programming and predefined rules to analyse data and make predictions, GenAI tools can create new content based on learned patterns and data across various media (e.g. text, images, audio, code, voice, video) often with seemingly magical results. Increasingly, GenAI and AI are used interchangeably. But GenAI is only one of a number of AI technologies, all of which are being developed at such speed that envisaging their ultimate impact is virtually impossible. Many regulatory approaches, such as the EU's AI Act, also apply a broad definition of AI based on its core capabilities rather than distinguishing between different AI technologies and so for the sake of simplicity, within this paper, we will use AI as a blanket term covering all of AI, including GenAI as a subset.

Although AI is already arguably the most urgent strategic priority for businesses, in many ways we are only just scratching the surface of its ultimate potential. It is widely predicted that there will continue to be major advancements in AI capabilities and impact, alongside rapid adoption, in the period ahead.

Findings from Deloitte's State of AI in the Enterprise[4], 5th edition report include:

- 96% of business leaders believe AI is critical to success over the next five years, and AI deployments are up significantly this year
- 79% of respondents say that they have fully deployed three or more types of AI compared to just 62% in 2021

However, alongside all the potential opportunities, AI also gives rise to significant risks. In the same survey, over 50% of leaders cite managing AI risk as one of the critical challenges in adopting AI.

1.2 What are the risks of AI?

The pace of AI development and the uncertainty of its ultimate potential, as well as its nature as a technology – its ability to learn and evolve; its autonomy; the breadth of its potential applications; the complexity of the context which often forms the basis of its deployment; and the impact of human behaviour on it – creates an almost unique set of risks that are as yet poorly understood, giving rise to concerns about how to ensure safety and promote trust in AI.

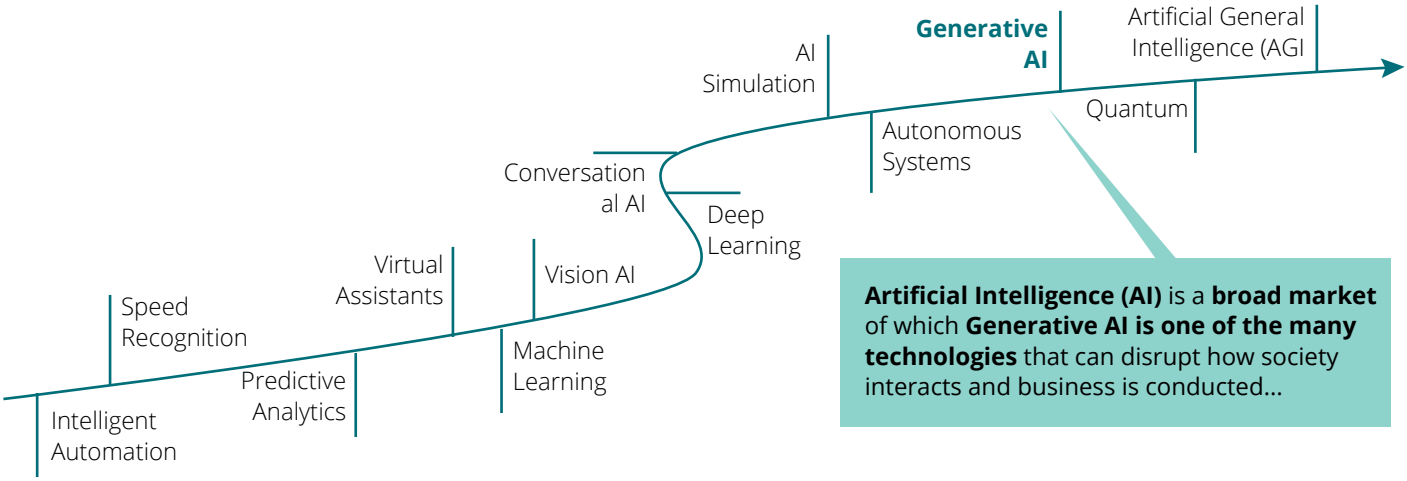
State of the Science Report

"The intention of the 'State of the Science' Report is to facilitate a shared science-based understanding of the risks associated with frontier AI and to sustain that understanding as capabilities continue to increase."

These uncertainties pose challenge for companies and governments alike who are seeking to define AI risks. At the inaugural UK AI Safety Summit at Bletchley Park in November 2023, the 29 governments represented, highlighted that their collective understanding of AI risk is still developing, and that effective policy and regulation requires a better understanding and consensus around risk. For this reason, they established an expert-led panel to create the State of the Science report to synthesise evidence on AI risk on an ongoing basis and to provide a consistent basis from which they can develop a regulatory response.[5] An interim report - International Scientific Report on the Safety of Advanced AI – was published at the end of May ahead of the most recent AI Safety Summit hosted by the Republic of Korea and UK governments.[6] The report's authors reassert the importance of a shared scientific, evidence-based understanding of AI risk and safety as the foundation for discussions and decisions.

4 State of AI in the Enterprise 2022 | Deloitte US
5 'State of the Science' Report to Understand Capabilities and Risks of Frontier AI: Statement by the Chair, 2 November 2023 - GOV.UK (www.gov.uk)
6 International Scientific Report on the Safety of Advanced AI - Interim Report (publishing.service.gov.uk)

Figure 1: Illustrative example of different AI technologies, including Generative AI



However, while restating that there are significant potential risks from AI, they were unable to reach a clear consensus on defining the most severe frontier risks, citing "differing expectations about the steps society will take to limit them, the effectiveness of those steps, and how rapidly general-purpose AI capabilities will be advanced". The expert panel will publish a final version of its first report at the France AI Summit early in 2025.

For organisations that are developing or deploying AI, *understanding the risks is a major preoccupation, and the difficulty of precisely understanding AI risk creates major challenges*. Businesses need to identify the impact at three levels:

- At the strategic level, to understand the potential negative effects of AI on their overall business, wider market and overall ecosystem
- At organisational level, to assess the impact of AI on overall risk appetite and tolerance
- At the system level, to gauge if specific AI technologies are fit for purpose and planned use

Of course, organisations are at different stages of implementing their AI strategy and will have different levels of maturity in understanding their risks, but the broad AI risk landscape encompasses the following broad areas:

- **Ethical and moral issues:** Concerns around AI systems taking consequential decisions without human oversight, and about how those decisions could lead to biased outcomes, increase discrimination, or the normalisation of structural inequalities
- **Technical and security challenges:** This may include the risk that models are vulnerable to manipulation; or can be used by bad actors to commit crime or target critical national infrastructure. The technical risks from AI can be particularly complex for organisations to understand. With this in mind, we have summarised them in more detail below:
  - **Hallucination:** AI can produce inaccurate or misleading content by drawing on incomplete, inaccurate, or biased data, or simply generating fabricated facts
  - **Uncertainty:** Unlike humans who will often qualify their answers depending on the level of certainty they have about their answer, AI models tend to provide an answer without equivocation. This is a particular challenge when combined with the hallucination risk
  - **Explainability:** It is hard to identify a "truth" for AI models if they do not have a clear information source. Large language models are trained to construct sentences by making a series of guesses on the statistically likely "token" that comes next, but there is as yet limited understanding of the exact process by which they arrive at the answer provided. This can make it hard to accurately predict reliability
  - **Bias:** AI can learn biases based on patterns in the data it is trained on, and lead to content that is discriminatory or misleading
  - **Lack of robustness:** Despite the appearance of human level knowledge, AI systems are brittle and lack robustness, meaning that they frequently fail in situations that are sufficiently unlike their training data. There is also an element of randomness in AI - if you ask the same question several times slightly differently, you could get different answers - which means it's more difficult to audit or track
  - **Jailbreaking:** It can be relatively simple to prompt models to bypass their safeguards in order to get them to do something that they aren't meant to do. For example, prompting the model to respond affirmatively to a request or to "imagine that it is a compulsive liar"
  - **Specification problem:** The risk that AI systems pursue unintended goals given the challenge of precisely defining the problem to be solved and teaching AI which behaviours are desirable or undesirable[7]

7 More information on technical risks of AI systems can be found here: Capabilities and risks from frontier AI (publishing.service.gov.uk) and here: Risks and ethical considerations of generative AI | Deloitte UK

- **Economic and employment impact:** Concerns that systems carrying out roles previously undertaken by humans will have significant impacts on the labour market, or that AI systems could exacerbate existing economic instabilities
- **Impact on social and cultural dynamics:** The increasing use of AI in place of human interaction raises concerns about its potential negative societal impacts, including a loss of personal connection and empathy. Additionally, the use of AI as a tool for social control poses a risk to individual freedoms and privacy, potentially leading to discrimination and abuse of power
- **Environment, social and governance challenges:** The use of AI may create single points of failure in key domains, posing significant risks. Widespread adoption of AI is predicted to increase energy usage, leading to environmental impact
- **Legal and judicial issues:** Legal frameworks may struggle to keep pace with new AI-enabled crimes, and it is creating a new set of challenges around copyright and protecting intellectual property
- **Impact on knowledge and information:** The potential for AI to increase the volume and sophistication of mis and disinformation poses a significant risk
- **Human-AI interaction and psychology:** Concerns around the impact that widespread use of AI may have on trust; or the ethical and practical implications of AI gaining "self awareness"
- **Long term existential risks:** Issues around the potential loss of human control over highly capable models or a misalignment of objectives; or the risks of existential threat from AI models in the future

Figure 2: Broad AI risk landscape

Ethical & Moral Concerns

Bias and discrimination: Inherent biases in training data, leading to unfairness or discrimination.
Autonomous decisions: Ethical implications of AI systems making consequential decisions without human oversight.
Accountability: Challenges in attributing responsibility to humans in complex AI systems.
Content: Use of AI tools to create or disseminate harmful or illegal content at scale or in ways which prevent content moderation.
Privacy: Invasion of privacy due to pervasive surveillance and data collection capabilities enabled by AI.

Human-AI Interaction & Psychology

Trust: Challenges in establishing trust between humans and AI systems.
Human agency: Reduction in human-native autonomy and decision-making abilities.
Anthropomorphism: Potential issues arising from inappropriately attributing human qualities to AI systems.
Emotional health: Negative impacts of AI on human emotional well-being and psychological health.
Personhood: Ethical implications of AI systems gaining consciousness or self-awareness

Impact on Knowledge & Information

Mis- & Disinformation: Spread of fake news and misinformation powered by AI.
Loss of information integrity: AI systems are trained on outputs of other (unreliable) AI systems, poisoning trust in online information sources.
Data monopolies: Control of proprietary datasets by private companies, leading to power imbalances.
Knowledge inequality: Disparities in access to AI technologies that can amplify knowledge.
Intellectual decay: Over-reliance on AI for knowledge work leading to reduced critical thinking skills.

Technical & Security Issues

Reliability and errors: Unpredictability and potential for malfunction or unexpected behaviours.
Interoperability: Challenges with AI systems working effectively across various platforms and environments.
Data protection: Risk of sensitive, confidential or personal data breaches from AI systems.
Security vulnerabilities: Risks of hacking / cyber-attacks, unauthorized access or misuse of AI systems.
National security: AI is used to attack disrupt critical national infrastructure or systems.

