

MAYER | BROWN

AI SUMMIT 2026

New York

Thought Leadership Materials

Panel Session 4: AI Governance

MAYER | BROWN

ARTIFICIAL INTELLIGENCE GOVERNANCE

DOMESTIC AND GLOBAL AI LAWS

Comprehensive AI Laws

- EU AI Act
- Colorado's Concerning Consumer Protections in Interactions with AI ("Colorado AI Law")
 - The Colorado AI Law may be overhauled based on a new proposed framework (TBD)
- South Korea's Artificial Intelligence Development and Establishment of a Foundation for Trustworthiness
- Vietnam Law on AI (No. 134/2025/QH15)

Narrower AI Laws in the US

- Texas Responsible AI Governance Act (prohibited practices)
- Frontier Model Laws (California and New York)
- Utah Generative AI Law (transparency)
- Illinois AI law amending Human Rights Act (transparency and avoiding discrimination)
- Illinois AI Video Interview Act (transparency and consent)
- New York City Local Law 144 (transparency and bias audit)
- California Generative AI: Training Data Transparency (AB2013) (transparency regarding training data)
- California AI Transparency Act (SB942) (AI detection tool and manifest and latent disclosures on AI-generated content)
- California AI in Healthcare Services (AB3030) (transparency regarding communication generated by AI)
- Chatbot Laws (transparency)

Interaction with Existing Laws

- Data privacy law considerations (state Attorneys' General and California Privacy Protection Agency)
 - **CCPA: New regulations on "Automated Decisionmaking Technology"**
- Employment laws (Equal Employment Opportunity Commission)
- Unfair and deceptive practices (Federal Trade Commission)
- Explaining credit decisions following adverse action (Consumer Financial Protection Bureau)

