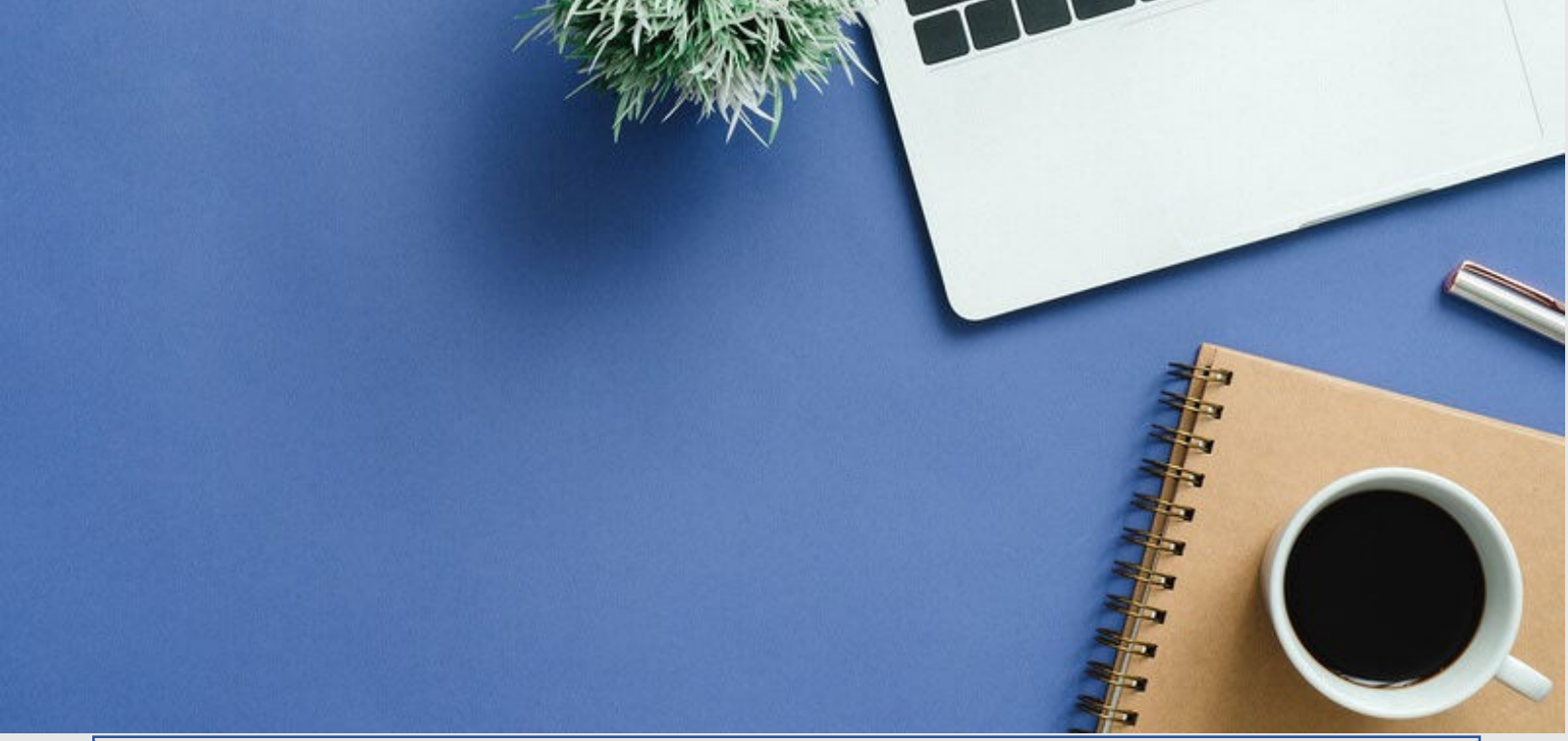
Effective AI Governance for a Trustworthy Innovative New Era

Whitepaper

Pamela Gupta

CEO Trusted AI

Contents



Overview

One thing is clear as we approach an AI Era.

We require a dramatic shift in our approach to AI Governance and risk management.

This **involves** integrating the stability and proven practices of traditional governance with the dynamic, often unpredictable nature of AI governance.

This **requires** not only a deep understanding of both realms but also a visionary approach to anticipate and shape future standards. Balancing these two aspects of governance is not just about compliance and risk management.