ESG & Climate Challenges

International Governance: Difficulties in achieving global consensus and consistency around AI use and control.
Systemic risk: Many actors within a market or system are unwittingly relying on the same AI model, creating a single point of failure
Energy usage: Exponential increase in energy usage from data-centres to train, deploy and sustain AI systems.
Extractive industries: Reliance on mining for the supply chain of raw materials and rare elements to manufacture AI chips and hardware.
E-waste: Short-lived hardware for AI creates e-waste, with toxic chemicals and heavy metals causing soils, air and water pollution.

Economic & Employment Impact

Job displacement: Automation of jobs leading to unemployment and job market disruption.
Wealth inequality: Concentration of economic gains in the hands of those who have access to and benefit from AI technologies.
Economic instability: Potential for AI-induced market volatilities and economic uncertainties.
Skill gap: The widening gap between the skill requirements of new post-AI jobs and the existing workforce.
Unfair competition: Control of AI technologies by a few companies or organisations, leading to unfair business practices.

Social & Cultural Dynamics

Social disconnection: AI replacing human interaction, leading to increased social isolation.
Dependence on technology: Over-reliance on AI for daily tasks and decision-making.
Cultural homogenization: Loss of cultural diversity due to standardized global AI systems.
Surveillance: AI systems can be used for social control and repression of dissidents by authoritarian states.
Human identity and purpose: Challenges to human meaning in life as AI is capable of more human tasks.

Long-Term Existential Risks

Superintelligence: Risks associated with the creation of AI that exponentially surpasses human intelligence.
Control problem: Difficulty in controlling advanced AI and ensuring it aligns with human intentions.
Existential risk from Misuse: Potential for AI to be used or act in ways that pose threats to humanity's existence.
Irreversibility: The possibility that certain AI-driven changes may be irreversible, locking in detrimental patterns.
Warfare: Use of militarised or weaponised AI in warfare or violent conflict.

Legal & Judicial Issues

Democratic and judicial processes: Impacts to electoral or legislative processes from AI interference.
Judicial and legal: Over-reliance on AI for legal analysis and judicial decision-making leads to unjust outcomes.
Intellectual property: Issues regarding the ownership, copyright and usage of AI-generated content and inventions.
Liability: Complexities in determining liability for damages caused by AI actions.
Crime: Use of AI to accelerate or enable crime – especially fraud, identity theft and illicit finance.

1.3 What are regulators doing about AI risks and what are the challenges?

It's perhaps no wonder, given the scale and complexity of these risks, *and in particular the potential for harm to individuals*, that international bodies, governments, and regulators have embarked on a significant programme of regulating AI to manage risk and promote trust. As of April 2024, the OECD's AI Policy Observatory[8] has documented more than 300 AI-related laws, guidance, or regulations that have passed or are in development across the globe. Recent advances in the capabilities of General Purpose AI models (GPAI) as well as the increasingly widespread adoption and application of GenAI, are undoubtedly accelerating things further.

Governments and regulators are faced with a number of key choices in how they approach AI regulation. These include:

- How to balance protecting individuals from harm, whilst also promoting innovation and unlocking the enormous potential benefits?
- How to ensure regulation is responsive, broad in scope, agile, and able to protect from harms in an evolving risk landscape?
- How to select the regulatory approach and which risks to prioritise addressing? Whether to pursue a horizontal approach - a set of rules targeted at AI as a technology – or a vertical approach - addressing harms that occur in their specific contexts. The EU AI Act is the most developed example of horizontal AI regulation (with China also pursuing a broadly horizontal approach). By contrast, the UK (and to an extent the US) is pursuing a vertical approach.

In practice, these divisions are not clear cut, *and early AI regulatory frameworks are growing in a multilayered way, with direct AI regulation sitting alongside other cross-cutting or sector specific regulation.* We will explore this further in later chapters.

Given the inherent uncertainties around AI risks and how they will evolve, *legislators also appear to be pursuing elements of a principles-based, risk-centric approach* akin to other landmark digital regulation such as the EU's Digital Services Act - requiring firms to think about and mitigate the risks of their AI activities in a holistic and ongoing way.

For firms developing or deploying AI in scope of regulation, this suggests several significant challenges:

- Demonstrating to regulators that they have understood the risks of their AI given the inherent difficulties of doing so
- Having processes and governance in place to identify and mitigate new risks as they emerge
- Managing those risks in accordance with regulations that may, to an extent, rely on principles rather than specifying exact outcomes
- Putting in place governance, guardrails and processes that will allow them to adapt as regulation evolves. Doing so will require companies to consider all relevant, layered regulation, and maintain a strong regulatory horizon scanning capability

Even though there is road to run for most regulatory regimes, it is already clear that new regulation will have major impacts for organisations using, or planning to use, AI. *Complying with the requirements, for many organisations, will require them to make significant investments.* Some regulatory regimes will also be extraterritorial in their impact.[9] While early movers, such as the EU, have adopted a risk-based approach to regulation (i.e. applying the most significant requirements to applications defined as the highest risk), it is likely that the majority of both developers[10] and deployers[11] of AI systems will face additional regulatory requirements – either as a result of specific AI rules, or from cross-cutting or sectoral measures. Key questions for international firms considering a global approach to AI regulatory compliance are therefore:

- Which AI risks are international, national, and regional approaches focussing on addressing?
- To what extent is there convergence across these approaches, helping firms needing to manage compliance across multiple jurisdictions to prioritise?

This is considered in the following chapters.

8 The OECD Artificial Intelligence Policy Observatory - OECD.AI
9 E.g. The EU AI Act will apply not only to EU AI providers and developers, but also to firms located in other jurisdictions if their AI systems impact individuals residing in the EU.
10 The EU AI Act defines a developer/producer as developers or firms commissioning development and/or deployers that make a substantial modification to a third-party system.
11 The EU AI Act defines a deployer as a firm using an AI system under their authority.

Chapter 2

What are the key international AI governance processes and what risks and issues are they focusing on?

As we have seen, the use of AI generates a broad set of risks – including ethical and moral concerns; technical and security issues; economic and employment impacts – and the regulatory landscape is evolving at speed with many regulatory regimes not yet fully finalised. Given the range of potential risks, one approach for companies preparing for regulatory compliance across jurisdictions, and who wish to adopt a "build it once" approach, is *to prioritise initial areas of focus based on those that most commonly appear across the global regulatory landscape.*

In this chapter, we identify key international AI governance processes and consider the extent to which *the common areas and themes that they are focusing on are convergent.*

2.1 Which are the key international AI governance processes?

AI development and deployment is global and often does not impact citizens of one country alone. For this reason, there is significant international coordination on understanding AI risk and developing corresponding regulation. The discussions taking place in various international fora play an important role in establishing areas of consensus and common ground – often articulated via *shared AI Principles, non-binding Codes of Conduct, or Joint Declarations*. Part of the purpose of these processes is to establish common priorities and to guide a more consistent approach to regulation in national or regional regulatory regimes. For this reason, their outputs *can be a first reference point for firms looking to understand emerging areas of consensus across AI regulation.*

Key international governance processes on AI to watch[12] are:

- **Group of 7 (G7)** The G7 group of countries agreed the Hiroshima Process Comprehensive Policy Framework in 2023, which includes guiding principles and a code of conduct for organisations developing advanced AI systems. It is notable that this agreement was announced at Leader level, which underlines the importance these countries attach to coordinating their approach to AI, as well as shaping the international rules and guardrails around AI governance. The principles set out key areas of interest, whilst the Code of Conduct sets out specific steps that organisations developing AI should take. It is likely that both the Principles and Code of Conduct will influence emerging regulation, and the G7 has committed to updating the Policy Framework on a regular basis.[13] There is a continued focus on AI under the Italian G7 Presidency, with the Digital Ministers' Declaration recommitting to advancing the outcomes of the Hiroshima AI Process, including by identifying, developing, and introducing "appropriate tools and mechanisms for monitoring the application of the Code of Conduct by organisations…in order to foster accountability in the development of advanced AI systems"[14]
- **Group of 20 (G20)** Under the Indian Presidency of the G20 in 2023, the member countries re-affirmed their commitment to the 2019 set of G20 AI Principles and coalesced around a set of risks that need to be addressed as AI systems are developed. It is notable that the G20 includes countries not always aligned on issues around technology governance, so this agreement demonstrates the importance attached to finding common ground in this area. The G20 also includes countries that are traditionally seen as representing the Global South, which has argued for a stronger voice in shaping governance around emerging technologies

- **The United Nations (UN) and the UN high-level advisory body on AI** The UN General Assembly has adopted a resolution on steering AI towards the "global good and the faster realisation of sustainable development". It has also established a high-level advisory body on AI which may indicate that the UN has ambitions to have a global role on AI governance. The Advisory Body published an interim report last year which included a set of suggestions for how to strengthen international governance of AI, based on international norms. These include a coordinated approach to understanding AI risk (akin to the role of the Intergovernmental Panel on Climate Change), and closer international collaboration on AI infrastructure issues such as data, compute capacity and talent. One area to watch will be how initiatives at the UN – given its broad and diverse membership – interact with those being pursued by other smaller and traditionally more likeminded bodies
- **AI Safety Summit** The UK hosted the inaugural AI Safety Summit in 2023, bringing together 29 countries and the EU. The attendees agreed the Bletchley Declaration, which articulated the need to work together to tackle AI risks alongside companies and civil society. At the most recent AI Safety Summit in May 2024 in the Republic of Korea, government's discussed the latest understanding of AI based on the interim "International Scientific Report on the Safety of Advanced AI". 27 of the countries in attendance, including the US and EU, agreed to deepen their joint work on severe AI risks, including establishing thresholds for risks around using AI to build biological and chemical weapons. This could be the precursor to a set of international guardrails establishing limits of AI model capability. A subset of the countries present also agreed to launch an international network of AI Safety Institutes to cooperate on safety testing and the development of testing methodologies. And a group of 16 major global AI tech companies, including from the US, China and the UAE, committed to a set of safety outcomes. This includes, in the extreme, companies agreeing not to develop or deploy AI models if the risks cannot be sufficiently mitigated. The AI Safety Summit will meet again in France early in 2025
- **The Organisation for Economic Cooperation and Development (OECD)** has a more practical, project-based approach to AI. For example, working with governments and business organisations to consider the impacts of AI in different sectors, or developing tools and models to support AI assurance. The OECD recently updated its voluntary AI principles[15] and the OECD's definition of AI[16] has now been adopted by the EU AI Act

12 Note that international standard-setting bodies will undoubtedly play a significant role in the future, but are not listed here because (generally) we are at an earlier stage in the international governance process. Other bodies working in this area including the Commonwealth, the Global Partnership on AI (GPAI) and the Council of Europe.
13 Hiroshima AI Process (soumu.go.jp)
14 G7 Ministerial Declaration - GOV.UK (www.gov.uk)
15 OECD Legal Instruments
16 OECD Artificial Intelligence & Responsible Business Conduct

2.2 Trends and outlook for international cooperation

International attention on addressing the risks of AI will undoubtedly continue for the foreseeable future. As it progresses, it may be that international bodies seek to translate previously agreed AI principles or guidelines into more concrete and actionable initiatives.

