RAMPARTS

Navigating the complexities of AI Governance:
A Comprehensive Guide

(First Published 30 October 2024 – updated March 2026)

Note: *This Guidance Note has been prepared with some assistance from generative AI tools including: Midjourney, Microsoft Copilot, Google Gemini, ChatGPT and Perplexity AI.*

AI generated image, Peter Howitt

1. Definitions

AI	The development of computer systems capable of performing, replicating or mimicing tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation.
AI Act	A European Union regulation that establishes a common regulatory and legal framework for AI within the EU. It aims to foster innovation and ensure AI systems are safe, ethical, and respect fundamental rights by categorising AI applications into different risk levels and imposing specific requirements.
AI Governance Committee	An independent body within an organisation responsible for overseeing the ethical development and deployment of AI systems.
AI Governance	The frameworks, rules, and standards that aim to balance the benefit of AI tools and systems with safety and ethical and societal values. It encompasses oversight mechanisms to address and minimise risks like bias, privacy infringement, and misuse while fostering innovation and trust.
AI Impact Assessment	A systematic evaluation of the potential benefits and risks of an AI System, including its impact on individuals, society, and the environment.
AI Tools	Software applications that use AI Algorithms to perform specific tasks and solve problems. These tools can automate tasks, enhance productivity, streamline processes, and gain insights from data and improve decision-making across various industries.
AI Systems	Machine-based systems that make inferences from input data to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. They can adapt and improve their performance over time based on new data or feedback.
Algorithms / Algorithmic	Step-by-step procedures or mathematical models that guide AI Systems to help it learn, make decisions, or solve problems.
Algorithmic Auditing	The process of examining AI algorithms to identify and mitigate biases, ensure fairness, and promote transparency.
Black Box AI	AI systems whose internal workings and decision-making processes are opaque and difficult to interpret.

Bias	The presence of systematic errors in AI systems that result in unfair outcomes for or unfair discrimination against persons or groups, often due to biased training data or flawed algorithms.
Data Lineage	Data lineage is the process of tracking and recording the flow of data throughout its lifecycle, from its origin to its final destination. It enables organisations to visualise and understand where data comes from, how it transforms over time, and where it's ultimately stored. Data lineage helps ensure transparency, validate data accuracy, and identify the root cause of errors in data processing or analysis.
Deepfakes	A sophisticated form of artificial intelligence-generated media that uses deep learning techniques to create or manipulate audio, video, or images to make them appear authentic.
Generative AI	A subset of AI that focuses on creating new content, such as text, images, or music, based on patterns learned from existing data. Examples include language models like GPT-4 and image generation models. Generative AI can also be used for tasks like translation, summarization, and creative content generation.
Data Protection or Data Privacy	Safeguarding personal data throughout its lifecycle, from collection to deletion. Protection of personal data from unauthorised access and ensuring that individuals have control over their own data and that it is secure.
GDPR	A regulation in EU law on data protection and privacy for individuals within the European Union and the European Economic Area. GDPR grants individuals certain rights regarding their data, such as the right to access, rectify, and erase their personal information. It also addresses the transfer of personal data outside the EU and EEA areas.
Ethical AI	The practice of designing and deploying AI systems in a manner that is fair, transparent, and respects human rights and values. Ethical AI strives to avoid harmful or unintended consequences.
Explainable AI	The ability to explain how an AI system makes decisions is critical to transparency and trust. Explainability is essential for building accountability and enabling humans to understand and challenge AI-driven decisions.

Model Drift	Model drift refers to the degradation of an AI model's performance over time as the real-world data it encounters begins to differ from the data it was originally trained on. This can occur due to changes in user behaviour, market conditions, or other factors that affect the underlying patterns in the data.
ML	Machine learning is a subset of AI that involves training algorithms on data to enable systems to learn from and make predictions or decisions based on that data without being explicitly programmed for specific tasks. ML is a key technique used to achieve AI, enabling systems to learn from data without being explicitly programmed.
Regulatory Sandbox	A framework that allows companies to test innovative products, and business models in a controlled environment under regulatory supervision.

Part 1 – Introduction

The term "AI" has been broadly applied to various technologies, many predating the recent surge of interest in generative AI. Algorithmic data processing and machine learning tools for pattern recognition and forecasting have been utilised for years, even before the introduction of prominent generative AI tools like ChatGPT, Gemini, Co-pilot, and Midjourney. This broad application of the term can lead to confusion in categorising and managing these technologies, as some might not strictly fit the definition of AI, with the term sometimes being used loosely for marketing purposes.

- AI is the overarching field, while Machine Learning (ML) is a specific approach within AI.
- Traditional algorithms are rigid and follow predefined rules, while AI systems (particularly those using ML) can adapt and learn from new data.
- AI systems often use complex algorithms, but they go beyond simple computational processes to mimic human thought and behaviour.

The Alan Turing Institute uses a functional definition of AI systems as computational systems employing statistical or mathematical methods to perform tasks typically associated with human intelligence, thereby aiding or replacing human decision-making in such tasks:

"*any computational system (or a combination of such systems) that uses methods derived from statistics or other mathematical techniques to carry out tasks that are commonly associated with, or would otherwise require, human intelligence and that either assists or replaces the judgment of human decision-makers in carrying out those tasks.*" [1]

Leaving aside these definition issues, there are a significant number of laws and regulatory frameworks that impact the use of AI systems and AI tools. One of the most relevant is the General Data Protection Regulation (**GDPR**). The European Union's framework governing how organisations handle personal data. As artificial intelligence (**AI**) technologies advance, organisations must align AI use with GDPR to ensure compliance, avoid penalties, and protect individual privacy. In addition we have a significant number of AI specific frameworks in China and the EU, whilst the US is also moving forward with Federal, state-level and Governmental AI policies and laws.[2]

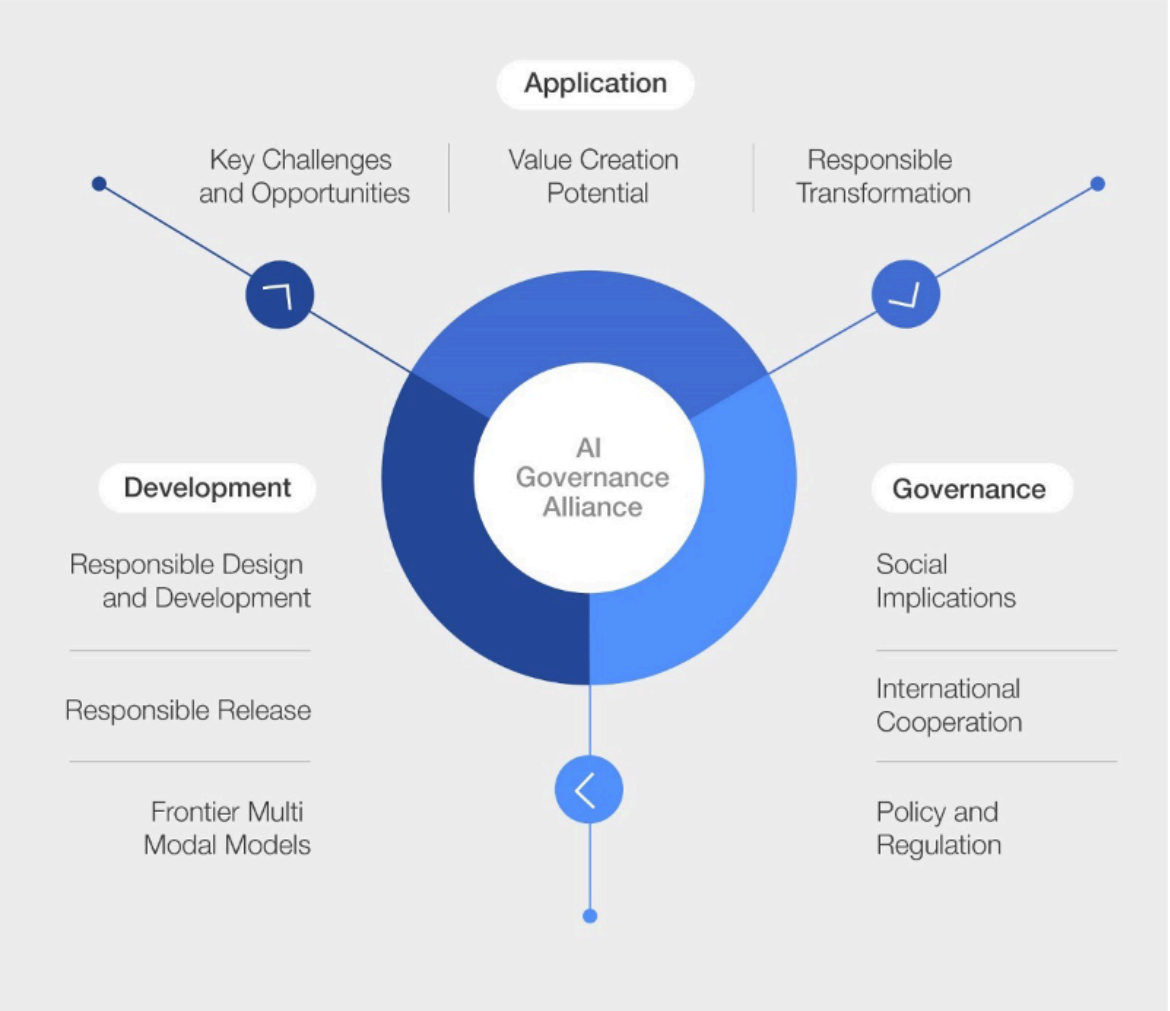
AI introduces unique challenges in data processing and decision-making due to its handling of large datasets, which often contain personal data and result in automated outputs (and potentially automated decision-making). While anonymisation tools and processes can mitigate these issues, the best use cases for AI tools and systems often involve personalised responses to user queries, content creation, and interaction. The real value of AI assistants for consumers will only be unlocked when the AI constructs know our interests, dislikes, abilities, weaknesses and loves and have persistent memory to help us in our life journey (and they understand our idiosyncrasies).

[1] AI Ethics and Governance in Practice: An Introduction
[2] See, for example, The State of Global AI Regulations in 2024

Additionally, AI systems must address concerns related to bias and fairness, ensure transparency and accountability, comply with other data protection requirements, and manage security risks associated with data breaches and misuse. Many aspects of AI present specific concerns for compliance and for the wider use of technology in an ethical and legal manner. Conversely, there are also concerns as to whether some Governments may be over-regulating this nascent industry[3] and undermining the ability to develop and deploy useful AI systems for everyone that will help us collectively make a great civilisational leap forward.

2. How Do Organisations Use AI?

AI has permeated various industries, transforming how businesses operate, from multinational corporations to small and medium-sized enterprises (**SMEs**).[4] The adoption and development of AI differ significantly between these two groups, driven by resources, scale, and strategic focus.



[5]

[3] Tech companies warn over EU AI regulation
[4] 101 real-world gen AI use cases from the world's leading organizations | Google Cloud Blog
[5] The AI Governance Alliance's debut report lays out strategies for equitable AI | World Economic Forum

2.1. Generative AI

One of the newest uses of AI tools is in respect of generative AI. Generative AI refers to AI systems that can create new content, such as text, images, music, and code, by learning patterns from existing data. Unlike traditional AI, which typically analyses or classifies data, generative AI can produce original outputs based on prompts. It uses techniques like deep learning and neural networks, particularly models like GPT or GANs, to generate content that mimics human creativity. However it has also given rise to significant concerns about the impact on human creativity, jobs and whether copyright and other intellectual property laws need to be updated to exclude or include such AI generated works.[6]

One key reason businesses adopt generative AI is efficiency. It significantly reduces the time and resources required for repetitive or creative tasks, enabling teams to focus on higher-level strategy and innovation. For instance, in marketing, AI-generated content like social media posts or product descriptions can be produced at scale, quickly adapting to customer trends. Generative AI can also provide personalised customer experiences. By analysing large datasets, it helps create tailored solutions, from chatbots to product recommendations, enhancing customer engagement.

We anticipate the growing use of generative AI in highly technical fields such as medicine, accounting, tax advisory, law (particularly in litigation), and financial analysis, as well as in sectors like engineering, architecture, and scientific research. These fields benefit from AI's ability to process vast amounts of complex, diverse data quickly and efficiently, enabling professionals to generate preliminary assessments, insights, or recommendations. As AI evolves, it will increasingly assist in automating data analysis, enhancing decision-making, and improving accuracy across a wide range of industries that rely on the synthesis and interpretation of large datasets.

2.2. Multinational Organisations

Multinationals typically have substantial resources and a global footprint, allowing them to invest heavily in AI research and development. They leverage AI across their vast operations to enhance efficiency, drive innovation, and gain a competitive edge.

Customer Service & Experience:

- AI-powered chatbots, virtual assistants and personalised videos and messaging tools are deployed to handle customer queries, provide personalised recommendations and even intervene when there are indications of problem behaviours.
- Use Cases:
    - McDonalds has developed automated ordering using IBM's Watson.
    - Amazon has integrated personalised recommendations and reorder prompts relying on AI enhanced tools.