AI FRAMEWORKS

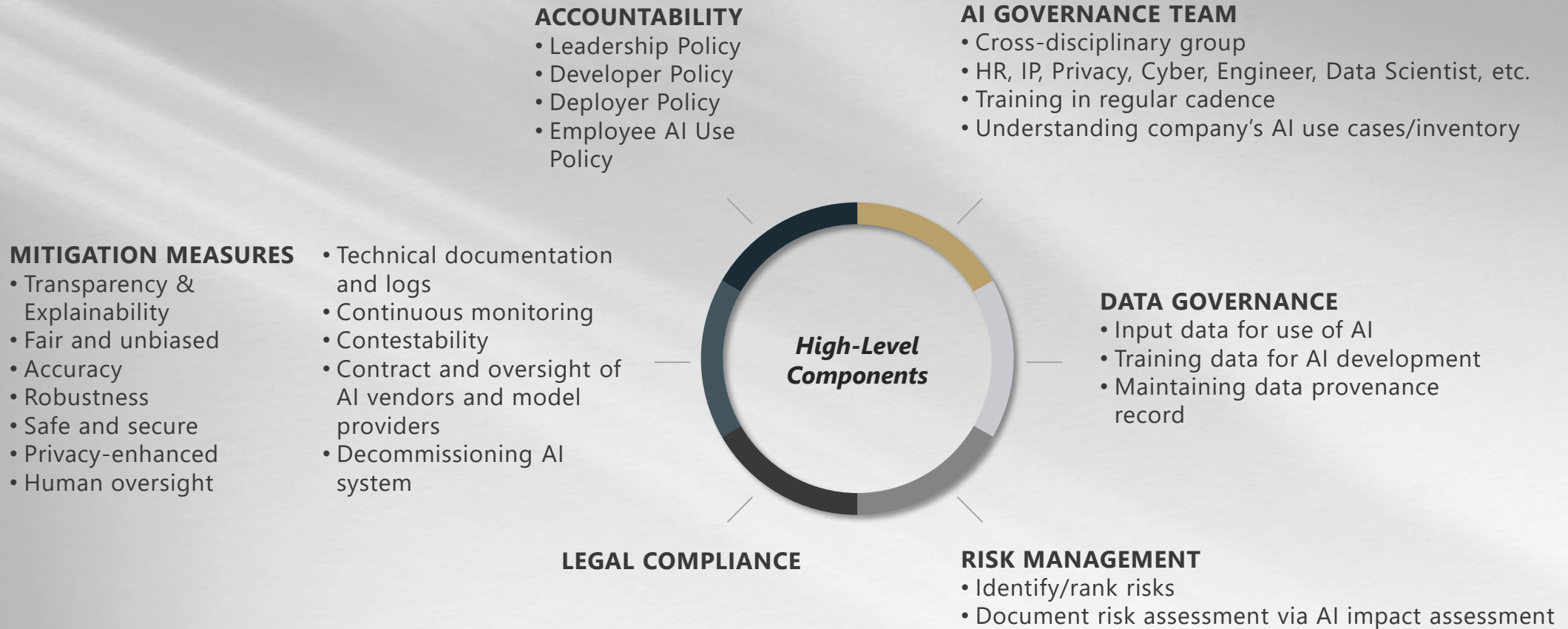
ISO/IEC 42001



NIST AI RMF



GLOBAL AI GOVERNANCE



WANT TO LEARN MORE?

Implementing A Global Artificial Intelligence Governance Program

🏠 Set as Home ★ Favorite

Select All Download | Email | +Workspace | Save for Later

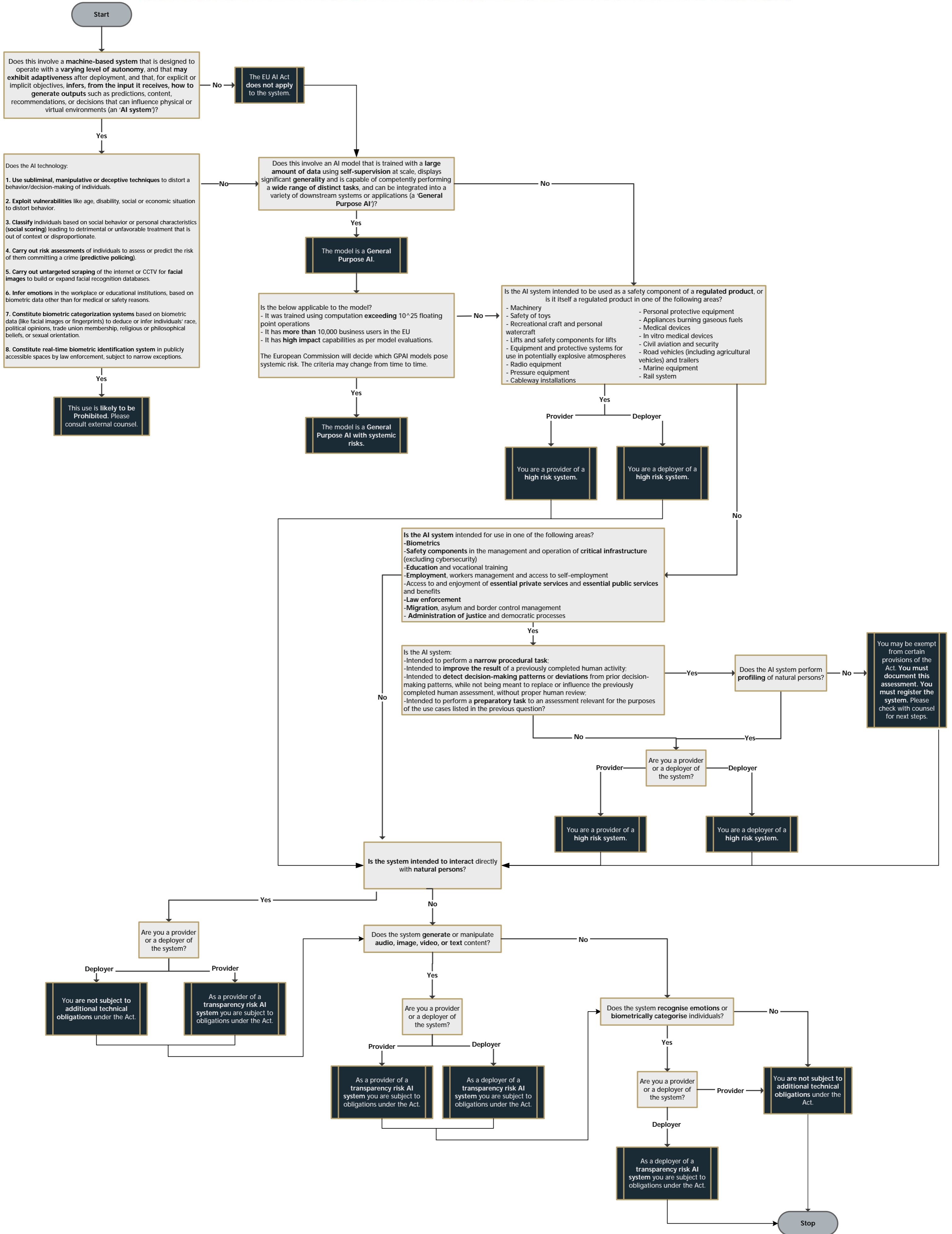
- Title Page
- Preface
- Introduction
- Chapter 1. Initial Scoping Topics To Consider
- Chapter 2. Background
- Chapter 3. Players In The AI Supply Chain
- Chapter 4. Implementing An AI Governance Program
- Chapter 5. Conclusion
- Appendix A. Artificial Intelligence Leadership Policy
- Appendix B. Artificial Intelligence Impact Assessment (AIIA)
- Appendix C. Artificial Intelligence Developer Policy
- Appendix D. Artificial Intelligence Deployer Policy