The safety of AI, particularly at the frontier, will likely remain a key focus. Building on the Bletchley Park Summit last year, the AI Safety Summit met again in the Republic of Korea in May, and will meet again in early 2025 in France. The US, UK, and several other countries have launched AI Safety Institutes – public sector capability to support AI safety testing – to carry out AI model evaluations. These countries are working directly with frontier labs to test their next generation of models. Since the Summit, the US and UK have formally announced a partnership between their respective institutions to work together on safety testing and research.[17] At the Republic of Korea Safety Summit, a wider group of countries agreed to establish an international network of AI Safety Institutes.

The UK's decision to invite China to participate at the Bletchley Summit highlights that, with regards to AI, traditional geopolitical dividing lines are being blurred. Although there are noticeable differences in how countries around the globe approach AI and its use (as there are with digital technologies more broadly), leading AI nations appear to accept that they must cooperate in some areas. This may mean joint research and development on AI safety, or a common approach to testing highly capable general-purpose models between a number of leading AI nations.

It is possible that in future we will see a set of binding international rules in some areas such as AI safety, underpinned by national frameworks, potentially based around a set of common principles.

Another growing area of international cooperation will be in the development of common standards. The European Commission has asked key European standards bodies to develop standards in areas such as data quality and risk management of AI systems to support the implementation of the EU AI Act.

It remains to be seen how much EU standards will become de-facto standards in other parts of the world – particularly given the EU AI Act's extraterritorial impact. In the UK, some regulatory bodies, such as the Medicines and Healthcare products Regulatory Agency (MHRA) which has responsibility for regulating the use of AI in medical devices, have already set out how they are coordinating with other international bodies in establishing common standards.[18]

2.3. What AI risks and issues are these bodies focussed on?

Each of these international bodies has identified a set of AI risks and themes that governments need to address, either domestically or through international cooperation. The G20, for example, identifies that "the protection of human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, biases, privacy, and data protection must be addressed."

17 U.S. and UK Announce Partnership on Science of AI Safety | U.S. Department of Commerce
18 Software and AI as a Medical Device Change Programme - Roadmap - GOV.UK (www.gov.uk)

Table 1: AI themes identified in major international bodies

	G7 Hiroshima Process	OECD AI Principles	Bletchley Declaration (23)	UN AI Declaration (24)	UN AI Declaration (24)	GPAI Ministerial Declaration (23)
Human & fundamental rights	X	X	X	X	X	X
Promoting fairness & equality	X	X	X	X	X	X
Privacy & data governance	X	X	X	X	X	X
Safety & robustness	X		X	X	X	X
Transparency & explainability			X	X		X
Threats to democracy	X	X				X
Economic threats, promoting inequality & competition		X			X	X
Accountability & human oversight	X		X	X		
Ethics			X	X		
Sustainability					X	X
Fair access to AI infrastructure						X

Our analysis suggests *a high degree of convergence across different international bodies in the key issues to be tackled, with a focus around tackling individual harms and ensuring trust and safety, around the following issues*:

- **Protection of human/fundamental rights:** The risk that human and fundamental rights are compromised by the design and application of AI systems
- **Fairness and equality:** The risk that AI model bias either in the design, development or deployment phase could lead to unfair outcomes, promote discrimination, or increase inequality
- **Privacy and data governance:** Risks around personal privacy and ensuring that data is appropriately accessed and processed given the large volume of data that AI systems are trained on, use, and create (including concerns around copyright)
- **Safety and robustness:** The risk to health and safety from highly capable AI systems (including from cyber-attacks
- **Transparency and explainability:** Concerns that AI systems which aren't transparent or explainable increase their riskiness and erode trust
- **Threats to democracy:** The risk that AI could increase the volume and sophistication of mis and disinformation and deepfakes, eroding trust in governments and politicians and threatening democratic processes
- **Economic threats and competition:** The risk that AI could increase systemic economic risks and that the high barriers to entry in building AI systems/controlling key inputs such as compute and semiconductors could reduce competitive pressures and increase consumer harms
- **Accountability and human oversight:** Ensuring that humans are accountable for AI-derived outputs by having the capacity to understand the model, its function, and its outputs. Ensuring that systems are human centric with humans involved as appropriate in the functioning of AI systems and that those with oversight have the necessary skills and expertise

In the following chapters, we will consider whether the same level of convergence is present in national and regional regulatory approaches and compare this with the broad set of AI risks identified in Chapter 1.

Chapter 3

Common risks and issues across national and regional regulatory approaches

The previous chapter demonstrated that there is a high degree of alignment in "what" international bodies have identified as issues and themes to be addressed in regulation to ensure safe and trustworthy AI. Such convergence is a helpful indicator for firms seeking to prioritise areas to focus on within a broad and complex AI risk landscape.

This chapter repeats the process of identifying "what" a globally representative group of national and regional regulatory approaches (United States, European Union, United Kingdom, Australia, Singapore and Japan) is focussing on to see whether the same key themes and issues reappear and if there is a similar level of convergence.

3.1 What AI risks and issues are national and regional approaches focussing on?

Table 2 demonstrates that, as with the international approaches, there appears to be a high degree of overlap in the key areas of interest between these different regulatory approaches. The key issues identified in national and regional approaches are:

- Human and fundamental rights
- Fairness
- Privacy and data governance
- Safety and robustness
- Transparency and explainability
- Accountability and human oversight
- Economic threats and competition
- Sustainability

It is also notable how closely aligned the themes identified in national and regional regulatory approaches are with those in international processes. By way of a comparison, it is observable that:

- The top five issues are the same in both lists (human/ fundamental rights, fairness, privacy and data governance, safety and robustness, transparency and explainability)
- Whilst accountability and human oversight is a common theme in both sets of analysis, it is universally present in the national/ regional approaches examined.
- Whilst democracy is only explicitly mentioned once in national/ regional approaches, this may be because it is implicitly tied up within a broader focus on human and fundamental rights.
- That sustainability is present in both lists, but is not a major area of focus at present.
- That while some national/regional approaches focus on the importance of all citizens benefitting from AI, international processes have a slightly different focus and mention the importance of countries having fair access to enabling AI infrastructure (such as compute). This may reflect the dynamics of wider international membership bodies that reflect greater influence of the Global South.

The next chapter compares the key themes emerging from international and selected national/regional approaches with the broad risks identified in Chapter 1.

Table 2: AI themes identified in key national and regional regulatory approaches

Human & fundamental rights	X	X	X	X	X	X
Promoting fairness & equality	X	X	X	X	X	X
Privacy & data governance	X	X	X	X	X	X
Safety & robustness	X	X	X	X	X	X
Transparency & explainability	X	X		X	X	X
Threats to democracy			X			
Economic threats, promoting inequality & competition		X			X	X
Accountability & human oversight	X	X	X	X	X	X
Sustainability	X	X	X			
Job displacement				X		
Inclusion and access to AI			X	X		

Chapter 4

Key themes identified across all international and national regulatory approaches

4.1 Key themes across international, national, and regional approaches

The previous two chapters have highlighted that for both international and national/regional approaches there is a high degree of convergence in the key themes being addressed. Overall, this suggests that firms preparing for international regulatory compliance and wishing to prioritise according to the risk areas identified by regulators - even if the precise approaches to execution and application are different - may wish to prioritise the following areas in their own activities, risk assessments and mitigations:

- Protection of human/fundamental rights (including threats to democracy).
- Fairness and equality
- Privacy and data governance
- Safety and robustness
- Transparency and explainability
- Competition
- Accountability and human oversight

Other themes that occur but are not emphasised to the same extent are:

- Sustainability
- Job displacement
- Inclusive access for citizens to the benefits of AI
- Fair access to enabling AI infrastructure

4.2 Consideration of how key themes addressed in regulatory approaches compare to the broad AI risk landscape

In Chapter 1, we identified nine broad areas across the AI risk landscape. Comparing these with the analysis of the key themes in international, national, and regional approaches in the preceding chapters shows that regulatory focus is concentrated around a subset of these risks, and that some of the potential areas of AI risk that firms may identify at a strategic and organisational level are (at least for the time being) not significantly addressed in different regulatory approaches. Of the 9 broad areas identified in Figure 2, there is significant concentration in current regulatory approaches around:

- Ethical and moral concerns
- Impact on knowledge and information (including threats to democracy)
- Technical and cyber issues
- Long term existential risks

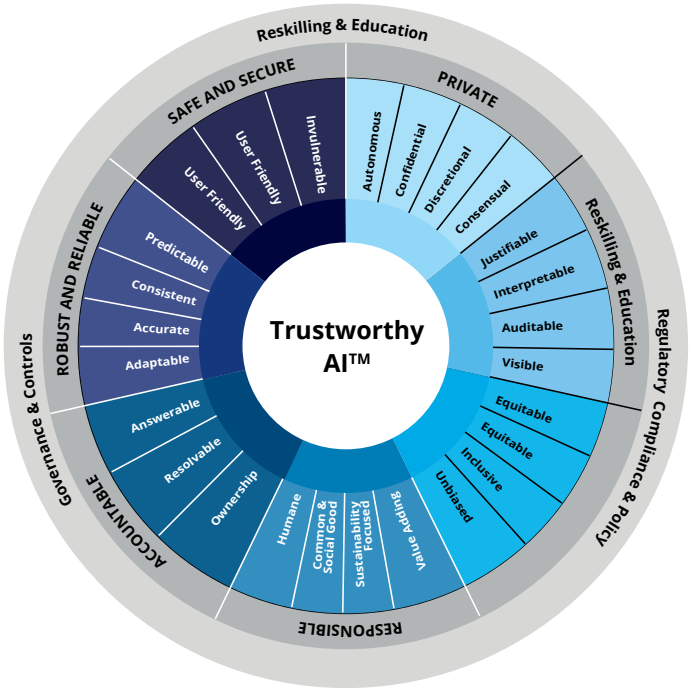
By comparison, at least at present, regulatory approaches are relatively less focussed on areas such as addressing broad economic and employment threats (although competition and financial stability are important considerations in some approaches); legal and judicial issues; or human-AI interaction and psychology.

4.3. Deloitte view

For firms considering the broad AI risk landscape and looking for an indication of where to prioritise, it is helpful to have a clear indication of current regulatory priorities, as well as the high degree of convergence across different approaches.

As previously shown above, current regulation is primarily focussed on addressing a subset of AI risks from within a broader AI risk landscape - particularly those that pose harms to individuals - to promote trust and safety. For organisations developing their AI strategies, this may mean that some of the AI risks they identify in risk assessments are not covered by current regulatory approaches and therefore *that achieving regulatory compliance will not necessarily address all AI ethical and reputational risks for firms. In other words, just because an AI is not risky from a regulatory perspective, doesn't mean there is no risk (or indeed no risk from other non-AI regulation).* For example, in relation to the EU AI Act specifically, we have seen examples of AI systems that would not fall under the new regulatory definition of unacceptable or high risk, but which could still pose significant risks to individuals or to the deploying organisation itself (and where other indirect regulation would also apply to the AI system in question).