[6] Copyright, AI and Generative Art - Ramparts

  - YouTube, Spotify and many other social and entertainment platforms use AI to personalise recommended content.
  - Online gambling companies are experimenting with the use of AI videos and messaging to provide personalised intervention when they see potential problem gambling behaviours.[7]

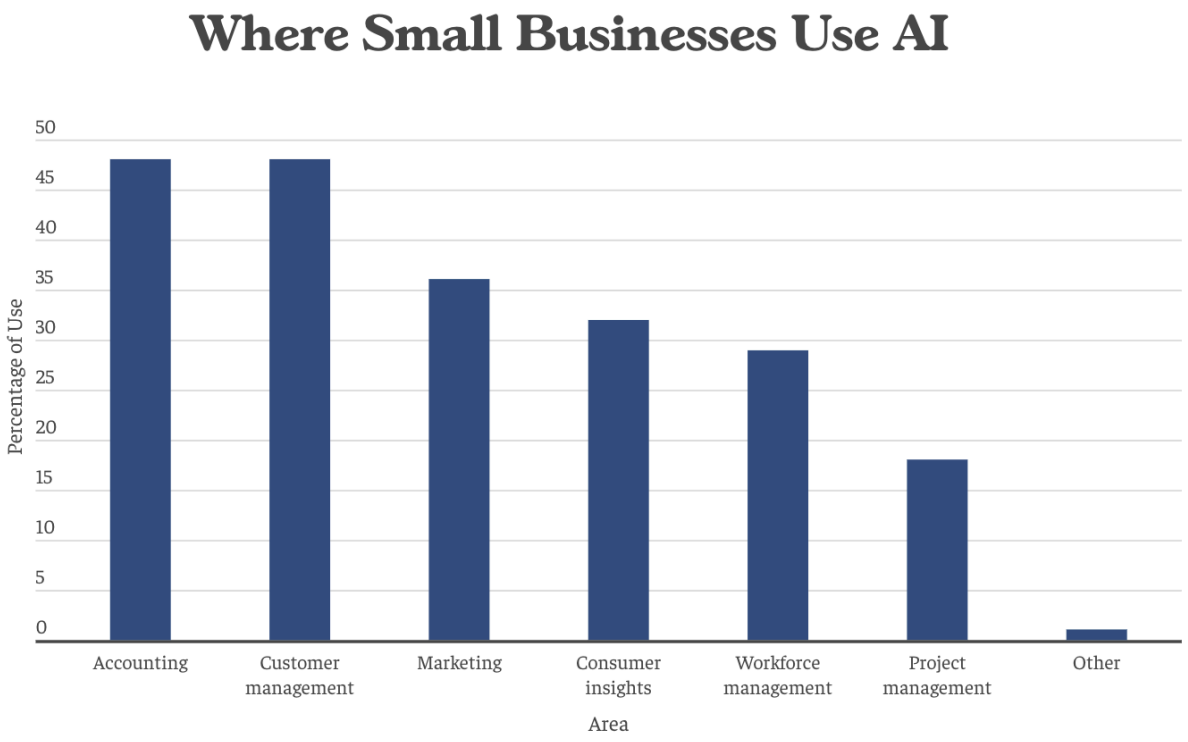
- **Supply Chain Optimisation:** Machine learning algorithms analyse vast amounts of data to predict demand, optimise inventory levels, and streamline logistics, resulting in cost savings and improved operational efficiency.
- Use Cases:
  - Walmart deploys AI technologies to manage inventory levels more effectively and enhance customer service. AI systems predict product demand to optimise stock levels. Walmart has experimented with AI-driven robots to assist in inventory management and customer service.[8]
- **Product Development & Innovation:** AI is used to analyse market trends, consumer preferences, and competitor activities to accelerate product development and identify new market opportunities. Perhaps the most transformative uses of AI so far have been in pharmaceutical R&D and in the exploration of new chemical structures.
- Use Cases:
  - Google's DeepMind AI algorithms can design entirely new chemical structures with target properties,[9] unlocking the potential for novel therapeutics that may not be discovered through traditional approaches.[10]
  - Roche and other pharmaceutical companies are using AI to analyse much more data and to predict the effectiveness of different compounds, enhancing and accelerating drug discovery and time to market for new drugs.
- **Risk Management & Fraud Detection:** AI algorithms identify patterns and anomalies in financial transactions, helping detect fraud and mitigate risks effectively.
- Use Cases:
  - Financial intermediaries and institutions like MasterCard, Visa, Amex, HSBC, and JP Morgan use AI tools to analyse, monitor and even intervene in real time to prevent potential fraudulent transactions.

[7] Concern as the gambling industry embraces AI
[8] 40 Detailed Artificial Intelligence Case Studies [2024] - DigitalDefynd
[9] Millions of new materials discovered with deep learning - Google DeepMind
[10] Google's DeepMind team utilised AI to predict the distribution of electrons within molecules which is a crucial step in understanding chemical bonding and reactions. Their model 'DM21" achieved unprecedented accuracy in predicting electron density, opening new possibilities for drug discovery and for materials design.

- **Talent Acquisition & Management:** AI-powered tools screen resumes, identify suitable candidates, and provide insights into employee performance and engagement, enhancing the recruitment and retention process.[11]

2.3. Small and Medium sized businesses (SMEs)

While SMEs may have limited resources compared to multinationals, they are increasingly adopting AI to optimise operations, improve customer experiences, and compete with larger players in all of the areas in which those larger organisations use AI.



In an interesting recent report published in April 2024[12], the Bipartisan Policy Center highlighted trends in use of AI for SMEs with accounting, customer management and marketing being the leading use cases followed by consumer insights, workforce management, project management and other.

SMEs tend to focus on easy to deploy, cloud based AI tools that provide a quick ROI and enable them to scale without continuous addition to their employee base or ever increasing outsourcing costs.

[11] 22 Examples of AI In HR and Recruiting to Know | Built In
[12] Small Businesses Matter: Navigating the AI Frontier | Bipartisan Policy Center

2.4. Development & Deployment Strategies

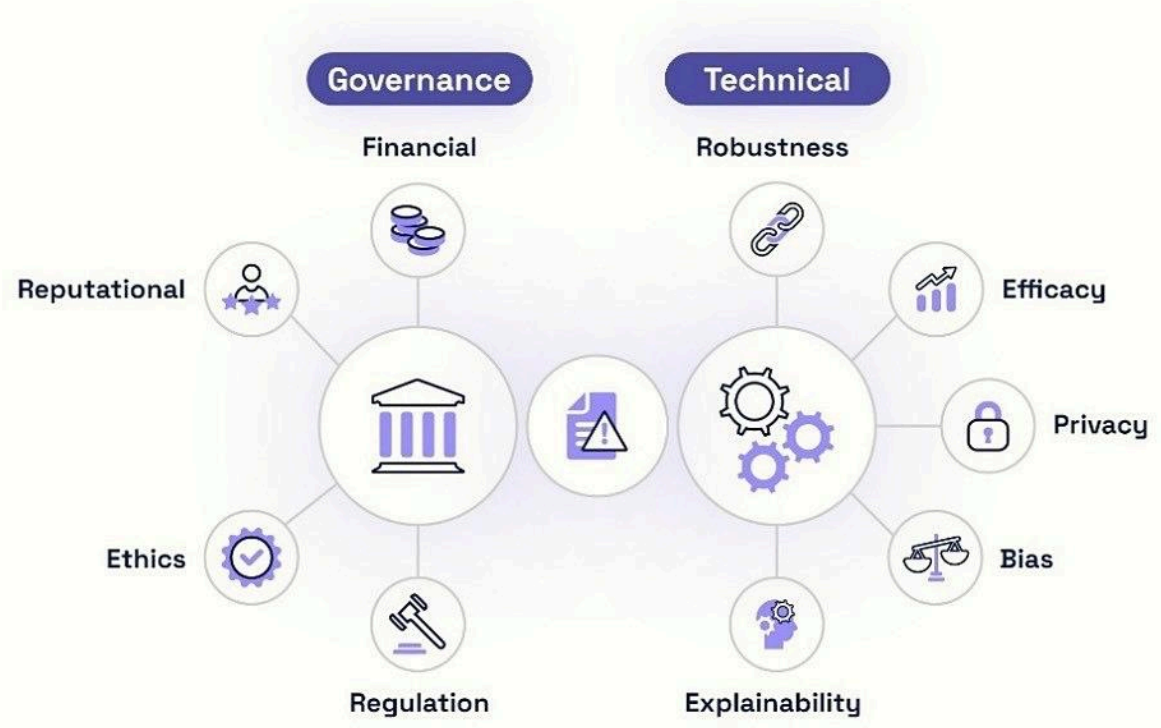
Multinationals typically have in-house AI development teams and invest in cutting-edge research to create proprietary AI solutions. They also collaborate with external AI startups and research institutions to stay at the forefront of technological advancements.

In contrast, SMEs often rely on third-party AI solutions and cloud-based platforms that are tailored to their specific needs and budgets. They prioritise AI applications that offer quick wins and tangible ROI, focusing on enhancing efficiency and improving customer experiences.

For those wishing to design their own AI systems and tools there are numerous sources of guidance on development and deployment processes to ensure the key ethical, legal and practical issues have been considered:

"*There are many ways of carving up the AI/ML project lifecycle*. However, it is important to work from a shared heuristic (an exploratory model) of the AI/ML project lifecycle that is optimally generalisable and practice-centred, while remaining algorithm neutral (i.e. that can apply to many different AI/ML techniques). It is also important that this approach prioritises the identification and mitigation of potential harms by taking into account the interwoven nature of its technical, social, and ethical aspects.*"

2.5. Why Should We Care About AI Governance?



13

[13] https://www.weforum.org/agenda/2023/03/why-businesses-should-commit-to-responsible-ai/

The rapid advancement of AI brings numerous ethical implications, particularly in areas like digital amplification, cybersecurity, bias, job displacement, data privacy, the 'digital divide' and in modern warfare.

2.5.1. Digital Amplification

Digital amplification refers to AI's ability to enhance the reach and influence of digital content, often through algorithms that prioritise certain information, shape public opinion, and amplify specific voices. This phenomenon raises ethical concerns about fairness, transparency, and potential misinformation. To counteract negative effects, businesses can encourage diverse participation in data collection and decision-making, promote open dialogue, and regularly review AI systems for fairness.[14]

2.5.2. Digital Divide

The digital divide refers to the gap between those who have access to modern information and communication technology and those who do not. AI can exacerbate this divide, as access to AI technologies often requires significant resources. For instance, advanced AI tools and education are more accessible in developed countries, leaving developing nations at a disadvantage. This disparity can lead to unequal opportunities in education, healthcare, and economic growth. Efforts to bridge this divide include initiatives like Google's AI for Social Good,[15] which aims to make AI technologies more accessible and beneficial to underserved communities.

2.5.3. Job Displacement

AI's ability to automate tasks traditionally performed by humans raises significant concerns about job displacement. For instance, in manufacturing, robots and AI systems can perform repetitive tasks more efficiently than humans, leading to reduced demand for human labour. A notable example is Amazon's use of AI-driven robots in warehouses, which has streamlined operations but also led to concerns about job losses. While AI can create new job opportunities, the transition period can be challenging for workers needing to reskill.

2.5.4. Bias and Discrimination

The risk of perpetuating existing unfair human bias is high. For example, in 2018, Amazon developed an AI-powered recruiting tool that showed bias against female candidates, highlighting how AI can perpetuate existing biases in hiring processes.[16] Generative AI Systems built on preexisting human data and decisions could become a high-tech echo chamber for our historic prejudices.

---

[14] 5 Ethical Considerations of AI in Business
[15] Google AI and Social Good
[16] Amazon's sexist hiring algorithm could still be better than a human - IMD business school for management and leadership courses

2.5.5. Cybersecurity

AI plays a dual role in cybersecurity, both mitigating and potentially promoting cybersecurity and spyware issues.[17]

- **Mitigation:** AI enhances cybersecurity by enabling real-time threat detection and response. Machine learning algorithms can analyse vast amounts of data to identify patterns and anomalies indicative of cyber threats. For example, AI-driven systems can detect and respond to phishing attacks, malware, and unauthorised access attempts more swiftly than traditional methods. AI can also automate routine security tasks, freeing up human experts to focus on more complex issues. Companies like Darktrace use AI to create self-learning cybersecurity systems that adapt to new threats autonomously.
- **Bad Actors:** Conversely, AI can also be exploited by cybercriminals. AI-powered tools can automate and enhance the sophistication of cyberattacks. For instance, AI can be used to develop more effective phishing schemes, create adaptive malware, and conduct large-scale attacks like Distributed Denial of Service (DDoS) and deepfakes more efficiently.[18] Additionally, AI can be used to bypass traditional security measures by learning and mimicking legitimate user behaviour or by hidden attacks on the very AI systems that are being used to manage email and data storage cybersecurity risks.[19] Balancing these aspects requires robust AI governance and ethical guidelines to ensure AI technologies are used responsibly and effectively to protect against cyber threats while minimising the risk of misuse.[20]

2.5.6. Privacy Concerns

AI systems often rely on vast amounts of data to function effectively, which can lead to privacy issues. For example, facial recognition technology used by law enforcement agencies can enhance security but also raises concerns about surveillance and the potential misuse of personal data.

2.5.7. Political Actor Misuse

The Cambridge Analytica scandal[21] is a prominent case where AI algorithms were used to harvest and exploit personal data from millions of Facebook users without their consent, highlighting the need for stringent data protection regulations and the risk of misuse by bad political and governmental agencies.

---

[17] AI and Cybersecurity: The Dual Role of Automation in Threat Mitigation and Attack Facilitation (itprotoday.com)
[18] AI and cybersecurity: Navigating the risks and opportunities | World Economic Forum (weforum.org)
[19] 4 use cases for AI in cyber security
[20] NIST Identifies Types of Cyberattacks That Manipulate Behavior of AI Systems | NIST
[21] POLITICO AI: Decoded: How Cambridge Analytica used AI — No, Google didn't call for a ban on face recognition — Restricting AI exports

2.5.8. AI In Warfare

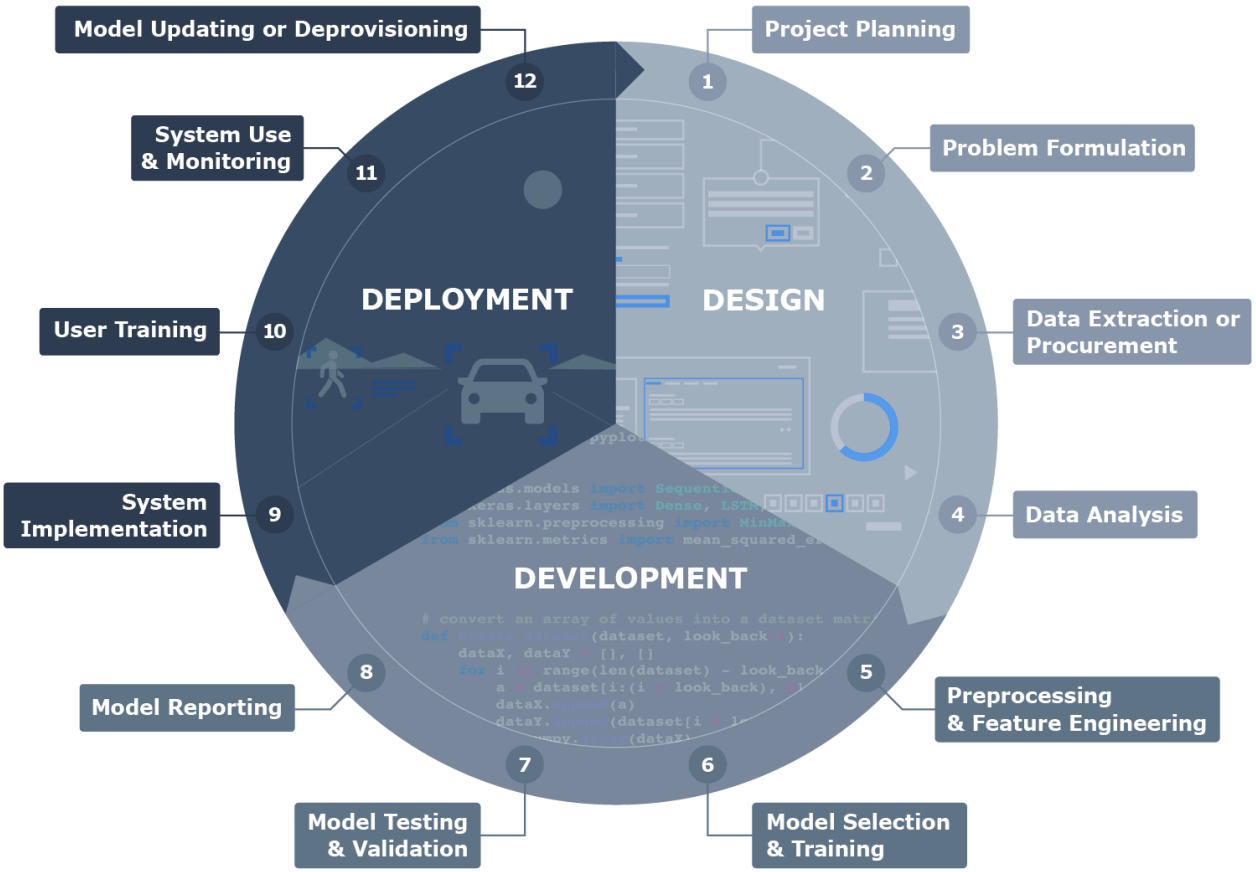
AI is increasingly integrated into military operations, offering both significant benefits and notable risks.