It is about steering the Organization in a way that respects ethical boundaries while embracing technological innovation. It's a delicate tightrope walk between the known and the unknown, the established and the emerging, the safe and the revolutionary.

This whitepaper lays out why we need to rethink our AI Strategy, AI Governance, and the impact of regulations, namely the differences in regulation approach that will heavily impact the approach organizations need to prepare for today.

I also talk about bridging the gap between requirements, planning and actionable strategy.

AI Governance

why we need a new approach, how can we do this right the first time around?

To understand how to create effective approach let's look at the Dilemmas in Effective AI Governance

To highlight the overview statement of this whitepaper - The core of the dilemma lies in integrating the stability and proven practices of traditional governance with the dynamic, often unpredictable nature of AI governance.

This requires not only a deep understanding of both realms but also a visionary approach to anticipate and shape future standards. Balancing these two aspects of governance is not just about compliance and risk management; it's about steering the company in a way that respects ethical boundaries while embracing technological innovation. It's a delicate tightrope walk between the known and the unknown, the established and the emerging, the safe and the revolutionary.

Balance dilemma #1 Regulatory Adherence vs. Pioneering New Standards:

Traditional Governance: Here, the path is clear with established laws and regulations. Companies know the rules and how to play by them.

AI Governance: The challenge is navigating uncharted waters. With AI, there's a constant tension between adhering to existing regulations and pioneering new ethical standards in areas yet to be regulated.

Balance dilemma #2 Risk Management: Predictable vs. Unpredictable Risks:

Traditional Governance: Focuses on managing predictable risks, like financial uncertainties and operational inefficiencies.

AI Governance: Confronts unpredictable and often unprecedented risks, such as algorithmic biases or unintended consequences of AI decision-making.

Balance dilemma #3 Stakeholder Engagement: Conventional Expectations vs. Expanding Concerns:

Traditional Governance: Engagement revolves around well-understood issues like financial performance and compliance.

AI Governance: Requires engaging with a broader spectrum of internal and external stakeholders to focus on additional areas of ethical AI use, societal impacts, and public trust in technology.

Balance dilemma #4 Ethical Obligations: Corporate Responsibility vs. Societal Impact:

Traditional Governance: Ethics are focused on corporate conduct and social responsibility within a known framework.

AI Governance: Involves grappling with the broader societal impacts of technology, raising questions about privacy, human rights, and equitable benefits of AI.

Balance dilemma #5 Expertise: Generalist vs. Specialist:

Traditional Governance: Requires a broad understanding of business processes and general technologies.

AI Governance: Demands specialized knowledge in AI, data science, and areas such as data provenance which may not have existing teams and roles with accountability.

In summary, while traditional governance and AI governance share common goals of risk management, and compliance, AI governance requires a more specialized focus on the unique challenges posed by AI technology, including its ethical, economic, social, and regulatory aspects. This requires a forward-thinking approach and a willingness to engage with emerging and often complex issues that are intrinsic to AI technology.

Further, there are roles and responsibilities that don't exist in the Organizations which will either slow down risk identification, accountability and/or mitigation.

How do we bridge this gap, can we bridge the gap and address the risks?

Trusted AI understands these challenges. We are helping clients across verticals to harness the value of AI by De-Risking AI early in their adoption journey.

Creating Contextual Solutions

Implementing AI risk management and transparency to align with comprehensive and fluid regulations can be challenging for organizations of any size.

They will need to implement a risk management process, conform to higher data standards, more thoroughly document the systems, systematically record its actions, provide information to users about its function, and enable human oversight and ongoing monitoring. Some of these requirements have likely already been implemented by sectoral regulators who have been dealing with AI for a long time, but most are likely new. The result is that AI systems within regulated products will need to be documented, assessed, and monitored on their own, rather than just evaluating the broader function of the product. This sets a new and higher floor for considering AI systems in products.