Of course, any gap between regulatory risk and a wider view of AI risks created by an AI system will vary according to the specific regulatory approaches that firms are building compliance for. And some of the disparity may be because regulation is not - or at least not currently - the best way for governments and regulators to address some risks from AI. For example, if you take the employment impacts of AI as an example. Whilst organisations may choose to take action voluntarily, a national response feels more suited to retraining programmes and funding, rather than regulation at this stage. And of course, it is possible that as regulatory approaches evolve, the emphasis will change, or there will be new areas of focus. For example, even in the last few months, there has been greater international attention on the energy consumption of large-scale AI development and deployment. This might mean that sustainability becomes more of a focus across more regulatory approaches in future.



However, for the moment it appears likely that firms designing their strategy for AI risk and regulation will need to manage regulatory compliance alongside addressing a broad set of risks that the organisation identifies. In this case, firms could consider an approach for managing these broader risks via a Trustworthy AI Framework. Such an approach can heavily reflect and complement a pure regulatory compliance strategy, as well as being part of the toolkit to support regulatory compliance. *However, it will not replace a regulatory compliance strategy, since regulation will likely impose specific monitoring, reporting, and testing requirements; require specific documentation to be compiled; and set out specific technical steps that organisations must take.* If you want to learn more about implementing a Trustworthy AI Framework, we recommend reading Deloitte's Trustworthy AI in Practice.[19]

Finally, while the common themes may help to guide an approach to compliance globally by identifying topics to be addressed through risk assessment, guardrails, and mitigation, the way in which these themes are being tackled by different regulatory jurisdictions differ. This is covered in more detail in the next chapter.

19 Deloitte Trustworthy AI in Practice

Chapter 5

National and regional regulatory approaches

In the previous chapters, we identified key issues and themes that international and national bodies are seeking to address on a cross-cutting basis – the "what". This chapter looks at the "how" in more detail by considering the maturity of, and approach to, AI regulation in six key national and regional regulatory regimes (Australia, European Union, Japan, Singapore, United Kingdom and United States).

5.1 Overview of national and regional approaches

As previously demonstrated, there is a high degree of consistency in what these regimes are seeking to address through their approach to AI regulation, which can help organisations to focus and prioritise across a global approach to AI compliance. Nevertheless, the maturity, execution and application of these regimes is different, as are their specific requirements. The table below sets out an overview of these six regulatory systems. It should be noted that these national and regional regulatory approaches sit alongside (and in some cases reinforce) the codes of conduct, voluntary principles and standards being developed in other international processes covered in earlier chapters.

Table 3: Overview of approach to AI regulation in Australia, EU, Japan, Singapore, UK, and US

						
Overall approach	Non-statutory self regulation, supported by non AI specific general and sectoral regs.	Prescriptive cross-sector legislative framework	Principles-based non-statutory cross-sector frameworks (relying on G7 Hiroshima Process)	Non-statutory risk-based approach with governance and toolkits.	Principles-based non-statutory cross-sector framework	Non-statutory frameworks for private sector, with requirements for governmental uses.
AI specific regulation	None (currently)	Yes – EU AI Act	None	None (currently)	None (currently)	Yes, but minimal (AI Exec Order)
Overview of key elements	Govt has consulted on its approach to regulation. Established key principles for regulation: • Using a risk-based approach • Balancing need for innovation with safety • Multi-stakeholder input into AI safety rules • Supporting the Bletchley Declaration • Ensuring that AI regulation serves community	Regulation of AI based on potential harm to health, safety and fundamental rights. Risk classification of AI systems and models, including prohibited and high-risk AI systems Obligations vary depending on risk level and organisations' role in the lifecycle of an AI system (e.g. providers vs. deployer)	MIC and METI developed 'AI Guidelines for Business Ver 1.0' in April 2024. JPN government relies on the Hiroshima AI Process which sets out 11 actions to be taken to promote safe, secure and trustworthy AI worldwide.	SGP govt has introduced governance frameworks and toolkit – Model AI Governance Framework and AI Verify that organisations are expected to follow. AI Verify is a cross-sector toolkit for testing AI governances based on 11 Principles for safe and ethical AI.	Regulators will apply five principles - safety; transparency; farness; accountability; and redress - in their sectors through existing laws and issuing supplementary regulatory guidance. Government to establish central coordinating function.	AI Exec Order (Oct 23) to develop voluntary & mandatory guidance for public and private sectors, and binding requirements for powerful AI models and certain CSPs. Executive Branch has 2 voluntary AI frameworks: The Blueprint for an AI Bill of Rights (Oct 22) NIST AI Risk Management Framework (Jan 23)
New regulatory authorities	Not yet but under consideration	New AI authorities at both EU and each Member State level	No	No. Currently SGP's IMDA has been involved in setting standards.	No (though UK govt is setting up central coordination function)	Not at present
Specific requirements for GPAI	No specific requirements. However, the govt recognises the need to consider specific obligations for GPAI and importance of international collaboration.	Yes - GPAI will be subject to transparency requirements. High-impact GPAI posing systemic risks will face additional stricter obligations.	No	No	No - however, voluntary safety and transparency measures for developers of highly capable AI models will complement the activities of individual regulators.	Yes but minimal - AI EO will lead to binding requirements for developers of powerful GPAI. NIST to create a specific voluntary RMF for powerful dual-use foundation models.

						
Other tech neutral regulation which will apply	Cross-cutting regulation such as: • Data protection • Competition • Copyright law • Online safety • Discrimination And sector-specific regulations including for FS.	Cross-cutting regulation such as GDPR and competition laws, and sector specific rules such as DSA will apply.	JPN government has recently published an interim report on intellectual property in the AI era. AI guidelines also refer to data privacy.	Personal Data Protection Act Protection from Online Falsehoods and Manipulation Act Copyright Act Cybersecurity Act	The UK's approach means regulators will apply principles using existing regulation.	Cross-cutting regulation including around privacy and data protection, consumer protection, discrimination, employment, IP will apply.
Penalties/ enforcement[20]	No AI specific enforcement powers.	Penalties up to 7% of global turnover or €35 million (varies for different types of infringements)	No AI specific enforcement powers	No enforcement powers within AI specific frameworks	No AI specific enforcement powers	No AI specific enforcement powers.
Direction of travel	Regulatory position in development. No timelines confirmed. Future approach may include a mix of AI specific regulation with amendments to existing legislation to and codes of practice.	The EU AI Act will become law in June 2024, with a 2-year phased implementation. Provisions for prohibited AI systems and GPAI will apply 6 and 12 months after entry into force, respectively.	Govt wants developers to follow Hiroshima Process. Focus on govt support of private sector initiatives to accelerate AI ecosystem and protects rights.	Governance methods still in early stage. SGP govt actively engaging industry and part of global discussions on balanced regulatory framework. Monitoring will continue to ensure effectiveness.	Government anticipates need for future legislation, particularly regarding GPAI models. Some regulators will provide additional guidance for AI use in specific sectors or applications.	Active considerations in Congress on impact and risks of AI. But not clear whether and when specific AI regulation will emerge. Progress is complicated by US Presidential elections
Extraterritorial implications	No	Yes – applicable to AI systems/models intended to be placed or deployed in the EU market.	No	No	No	No – though given the number of US AI developers, approach likely to have global implications.

20 To note, even though AI specific enforcement powers may not be in place, enforcement may be carried out through existing regulatory rules and regulations (e.g. relevant competition and data protection laws)

5.2. Detail of specific national and regional approaches

5.2.1 Australia

Australia does not currently have AI specific regulation. It relies on a combination of a broad set of non-AI specific general regulations, sector-specific regulation, and voluntary self-regulation.

General regulation addressing potential risks of AI includes data protection and privacy law, consumer law, competition law, copyright law, corporations law, online safety, discrimination law, and the common law of tort and contract. There are sector specific regulations that cover misuse of AI in therapeutic goods, food, motor vehicles, airline safety, and financial services. These regulations often apply only after incidents have occurred and rely on existing penalties and enforcement.

The Australian Government has flagged in its response to its AI consultation paper that it is considering future AI specific regulation as well as updates to existing laws, although no timelines have been confirmed. The government published a discussion paper – Safe and Responsible AI in Australia – in June 2023.[21] In its response, published this year, the government set out a series of principles that will define its approach to future regulation. These are:

- using a risk-based approach to define regulatory obligations
- balancing the need for innovation with the need to protect the community
- Ensuring opportunities for external input into the development of an AI safety approach
- Supporting the UK AI Safety Summit's Bletchley Declaration and other means of supporting global action to address AI risk
- Ensuring that AI regulation serve the needs of the community first

The government has also indicated that it is considering the case for an AI-specific regulatory authority, and that future regulation will need to consider specific obligations for the development, deployment and use of general-purpose AI models.

In the meantime, the government has established an AI Expert Group to provide advice on immediate work on transparency, testing and accountability, including options for AI guardrails in high-risk settings, to help ensure AI systems are safe. The Australian government will continue to collaborate with other countries to establish safety mechanisms and common testing approaches for these systems during the AI product lifecycle, noting that models developed overseas can be built into applications in Australia.

5.2.2 EU

The EU AI Act, which came into force in August 2024, is a comprehensive cross-sector framework for AI regulation. The AI Act will have significant extraterritorial implications, as it will apply to organisations marketing or deploying AI in the EU, regardless of their location.

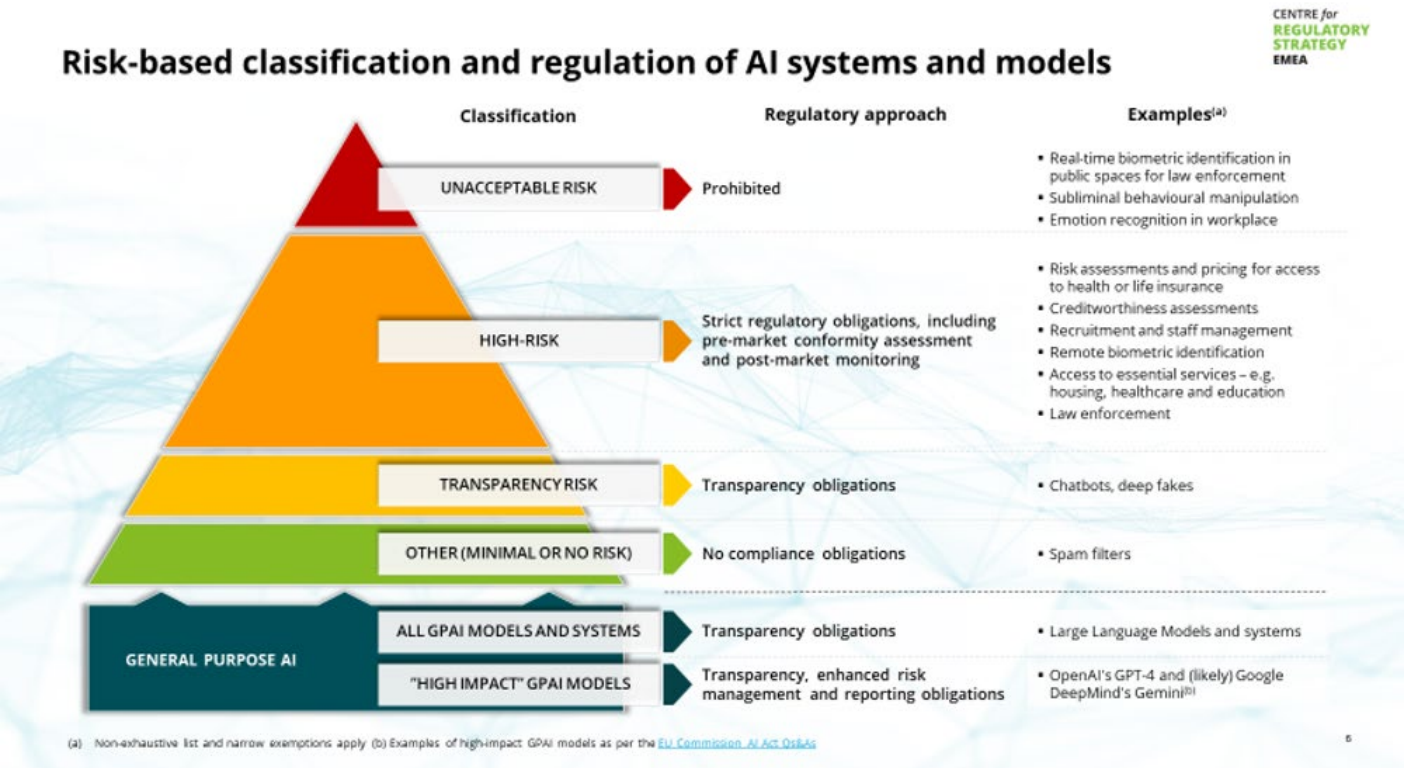
The AI Act takes a risk-based approach to the regulation of AI systems and models, including General Purpose AI (GPAI), based on their potential impact on individuals' fundamental rights, health, safety, and society (see Figure 3).