- **Uses and Benefits:** AI enhances military capabilities through improved decision-making, autonomous systems, and predictive maintenance. For example, AI-driven drones can conduct surveillance and reconnaissance missions, reducing the risk to human soldiers.[22] AI algorithms can analyse vast amounts of data to provide real-time intelligence, helping military leaders make informed decisions quickly.[23] Predictive maintenance, as used by the US Air Force, helps identify potential equipment failures before they occur, ensuring operational readiness.
- **Risks:** However, using AI in warfare also presents significant risks. Autonomous weapons systems, which can identify and engage targets without human intervention, raise ethical and legal concerns.[24] There is a risk of AI systems making errors in target identification, potentially leading to unintended civilian casualties. The weaponization of AI can lead to an arms race,[25] with nations developing increasingly advanced AI technologies to gain a strategic advantage. This could destabilise global security and increase the likelihood of conflicts. Balancing these benefits and risks requires robust international regulations and ethical guidelines to ensure AI technologies are used responsibly in military contexts.

Addressing these ethical implications requires a balanced approach, involving robust regulations, ethical guidelines, and initiatives to ensure that AI benefits all of society equitably.

[22] On the warpath: AI's role in the defence industry
[23] Algorithms of war: The use of artificial intelligence in decision making in armed conflict
[24] Militarization of AI Has Severe Implications for Global Security and Warfare | United Nations University
[25] Strategic competition in the age of AI: Emerging risks and opportunities from military use of artificial intelligence | RAND

Part 2 - AI Governance



26

3. What is AI Governance?

AI governance refers to the frameworks, rules, and standards that guide the development and use of AI systems to ensure they are safe, ethical, and aligned with societal values. It involves:

- Establishing oversight mechanisms to address risks like bias, privacy infringement, and misuse
- Fostering innovation and trust in AI technologies
- Engaging diverse stakeholders in the governance process
- Mitigating human biases and errors in AI development
- Implementing responsible and ethical AI practices
- Addressing risks through policy, regulation, and data governance
- Aligning AI behaviours with ethical standards and societal expectations

[26] Alan Turing Institute: AI Ethics and Governance in Practice

The goal is to create a structured approach that maximises AI's benefits while minimising potential harms and ensuring accountability in AI development and deployment.[27]

4. Why AI Governance is Crucial for Organisations

As organisations increasingly adopt AI technologies, effective AI governance has become essential for ensuring responsible use and maximising benefits. Without robust AI governance, organisations risk facing operational, reputational, and legal challenges.

AI systems are not infallible. They can introduce biases, make incorrect decisions, or function in ways that are opaque to their operators. This lack of transparency can erode trust with stakeholders, from customers to regulators. For instance, AI algorithms used in hiring may unintentionally discriminate against certain groups if they rely on biased historical data. Similarly, AI in finance or healthcare can make life-altering decisions, such as approving loans or diagnosing diseases, where the stakes are extremely high. In such critical scenarios, poorly governed AI could lead to financial loss, ethical violations, or even physical harm to individuals.[28]

Furthermore, legal and regulatory compliance around AI is rapidly evolving. Governments and regulatory bodies are increasingly scrutinising how AI technologies are used, particularly regarding data privacy, discrimination, and accountability.[29] In the European Union, for example, the AI Act imposes stringent regulations on high-risk AI systems. Organisations that fail to meet these evolving legal requirements may face penalties, lawsuits, or forced discontinuation of AI-related projects.

4.1. Potential Risks of Inadequate AI Governance

- **Bias and Discrimination:** AI systems trained on biased data sets can amplify existing societal biases, leading to discriminatory outcomes. This can damage an organisation's reputation and cause legal liabilities.[30]
- **Product Liability:** The proposed AI Liability Directive (AILD) was formally withdrawn by the European Commission in February 2025, with the withdrawal confirmed definitively in July 2025, due to a lack of agreement among stakeholders and calls for regulatory simplification. AI liability in the EU is now primarily addressed through the Product Liability Directive 2024/2853 (which explicitly covers AI systems and software) and the AI Act's own compliance framework. New dedicated AI liability rules are not expected until after the AI Act is fully implemented.

[27] What Is AI Governance? | IBM
[28] AI in healthcare: what are the risks for the NHS? - BBC News
[29] Here's why organizations should commit to responsible AI | World Economic Forum
[30] In addition to existing laws, we note the proposed specific EU AI Liability Directive (AILD), introduced on September 28, 2022. AILD, if enacted, seeks to create harmonised rules across EU member states concerning damages caused by AI, focusing particularly on easing the burden of proof for claimants in liability cases. This was subsequently withdrawn.

- **Lack of Accountability:** Without clear governance structures, it may be difficult to hold anyone accountable when AI systems fail. Organisations may struggle to explain or rectify situations where AI decisions negatively impact individuals or society.
- **Data Protection and Privacy Violations**: Many AI systems rely on large amounts of personal data. If not properly governed, organisations may inadvertently violate data protection laws, such as GDPR, exposing themselves to hefty fines and reputational damage.
- **Human-centric AI and Human Rights**: AI should be developed and deployed in a way that is good for the end-users and society more generally. Respect for and protection of fundamental human rights, the desire for human autonomy and the right of human creators to benefit economically from use of their work should be integral to the AI Governance Framework.
- **Ethical Violations**: AI decisions that conflict with societal norms or ethical standards can cause public outrage. For example, using AI in surveillance or facial recognition can raise concerns about privacy and civil liberties.
- **Operational Risks**: Poorly governed AI systems can malfunction, make incorrect decisions, or disrupt critical operations. This could lead to costly mistakes, such as erroneous financial transactions, medical misdiagnosis, or faulty product recommendations.

4.2. Benefits of a Well-Structured AI Governance Framework

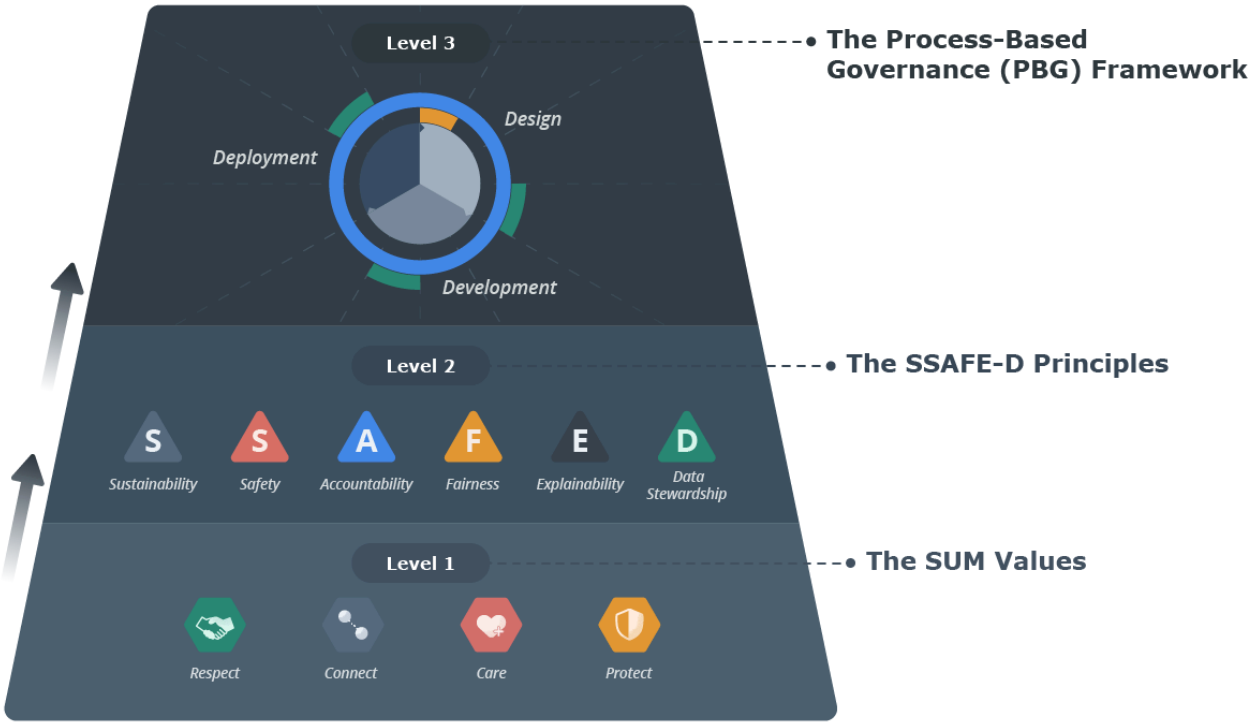
- **Risk Mitigation**: A well-structured AI governance framework helps organisations identify, assess, and mitigate risks associated with AI systems. This includes addressing biases, ensuring data privacy, and maintaining transparency in AI decision-making processes[4].
- **Legal Compliance**: AI governance frameworks ensure that AI deployments comply with existing regulations and emerging legal requirements. This reduces the risk of fines, lawsuits, and regulatory interventions, ensuring business continuity[5].
- **Enhanced Transparency and Accountability**: Governance frameworks provide transparency into how AI systems make decisions and who is responsible for overseeing them.
- **Improved Decision-Making:** Proper AI governance ensures that AI systems are aligned with organisational goals and values. It allows decision-makers to confidently use AI in critical areas like customer service, finance, and operations, knowing that risks have been minimised.
- **Reputation Management:** Ethical AI practices can enhance an organisation's reputation, building trust with customers and stakeholders. Companies that lead in AI governance are more likely to be seen as innovators and socially responsible, attracting customers, investors, and top talent.
- **Improved Innovation:** While governance is about risk management, it also enables innovation by providing a structured environment for AI development. A clear governance

framework encourages experimentation within defined ethical and legal boundaries, fostering AI innovations that can provide a competitive edge.

5. Governance and Accountability

In order to manage the various regulatory obligations arising from the use or development of AI tools it is essential that your organisation takes the following steps:

- **Establish an AI Governance Committee or**, for SMEs, identify a person within your organisation that can be the **AI Governance Officer**
- **Define the scope, roles and responsibilities** of the persons responsible for AI Governance. Ensure the identified persons are responsible for developing, maintaining and monitoring your AI Risk Management framework including policy and procedures
- **Ensure cross-functional collaboration** (legal, compliance, IT, engineering, etc.)
- Consider whether you can **identify/create an industry association** that will allow cross-organisational collaboration including know-how and resource sharing



[31]

6. AI Governance Frameworks

[31] Process-Based Governance (PBG) Framework

AI Governance frameworks need to be process driven. The Alan Turing Institute has helpfully set out a detailed guide to Process Based AI Governance in Action[32] and recommends the following key requirements:

- **Document ethical considerations and decisions:** AI ethics guides moral conduct in AI development and use. The key principles are: Sustainability, Safety, Accountability, Fairness, Explainability and Data Stewardship.
- **AI Project Lifecycle:** Outlines the stages where ethical questions arise and decisions are made, from project design and model development to system deployment and monitoring.
- **Risk Assessment and Mitigation:** Organisations should conduct context-based risk assessments (COBRA) to identify and mitigate potential ethical risks throughout the project lifecycle.
- **Stakeholder Engagement:** Engaging stakeholders is crucial for understanding the potential impacts of AI systems and ensuring that diverse perspectives are considered.
- **Operationalising Ethical Principles:** Organisations need to translate ethical principles into practical guidelines, policies, and procedures. Core attributes help specify and operationalize principles within the project context.
- **Bias Self-Assessment:** Conducting bias self-assessments helps identify and address potential biases in AI systems, particularly concerning fairness.
- **Data Protection and Intellectual Property:** Organisations must consider data protection, privacy, transparency, and intellectual property implications when developing and deploying AI systems.
- **Monitoring, Evaluation, and Communication:** Ongoing monitoring and evaluation are essential to ensure that AI systems remain aligned with ethical principles and to maintain public trust.

In addition, AI Governance frameworks need to consider the following key issues:

- Scope of the AI Governance Officer or the AI Governance Committee (including procedures for decision making)
- Permitted use of approved third party AI tools within your organisation
- Policy, procedure and rules for developing AI solutions internally and when working with suppliers
- Independent standards which your organisation may wish to deploy (e.g. ISO/IEC 42001:2023 and ISO 37000:2021)
- Policies and Procedures:
    - Need for continuous risk assessment processes

[32] Process Based Governance in Action - Guide

- Identification and analysis of risks (health, safety, fundamental rights)
- Risk evaluation (scored for severity and likelihood) and mitigation strategies
- Record-keeping (e.g. an AI Risk Register)
- Reporting requirements to the AI Governance Committee / executive governing body
- Periodicity of review

- Stakeholder Compliance:
  - Situating your AI Governance Framework within the various regulatory frameworks and sectoral guidance
  - Reporting requirements to regulators or capital markets.

7. AI Governance for SMEs

Small and medium enterprises will find it difficult to implement and maintain and full blown AI Governance Framework given the resources required. Therefore they should focus on the essential requirements of an AI Governance Framework detailing how the organisation will deploy AI, identify an AI Officer and the reporting lines and ensure there is a Risk Register in place.