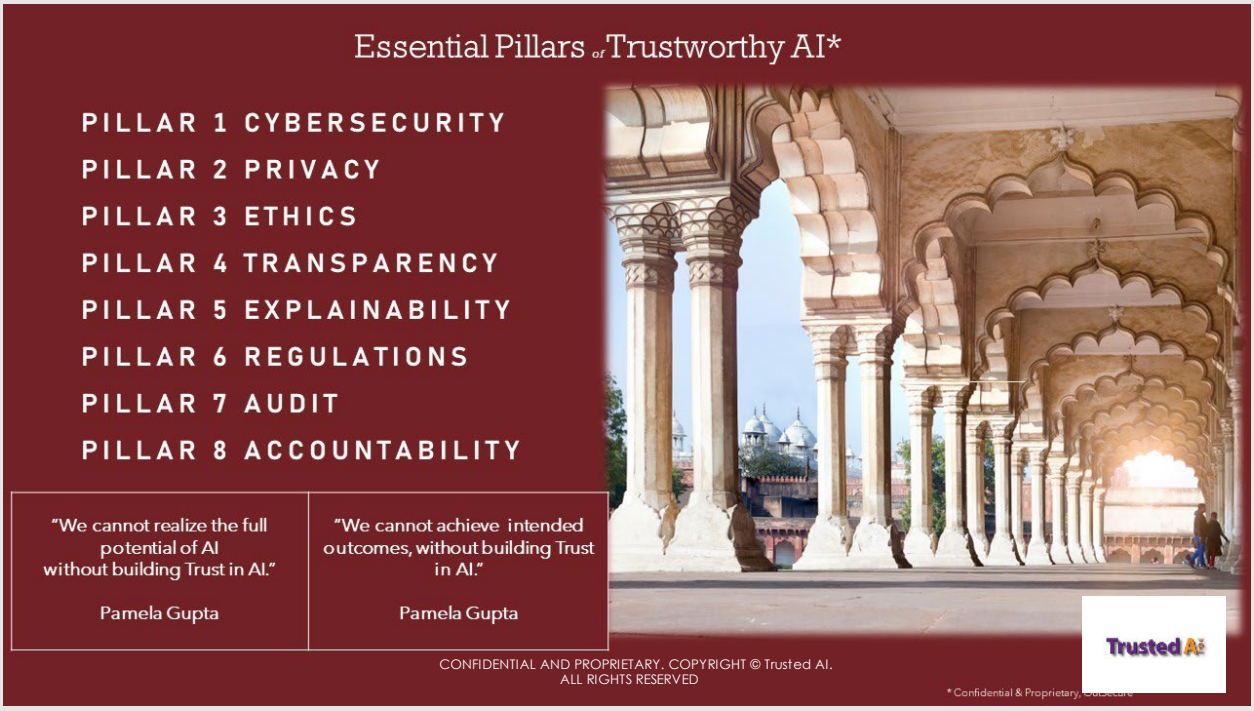
These include the risks of fines posed by current and new regulations, as well as Security, Privacy, Transparency, Explainability, Accountability and Trustworthy AI risks these guidelines are enacted to prevent. In addition to regulations, we are helping avoid unintended outcomes for society and the environment.

The biggest challenge in adopting Trustworthy AI/Responsible AI is the 'how', adopting and operationalizing the new approach for governance, integrating with existing framework is not easy, not defined and will vary by organization.

We created an approach based on years of creating holistic and strategic risk-based management programs at global firms operating in complex regulatory environments. Our approach for creating actionable and operational governance is called AI TIPS. It is designed for a Trustworthy AI implementation that is built intentionally and with the right stakeholders from inception to implementation.

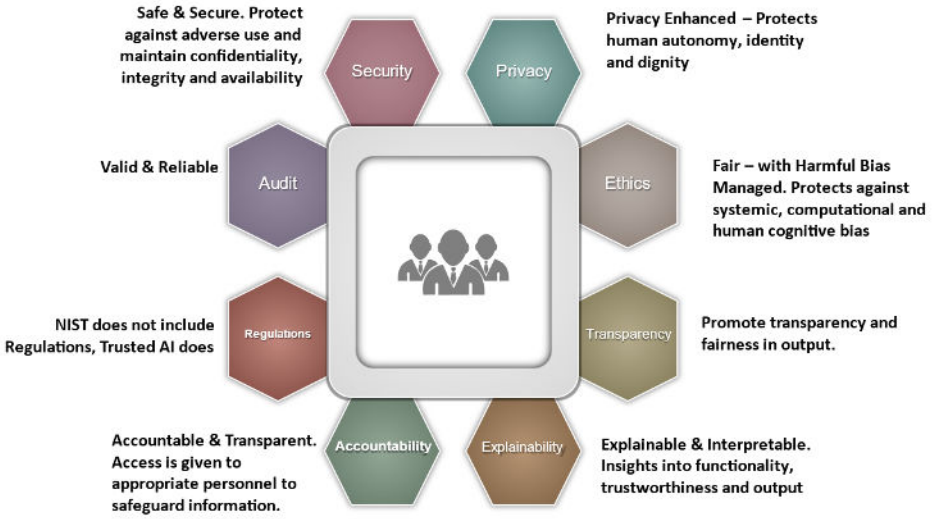
AI TIPS. Artificial Intelligence Trust Integrated Pillars for Sustainability is a unique model for operationalizing governance based on risk management.