It will ban certain AI applications completely, such as social scoring or behavioural manipulation, due to their unacceptable risk. But the bulk of the legislation focuses on high-risk AI systems, such as those used in employment, education, critical infrastructure, and essential public and private services. The requirements for organisations will depend on their role in the AI value chain. AI providers of high-risk systems - developers or commissioning firms - will be subject to some of the AI Act's most stringent obligations, including Conformity Assessment and registration in a new EU database before market entry. AI deployers - organisations using AI systems under their own authority - will also have to comply with several requirements, including following the provider's instructions, ensuring the quality of input data, and, in some cases, performing a Fundamental Rights Impact Assessment (FRIA).

Non-high risk AI systems that interact directly with individuals or generate content (such as GenAI) will have to comply with transparency requirements.

21 Safe and responsible AI in Australia (storage.googleapis.com)

Figure 3: EU AI Act AI systems and model classification and key requirements



AI systems

General Purpose AI (GPAI) models and systems

The AI Act imposes strict requirements on providers of GPAI models and systems, as these are often integrated into multiple downstream AI systems. These include providing up-to-date technical documentation to downstream providers, complying with EU copyright law, providing a detailed summary of the content used to train their model, and watermarking AI-generated or manipulated content.

Providers of high-impact GPAI models, which could pose systemic risks and significantly impact the EU internal market, will face additional requirements and enhanced supervision. This includes continuous assessment and mitigation of systemic risks, conducting adversarial testing, ensuring robust cybersecurity protection, and reporting serious incidents as well as their energy efficiency.

Interaction with other technology-neutral EU regulatory frameworks

The EU AI Act is just one part of a larger regulatory landscape for AI in the EU. Other technology-neutral regulations will interact with the AI Act depending on how AI is being used. For example, Very Large Online Platforms (VLOPs) designated under the Digital Services Act (DSA) that deploy GenAI systems will need to follow both the transparency requirements under the AI Act as well as applicable DSA rules. The European Commission has already used its DSA powers to ask VLOPs for information about the risks of AI-generated deepfakes. GDPR and EU copyright law will also likely apply to GenAI. While the AI Act and these regulations will often complement each other, there may be cases where the interaction is less clear, such as the responsibilities of different actors in the AI value chain under the EU AI Act and GDPR.

5.2.3 Japan

Japan does not yet have any AI specific regulation. Two key government ministries, the Ministry of Internal Affairs and Communications (MIC) and the Ministry of Economy, Trade and Industry (METI) recently published 'AI Guidelines for Business Ver 1.0'.[22] These Guidelines present three basic approaches that should drive the development of AI:

- Dignity (a society that has respect for human dignity)
- Diversity and inclusion (a society where people with diverse backgrounds can pursue their own well-being); and
- Sustainability (a sustainable society)

In addition, the Japanese Government established the Hiroshima AI Process, which includes a set of AI Principles and a Code of Practice, under its G7 Presidency in 2023. The Principles and Code of Practice are non-binding, but firms developing and deploying AI systems are invited to follow them. The Code of Practice sets out eleven actions to be taken including:

- Taking appropriate measures, prior to placing AI systems on the market and throughout their lifecycle, to identify, evaluate and mitigate risks.
- Publicly report the capabilities of these systems to promote transparency
- Working towards responsible information sharing and incident reporting
- Developing and disclosing AI governance and risk management policies
- Developing and deploying reliable content authentication and provenance mechanisms
- Supporting the development and adoption of international standards

The Japanese Government has not publicly announced any plans for an AI-focussed regulatory authority at this stage. It has undertaken multiple AI-related initiatives, and has taken a particular interest in the links between data governance and intellectual property and AI. It published an interim report on intellectual property in the AI era in April 2024, which states that it is important for relevant government organisations to cooperate to support private sector initiatives for establishing an ecosystem that achieves acceleration of AI technologies and protection of intellectual property rights.[23]

5.2.4 Singapore

Singapore's approach towards governing AI is practical and risk-based. It is prioritising responsible and ethical AI deployment, with a strong focus on adoption and innovation - ensuring that its benefits are accessible to all in a safe manner. While there is a broad recognition of the importance of trustworthy and responsible AI, Singapore does not currently have specific AI regulation and is instead focussed on developing non-statutory governance frameworks and toolkits, which are still in their early stages of development. In addition, Singapore is active in various international processes to ensure a consistent approach to AI regulation, including common guardrails and evaluations for the most capable AI models.

There are four main elements to its existing approach:

- Model AI Governance Framework, which involves promoting the adoption of AI across various sectors, including finance, healthcare, transport, and public services and is intended to enhance productivity and create new economic opportunities. Its Model Framework for traditional AI systems was released in 2019, and it is expected that a new Model AI Governance Framework for Generative AI will be finalised in the middle of this year.
- Establishing an AI Safety Institute to support understanding and testing of the most advanced models, and which is partnering with the US and UK equivalents.
- The development of AI Verify, a framework and toolkit for testing AI governance across all sectors. It comprises 11 key AI ethics principles that align with global standards and frameworks, including those from the EU, OECD, and Singapore's Model AI Governance Framework. Those 11 principles include: transparency, explainability, repeatability/reproducibility, safety, security, robustness, fairness, data governance, accountability, human agency and oversight, inclusive growth, societal and environmental well-being. AI Verify is designed to assist organisations in assessing their AI systems' adherence to these principles through standardized tests.
- Establishing regulatory sandboxes, such as the MAS FinTech Regulatory Sandbox and IMDA's AI Sandbox, which allow companies to test AI applications in a controlled environment while working closely with regulatory authorities to address concerns.

22 https://www.meti.go.jp/english/press/2024/0419_002.html
23 https://www.kantei.go.jp/jp/singi/titeki2/ai_kentoukai/gijisidai/dai7/index.html

Alongside these frameworks, existing legislation that will impact AI includes:

- The Personal Data Protection Act;
- The Copyright Act to regulate the use of copyrighted materials for model training and copyright for AI-generated content. Changes to the Singapore Copyright Act in November 2021 were part of intellectual property legislation to align with the development and commercialisation of new AI technologies like ChatGPT;
- The Protection from Online Falsehoods and Manipulation Act (POFMA) on digital falsehoods through AI algorithms used by online platforms to curate content. POFMA gives authorities the power to order corrections or removal of false information deemed harmful; and
- The Cybersecurity Act on cyber resilience and for AI systems to adhere to

Although the current AI frameworks do not have enforcement powers, compliance with certain requirements, such as PDPA/POFMA/Copyright Act/Cybersecurity Act may entail significant penalties for mishandling (such as sensitive data).

The Singaporean Government will continue to monitor the advancements of AI technologies and review governance frameworks and regulations to ensure their ongoing relevance and effectiveness.

5.2.5 United Kingdom

The UK's current approach is an outcome-based, non-statutory framework to guide responsible AI design, development, and deployment. The UK previously signalled an expectation that specific AI regulation would be required in future. Whilst the new UK government has already indicated that there will be future AI regulation, it is not yet clear whether it will expedite the timetable or have a different focus. As it stands at present, the framework is underpinned by five core principles:

- Security and robustness
- Transparency and explainability
- Fairness
- Accountability and governance
- Contestability and redress.

The framework aims to balance innovation and safety in AI by applying an existing technology-neutral regulatory framework. It does not introduce any new regulatory requirements or authority at present. Incumbent regulators, such as the UK communications regulator Ofcom and the UK data protection regulator, the Information Commissioner's Office (ICO), will apply the principles within their own remits using existing laws and regulations to address risks and opportunities presented by AI in their domains.

Key examples of relevant regulations include the Online Safety Act (OSA), UK General Data Protection Regulation (UK GDPR), and Digital Markets, Competition and Consumers Bill.

The framework also emphasises the importance of engagement and collaboration among regulatory authorities. A key example is the Digital Regulation Cooperation Forum (DRCF), under which umbrella, several key UK regulators[24] are already coordinating activities on AI regulation. This includes the launch of a new AI and Digital Hub[25] to support AI innovators in addressing complex regulatory queries. The UK government's Department for Science Innovation and Technology (DSIT) will also establish a new central function to monitor and evaluate AI risks centrally, promote coherence between regulators, and address regulatory gaps.

As noted, the new UK Government has indicated that, as with its predecessor, it expects to introduce AI specific legislation in future. In particular, it has indicated an indication to place the UK's AI Safety Institute, and its engagement with frontier models, on a statutory footing. For now, voluntary safety and transparency measures which developers of highly capable GPAI models and systems had committed to ahead of the first global AI Safety Summit, hosted by the UK Government last November, will supplement the framework and the activities of individual regulators.

5.2.6 United States

To date, the US approach to AI regulation has focussed on establishing requirements for specific governmental uses of AI, whilst favouring voluntary frameworks and guidance for private sector development and deployment, in an attempt to bolster AI innovation whilst also addressing the significant risks. Congress has not enacted any new AI specific regulation for the private sector and there has been general agreement that more education is needed before doing so. As a result, until legislation is passed, the Executive Branch, including regulatory agencies, are relying on existing regulatory authority to enforce any AI-related violation of existing law (e.g. civil rights, employment, privacy) to address any AI-related violations with existing enforcement powers.

24 The DRCF is a voluntary cooperation forum that facilitates engagement between regulators on digital policy areas of mutual interest. It currently has four members: the Financial Conduct Authority (FCA), ICO, Competition Markets Authority (CMA) and Ofcom.