It is essential that SMEs look for freely available guidance and tap into similarly resource constrained organisations within their sector that face the same technology and regulatory compliance challenges.

Fundamental AI compliance steps for SMEs:

- Identify an AI Governance expert and overall AI risk controller: choose the best person or group within the organisation to develop and manage your framework and report to the Board.
- Identify Regulations: Begin by understanding the specific regulations they must comply with, such as GDPR, HIPAA, AML, or industry-specific rules.
- Undertake the Risk Assessment: Evaluate which areas of the business have the highest compliance risks (e.g., data privacy, financial reporting, customer due diligence).
- Create and maintain the Risk Register: start to identify, monitor and manage AI risks. This should include the risks of issues such as model drift when AI systems can deviate from their intended performance over time as real-world data evolves but the underlying AI system training data is not keeping up, leading to incorrect predictions, biases, or other risks.
- Funding: see whether there is any SME funding available for AI use and compliance
- Industry Support: evaluate industry and sector specific support (e.g. industry associations)
- Identify third party AI compliance solutions that:

- Meet your core requirements
- Are regularly updated for changes in law and compliance standards
- Have the highest levels of data security and comply with key certification standards
- Can integrate with other databases (HR, sales, customer service, reporting)
- Have a good user interface (so can be used by users of varying technical ability) and provide staff training
- Do not lock you into their environments (i.e., will not be difficult to migrate from or substitute)

8. Additional AI Governance Tools

- OECD AI Principles:[33] These are the first intergovernmental standards on AI promoting innovative, trustworthy AI that respects human rights and democratic values. Composed of five values-based principles and five recommendations that provide practical and flexible guidance for policymakers and AI actors.
- AI Fairness 360:[34] IBM's open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle.
- Explainable AI: IBM, Google[35] and others provide toolkits on explainable AI.
- AI Governance Resources & Toolkits: the UK AI Standards Hub,[36] Google,[37] the European Commission,[38] IBM,[39] Microsoft,[40] Stanford University[41] and the Alan Turing Institute[42] (as well as many other organisations and commercial advisors) provide detailed resources, toolkits and/or training on AI and Ethics, Responsible AI Development and AI Governance.

9. AI Risk Register

Regulated organisations have experience of monitoring, managing and mitigating risks as part of their regulatory obligations to ensure they understand and manage risks appropriately. They will usually maintain a risk register that identifies known risks and potential risks and evaluate their likelihood and impact. Risk registers will usually include:

- The cause of the risk

[33] AI principles | OECD
[34] AI Fairness 360 (ibm.com)
[35] Explainable AI | Google Cloud
[36] AI Standards Hub
[37] Responsible AI: Applying AI Principles with Google Cloud
[38] Implementing AI Governance: from Framework to Practice | Futurium (europa.eu)
[39] What Is AI Governance? | IBM
[40] Empowering responsible AI practices | Microsoft AI
[41] AI Index
[42] Process Based Governance in Action | The Alan Turing Institute

- A description
- Action taken (if any) to manage the risk (controls)
- Pre-control and post-control:
    - Impact score
    - Likelihood score
- Who is responsible for controlling the risk identified?
- What KPIs or reporting processes will be put in place to measure the success of the Framework?

See below a simplified example of a risk register (with 1 being low and 6 high).

AI Use Case	Identified Risks	Impact (1-6)	Likelihood (1-6)	Mitigation Strategies
Customer Service Chatbot	Data Privacy Breach	3	5	Implement data encryption and access controls
Predictive Analytics Tool	Bias in Decision Making	6	2	Regular audits of algorithms for fairness

10. Independent Standards and Guidance

There are a number of independent standards which organisations can adopt to help ensure they have a comprehensive and methodical approach to AI Governance:

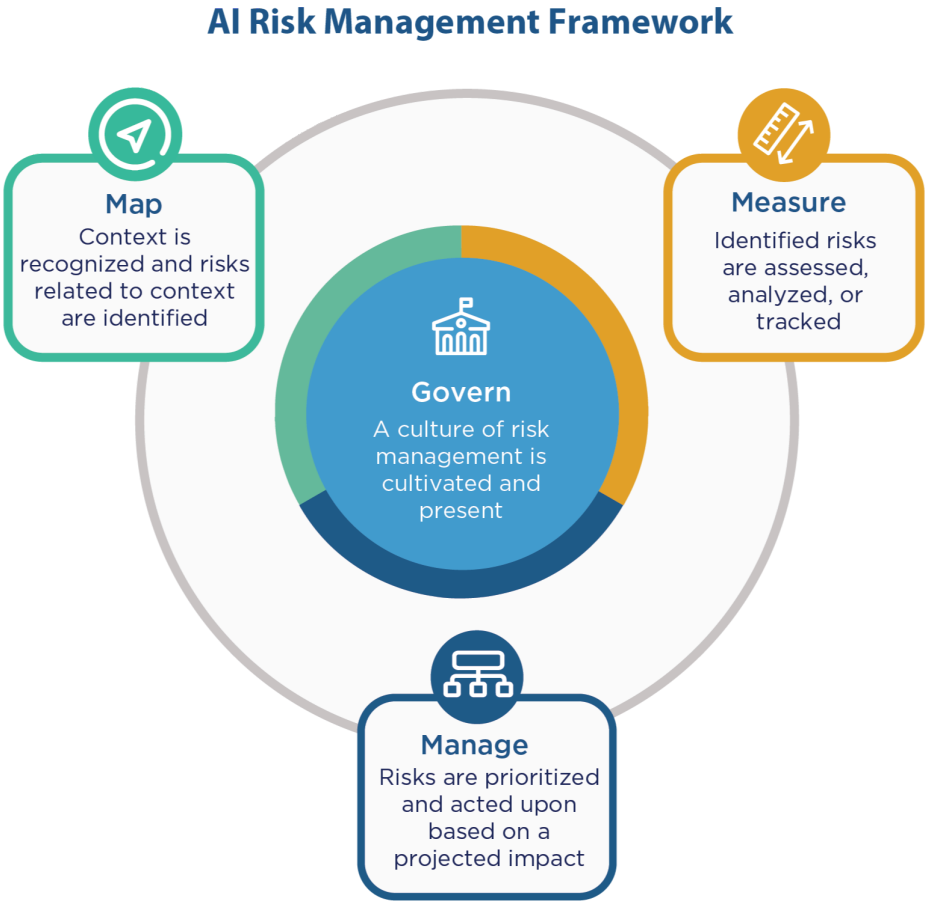
10.1. ISO/IEC 42001:2023 Framework for AI Management Systems:[43]

- **ISO/IEC 42001:** this establishes a comprehensive framework for organisations to implement, maintain, and improve AI Management Systems (AIMS). This standard promotes ethical, secure, and transparent practices in AI development and deployment across various industries.
- **Holistic Approach:** Unlike technical standards that focus solely on specific applications, ISO/IEC 42001 offers a holistic approach to managing AI-related risks and opportunities. It emphasises integration with existing organisational processes and continuous improvement.
- **Key Principles:** The standard outlines essential principles such as ethical AI practices, risk management, data governance, and stakeholder communication. It requires organisations to

[43] ISO/IEC 42001:2023 - AI management systems

conduct AI risk assessments and AI impact assessments to understand the broader consequences of AI deployment.

- **Complementary Standards:** ISO/IEC 42001 is supported by other standards like ISO/IEC 38507 (Governance Implications of AI) and ISO/IEC 23894 (AI Risk Management), which provide additional guidance on specific aspects of AI governance as well as the more general ISO 37000:2021 (Governance of Organizations).
- **Certification Benefits:** Achieving certification under ISO/IEC 42001 demonstrates an organisation's commitment to responsible AI governance. It can enhance trust among stakeholders, mitigate risks, and provide a competitive advantage in the marketplace.
- **Regulatory Approval:** ISO 42001 is the world's first certifiable AI management system standard — unlike NIST RMF, it can lead to formal third-party certification. It aligns closely with the EU AI Act's risk-based approach, making it a natural compliance bridge for Gibraltar firms with EU market exposure. Certification involves Annex A controls covering AI impact assessments, bias mitigation, audit trails, and third-party supplier oversight



10.2. NIST AI Risk Management Framework (AI RMF):[45]

- **NIST AI RMF**: provides a structured approach to identifying, assessing, managing, and mitigating risks associated with AI technologies. It emphasises the importance of integrating risk management into the entire lifecycle of AI systems.
- **Focus on Ethical Considerations:** Similar to ISO/IEC 42001, the NIST framework addresses ethical considerations in AI deployment. It encourages organisations to prioritise fairness, accountability, transparency, and privacy in their AI initiatives.
- **Guidance on Implementation:** The NIST framework offers practical guidance for organisations to implement effective risk management strategies tailored to their specific contexts and needs. This includes establishing governance structures that align with organisational objectives and regulatory requirements.
- **Alignment with International Standards:** The NIST AI RMF aligns with international standards like ISO/IEC 42001, reinforcing the importance of global best practices in AI governance. Organisations can leverage both frameworks to enhance their overall governance strategies.

10.3. Benefits of adhering to industry standards:

- **Impact on AI Governance Enhanced Accountability:** Both ISO/IEC 42001 and the NIST AI RMF promote accountability by establishing clear guidelines for ethical conduct in AI development and deployment.
- **Risk Mitigation:** These standards provide structured methodologies for identifying and mitigating risks associated with AI technologies, ensuring safer operations.
- **Stakeholder Trust:** By adhering to these standards, organisations can build trust with stakeholders by demonstrating their commitment to responsible practices.
- **Compliance with Regulations:** Implementing these frameworks helps organisations comply with emerging regulations related to AI governance, such as those being developed in the EU and other jurisdictions.
- **Continuous Improvement:** Both frameworks emphasise the need for continuous monitoring and improvement of AI systems to adapt to evolving challenges and technological advancements.

[44] NIST AIRC - AI RMF
[45] AI Risk Management Framework | NIST

10.4. Comparing different standards

IST AI RMF is highly adaptable and risk-focused, making it suitable for organisations prioritising AI risk management, whereas ISO/IEC 42001 provides a structured approach to ethical AI management, suitable for organisations emphasising transparency and compliance.

11. Third Party AI Compliance Tools

- **Automated monitoring and reporting:** Use affordable AI compliance software that automates monitoring and reporting tasks.
- **Cloud Solutions:** Consider cloud-based solutions that offer scalability and reduce infrastructure costs while providing necessary compliance features.
- **Bespoke risks**: consider, for example, whether you intend to use AI in a manner that risks copyright, design right or other intellectual property infringement and whether the proposed AI service provider includes insurance or other forms of support in the event of a claim of infringement.[46]

[46] Further reading: Copyright, AI and Generative Art - Ramparts

Part 3 - Key Legislative Requirements

12. Overview

The US, China and the EU have taken the lead with federal-level laws regulating AI development. The EU recently passed a new AI specific law and the UK is developing its own initiatives.

Developments have also being taking place at UN,[47] OECD[48] and G7[49] levels. This diversity of jurisdictional approaches – in addition to sector-specific guidance (e.g. financial services, medical devices, healthcare) – makes it challenging for organisations to understand how to navigate the compliance requirements and various AI standards when seeking to develop or implement AI solutions.

12.1. US AI Regulation

The United States has seen increasing federal and state initiatives to regulate AI:

- At the federal level, the National Institute of Standards and Technology (**NIST**) released its AI Risk Management Framework in January 2023 (**AI RMF 100-1**),[50] providing voluntary guidelines for organisations to assess and manage AI risks and subsequently the Generative Artificial Intelligence Profile (AI 600-1).[51] This has since been updated in 2025:
    - Expanded taxonomy of AI-specific threats: data poisoning, evasion attacks, model extraction, and hallucination risks
    - New guidance on generative AI and LLM-specific vulnerabilities
    - Stronger emphasis on supply chain and third-party model risk — particularly relevant for firms relying on external AI APIs or open-source models
- **The Blueprint for an AI Bill of Rights** (October 2022): Issued by the White House Office of Science and Technology Policy (**OSTP**)[52], this framework outlines five principles to protect the public from harms associated with AI.
- **The Algorithmic Accountability Act** of 2022, if passed, would require the Federal Trade Commission to require large tech companies to conduct AI impact assessments for bias and effectiveness of their automated decision systems.
- The **Artificial Intelligence Research, Innovation, and Accountability Act of 2023 (AIRIA)**,[53] introduced in November 2023, is a bipartisan legislative proposal aimed at establishing a governance framework for AI in the U.S. Its key goals are to foster innovation

[47] https://www.un.org/en/information-integrity/code-of-conduct
[48] OECD AI Principles overview and the The Global Partnership on Artificial Intelligence (GPAI): Global Partnership on Artificial Intelligence | OECD
[49] G7 Leaders' Statement on the Hiroshima AI Process | Shaping Europe's digital future
[50] AI RMF Development | NIST
[51] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile
[52] Blueprint for an AI Bill of Rights | OSTP | The White House
[53] S.3312 - 118th Congress (2023-2024): Artificial Intelligence Research, Innovation, and Accountability Act of 2023

while ensuring transparency, accountability, and security in AI applications, particularly in high-risk areas like critical infrastructure.

- The **National AI Initiative Act** of 2021[54] established a coordinated federal strategy for AI research and development which includes federal investment in research and development.[55]
- The U.S. Department of Defense's **Joint Artificial Intelligence Center (JAIC)** and the **National Security Commission on AI (NSCAI)** have been advancing AI research and application for defence purposes. The 2021 report from the NSCAI emphasises the need for U.S. leadership in AI, particularly in military and national security contexts.[56]
- At the state level, the California Privacy Protection Agency (**CCPA**) has taken a leading role under its **Consumer Privacy Act**, which includes draft regulations related to automated decision-making.[57] New York City implemented regulations (NYC-144) on using AI in employment decisions.[58]
- In addition, various federal agencies are tasked with supporting the AI Initiatives. For example, the Equal Employment Opportunities Commission (**EEOC**) has issued technical assistance to provide guidance on algorithmic fairness and the use of AI in employment decisions.[5960] The United States Patent and Trademark Office (**USPTO**) is also supporting AI innovation in intellectual property.[61]
- Some of the developments in AI legislation and guidance are a result of Presidential Executive Orders, such as **The President's Executive Order (EO) on Safe, Secure, and Trustworthy Artificial Intelligence** (14110) issued on October 30, 2023[62] which charges multiple agencies – including NIST – with producing guidelines and taking other actions. This was subsequently revoked by Trump. December 11, 2025 EO — **"Ensuring a National Policy Framework for Artificial Intelligence"** — which established a DOJ AI Litigation Task Force to challenge state AI laws and directed the Commerce Department to identify and report on state laws obstructing national AI policy.