AI TIPS has 8 Essential Pillars of Trust



Our Essential Pillars for Trust include those in the NIST AI RMF plus one, namely Regulations. In our holistic risk based strategic approach for AI risk management and governance we combine people, process and technology and the operational environment.

Trusted AI Trust Pillars (8) align NIST AI RMF Trust areas (7)



Role of AI Regulations

Enterprises will need to earn the trust of customers and regulators globally. AI is unique from a risk and governance perspective in that we need to understand the two How(s):

1. How did it come to that conclusion or answer?
2. How was it Built?

Regulations such as the EU AI Act will help garner trust that the AI Systems are Safe, Secure, Reliable and Trustworthy.

Critical for Trust in AI: Data

1. **What data is used?**
2. **How is the data being used?**
3. **Where did it come**

Note: the EU AI Act will be applicable to those in the EU and outside of the EU.

Chinese AI Regulations

China has been leading the way when it comes to AI Governance. Chinese laws addressing data privacy and security are raising critical questions for businesses operating inside and outside of China.

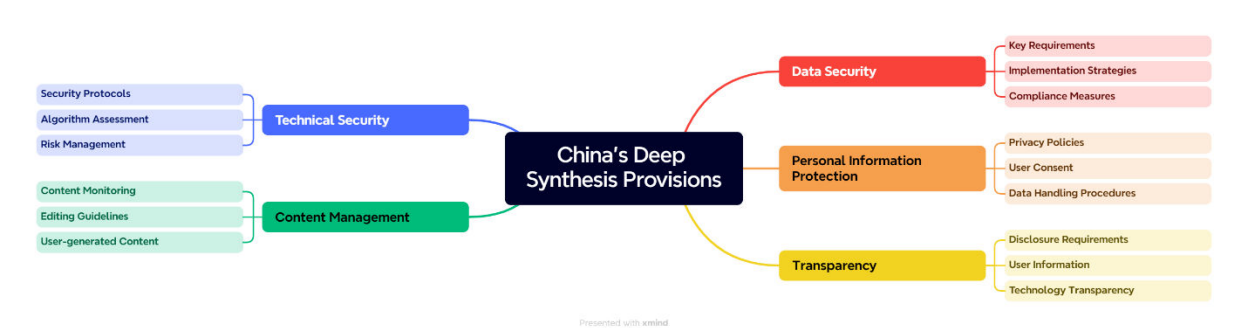
There are three significant AI Algorithms regulations – all with requirements for developers to file their algorithms in a central government Algorithm registry:

2021 **Algorithm recommendation regulations**

2022 **Rules on deep synthesis aka synthetic content aka deepfakes**

2023 **Rules on Generative AI.**

These regulations target recommendation algorithms for disseminating content, synthetically generated images and video, and generative AI systems like OpenAI's ChatGPT. The rules create new requirements for how algorithms are built and deployed, as well as for what information AI developers must disclose to the government and the public.