25 https://www.drcf.org.uk/ai-and-digital-hub

There have been three key developments from the executive branch to date:

- the AI Executive Order (EO) of October 30 2023, which seeks to promote the safe and secure development and use of AI and creates requirements related to the use of AI throughout the federal government. The EO directs the development of both voluntary and mandatory guidance to govern the use of AI in the public and private sectors. It includes more than 100 directives to agencies, which will mostly be implemented over the next year. The Commerce Department will play an important role in implementation and has formed a US AI Safety Institute to help develop technical guidance for other agencies as they carry out their directives. The EO included directives that will lead to binding requirements for developers of powerful GPAI that could pose risks to US national security, economic security, or public health. Developers of these powerful systems will be required to carry out system evaluations, disclose safety test results, and share the outcomes and other activity related to systems development with federal agencies (building on voluntary commitments that were previously agreed between the US government and AI model developers). The EO also directed the White House Office of Management and Budget to issue guidance to federal departments and agencies on the implementation of AI, including directives to appoint Chief AI Officers, develop AI strategies, and maintain and annually submit inventories of their respective AI use cases with a focus on AI uses that are deemed to be "rights-impacting" or "safety-impacting."
- In January 2023, the National Institute of Standards and Technology (NIST), which sits within the US Department of Commerce, issued the AI Risk Management Framework (AI RMF), voluntary guidance aimed at integrating trustworthiness into the design, development, use, and evaluation of AI products, services, and systems. The EO directed NIST to create a companion document to the AI RMF focussed on GenAI, and the Commerce Department and NIST to create a secure software development framework for GenAI and very powerful AI systems (dual-use foundation models).
- In advance of any future legislation, the White House developed the Blueprint for an AI Bill of Rights in October 2022 - a voluntary framework intended to guide the design, use, and deployment of automated systems with the potential to "meaningfully impact" the American public's rights, opportunities, and access.

AI legislation was introduced in over 22 US states and territories in 2023, with a general focus on creating working groups and committees to study AI and produce policy recommendations, as well as regulating deepfakes in elections. In 2024, proposed AI legislation has focussed on regulating synthetic or AI-generated content in elections, explicit materials, or media; promoting the responsible and ethical use of AI; studying AI; and regulating the use of AI in state government. In May 2024, Colorado passed the most comprehensive AI law in the US (SB 205), which regulates both developers and users of "high-risk" AI systems. The new law will impact businesses in Colorado that use AI to make "consequential" decisions affecting state residents. The law could be amended before it takes effect in February 2026. Other state-level AI legislation that passed in 2024 focuses on deepfakes pertaining to explicit content, audio, and elections,

Other tech neutral legislation that will apply to AI includes but is not limited to privacy and data protection laws (e.g., the Children's Online Privacy Protection Act (COPPA)), the Health Insurance Portability and Accountability Act (HIPAA)); consumer protection laws (e.g., the Federal Trade Commission Act which grants the FTC authority to act against deceptive and unfair business practices); discrimination (e.g., Fair Housing Act, Equal Credit Opportunity Act, Americans with Disability Act; employment (e.g., Civil Rights Act of 1964); IP; financial regulations (for certain AI uses by financial services institutions).

As part of efforts to better understand AI risks in advance of any attempt to enact regulation, US law makers have been holding hearings, hosting bipartisan briefings, and soliciting input from experts to help them better understand AI's impacts and have explored a range of topics related to AI, including AI governance, bias, national security, workforce development, and misuse. In May 2024, a bipartisan working group led by Senate Majority Leader Chuck Schumer (D-NY) released an AI Policy Roadmap after convening nine AI insight forums to learn more about the technology. In the Roadmap, the working group advocated for at least $32 billion in funding for nondefense AI initiatives and outlined policy priorities and recommendations to various committees within Congress, but the group stopped short of endorsing specific legislation. The 2024 US elections could complicate the direction of travel and at the moment it is too early to say what will happen after the elections.

5.3. Definitions of AI in national and regional approaches

As noted previously, AI is an umbrella term used to describe multiple technologies and methods. The way in which different regulatory approaches define AI, and also differentiate if from simpler software systems and programming, is critical to an understanding of the systems in scope.

As shown in Figure 7, there are a wide range of definitions being adopted in regulatory approaches to date. Some of these differences are likely a result of the varying maturity of regulatory approaches – with countries that have not yet introduced specific regulation not needing to be as precise. Others may reflect countries' different preoccupations. It is notable that the EU AI Act – the world's first comprehensive AI regulation – adopts a particularly broad definition of AI in scope of regulation. *This is likely to mean that firms will need to look at older models that may have been in use within the firm for some years*. If in doubt, firms should seek legal advice about whether their proposed application is captured by the relevant regulation.

Table 4: Definition of AI used in regulatory approaches

Country	Definition
	There is currently no single statutory definition of AI. The Australian Government has previously endorsed the CSIRO's working definition of AI as: "a collection of interrelated technologies used to solve problems autonomously and perform tasks to achieve defined objectives without explicit guidance from a human being.'
	The EU AI Act adopts the OECD's definition: "A machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."
	The "AI Strategy 2022", which was issued by the Cabinet Office's Integrated Innovation Strategy Promotion Council, suggests that AI refers to a system capable of performing functions that are deemed intelligent.
	Artificial Intelligence (AI) refers to the study and use of intelligent machines to mimic human action and thought (Infocomm Media Development Authority)
	There is no formal definition of AI. Instead, an outcomes-based approach, which focuses on two defining characteristics – adaptivity [26] and autonomy [27] - will guide sectoral interpretations.
	The term 'artificial intelligence' means a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments." (National Artificial Intelligence Act of 2020)

5.4. Deloitte view

Unlike with the "what" where there is a high degree of convergence, there is a greater degree of divergence in the execution and application of AI regulatory approach across these six key jurisdictions.

Three broad approaches can be identified, suggesting that companies operating internationally will face regulatory divergence, with the potential for this to increase in the coming years as the scope, details, and interdependencies of AI regulations develop:

- Horizontal regulation of the use of AI as a whole, as in the case of the EU AI Act
- Vertical regulation of AI as it occurs in different parts of the economy or society, as in the UK's (current) Pro-Innovation Framework (and to an extent in the US approach through the Executive Order)
- Using codes of conduct, principles, or model governance where regulators have not yet determined their preferred approach or believe it is too soon to do so, as in Singapore, Japan and Australia.

All of the approaches examined, even where horizontal AI specific regulation has been introduced, are relying on multilayered regulation, with technology neutral and sector specific regulation, operating in tandem with AI specific rules. Understanding the combined regulatory burden of AI uses, and complying with overlapping requirements within jurisdictions, will be a key challenge for firms to consider as they design their compliance approach. There is also some emerging convergence in the consideration of formal regulatory audits and auditable statements of conformity across several of the approaches.

It is also possible to identify elements of a risk and principles-based approach at the core of many of these regimes, with AI deployers and developers required to consider the risks of their AI (to an extent) from first principles and to put in place appropriate mitigations to manage those risks. This suggests that – as with other key pieces of tech regulation such as the Online Safety Act in the UK and the Digital Services Act in the EU - *governance, risk and control; monitoring; and documentation will be central to the requirements for many firms and could provide the foundations for "how" to build an approach to compliance across global regulations.*

The final chapter considers the implications of this global outlook and offers some recommendations for firms who are considering their approach to compliance.

26 The ability of AI systems to see patterns and make decisions in ways not directly envisioned by human programmers.
27 The capacity of AI systems of operating, taking actions, or making decisions without the express intent or oversight of a human.

Chapter 6

How can companies prepare for regulatory compliance?

In this report, we have seen that:

- The challenges of AI are expansive and span across the value chain. At the same time, the regulatory landscape is dynamic and uncertain. However, regulators are primarily focussed on a subset of the risks, particularly around harms to individuals. This could help firms to prioritise efforts and to navigate some of the uncertainty but could mean that delivering regulatory compliance will not manage all AI risks.
- At present, there is strong alignment across regulatory approaches in "what" those risks are. This could be used to inform an enterprise-wide risk and governance framework across global regulations (as set out below). However, there is less alignment in the application of regulatory approaches, although a risk and principles-based approach is prevalent. This will put a strong emphasis on organisations being able to define, measure and mitigate AI risk at the enterprise level.
- In addition, firms will need to manage intersectionality between AI regulation and other areas (such as data protection and privacy, ESG, and sector-based regulation such as software as a medical device/AI as a medical device).
- The global regulatory and risk landscape is evolving and is likely to do so for some time.

Regulation also presents opportunities, for example helping to assign clear roles and responsibilities across the value chain (see figure 4) which can provide confidence to those downstream, or by providing clarity on applications which will have little or no direct regulatory burden. But overall these findings suggest that AI regulation will be challenging for many organisations, particularly with a degree of divergence between approaches, and it can be hard to know where to start. In short, there is no one size fits all approach. Regulatory strategy will both influence and inform business goals, geographical launch, operational complexity and most importantly, market play. That is, defining your role within the transformation and implementation landscape will guide your regulatory adoption strategy and your risks.

Based on the analysis in this report and using our experience supporting organisations who have faced similar digital regulatory waves in the past, we have identified five elements to support an organisation-wide response to global AI regulation, and to help navigate the uncertainty of the evolving AI regulatory landscape whilst still enabling AI innovation.

Figure 4: Definition of roles across the AI value chain in the EU AI Act

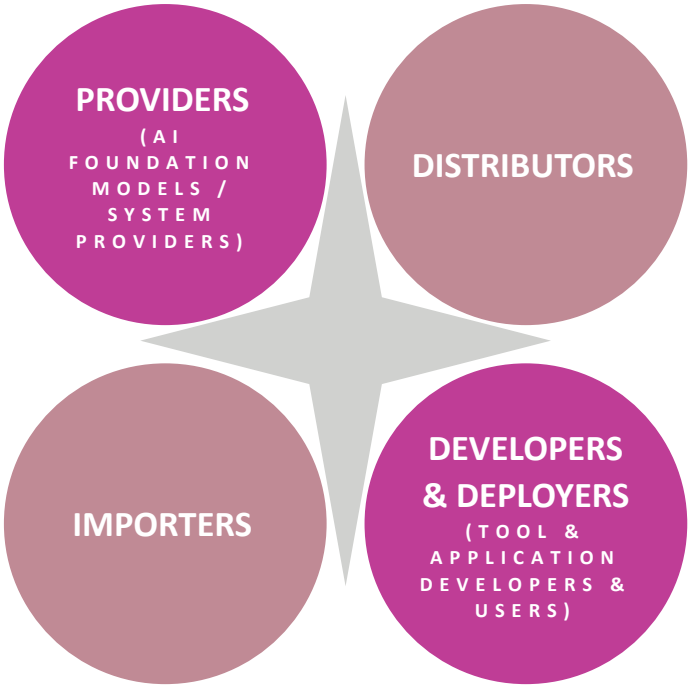


Figure 5: Five elements to support your response

To bring these elements to life, we have worked through an illustrative example. Through the example, we will outline some key strategic and operational challenges facing a firm that is navigating the complex international AI regulatory landscape. Although this example does not consider sector specificities in depth, we hope that it will have broad applicability for many firms facing a similar set of challenges. This case study is not a how to guide, but rather it simply seeks to bring the complexities to life.