12.2. China's AI Regulations

China has various data privacy laws and AI specific regulations.

12.2.1. Data Privacy

- **Cybersecurity Law (CSL)**

[54] The White House Launches the National Artificial Intelligence Initiative Office
[55] National Artificial Intelligence Research and Development Strategic Plan 2023 Update | The White House
[56] National Security Commission on Artificial Intelligence
[57] Draft Automated Decisionmaking Technology Regulations (ca.gov)
[58] New York City Adopts Final Regulations on Use of AI in Hiring and Promotion | Littler Mendelson P.C.
[59] Artificial Intelligence and Algorithmic Fairness Initiative | U.S. Equal Employment Opportunity Commission
[60] AI and the Workplace: Employment Considerations | Insights | Skadden, Arps, Slate, Meagher & Flom LLP
[61] Artificial Intelligence | USPTO
[62] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence | The White House

- The foundation of China's cybersecurity framework
- Requires network operators to secure networks & protect user data
- Addresses critical information infrastructure & cross-border data transfers

- **Data Security Law (DSL)**
  - Focuses on data security & management
  - Categorizes data by importance, with stricter rules for sensitive data
  - Emphasises data localization & security assessments for cross-border transfers

- **Personal Information Protection Law (PIPL)**
  - China's comprehensive data protection law
  - Sets strict rules for personal data handling: consent, impact assessments, security
  - Closely aligned to the EU's GDPR in protecting individual privacy

12.2.2. AI Regulations

- **Algorithm Regulations**
  - Regulates recommendation algorithms used by internet information service providers.
  - Requires algorithmic transparency and explainability, especially for those affecting public opinion or social mobilisation
  - Prohibits algorithmic discrimination and the use of algorithms to manipulate user choices or induce addiction
  - Mandates user control, allowing users to opt out of personalised recommendations or access explanations for algorithmic decisions
  - Imposes compliance obligations, including algorithm filing and regular security assessments

- Provisional Provisions on Management of Generative Artificial Intelligence Services (**GAI Measures**)
  - Specifically governs generative AI services that generate text, images, audio, video, or other content.
  - Emphasises adherence to socialist core values and avoidance of content that endangers national security, undermines social stability, or violates laws and regulations

- Requires providers to implement measures to prevent the generation of illegal or harmful content, including content labelling and user real-name registration
- Places responsibility on providers for the content generated by their services and mandates them to establish mechanisms for user feedback and complaint handling

- Provisions on Management of Deep Synthesis in Internet Information Service (**Deep Synthesis Provisions**)
  - Regulates the use of deep synthesis technologies, such as deepfakes, that generate or manipulate realistic images, audio, or video.
  - Requires clear labelling of deep synthesis content and prohibits its use for illegal or harmful purposes
  - Mandates consent from individuals whose likeness is used in deep synthesis content
  - Holds service providers accountable for preventing the misuse of deep synthesis technology and addressing user complaints

12.2.3. AI Governance

- China's approach to AI governance focuses on social order and stability and economic development. Key aspects:
  - **Social Stability:** China prioritises using AI to maintain social order and stability. This includes deploying AI in surveillance and public security systems. The government emphasises the need for AI systems to be transparent, fair, and free from bias to prevent social unrest.[63]
  - **Economic Development:** AI is seen as a critical driver of economic growth. China aims to become a global leader in AI by investing heavily in AI research and development, fostering innovation, and supporting AI startups. The government encourages integrating AI across various industries to boost productivity and competitiveness.[64]

12.2.4. The Cyberspace Administration of China (**CAC**)

CAC is central to AI regulation. It is responsible for:

- **Policy Formulation:** The CAC drafts and enforces regulations related to cybersecurity, data protection, and AI. It has issued guidelines on the ethical use of AI, emphasising the need for AI systems to align with socialist values and ethical norms.[65]

[63] China's AI Regulations and How They Get Made - Carnegie Endowment for International Peace
[64] China vs US Approaches to AI Governance – The Diplomat
[65] AI Governance in China: Strategies, Initiatives, and Key Considerations

- **Regulatory Oversight:** The CAC oversees the implementation of AI regulations, ensuring compliance by companies and organisations. It conducts audits and assessments to monitor the use of AI technologies and their impact on society.
- **International Collaboration:** The CAC engages in international dialogues on AI governance, contributing to the development of global AI standards and best practices. It aims to position China as a key player in the global AI regulatory landscape.

12.3. UK AI Governance

The National AI Strategy, launched in September 2021, outlines the UK's vision for AI over the next decade. It is built on three key pillars:[66]

- **Investing in the AI Ecosystem:** Ensuring sustained investment in AI research and development to maintain the UK's leadership in AI. Attracting and retaining top AI talent through education and training programs.
- **Transition to an AI-enabled Economy:** Promoting the integration of AI across various sectors to drive productivity and economic growth and ensuring that the benefits of AI are distributed across all regions and sectors of the UK economy.
- **Effective AI Governance:** Developing a pro-innovation regulatory framework that protects public values and encourages ethical AI use and engaging in global dialogues to shape international AI standards and practices.

Recent changes:

- Data (Use and Access) Act 2025, meaningfully relaxes the UK GDPR's Article 22 automated decision-making rules (a key area for AI governance) and raises PECR fines to £17.5 million or 4% of global turnover.
- A dedicated UK AI Bill is not expected to be introduced until the second half of 2026.

12.3.1. The Information Commissioner's Office (ICO)

Provides guidance on the use of AI for data protection purposes[67] and the Financial Conduct Authority has issued an update[68] on its measures to support the pro-innovation Govt White Paper on the use of AI[69] although we await a more detailed understanding of how this will be developed.[70] The ICO is expected to update the AI and data protection guidance in 2026.

12.3.2. The Alan Turing Institute

[66] National AI Strategy - GOV.UK
[67] Artificial intelligence | ICO
[68] Artificial Intelligence (AI) update – further to the Government's response to the AI White Paper | FCA
[69] A pro-innovation approach to AI regulation - GOV.UK
[70] AI in principle: the FCA's approach to the regulation of AI within financial services

The Institute has been a significant enabler of the UK AI Strategy and provides a range of research, training and development tools in AI Governance[71] including a guide to public sector AI development and deployment.[72]

12.3.3. Multi-Regulator Sandbox

An AI and Digital Regulation Cooperation Forum (**DRCF)** has been established with a multi-regulator sandbox pilot (**the AI and Digital Hub**) to support AI innovation and address cross-regulatory issues including in respect of financial services, telecoms and communications.[73]

12.4. Other Regions

12.4.1. Gibraltar

Gibraltar has no dedicated AI Act or sector-specific AI legislation as of March 2026. The Government of Gibraltar has signalled an intention to develop a regulatory framework for AI — the Chief Minister referenced this in late 2023 — but no formal legislation has followed. In the interim, AI governance for regulated firms is addressed through the existing principles-based regulatory framework administered by the Gibraltar Financial Services Commission (GFSC).

GFSC Policy Statement on the Use of Artificial Intelligence (January 2026)

On 29 January 2026, the GFSC published its first formal Policy Statement on the Use of Artificial Intelligence. This is the operative regulatory guidance for all GFSC-regulated firms deploying or considering deploying AI systems and should be treated as the baseline compliance reference for Gibraltar-based clients.

The GFSC's position is explicitly permissive and principles-based:

> *"The GFSC supports the responsible adoption of AI and does not intend to introduce prescriptive AI-specific rules at this stage. Firms are not required to seek prior approval before deploying AI, nor are they required to certify to any particular AI standard."*

Regulated firms should nevertheless approach AI adoption with the same rigour as any other material technical or strategic project. The GFSC expects firms to apply existing governance obligations, in particular:

- The **Finance Sector Code of Corporate Governance** — governing boards must be able to demonstrate oversight and accountability for AI systems, including an understanding of material risks, and should not treat AI as purely an IT or operational matter

[71] AI Ethics and Governance in Practice | The Alan Turing Institute
[72] Understanding artificial intelligence ethics and safety | The Alan Turing Institute
[73] AI and Digital Hub | DRCF

- The **Minimum Criteria for Licensing (MCL)** — fitness and propriety, adequate risk management, and the requirement for systems and controls to remain appropriate to the nature and scale of the business apply regardless of whether risk is introduced by human or automated processes

While the GFSC does not mandate adherence to any particular AI governance standard, the Policy Statement identifies three voluntary frameworks that regulated firms may find useful as structured approaches to AI risk management:

1. **NIST AI Risk Management Framework (AI RMF)**
2. **ISO/IEC 42001:2023** (AI Management Systems)
3. **NCSC AI Cybersecurity Guidelines**

The alignment between the GFSC's recommended frameworks and those covered in this guide means that firms following the governance approach set out in this document are well-positioned relative to regulatory expectations in Gibraltar.

The GFSC Policy Statement explicitly names all of the following as within scope of its guidance:

- Traditional **machine learning** and statistical models
- **Large Language Models (LLMs)** and generative AI systems
- **Agentic AI** — systems capable of autonomous, multi-step task execution
- AI used in **customer-facing**, **back-office**, and **decision-making** contexts

Firms should not interpret the absence of sector-specific AI rules as removing the need for governance. The GFSC makes clear that where AI introduces or amplifies material risk — whether to consumers, market integrity, or the firm's own resilience — existing regulatory obligations are engaged.

Regulated firms that encounter regulatory friction when deploying novel AI systems — for example, where existing rules were not designed with AI in mind — may engage the GFSC through established innovation pathways:

- **GFSC Innovation Sandbox** — allows firms to test products and services in a controlled environment with regulatory oversight
- **Pilot waivers** — time-limited exemptions from specific rule provisions
- **Rule redrafting engagement** — firms may raise proposed rule amendments directly with the GFSC where existing frameworks are poorly suited to AI-driven business models

Although Gibraltar is not part of the European Union, firms operating in or providing services into the EU market — including financial services firms passporting into EU Member States — remain subject to the **EU AI Act** to the extent they place AI systems on the EU market or provide AI-enabled services to EU users. The GFSC Policy Statement does not address the EU AI Act

directly; firms with EU exposure should cross-refer to Section 13 of this guide for a full analysis of applicable obligations.

12.4.2. Developing Countries

- AI governance in developing countries presents both significant challenges and opportunities. As AI becomes more prevalent, developing countries face the dual task of harnessing AI's potential for growth and ensuring its ethical and responsible use.
- Many developing countries rely on AI technologies developed by international companies or foreign governments, which could lead to issues with data sovereignty, privacy, and regulatory control. These countries may lack the bargaining power to shape AI technologies in ways that align with local laws and ethical standards.
- However, developing countries also have the potential to leapfrog stages of development by embracing AI technologies, bypassing traditional methods of economic growth (as we have seen in e-payments and telecoms). AI can help improve efficiencies in agriculture, healthcare, education, infrastructure, and public service delivery, providing a fast-track to modernisation.
- With the right governance frameworks and partnerships, these countries could benefit from sustainable growth, tackling poverty, and solving infrastructure issues through AI-driven solutions. Many developing countries are starting to engage in international collaborations for AI development and governance. Organisations like the United Nations, the World Bank, and regional bodies like the African Union have begun formulating strategies and offering support for AI governance.
- The **United Nations Sustainable Development Goals (SDGs)**[74][75] align closely with AI's potential for positive impact. AI can help developing countries achieve these goals more efficiently, such as through smart agriculture for food security, predictive health systems for disease outbreaks, and urban planning to address rapid urbanisation.
- AI governance frameworks focused on the SDGs would promote sustainability, equity, and long-term growth, allowing developing countries to harness AI for social good.

12.4.3. Canada

Artificial Intelligence and Data Act (AIDA): Canada has introduced the AIDA, which forms part of Bill C-27 (Digital Charter Implementation Act, 2022) which seeks to regulate high-impact AI systems, promoting responsible AI development and deployment. It emphasises transparency, fairness, and measures against algorithmic bias, with strong privacy protection rules. The act will also establish oversight bodies to ensure AI compliance.

[74] https://sdgs.un.org/goals
[75] AI for social good in sustainable development goals | McKinsey

12.4.4. Singapore

Singapore has taken a proactive stance with its voluntary guidelines through the **Model AI Governance Framework**, launched in 2019.[76] It focuses on four areas: internal governance, risk management, transparency, and accountability in AI systems.

- **Singapore AI Verify:** The government launched AI Verify,[77] a self-assessment framework that helps organisations ensure their AI systems meet ethical and transparency standards.
- Singapore's **Personal Data Protection Commission** (PDPC)[78] plays an active role in integrating AI governance into broader data protection and digital economy strategies.

12.4.5. United Arab Emirates (UAE)

- **AI Strategy 2031**:[79] The UAE has positioned itself as a global AI leader, with a national strategy that aims to integrate AI across various sectors, from healthcare to education. The country has appointed a **Minister of AI**[80] to oversee AI development and regulation.
- **Ethical AI Guidelines**: While the UAE's regulation of AI is still developing, it has laid down Principles & Guidelines[81] promoting fairness, accountability, and the responsible use of AI.