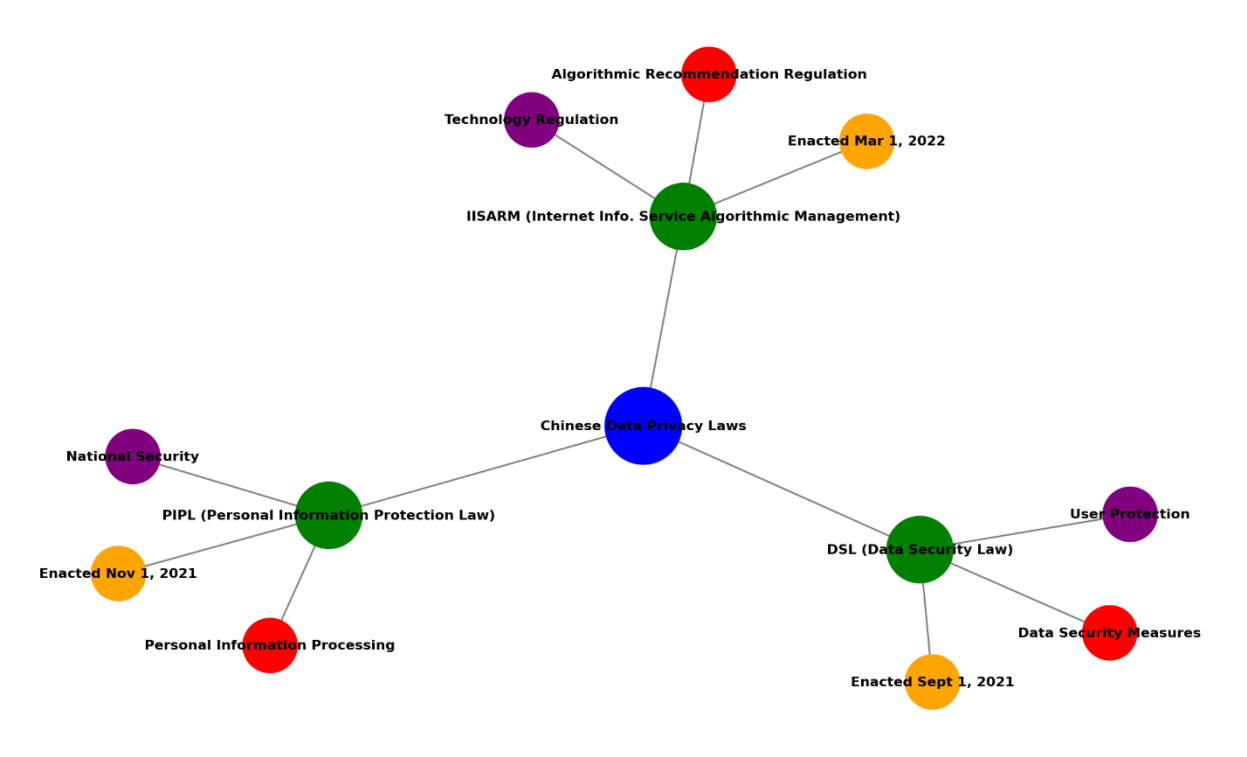


There are three major Chinese data privacy laws:

1. **Data Security Law (DSL),** It requires business data to be categorized by different levels of importance and puts new restrictions on cross-border transfers. These regulations will have a significant impact on how companies collect, store, use and transfer data.
2. **Personal Information Protection Law (PIPL).** In September 2021, the CAC announced a three-year plan to regulate predictive algorithms used by online content providers. The draft rules prohibit algorithms that encourage online addiction, a main issue in China. The proposed regulations also require that users be told about algorithmic recommendation services and be given a way to switch them off. Because these regulations are enabled by the PIPL, they can impact foreign businesses as well as Chinese companies, and
3. **Internet Information Service Algorithmic Management (IISARM) regulations.** These new regulations retroactively apply to previously enacted data privacy laws in China, namely the Data Security Law (DSL) and the Personal Information Protection Law (PIPL). However, the IISARM regulations are also meant to safeguard national security in China as well as the social and public interest. Under the new regulations, algorithms cannot be used to influence online public opinions.

Thus, the IISARM regulations could have broad implications on algorithm service providers, users, and the balance between user autonomy and national security. There are already precedents on the effect on discovery in litigation matters in the United States.

This visual representation shows the relationships and key features of each law, including their enactment dates, primary focus areas, and overarching themes like user protection, national security, and technology regulations.



EU Digital Legislation

The EU has enacted an extensive body of digital legislation, mostly under the Digital Agenda and the Digital Single Market (DSM) initiatives.