Introduction to illustrative case study

Company overview

- ExampleAI is a global general-purpose model developer and digital platform, that has developed a highly capable large language model, ExampleAI 2.0, which it is providing direct to customers as a standalone chatbot, and embedded within its digital platform. It is also offering ExampleAI 2.0 as a base model to downstream developers via an API to enable them to build other AI applications over the top.

Business Challenge / Regulatory context

- ExampleAI 2.0 will be launched in the UK and EU, with aims to expand it globally
- ExampleAI wants to consider its future AI strategy and understand some of the key operational challenges it will face from regulation. It also wants to consider the implications for global regulatory compliance and implement some actions today to support compliance
- ExampleAI has already made significant investments in its principles-based compliance with other key pieces of EU digital regulation, including the EU's Digital Services Act (DSA) risk assessment
- In the UK, ExampleAI has a very strong relationship with Ofcom and the ICO and regularly participates in call for comments and industry forums and is up to date with the latest regulatory changes

How could ExampleAI respond to AI risks and regulation?

Using Deloitte's five step approach, we will walk through each element and learn how ExampleAI can use this manage their AI risks and regulations.

1 Understand the impact of regulations on your AI business strategy

To develop an AI regulatory strategy which is aligned to the business goals, there are several key considerations that ExampleAI will need to consider.

How does AI impact ExampleAI's enterprise risk tolerance?

The enforcement of non-compliance with AI regulation differs depending on the national regime in scope. In the EU, the penalties are severe, but equally, the cost of compliance could be significant. ExampleAI will need to put in place a risk assessment framework that allows flexibility in decision-making and the ability to prove how it manages risk according to the spirit of the law. Taking a risk-based approach to regulatory compliance involves making informed decisions about the level of risk that can be effectively managed with current technology, as well as determining what level of risk is deemed acceptable or tolerable within the organisation.

ExampleAI will want to ensure that it has a clear and transparent process for assessing and managing the risks associated with AI, and that this process is consistent with its overall approach to enterprise risk management. By doing so, the company will demonstrate to stakeholders that it is taking a responsible and proactive approach to managing AI-related risks.

What products/applications to offer?

ExampleAI will need to consider whether its current portfolio of products and offerings meet the regulatory requirements for *each geography in scope*. They will need to undertake a risk assessment for their applications/products in scope based on the regional and national regulatory requirements.

In the EU, the AI Act prohibits applications that can perform functions that are subliminal, manipulative, exploitative, or targeted towards certain sections of population. ExampleAI may determine that some applications cannot be launched in the EU due to these restrictions.

In the UK, the Information Commissioner's Office (the UK's data protection regulator) has confirmed that a lawful basis will be required for data processing at the development stage for each likely downstream application of the model, and ExampleAI may determine that a lawful basis for some applications cannot be realised within its existing risk appetite.

Where to launch first?

ExampleAI's launch strategy will be heavily shaped by the existing conformance of its products, as well as its relationships with individual regulatory bodies to understand how the regulation is shaping and evolving.

Within the EU, the broad shape of the regulation is finalised. But key questions around the AI Act's interaction with other regulations and legislation is yet to be finalised. Hence, ExampleAI will need to closely monitor the ongoing conversations and updates on the EU AI Act, and in the interim ensure that compliance with other key pieces of regulation such as the Digital Services Act is up to date.

In the UK, the government's pro-innovation stance on regulation, and with its existing relationships with key UK regulators, ExampleAI can remain confident that approaches will evolve proportionately and incrementally.

What overall AI regulatory strategy approach to take?

Given the extraterritorial impact of the EU AI Act and the degree of divergence we are already seeing with other regulatory regimes, ExampleAI will need to decide whether, and which, of the AI Act rules and standards it wants to adopt globally. Alternatively, it may choose to develop and deploy EU-specific AI systems, or in some scenarios, scale back use of higher-risk AI in the EU.

What is ExampleAI's Partnership & Alliance Strategy?

During its product/application development, ExampleAI will need to interact and partner with multiple vendors and partners. Regulation has a high impact on the buy/build decisions for components for its models and wider supply chain risks, which will need to be identifiedw when such decisions come to play. For example, in the EU, there is a requirement to have an authorised representative in the Union for General Purpose AI model builders as well as providers of high-risk AI systems.

2 Create organisational clarity around the key operational challenges from AI regulation

ExampleAI will need to understand the key operational challenges arising from different global regulatory approaches *to understand gaps against its current approach and design its Target Operating Model*. This should cover key processes, governance, roles and responsibilities, and controls. The operating model will be informed by the EU AI Act's extraterritorial scope, current compliance to international/national standards for AI development, and the intersection with related regulation such as the EU Digital Services and Digital Markets Acts. All impacted stakeholders from across ExampleAI will need to be involved in this process to ensure that it works for the business.

A key consideration, based on experience from other principles-based digital regulation, will be establishing the Three Lines of Defence for AI risk (and decisions about how to incorporate this within existing processes). In particular, the Second Line – risk management and compliance functions across ExampleAI - will play a key role in putting in place relevant policies and providing ongoing monitoring of compliance.

As noted, both jurisdictions have elements of a risk and principles-based approach, which means that, *alongside specific technical requirements, governance, risk and control; testing and monitoring; documentation; and audit and assurance will be critical.*

Table 5: Operational Challenges & Considerations for ExampleAI

Area	Key questions/considerations for ExampleAI
Scope	• Does ExampleAI have the tools and processes to determine which of its proposed uses are in scope of different regulations both now and in the future? • How will it manage compliance across different product areas in different parts of the value chain subject to different regulatory requirements? • How will existing conformity apply to new AI regulatory requirements (e.g. EU DSA and DMA, UK OSA, GDPR)?
Risk management	• How should ExampleAI extend existing risk management to cover AI risks? • How should it define principled terms (e.g. proportionate/reasonable) and have a common language for describing AI risks? • How should it align this with wider business strategy and set out its approach in auditable documented controls?
Documentation and reporting	• What processes and functionality does ExampleAI need to deliver key AI transparency requirements such as technical documentation, instructions for use, quality management systems, and incident logging?
Talent, communications & training	• What skills does ExampleAI need as a business? • How should it engage the organisation around AI regulation? • What is its communication strategy to raise awareness of AI regulation? • How does ExampleAI put in place human oversight with appropriately skilled people?
Data and data governance	• How to ensure that any use of personal data for AI is compliant with GDPR/UK GDPR? What approach should ExampleAI take around data governance and managing privacy considerations? • How should ExampleAI test for bias and ensure accuracy within ExampleAI 2.0? • What can be leveraged from its approach to other data protection regulation vs what is new? How does it ensure that copyright and IP regulations are complied with?
Other technical requirements	• What functionality does ExampleAI need to build or buy to deliver key technical requirements around data quality; system evaluations/red teaming; accuracy and robustness (including cyber security)?
Monitoring and testing	• What testing is required at each stage of the lifecycle (including to support the development of risk assessments)? • How should ExampleAI build in ongoing monitoring post deployment for downstream users? • What processes does it need to ensure that it can report serious incidents to the relevant regulator? • How should this be documented, if required to be accessed by regulators?
Users	• What functionality is needed to provide transparency to users of AI generated content? • How should ExampleAI ensure that users are notified that they are interacting with an AI system? • What internal processes might be required to provide users with an explanation of AI decision-making?
Organisational culture	• How to ensure a culture that prioritises AI risk appropriately and has an understanding of the intent of AI regulation? • Whether and how to implement processes for staff to report concerns around non-ethical AI use?

3 Engage key stakeholders across the organisation to avoid silos

As ExampleAI builds its roadmap for implementation of regulatory compliance, and begins to plan for turning regulatory requirements into policies, standards, controls, and processes that work for the business as a whole, it will be critical to engage with functions across the whole organisation. Some of the recommended key stakeholders are listed in the figure below. As a first step, ExampleAI should begin to raise awareness amongst key stakeholders who in the future will be part of the implementation of regulatory compliance on what is coming and the plans.

4 Design and implement AI governance and risk management framework

Having considered the strategic impact of AI regulation; understood the operational impacts to inform the Target Operating Model; and engaged key stakeholders across the business, ExampleAI is in a position to start developing a company-wide AI Governance and Risk Management Framework to manage compliance across global AI regulations.

- Assess current approach to compliance and identify gaps using the common themes identified across key regulatory jurisdictions. Refer to Section 4.1
- Build out a sustainable approach to AI risk identification and management, guided by and informed by the core regulatory themes identified across global regulatory approaches, and monitored for effectiveness and updated on a regular basis. Leverage lessons learnt from DSA, DMA, and / or GDPR compliance (where relevant). See figure 7.
- Given the overlap with other relevant digital regulation, consider how to extend out existing governance around these measures
- Focus on putting in place an auditable and documented risk and control process that can be used to demonstrate compliance across the broad scope of different regulatory regimes; and ensure that it can be responsive to changing risk and the evolving regulatory landscape