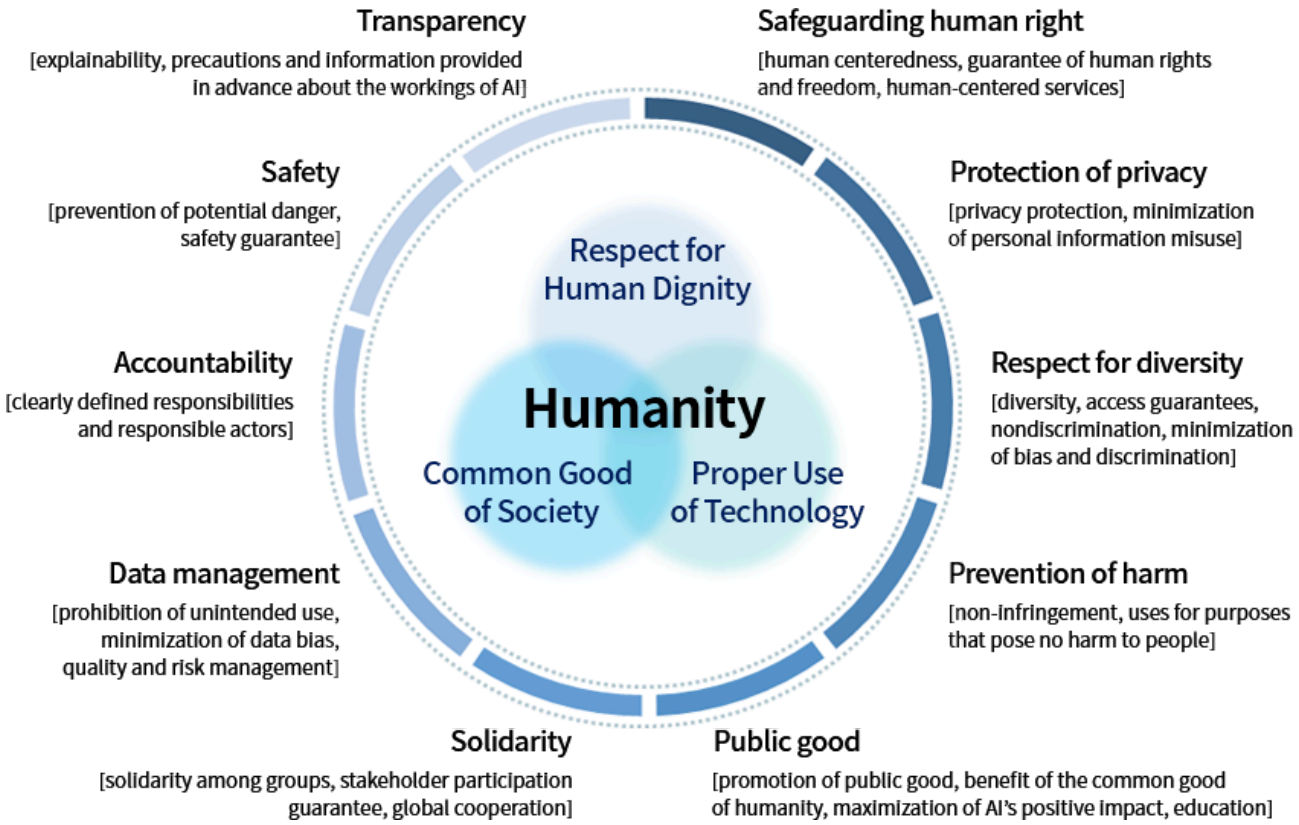
12.4.6. South Korea

The Framework Act on Intelligent Informatization was introduced in 2020, aiming to promote the ethical and responsible development of AI.[82] The act emphasises creating AI infrastructure and safety measures, fostering innovation while protecting data privacy.

- **South Korea's AI strategy**: includes the "National Strategy for Artificial Intelligence," created in 2019 which outlines plans for AI development and regulation.
- **AI Ethics Standards**: The government released AI ethics guidelines in 2020,[83] built around the principles of fairness, human rights, and transparency. South Korea is working towards more concrete regulations and frameworks for responsible AI use, particularly in industries like healthcare, education, and finance.
- **Korean New Deal**: As part of a national plan to become a global AI leader, South Korea has also launched initiatives to create a data-driven economy and develop national AI capabilities to ensure it in the top 3 AI jurisdictions in the world by 2027.[84]

---

[76] PDPC | Singapore's Approach to AI Governance
[77] What is AI Verify
[78] PDPC | Advisory Guidelines on use of Personal Data in AI Recommendation and Decision Systems
[79] UAE National Strategy for Artificial Intelligence 2031
[80] About | Artificial Intelligence Office, UAE
[81] AI ETHICS
[82] Legislative Trends in AI: National AI Ethical Standards and Amendment to Enforcement Decree of Framework Act on Intelligent Informatization - Kim & Chang
[83] The National Guidelines for AI Ethics | INTRODUCTION
[84] Korea aims to rank among top 3 global AI powerhouses



85

12.5. Sector-Specific Regulations

12.5.1. Healthcare

- **FDA Guidelines (USA):** The U.S. Food and Drug Administration (**FDA**) regulates AI in healthcare through its "Software as a Medical Device" (**SaMD**) framework. This framework ensures AI-based medical devices meet safety and efficacy standards.[86] The FDA also provides guidelines for the development and deployment of AI in healthcare, focusing on transparency, reliability, and patient safety.[87]
- **Japan:** Japan's Pharmaceuticals and Medical Devices Agency (**PMDA**) regulates AI in healthcare, ensuring that AI-based medical devices follow safety and efficacy standards. The PMDA also provides guidelines for developing and using AI in healthcare.

12.5.2. Finance

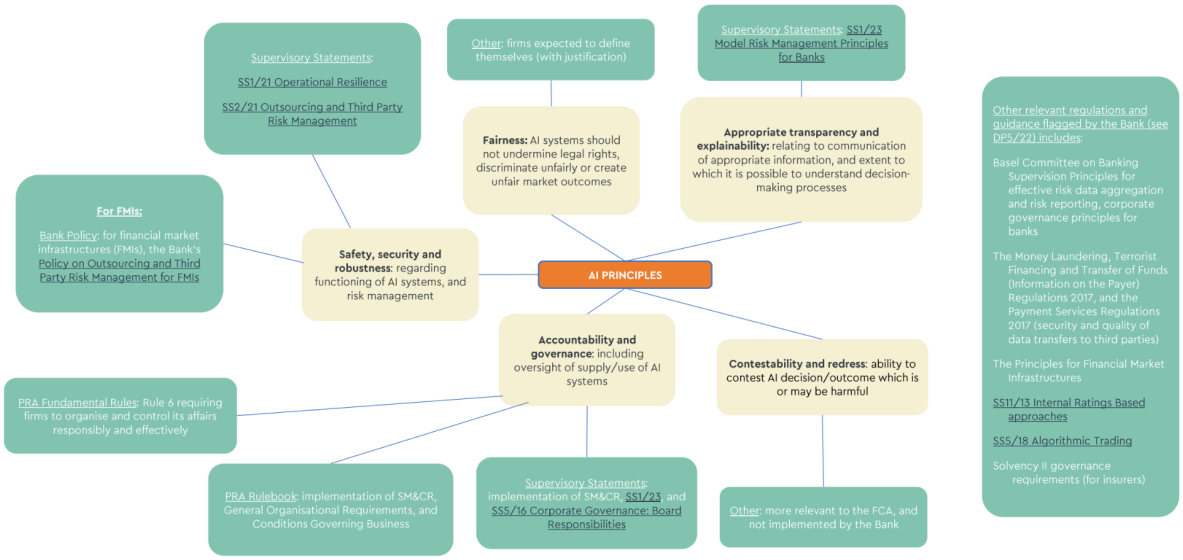
- **SEC Regulations (USA):** The U.S. Securities and Exchange Commission (**SEC**) oversees the use of AI in finance, focusing on transparency, fairness, and investor protection. The

[85] "Basic and comprehensive standards that should be followed by all members of society to implement human-centered AI": The National Guidelines for AI Ethics | INTRODUCTION
[86] How FDA Regulates Artificial Intelligence in Medical Products | The Pew Charitable Trusts
[87] Artificial Intelligence and Medical Products

SEC requires financial institutions to ensure that AI algorithms do not introduce biases or unfair practices.

TRAVERS SMITH

Dual-regulated firms (and FMIs): starting point for understanding the regulatory environment applicable to your use of AI



88

- The **UK Bank Of England** and the Financial Conduct Authority (**FCA**): The latest guidance from the Bank of England[89] and the FCA[90] emphasises a strategic approach to the use of AI in financial services. Released in April 2024, this guidance outlines the necessity for firms to clearly articulate their AI usage and associated risk management strategies to regulators. Firms are expected to have a comprehensive understanding of how AI is integrated at all operational levels, including third-party suppliers.
- Key aspects of the guidance include:
    - **Regulatory Expectations:** Firms must explain their AI deployment, focusing on risk identification, assessment, and management. The FCA stresses that a compelling narrative regarding AI risk management is essential for compliance.
    - **Consumer Protection:** The FCA's principles-based approach highlights the importance of fairness, accountability, and governance in AI systems to prevent consumer harm and ensure legal rights are upheld.
    - **Future Adaptations:** The FCA is open to evolving its regulatory framework as necessary, particularly in light of advancements in AI technologies like large

[88] The FCA and Bank of England's "strategic approach" to AI – what it means for regulated firms | Travers Smith
[89] DSIT-HMT letter - London
[90] AI Update | FCA

language models. It also plans to explore innovative regulatory engagement through an AI sandbox to test and assess the impact of AI in financial markets.

- **MAS Guidelines (Singapore)**: The Monetary Authority of Singapore (MAS) has issued guidelines for the use of AI in finance, emphasising fairness, ethics, accountability, and transparency (FEAT).[91] These guidelines aim to ensure that AI systems in finance are used responsibly and ethically. They have also issued an AI toolkit to assist with the responsible use of AI.[92]

12.5.3. Autonomous Vehicles

- **Japan:** Japan has established regulations for autonomous vehicles, focusing on safety, cybersecurity, and ethical considerations. The country aims to integrate autonomous vehicles into its transportation system while ensuring public safety.
- **South Korea:** South Korea's regulations for autonomous vehicles include safety standards, testing protocols, and guidelines for ethical AI use. The country is also investing in infrastructure to support the deployment of autonomous vehicles.

12.6. Regional Differences

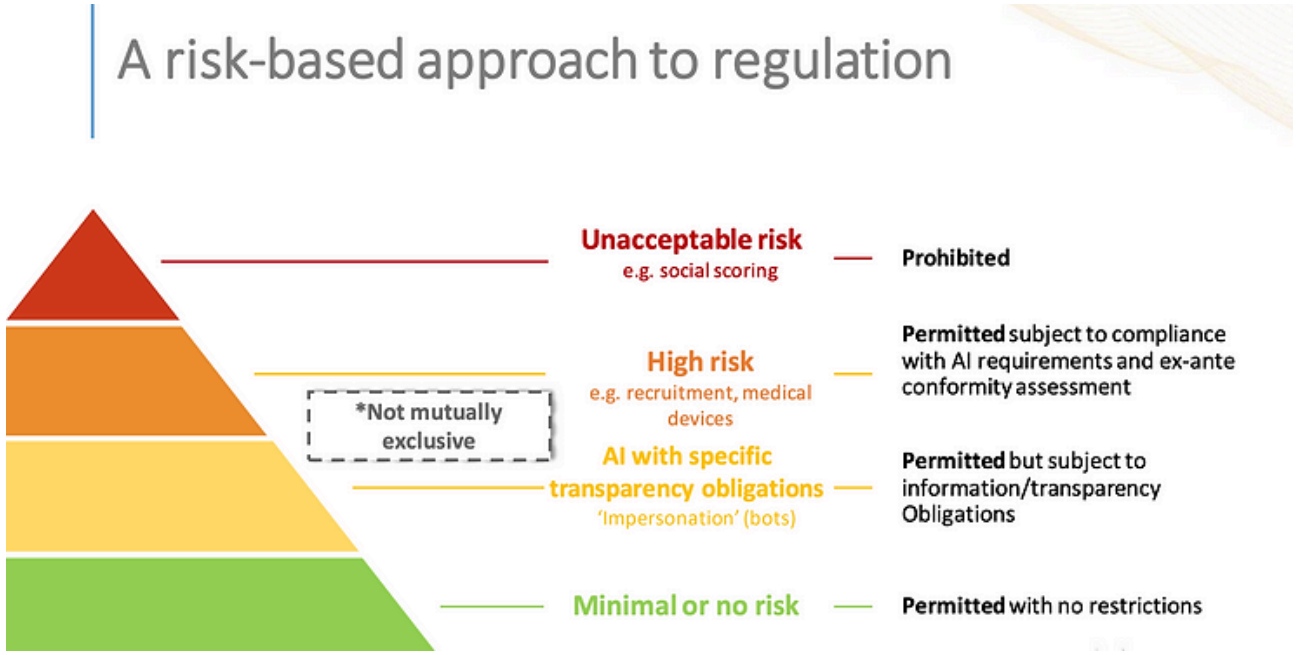
- The differences in approach to AI between the US and the EU particularly have led to concerns that whilst the US also seeks to encourage its development the EU is making it difficult for organisations to develop and deploy AI tools.[93] The UK is likely to seek a middle ground between the EU and US approaches.
- There are many sources of information on different laws and regulatory requirements across the world and on a sectoral specific basis. In addition, leading think-tanks and institutions public white papers, policy guidance and research on the state of play for AI technologies. In addition, many major law firms, accountancy practices and technology consultancy businesses maintain their own knowledge hubs.
- We have created an AI Law and Compliance Hub where we keep track of much of this information (including from other professional service providers) to make it easier for organisations (and particularly SMEs and service providers) to get up to speed with latest developments and best practice: AI Law & Compliance Knowledge Hub - Ramparts

13. The EU AI Act

The EU **AI Act** is a landmark regulatory framework designed to address the risks associated with the development, deployment, and use of AI across the European Union. It is part of the European

[91] Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector
[92] MAS-led Industry Consortium Releases Toolkit for Responsible Use of AI in the Financial Sector
[93] Big Tech criticises EU's AI regulation – is it justified?

Commission's broader strategy to position the EU as a leader in trustworthy AI while ensuring that the technology is safe and respects fundamental rights.



94

13.1. Risk-Based Approach

The AI Act categorises AI systems into four distinct risk levels:

- **Unacceptable Risk:** AI systems that are deemed to pose a significant threat to fundamental rights, such as social scoring by governments or real-time biometric identification in public spaces (except in certain narrowly defined cases). These are prohibited under the Act (Art. 5).
- **High Risk:** AI systems that have a significant impact on safety, security, or fundamental rights. This includes AI used in critical infrastructures (e.g., transport, health), educational systems, law enforcement, and employment processes. High-risk AI systems are subject to strict obligations, including:
    - Conformity assessments before market deployment
    - Risk management processes
    - Data quality requirements
    - Record-keeping and traceability
    - Human oversight and transparency obligations

94 The EU AI Act Explained: Tracking Developments for Responsible AI: image source https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf?

- **Limited Risk**: These systems pose moderate risks and are subject to transparency requirements, such as informing users they are interacting with AI. Examples include chatbots and deepfakes.
- **Minimal or No Risk:** These are AI applications that pose little to no risk, such as spam filters or AI used in video games. They are largely unregulated by the Act.

13.2. Transparency & Accountability Requirements

- Systems interacting with humans or manipulating content (e.g., deepfakes) must disclose their nature as AI.
- Developers and users of high-risk AI systems are required to maintain logs and documentation on their systems' development, deployment, and outcomes. This ensures traceability and allows authorities to audit compliance with the Act.
- Human Oversight: High-risk AI systems must be designed to enable human oversight, ensuring that decisions made by AI can be overridden or intervened upon in cases of malfunction or unintended outcomes.
- Data and Algorithm Governance: The AI Act mandates that datasets used to train high-risk AI systems must be:
    - Relevant and representative
    - Free from bias and discrimination
    - Of high quality to minimise risks and ensure accuracy
    - Proper documentation and traceability are required to track the dataset's origin and use.