In the current legislative term (2019 - 2024), major new measures relevant to digitalization have all either been enacted or are in the legislative process with prospects of being enacted in the coming months.
- The Digital Markets Act (DMA) aims to regulate the behavior of large digital platforms to ensure fair competition.
- The Digital Services Act (DSA) intends to establish rules for online services providers to enhance user safety and accountability.
- The Data Act focuses on rules governing data access and use in the digital environment.
- The Artificial Intelligence Act (AI Act) seeking to set standards for AI systems to ensure they are safe, transparent, and accountable.
- The Data Governance Act (DGA) aims to facilitate data sharing and access while safeguarding data privacy and security.
- The European Health Data Space (EHDS) aiming to create a secure and unified ecosystem for health data sharing across the EU.
- An update to the regulation on electronic identification and trust services (eIDAS 2) seeks to enhance the security and interoperability of electronic identification and trust services in the EU.
- The measure to strengthen the cybersecurity of critical infrastructure (NIS2) aims to enhance the cybersecurity resilience of critical infrastructure entities in the EU.

The initiatives are categorized based on whether they primarily relate to:
(1) research and innovation; (2) industrial policy; (3) connectivity; (4) data and privacy; (5) cybersecurity; (6) law enforcement; (7) trust and safety; (8) e-commerce and consumer protection; (9) competition, (10) IPR and media; and (11) finance.

EU AI Act

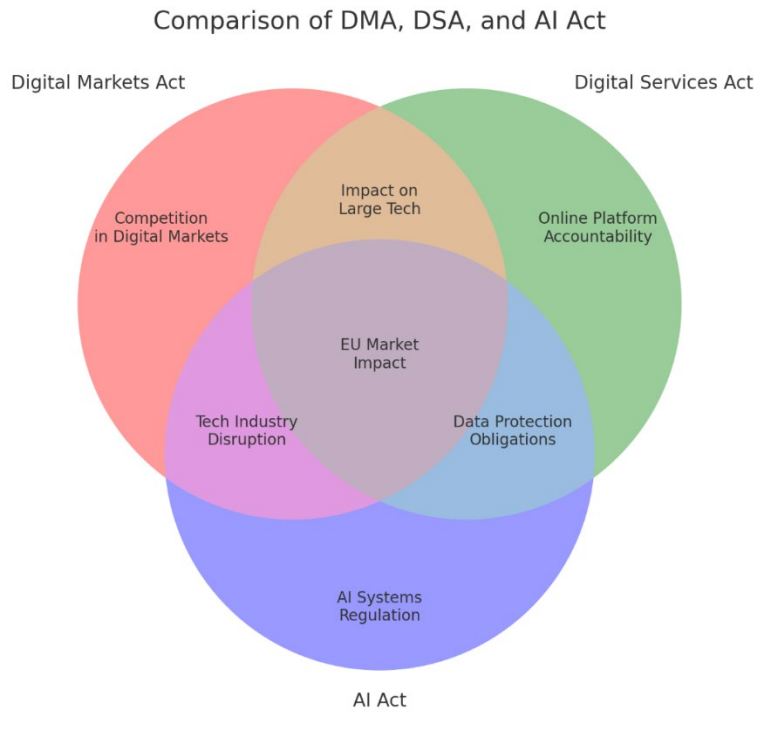
Origin and Scope: Proposed by the European Commission in April 2021, the EU AI Act is part of the European approach to digital transformation and aims to regulate AI use within the European Union.

Risk-Based Classification: Central to the EU AI Act is a risk-based classification system for AI systems, categorizing them into unacceptable risk, high risk, limited risk, and minimal risk.

Regulations for High-Risk AI: For AI systems classified as high-risk, there are stringent compliance requirements, including transparency, human oversight, robustness, and accuracy.

Bans Certain AI Practices: It proposes bans on certain AI practices deemed as a clear threat to the rights and safety of individuals, such as social scoring by governments.

Focus on Fundamental Rights: The Act emphasizes the protection of fundamental rights, particularly regarding surveillance and biometric identification.



White House AI Blueprint and Executive Order

On 30 October, President Biden issued an executive order ("Executive Order" or EO) on artificial intelligence (AI) with the goal of promoting the "safe, secure, and trustworthy development and use of artificial intelligence." This Executive Order represents a significant contribution to the subject of accountability in how AI is developed and deployed across organizations in all sectors of the economy, from the most mature AI implementers to first-time adopters. The Executive Order's definition of AI systems is also broad; it is not limited to generative AI or systems leveraging neural networks but is inclusive of predictive AI systems.

Also, it will require Organization's to carefully evaluate the extent to which the EO affects not only an entity's own use of AI, but also the extent to which its products and services incorporate or are reliant on third-party vendors' AI-enabled capabilities.