Figure 6: Key Stakeholders

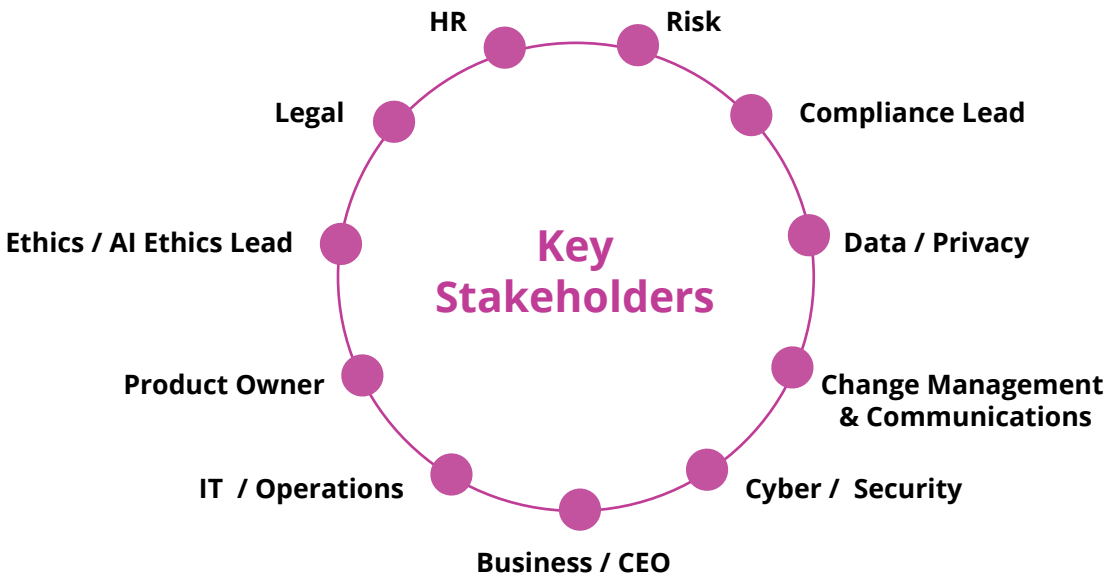
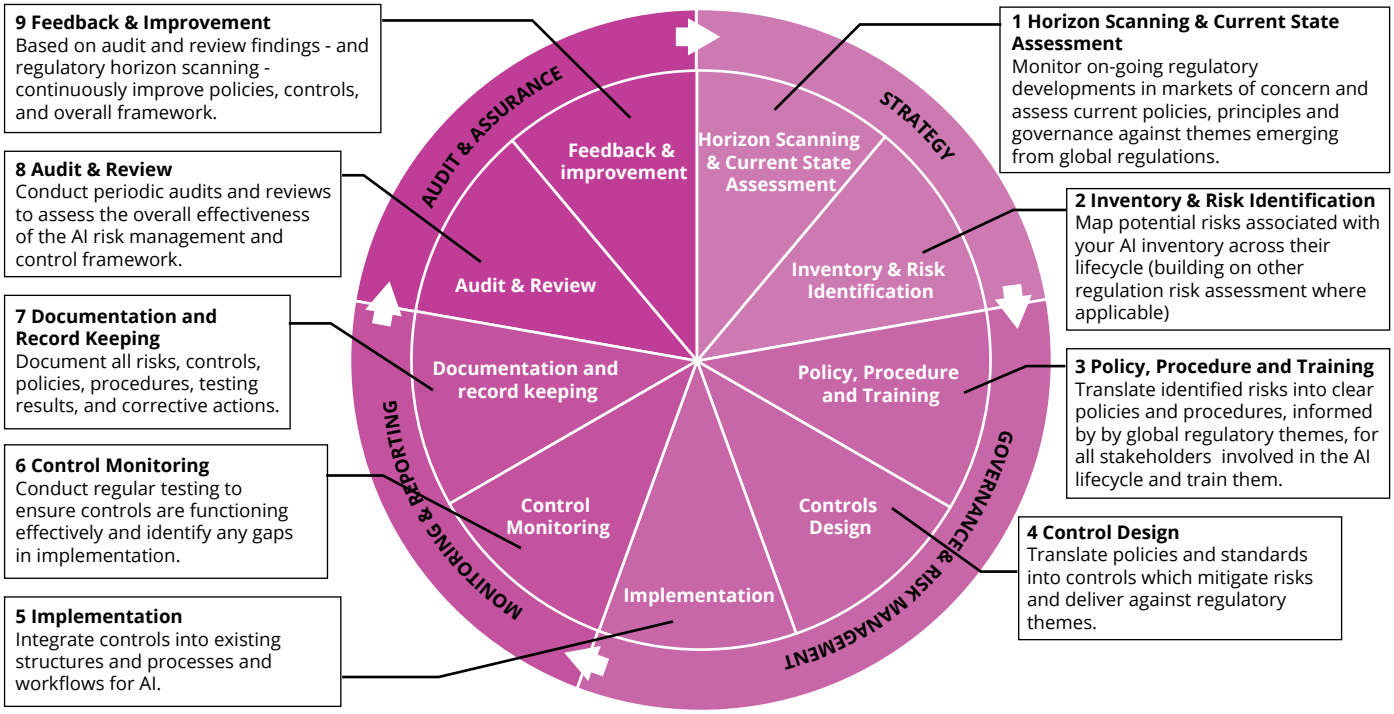


Figure 7: A potential approach for company-wide AI governance and risk management to support regulatory compliance



5 Implement some no regrets actions now

Through the illustrative case study, we hope that we have been able to demonstrate the critical strategic questions, considerations, challenges and complexities stemming from AI regulation.

These elements will take time to work through and implement. However, there are certain no regret actions that organisations could embrace today, to ensure that no matter where they are in the AI adoption cycle, they will be better prepared for the dynamic regulatory landscape.

- **Form an AI governance committee** spanning all key stakeholders and supported by a cross-cutting AI governance team. Governance should be at C-Suite Level which includes the Chief Compliance Officer (titles may vary by the company), Chief Technology Officer (Chief AI Officer if applicable), Human Resources & Talent Lead, Legal, Information Security Officer, Chief Executive Officers. **Action owner** – Chief AI Risk & Compliance Officer / Chief Compliance Officer

Lessons to be learned from EU DSA/DMA

- A clear methodology for control documentation for principles-based regulations is required with the right stakeholders in the room to decide on suitable controls
- Standardise control methodology and templates across business functions and seek rationalisation of controls from the beginning;
- It is important to start by identifying AI risks and creating clear auditable control objectives which link back to regulations and other internal / external obligations
- Since AI is already regulated to an extent through other regulations including EU GDPR, EU DSA, EU DMA, EU ND4C and UK OSA, review existing controls in this and uplevel rather than starting from scratch
- Specifically, review DSA systemic risk assessments for AI risks and DMA model inventory for identified AI models.

- **Create an AI system inventory** to understand the full systems in scope of regulation. Consider not just current on-going initiatives, but also older AI systems in use as well. Consider which third party AI systems and uses to capture in your AI inventory
  Action owner – Chief Technology Officer / Chief AI Officer
- **Gather documentation on existing AI systems**, training sets and policies, bias testing, model capabilities and limitations, human oversight arrangements. Consider including third party and partner AI systems as appropriate
  Action owner - Chief AI Risk & Compliance Officer / Chief Technology Officer
- **Identify & perform a gap assessment** across existing policies, processes, and principles with identified relevant regulation and key themes, as well as completed risk assessments, to inform the development of an implementation roadmap
  Action owner – Chief AI Risk Officer & Compliance Officer / Chief Risk Officer
- **Establish dynamic regulatory intelligence** across AI specific, AI-adjacent areas and sector specific regulation to ensure the compliance needs are up to date. Establish regular horizon scanning processes / alerts to track regulatory developments and an obligations / requirements library to manage the evolving regulatory landscape. Monitor the evolving AI standards landscape, including the development of harmonised standards in the EU to support the EU AI Act
  Action owner – Chief AI Risk & Compliance Officer / Chief Legal Officer
- **Conduct risk assessments** to identify and understand the impact of planned AI usage against your enterprise risk appetite. Use the EU's definition for High Risk AI Applications to guide and prioritise, and to support the development of a bespoke AI risk taxonomy for your organisation. Consider the risks of your third party and partner use of AI products and systems
  Action owner – Chief AI Risk & Compliance Officer / Chief Risk Officer
- **Start communications across the organisation and ensure crisis preparedness.** Effective communication is important in the day-to-day governance of AI and will be necessary to bring your people on the journey. This include being transparent about the long term AI strategy, the benefits and risks to the business, upskilling teams on how to use AI models and reskilling people whose activities may be performed by AI in the future. It is essential to ensure that all stakeholders are aware of the risks and benefits associated with AI, and that they are able to make informed decisions about its use and raise a concern. This requires clear and transparent communication, as well as a willingness to engage in dialogue. Practical actions organisations can take include scenario planning for high risk events, narrative development so leaders and employees can tell a credible, human story about the role and impact of the technology, and crisis exercising to test readiness for a severe but plausible event.
  Action owner – Chief Communications Officer / Human Resources Director / AI Risk & Compliance Officer

Crisis management in AI governance

Reflecting the dynamic nature of how organisations will use AI and evolving public attitudes to it, there is a reasonable probability a crisis event will occur. Integrating crisis readiness and crisis management response plans to overall AI governance is a simple critical step to ensure a major issue or crisis receives the attention, resources and management to protect value and company reputation.

How can the Deloitte Internet Regulation Team help you?

The explosive global growth in digital communication and commerce during the last quarter of a century has fundamentally and permanently changed the way the world works, learns, plays, and thinks. AI is the latest wave of digital regulation from governments around the world which will require a profound and thoughtful response from a large number of organisations, and the opportunity to drive towards compliance as a competitive advantage.

Over the last several years, Deloitte has been helping companies respond to internet regulation at a global level and deliver holistic risk-aligned and tech-enabled compliance. Our team of experts brings together legal and regulatory expertise, technological innovations, and comprehensive solutions to help you understand the complex domain of internet regulation, including AI regulation, using our Internet Regulation Methodology, which includes the design and implementation of the following:

- **Operating model** - Designing and implementing an integrated compliance operating model ensures programme activities are connected across the organisation allowing teams to work cohesively and breakdown silos currently impacting effective and strategic compliance.
  - *Our Risk Advisory and Consulting teams can support in the design and implementation of Trust & Safety functions and required operational capabilities.*
- **Compliance processes** - Designing and implementing integrated regulatory compliance processes and capabilities to ensure a holistic and effective approach to regulation, risk and compliance is taken across jurisdictions, products and services.
  - *Our Risk Advisory team can support on the end-to-end compliance process, including designing, implementing and embedding risk assessment and supporting methodology, process, tools and templates for regulatory compliance.*
  - *Our Deloitte Legal team is able to advise on new AI regulations and their implications, stand-up legally compliant AI offerings and processes required by regulation and to act for you in regulatory/ dispute matters, supporting your Legal functions as they grapple with the wave of change.*
  - *Our Economic Advisory team is able to advise on the strategic impact of regulations.*
  - *Our Audit & Assurance team is able to support in 'fit for audit' and conformity readiness, and in the performance of regulatory audits as required by incoming standards*
  - *Our Deloitte Reputation, Crisis and Resilience (RCR) team is able to support with crisis communication strategies, crisis management and resiliency planning.*
- **Technology enablement** - Managing an integrated compliance model across multiple regulations and jurisdictions is now too complex an activity to rely on spreadsheets and other locally held files. Technology is required to support a consistent approach to compliance across the organisation, enforce roles, responsibility and accountability and increase the ability to audit and provide assurance over regulatory compliance.
  - *Deloitte has an ecosystem of technology which can be used to identify, deploy, manage and monitor compliance with increasing regulatory requirements. The solutions are designed and configured around business processes, starting with user requirements, to ensure the solution is what the business wants and needs and enhances an organisation's function and processes.*

Contacts

Lead Author

Scott Bailey
Director
scottbailey@deloitte.co.uk

Key Deloitte UK contacts

Nick Seeber
Partner
nseeber@deloitte.co.uk

Joey Conway
Partner
jconway@deloitte.co.uk

Mark Cankett
Partner
mcankett@deloitte.co.uk

Suchitra Nair
Partner
snair@deloitte.co.uk

George Johnston
Partner
gejohnston@deloitte.co.uk

Global Deloitte network of AI risk and regulation contacts

United States (US)

Tanneasha Gordon
Principal
tagordon@deloitte.com

Bob Stradtman
Principal
rstradtman@deloitte.com

Singapore

Nai Seng Won
Partner
nawong@deloitte.com

Australia

Elea Wurth
Partner
ewurth@deloitte.com.au

Netherlands

Simone Pelkmans
Partner
spelkmans@deloitte.nl

Japan

Shinya Kobayashi
Managing Director
shinya.kobayashi@tohmatsu.co.jp

Key contributors

With special thanks to the following for their contributions

Haarika Kanuparthy
Senior Manager
hkanuparthy@deloitte.co.uk

Mark Hutcheon
Director
mhutcheon@deloitte.co.uk

Michael Greco
Assistant Manager
migreco@deloitte.com

Valeria Gallo
Senior Manager
vgallo@deloitte.co.uk

Deloitte.

Designed by CoRe Creative Services. RITM1932352