13.3. EU AI Code Of Practice

- The AI Act also envisages an AI Code of Practice drafted by 13 AI experts[95] that will be designed to help providers of general-purpose AI (**GAI**) models comply with the AI Act.
- The EU AI Code of Practice will detail the rules for GAI providers, including transparency and copyright-related rules, as well as systemic risk taxonomy, risk assessment, and mitigation measures.[96] Compliance with the Code of Practice will be an important tool for demonstrating compliance with the AI Act.

13.4. Market Surveillance and Enforcement

- The AI Act sets up a European Artificial Intelligence Board (**EAIB**)[97] to oversee its implementation and ensure compliance across member states.

[95] European Commission appoints 13 experts to draft AI Code | Euronews
[96] AI Act: Participate in the drawing-up of the first General-Purpose AI Code of Practice | Shaping Europe's digital future
[97] European Artificial Intelligence Act comes into force

- Each member state will designate national authorities responsible for enforcement, market surveillance, and risk assessments.
- The maximum penalties for non-compliance depend on the nature of the infringement:
    - For non-compliance of the prohibited activities (as per Article 5): administrative fines of up to €35m or, if the offender is an undertaking, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher.
    - For other non-compliance: administrative fines of up to €15m or, if the offender is an undertaking, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher
    - The supply of incorrect, incomplete or misleading information to notified bodies or national competent authorities in reply to a request shall be subject to administrative fines of up to €7.5m or, if the offender is an undertaking, up to 1% of its total worldwide annual turnover for the preceding financial year, whichever is higher.
- **Extraterritorial Impact:** The AI Act applies not only to organisations operating within the EU but also to those outside the EU if their AI systems affect individuals within the EU. This ensures that global companies deploying AI in or for Europe must comply with the regulation.
- **Innovation and Regulatory Sandboxes:** The Act promotes AI innovation by encouraging the creation of regulatory sandboxes, where organisations can develop and test AI technologies under the supervision of national authorities without facing the full weight of regulatory requirements. This promotes innovation while managing risks.

13.5. Alignment with GDPR and Other EU Laws

The AI Act is designed to work alongside other major European regulations, such as the General Data Protection Regulation (GDPR), ensuring that AI systems respect privacy rights, data security, and human dignity.

13.6. Implementation

The AI Act entered into force on 1 August 2024 effective from 2 August 2026 (Art. 11), except for the following specific provisions listed in Art. 113:[98]

- Enforcement of Chapters I and II (general provisions, definitions, and rules regarding prohibited uses of AI): 2 February 2025 (Art. 113)
- Enforcement of certain requirements (including notification obligations, governance, rules on GPAI models, confidentiality, and penalties (other than penalties for providers of GPAI models)): from 2 August 2025 (Art. 113).
- Providers of GPAI models placed on the EU market before 2 August 2025 are grandfathered for compliance purposes until 2 August 2027 (Art. 111)

[98] Long awaited EU AI Act becomes law after publication in the EU's Official Journal

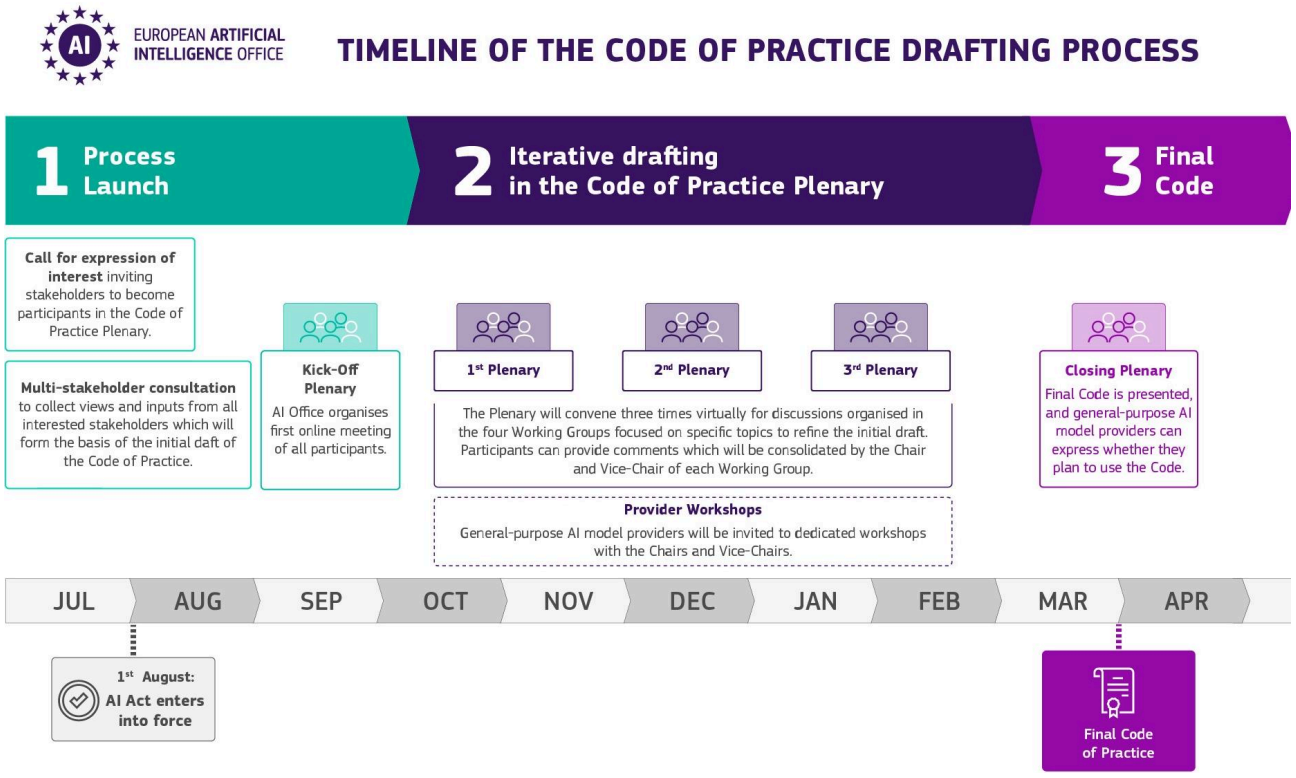
- Enforcement of Art. 6 (and the corresponding obligations regarding high-risk AI systems) commences on 2 August 2027 (Art. 113)

Update from 2024 guide:

- August 2, 2025 : GPAI model obligations and the EU AI Office became operational August 2, 2026: Now limited to transparency obligations (Article 50) and Member State sandbox requirements; no longer the high-risk deadline
- December 2, 2027: New deadline for Annex III high-risk AI systems (recruitment, biometrics, critical infrastructure, etc.) — agreed by the EU Council on March 16, 2026
- August 2, 2028: New deadline for Annex I high-risk AI embedded in regulated products (medical devices, vehicles, machinery)

13.7. High Level Expert AI Group[99]

The EU has also established a new group of AI experts to advise on its AI strategy.



100

14. GDPR

[99] High-level expert group on artificial intelligence | Shaping Europe's digital future

[100] AI Act: Participate in the drawing-up of the first General-Purpose AI Code of Practice | Shaping Europe's digital future

14.1. Overview

The use of AI systems is particularly challenging under the data protection legislation, especially under the GDPR.[101] This regulation imposes strict requirements on the handling of personal data, which can conflict with the operational needs of AI technologies.

One major difficulty is the complexity of data flows inherent in AI systems. These systems often require large datasets collected from diverse sources, making it challenging to ensure compliance with GDPR's principles of data minimization and purpose limitation. Organisations must meticulously manage how personal data is collected, processed, and shared, which can be cumbersome given the dynamic nature of AI.

Another significant challenge is the lack of transparency associated with many AI models. Often described as "black boxes," these systems can make decisions based on algorithms that are not easily interpretable by humans. This opacity complicates compliance with GDPR provisions that grant individuals the right to understand how their data is used and to contest automated decisions affecting them.

Informed consent is also a critical concern. The complexity of AI systems can hinder organisations from providing clear information about data usage, making it difficult for individuals to give truly informed consent. This is compounded by GDPR's requirement that consent must be specific, informed, and unambiguous.

Moreover, **AI systems can inadvertently perpetuate bias and discrimination** by relying on historical data that may reflect societal biases. This raises ethical concerns and compliance issues, as GDPR prohibits discriminatory practices based on sensitive personal data.

Finally, **cross-border data transfers pose additional hurdles**. AI applications often involve transferring personal data across jurisdictions, necessitating adherence to varying legal frameworks, which can complicate compliance efforts. In conclusion, while AI offers transformative potential, navigating its complexities within the stringent framework of GDPR remains a formidable challenge for organisations.

14.2. GDPR Core Principles

- **Lawfulness, Fairness, Transparency**: Process data lawfully, fairly, and transparently.
- **Purpose Limitation**: Collect data for specific, explicit, legitimate purposes only.
- **Data Minimisation**: Collect only necessary data.
- **Accuracy**: Keep data accurate and up-to-date.
- **Storage Limitation**: Retain data only as long as needed.
- **Integrity, Confidentiality**: Ensure data security.
- **Accountability**: Demonstrate GDPR compliance.

[101] The impact of the General Data Protection Regulation (GDPR) on artificial intelligence

14.3. GDPR Challenges in AI Applications

- **Automated Decision-Making & Profiling**:
    - AI systems frequently rely on automated decision-making, which can have a significant impact on individuals.
    - Article 22 of GDPR addresses this by giving individuals the right not to be subjected to decisions based solely on automated processing, unless certain conditions are met (e.g., explicit consent or necessity for contract performance).
    - Note, California has also proposed regulations on automated decision-making technology (ADMT) under the California Consumer Privacy Act (CCPA). The proposed regulations include granting consumers (and employees and business contacts) the right to receive pre-use notice regarding the use of ADMT and to opt out of certain ADMT activities.[102] California SB 1047 (the AI safety bill) was vetoed by Governor Newsom in September 2024.
    - Organisations using AI systems for profiling or automated decision-making must ensure they meet these legal thresholds.
    - Organisations should conduct AI impact assessments and algorithmic auditing to evaluate the potential impacts of automated decision-making systems. Assess various risk factors and mitigation measures related to the system's design, algorithm, decision type, impact, and data used.
- **Data Minimisation vs. AI's Data Needs**: Balancing GDPR's data minimisation principle with AI's need for large datasets is challenging. GDPR mandates that organisations collect only the minimum amount of data necessary to achieve their specific purpose. This can be challenging for AI-driven organisations, which may need extensive datasets to train their models.
- **Purpose Limitation**: organisations must ensure that any further processing of personal data aligns with the original purpose or obtain fresh consent from data subjects. AI developers may process data for unforeseen purposes, requiring careful alignment with GDPR.
- **Transparency and Explainability**: Complex AI models can be opaque, hindering GDPR's transparency requirement. It is difficult to explain or even ascertain precisely AI models based on deep learning and backpropagation make specific decisions (Black Box AI). GDPR requires that data subjects are informed about how their data is being processed in a clear and understandable way. This demand for transparency can be hard to meet with certain AI models and requires strong Data Lineage processes.[103]

See further: Guidance on AI and data protection | ICO

[102] Skadden: AI in 2024: Monitoring New Regulation and Staying in Compliance With Existing Laws | Insights

[103] See for example, KPMG: AI Regulations: Present & Future

14.4. Best Practices for GDPR-Compliant AI

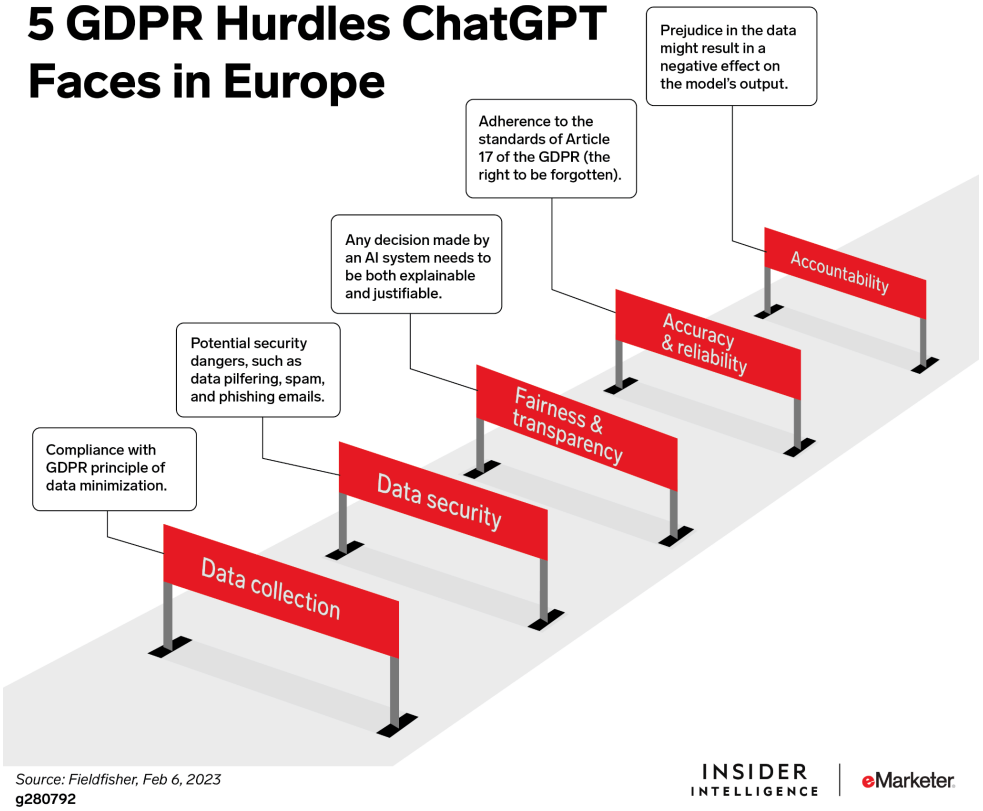
- **Data Governance:** Establish clear frameworks, designate a Data Protection Officer (DPO) if needed, and map data flows.
- **Data Mapping:** Organisations should conduct comprehensive data mapping exercises to understand what data they hold, where it is stored, and how it is processed. This is particularly crucial for AI systems that may process personal data in multiple locations or for a variety of purposes. Strong Data Lineage management is key.[104]
- **Consent Management:** Obtain explicit, informed consent for data use, especially for sensitive data and automated decision-making.
- **Data Limitation, Minimisation and Anonymisation:** Limit the data processed as much as possible. Anonymise data where possible, limit collection, and explore techniques like synthetic data or federated learning.
- **Accountability and Transparency:** Ensure AI explainability, maintain documentation of data processing, and implement human oversight for automated decisions. Explainable AI (XAI) techniques[105] can help address this challenge.
- **Human oversight:** For AI systems involved in automated decision-making, organisations should incorporate mechanisms that allow for human intervention where necessary. This helps ensure that individuals are not subject to decisions based solely on automated processing.
- **Regular Audits and Testing:** Regularly audit AI systems for GDPR compliance, Model Drift, bias, fairness, and accuracy.
- **Cybersecurity:** Enhance security measures to protect personal data used in AI, guarding against model attacks and ensuring data integrity.
- **Risk Framework:** Organisations should prepare and maintain detailed AI Risk Management frameworks. AI systems may be vulnerable to attacks like model poisoning or data theft. Organisations must implement stringent security and change process protocols to guard against these risks and ensure data security.