The Executive Order is guided by eight principles and priorities:

1. AI must be safe and secure by requiring robust, reliable, repeatable and standardized evaluations of AI systems, as well as policies, institutions, and, as appropriate, mechanisms to test, understand, and mitigate risks from these systems before they are put to use.
2. The US should promote responsible innovation, competition and collaboration via investments in education, training, R&D and capacity while addressing intellectual property rights questions and stopping unlawful collusion and monopoly over key assets and technologies.

3. The responsible development and use of AI require a commitment to supporting American workers though education and job training and understanding the impact of AI on the labor force and workers' rights.
4. AI policies must be consistent with the advancement of equity and civil rights.
5. The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected.
6. Americans' privacy and civil liberties must be protected by ensuring that the collection, use and retention of data is lawful, secure and promotes privacy.
7. It is important to manage the risks from the federal government's own use of AI and increase its internal capacity to regulate, govern and support responsible use of AI to deliver better results for Americans.
8. The federal government should lead the way to global societal, economic, and technological progress including by engaging with international partners to develop a framework to manage AI risks, unlock AI's potential for good and promote a common approach to shared challenges.

Importantly, the National Institute of Standards and Technology (NIST) will be foundational in the development of guidelines and best practices for "developing and deploying safe, secure and trustworthy AI systems," and companies may consider evaluating their existing AI risk management frameworks against the NIST AI Risk Management Framework to develop a baseline and prepare for additional guidance to be released from relevant agencies and regulatory bodies.

The EO directs the NIST, working with the Department of Commerce, to develop two sets of guidelines within 270 days (about 9 months):

- NIST is directed to establish guidelines and best practices for "developing and deploying safe, secure, and trustworthy AI systems."
- NIST is called upon to develop standards and procedures for developers of AI (outside of national security applications) to conduct AI red-teaming tests (structured testing to identify potential flaws and vulnerabilities).

NIST AI Risk Management Framework, RMF

The goal of the AI RMF is to offer a resource to the organizations designing, developing, deploying, or using AI systems to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems. The Framework is intended to be voluntary, rights-preserving, non-sector-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors and throughout society to implement the approaches in the Framework.

AI risk management is a key component of responsible development and use of AI systems. Responsible AI practices can help align the decisions about AI system design, development, and uses with intended aim and values. Core concepts in responsible AI emphasize human centricity, social responsibility, and sustainability. AI risk management can drive responsible uses and practices by prompting organizations and their internal teams who design, develop, and deploy AI to think more critically about context and

potential or unexpected negative and positive impacts. Understanding and managing the risks of AI systems will help to enhance trustworthiness, and in turn, cultivate public trust.

Key differences in EU and US Governance Focus

Focus: The EU AI Act is more regulatory and protective, focusing on risk management and fundamental rights, while the U.S. Executive Order emphasizes AI development, innovation, and competitiveness.

Regulatory Approach: The EU approach is more prescriptive, with specific rules for different risk categories of AI, whereas the U.S. approach is more about fostering growth and leadership in AI.

Aspect	U.S. Executive Order on AI	EU AI Act
Purpose and Focus	To promote the safe, secure, and trustworthy development and use of AI, balancing innovation with societal protections.	To ensure AI systems are safe and respect EU laws and values, preventing harmful AI practices.
Scope and Applicability	Federal government approach, potentially influencing private sector and academia.	Applies across all EU member states, with implications for global businesses operating in the EU.
Safety and Security	Emphasizes robust evaluations and risk mitigation for AI systems, addressing national security concerns.	Focuses on high-risk AI systems, requiring rigorous assessment and compliance with safety standards.
Innovation and Competition	Encourages innovation and competition in AI, supporting education and training, and addressing IP and market competition issues.	Prioritizes a balanced approach to foster innovation while ensuring AI systems are trustworthy.
Workforce and Employment	Addresses the impact of AI on workers, promoting job training and education to adapt to AI-induced changes in the job market.	Less explicit focus on workforce; however, the emphasis on safety may indirectly impact job training.
Equity and Civil Rights	Dedicated to preventing AI from exacerbating discrimination, ensuring AI advances equity and justice.	Emphasizes the need for AI to comply with fundamental rights, including privacy and data protection.
Consumer Protection	Aims to enforce consumer protection laws in the age of AI, safeguarding against fraud and bias in AI applications.	Includes provisions for consumer safety, particularly in high-risk AI applications.
Privacy	Stresses the importance of privacy and data protection in AI development and use.	Strong focus on data protection and privacy, consistent with GDPR and other EU privacy regulations.
Government Use of AI	Focuses on improving the federal government's use of AI and developing public sector AI expertise.	Less emphasis on government use of AI, with more focus on the private sector and broad societal impacts.
International Collaboration	Encourages global engagement and collaboration on AI, promoting responsible AI use internationally.	Seeks to establish EU leadership in setting global AI standards, encouraging international cooperation.

In summary, while the EU AI Act is a comprehensive regulatory framework with a strong focus on risk and rights protection, the U.S. Executive Order on AI is oriented towards promoting AI development and maintaining technological leadership.

FTC US AI Regulator

On November 21, 2023, the Federal Trade Commission (FTC) approved a resolution to facilitate the issuance of Civil Investigative Demands (CIDs) in AI-related investigations. This 10-year resolution

allows streamlined issuance of these subpoenas, while the FTC retains ultimate control over their use. This action aligns with the FTC's broader strategy under Chair Lina Khan to enhance its regulatory authority in the AI sector.

The FTC's move reflects a growing focus on the potential risks AI poses to consumers, acknowledging both its benefits and possible misuses, such as deception or discrimination further committing to protect individual rights against violations through advanced technologies.

The Biden administration, in the EO on “safe, secure, and trustworthy development and use of artificial intelligence", encouraged the FTC to expand its regulatory activities in AI. The FTC's new resolution positions it to investigate AI companies more vigorously for illegal or unfair practices, underlining its ambition to become a leading AI regulator. The FTC has already initiated such actions, as evidenced by its investigation of OpenAI, the developer of ChatGPT.

Key Takeaways

- Artificial Intelligence is unprecedented from any other technology in terms of the potential for negative and positive impact.
- All the existing trustworthy pillars such as security, privacy, accountability, audit require a recalibration, in other words the requirements for them for conventional systems and for AI systems are vastly different. We can create existing principles but only partially at best.
- Example:
  The Federal Trade Commission has reached a settlement with Rite Aid Jan 2024, banning the pharmacy chain from using facial recognition technology for the next five years, following charges that the company misused the biometric tech in hundreds of its stores.
  The FTC said in a Dec. 19 announcement that from 2012 to 2020, Rite Aid used artificial intelligence-based surveillance to identify customers who have engaged in shoplifting and other problematic behavior in its stores. The tech erroneously identified some customers as shoplifters, prompting Rite Aid employees, in some cases, to follow customers around the store, search them, call police, and accuse them of shoplifting, the FTC said, adding that the false identifications disproportionately impacted people of color.
  “Rite Aid's reckless use of facial surveillance systems left its customers facing humiliation and other harms, and its order violations put consumers’ sensitive information at risk,” said Samuel Levine, Director of the FTC’s Bureau of Consumer Protection. “Today’s groundbreaking order makes clear that the commission will be vigilant in protecting the public from unfair biometric surveillance and unfair data security practices.”

References

1. Overview of EU legislations in the Digital Sector Bruegel - Improving economic policy
2. China's AI Regulations and How They Get Made MATT SHEEHAN
3. NIST AI RISK MANAGEMENT FRAMEWORK
4. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence
5. Supermarket News Rite Aid now banned from using facial recognition by FTC for next five years

Holistic AI Governance, Risk Management for the Earth, Society and Democracy
About Trusted AI: We are a Responsible AI/Trustworthy AI consulting company.

Implementing the correct Trust AI principles can be complex. Companies are partnering with Trusted AI experts who can provide relevant guidance. We are helping clients answer:

How can Trustworthy AI be instrumental in helping my organization gain a competitive edge?

Promote better business outcomes, including accelerated innovation with AI?

To request an AI Risk assessment:

Create a Responsible/Trustworthy AI Adoption Plan

De-Risk your AI initiatives contact us at https://www.trustedai.ai/strategic-ai-risk-governance-compliance/

Pamela Gupta
CEO Trusted AI
2 Trap Falls Rd, Shelton CT 06604
Pamela.Gupta@TrustedAI.AI
Schedule Meeting