14.5. Implications for AI Governance

- **Compliance Strategy:** Organisations must establish a compliance strategy to ensure that their AI systems, especially high-risk applications, meet the requirements of the EU AI Act. This includes conducting risk assessments, implementing human oversight mechanisms, and maintaining thorough documentation.

[104] The State of Global AI Regulations in 2024
[105] What is Explainable AI (XAI)? | IBM

- **AI Risk Governance:** Governance structures must be set up to oversee AI systems, particularly those categorised as high-risk. This involves ensuring accountability, monitoring data quality, and providing adequate human oversight.
- **Transparency and Trust:** Transparency measures should be embedded into the organisation's AI systems to meet the EU AI Act's requirements. Informing users when interacting with AI (particularly in sensitive contexts) is essential for maintaining trust and legal compliance.
- **Monitoring and Reporting:** Organisations must establish continuous monitoring and reporting mechanisms for their AI systems, including logging key decisions, data usage, and potential risks.
- **Cross-Border Considerations:** Organisations that operate internationally must assess the Act's extraterritorial implications and ensure that their AI systems used in or affecting the EU comply with the regulation, regardless of where the organisation is based.



[106]

14.6. Additional Considerations

- **Data Protection Impact Assessments (DPIAs):** Conduct DPIAs to identify and mitigate risks associated with AI applications.

[106] GDPR and AI: The next frontier for digital privacy regulation

- **User Rights:** Ensure mechanisms are in place to facilitate user rights under GDPR, such as access, rectification, and erasure of data.
- **Training and Awareness:** Provide ongoing education and training for staff on GDPR and AI ethics to foster a culture of compliance and awareness.
- **Third-Party Management:** Ensure that third-party vendors and partners comply with GDPR when handling personal data for AI purposes.

14.7. Obtaining Consent Under GDPR

To obtain explicit consent for data use in AI applications organisations should:

- **Provide clear and specific information:** Explain in plain language what data will be collected, how it will be used in AI systems, and the potential implications for individuals.
- **Highlight the use of sensitive data:** If the AI application involves sensitive data (e.g., health, race, religion), be particularly transparent about its use and the safeguards in place.
- **Offer granular consent options:** Allow users to consent to specific types of data processing or AI functionalities, giving them more control over their data.
- **Ensure consent is freely given:** Avoid making consent a precondition for service, as this may invalidate consent under GDPR.
- **Obtain separate consent for different purposes:** If data will be used for multiple purposes, obtain separate consent for each purpose.
- **Keep consent records:** Maintain records of consent obtained, including the time, method, and information provided to the individual.
- **Allow easy withdrawal of consent:** Provide a simple mechanism for individuals to withdraw their consent at any time.

14.8. Use of Sensitive Personal Data

Referred to as "special categories of data" under GDPR, includes data such as racial or ethnic origin, political opinions, religious beliefs, trade union membership, genetic and biometric data, health information, and data related to a person's sex life or sexual orientation. Article 9 of GDPR explicitly prohibits the processing of these special categories unless certain conditions are met, one of which is explicit consent. However, obtaining consent can still be unlawful in the following circumstances:

- **Imbalance of Power:** Consent must be freely given, meaning that it cannot be obtained in situations where there is an imbalance of power, such as between an employer and an employee. If there is a power imbalance, the individual may not feel they have a real choice in giving consent, which would invalidate it under GDPR. Consent given under pressure or coercion is not valid.

- **Specific and Informed Consent:** Consent must be specific to the processing purpose and informed. If an organisation obtains vague or broad consent or fails to provide clear information on how the data will be used, this consent can be considered invalid. The consent must also detail the specific sensitive data being processed and for what purpose.
- **Lack of Other Legal Bases:** In some cases, there are other more appropriate legal bases for processing sensitive data. For instance, the processing may be required to protect a person's vital interests (such as life-threatening situations) or to comply with other legal obligations. If consent is used inappropriately when another legal basis would be more appropriate, the processing could still be unlawful.
- **Inadequate Security Measures:** Even with consent, processing sensitive data must comply with GDPR's security requirements (Article 32). If an organisation fails to implement appropriate technical and organisational measures to protect sensitive data, the processing may be unlawful even with consent.

While explicit consent can make processing of sensitive data lawful under GDPR, it must be freely given, informed, specific, and capable of being withdrawn at any time. The organisation must also ensure that other GDPR principles, such as data minimisation, transparency, and security, are followed.

15. Compliance with Other International Laws

Navigating the complex and rapidly evolving landscape of global AI regulations requires a strategic, risk-based approach. This strategy prioritises compliance efforts based on the potential impact and likelihood of regulatory risks, ensuring that resources are allocated efficiently and effectively.

- **Identify and Assess Risks:** Begin by identifying the jurisdictions where your AI systems operate and the relevant regulations, such as GDPR in the EU, CCPA in California, and the AI Act in the EU. Assess the potential risks associated with non-compliance, considering factors like fines, legal actions, and reputational damage. Use a risk matrix to categorise these risks based on their severity and likelihood.
- **Prioritise High-Risk Areas:** Focus on the highest-risk areas first. For instance, regulations that impose significant penalties or those in jurisdictions where your company has substantial operations should be prioritised. This ensures that the most critical compliance issues are addressed promptly.
- **Implement Robust Compliance Controls:** Develop and implement controls tailored to mitigate identified risks. This includes data protection measures, transparency in AI decision-making, and regular audits. Ensure these controls are adaptable to accommodate changes in regulations. Consider cross-border data transfer mechanisms.
- **Continuous Monitoring and Adaptation:** Given the speed at which AI laws evolve, continuous monitoring is essential. Establish a dedicated compliance function to track

regulatory changes and update compliance programs accordingly. Utilise AI tools to automate monitoring and reporting processes, enhancing efficiency.

- **Stakeholder Engagement and Training:** Engage with stakeholders across the organisation to ensure a comprehensive understanding of compliance requirements. Provide regular training to employees, emphasising the importance of compliance and the role they play in maintaining it.
- **Documentation and Reporting:** Maintain thorough documentation of compliance efforts and decisions. This not only aids in demonstrating compliance during audits but also helps in identifying areas for improvement.
- **Wider Consultation:** Consult with regulatory bodies, associations and industry experts
- **Industry Specific Issues:** Follow, incorporate and adhere to industry-specific standards and guidance.

By adopting a risk-based strategy, organisations can effectively manage the complexities of global AI regulations, ensuring compliance while minimising disruptions to their operations.

16. Ethical Considerations

The world is still grappling with the ethical implications of the use of AI including for social welfare, warfare, generative art and literature and how we structure economies for the future. One of the biggest concerns is that AI will lead to a loss of employment opportunities in many sectors. In addition, there are concerns about automated decision making and the risk of people being excluded from services (such as insurance) based on deep data analysis.

Perhaps the greatest risk is of AI tools reflecting and perpetuating existing human prejudices, stereotypes and social and economic inequities (particularly if the most advanced tools are in the hands of a few wealthy billionaires, tech companies and Governments).

As a reminder, any AI Governance Framework must take account of the ethical issues involved with AI including:

- Ethical guidelines for AI development and deployment
- Bias and discrimination detection and mitigation
- Fairness, accountability, and transparency principles
- Risk Management
- Impact of the AI tools being deployed or developed on existing staff, stakeholders, clients and the wider community.

17. Emerging Trends

17.1. Risk Based Compliance

There has been a shift towards risk-based approaches, where governance frameworks prioritise high-risk AI Systems and AI Tools that could significantly affect individuals or society. This allows for more targeted and efficient regulation while fostering innovation in lower-risk areas.

17.2. AI Arms Race

The arms race is accelerating and this ensures that only the very best AI system cybersecurity providers will be able to protect organisations, governments and people from increasingly sophisticated spyware, hacking, phishing and other attack forms.[107] Newer AI systems look to exploit both human weaknesses and AI systemic vulnerabilities.[108]

17.3. Non-Human Forms of Reasoning

AlphaZero showed us that training AI to emulate human knowledge, experience and judgement is not the best way to get extraordinary results. New forms of AI Systems will likely use similar non-human centred means to evaluate information and find optimal routes to success with deeper, more unusual decision-trees.[109]

17.4. Hybrid AI

New AI systems are adopting a hybrid approach to enhance accuracy and explainability. In this approach, machine learning and non-LLM AI handle complex data analysis, while LLM-based generative AI focuses on interpreting and communicating the findings in a clear, understandable manner. This method reduces the risk of "hallucinations" or inaccurate outputs, which is particularly crucial in fields that demand high levels of precision and trust, such as science, healthcare, law, and accounting. For example in medical diagnosis, a hybrid AI system could analyse patient data (e.g., medical history, test results) using machine learning algorithms to identify potential health risks. Then, the generative AI component could explain these findings to the physician (or even a patient) in a concise, understandable report, highlighting key factors and potential treatment options.

17.5. Agentic AI

Agentic AI — systems that autonomously plan, execute multi-step tasks, and interact with external tools and services without continuous human input — has become the dominant AI deployment trend in 2025–2026. It raises distinct governance challenges around accountability, human-in-the-loop requirements, and the boundaries of automated decision-making under frameworks like the EU AI Act and UK GDPR. The FCA has already published proactive guidance on agentic AI in financial services.

17.6. Transparency & Explainability

Another emerging trend is the emphasis on transparency and explainability in AI systems.

[107] 'It's going to take AI to catch the darker side of AI' — Nvidia CEO
[108] Could AI be your company's Achilles heel? - Raconteur
[109] AI begins its ominous split away from human thinking

Policymakers and organisations are increasingly recognizing the importance of understanding how AI makes decisions, leading to the development of tools and methodologies for AI auditing and interpretation.

17.7. Collaboration

International collaboration is gaining traction as countries recognise the global nature of AI development and deployment. Efforts to establish common standards and principles for AI governance are underway, aiming to create a more cohesive regulatory landscape.

17.8. Ethical revolution & AI For Social Good

Ethical considerations will remain at the forefront of AI governance discussions. Addressing issues such as bias, fairness, and the potential existential risks of advanced AI systems will require ongoing dialogue and collaboration between stakeholders. The development and use of AI will ultimately challenge many of our deepest held prejudices and assumptions and so it is likely to lead to a philosophical and ethical revolution over time.

AI for Social Good involves using AI to address societal challenges like climate change, healthcare access, and poverty. This is gaining traction, with initiatives from governments and organisations and is a key area where new technology may provide innovative solutions that go beyond our current abilities and conceptual frameworks.

17.9. Innovation Risks

Another significant challenge will be balancing innovation with regulation. As AI becomes more pervasive, consumers and business leaders will expect policymakers to strike the right balance between fostering technological progress and ensuring adequate safeguards are in place.

17.10. Pace & Complexity

The rapid pace of AI advancement continues to outstrip regulatory efforts, creating a persistent gap between technology and governance. This dynamic environment will require adaptive and flexible frameworks that can evolve alongside AI capabilities. The increasing complexity of AI systems, particularly foundation models and large language models, will pose challenges for transparency and accountability. Developing effective governance mechanisms for these sophisticated systems will be crucial for maintaining public trust and ensuring responsible AI development.

17.11. AI Compliance Systems

AI systems may play an increasingly important role in ensuring regulatory compliance. These tools could automatically monitor and analyse vast amounts of data to identify potential violations, reducing human error and improving efficiency. For example, AI could be used to detect patterns indicative of money laundering or fraud in financial transactions, helping organisations stay compliant with anti-money laundering regulations.

17.12. Blockchain technology

Blockchain tools may be integrated into AI governance frameworks to enhance transparency and traceability. By recording key decisions, data sources, and model updates on a blockchain, organisations could create an immutable audit trail of AI operations. This could help address concerns about AI transparency and accountability, particularly in sensitive applications like healthcare or criminal justice.

17.13. Personalised AI

AI systems and tools will become increasingly linked to each of our hopes, dreams, fears, skills, weaknesses and interests. These AI tools will become 'friends' and trusted advisors for life and for all major life decisions (university, potential partners or lovers, house, car, choice of doctor, investment decisions etc). Even meeting strangers casually in the future will involve automatic due diligence using available public sources. These tools will also enable us to feel like we can speak with the dead again.[110] The risks will increase exponentially as our reliance on AI systems continues to grow.

17.14. AI Augmented Humans

Google appears to have been early but not wrong with its Glasses project (which may well be revived in another form).[111] Recently Meta has partnered with Ray-Ban to offer smart glasses[112] and this is an area where Elon Musk is likely to have a major impact given his interests in neural augmentation[113] and AI systems. It is inevitable that humans will be augmented with AI technologies and that will likely not be with handheld phones. A future where humans use AI to enhance every part of their day to day lives is inevitable.

17.15. Quantum Computing Impact

The advent of quantum computing could significantly impact AI governance. Quantum computers may be able to process complex AI models much faster than classical computers and with greater inherent security from a cybercrime perspective, potentially accelerating AI development and deployment. This could necessitate new governance frameworks to keep pace with rapid advancements. Additionally, quantum computing could pose new security challenges, requiring updated encryption methods to protect AI systems and data.

18. Conclusion

In conclusion, AI governance is crucial for organisations to ensure responsible AI use and mitigate risks. Organisations should establish clear AI governance frameworks, including risk management processes, ethical guidelines, and compliance with regulations such as the EU AI Act and GDPR.

---

[110] Deepfakes of your dead loved ones are a booming Chinese business | MIT Technology Review
[111] Project Astra - Google DeepMind
[112] Ray-Ban Meta Smart Glasses Review: Fine Audio and Video, Privacy Issues | WIRED
[113] Neuralink

By proactively addressing the challenges and opportunities of AI, organisations can harness its benefits while upholding ethical standards and ensuring long-term success.

AI generated image, Peter Howitt