Check out Arsen Kourinian's e-book on Bloomberg Law:
<https://www.bloomberglaw.com/product/blaw/browser/105.564794>



CLASSIFICATION OF AI SYSTEMS AND MODELS UNDER THE EU AI ACT

22 June 2025



Implementing a Global Artificial Intelligence Governance Program

2025 Update

**Author
Arsen Kourinian**

**Bloomberg
Law**

Copyright © 2025
Bloomberg Industry Group, Inc.

The materials contained herein represent the opinions of the authors and editors and should not be construed to be those of Bloomberg Industry Group, Inc. Nothing contained herein is to be considered as the rendering of legal advice for specific cases, and readers are responsible for obtaining such advice from their own legal counsel. These materials and any forms and agreements herein are intended for educational and informational purposes only.

Published by Bloomberg Law
1801 S. Bell Street, Arlington, VA
22202

e-ISBN: 978-1-68267-943-2
Printed in the United States of America

Summary Table of Contents

Preface	i
Introduction	iii
Chapter 1. Initial Scoping Topics to Consider	2
Chapter 2. Background	4
Chapter 3. Players In the AI Supply Chain	20
Chapter 4. Implementing An AI Governance Program	32
Chapter 5. Conclusion	72
Appendix A. Artificial Intelligence Leadership Policy	74
Appendix B. Artificial Intelligence Impact Assessment (AIIA)	78
Appendix C. Artificial Intelligence Developer Policy	100
Appendix D. Artificial Intelligence Deployer Policy	106
Appendix E. Employee Artificial Intelligence Use Policy	112
Appendix F. Incorporating Artificial Intelligence Regulatory Requirements in Contracts	118

Detailed Table of Contents

Preface	i
Introduction	iii
Chapter 1. Initial Scoping Topics to Consider	2
1.I. OVERVIEW.....	2
Chapter 2. Background.....	4
2.I. BACKGROUND.....	4
2.II. AI DEFINED	4
2.III. INCIDENTS, HAZARDS, AND HARMS	9
2.III.A. The OECD and EU AI Act.....	9
2.III.B. US AI Laws.....	10
2.III.C. The Harm Categories Under the National Institute of Standards and Technology (NIST)	11
2.III.D. The OECD Classification of AI Impacts	12
2.III.E. Examples of AI Incidents	13
2.IV. STRONG/BROAD V. WEAK/NARROW AI.....	14
2.V. MACHINE LEARNING, DEEP LEARNING, AND NEURAL NETWORKS.....	15
2.VI. METHODS OF AI TRAINING	15
2.VII. GENERATIVE V. DISCRIMINATIVE MODELS	16
2.VIII. MULTI-MODAL MODELS	17
2.IX. AGENTIC AI	17
2.X. AI LIFECYCLE.....	17
Chapter 3. Players In the AI Supply Chain	20
3.I. OVERVIEW.....	20
3.II. AI Architecture Providers.....	20
3.II.1. Semiconductor Chips.....	21
3.II.2. Accelerators.....	21
3.II.3. Cloud Computing.....	21
3.III. FOUNDATION MODELS.....	22
3.IV. AI SYSTEM PROVIDERS.....	28
3.V. AI SYSTEM DEPLOYERS	28
3.VI. OTHER ACTORS IN THE AI SUPPLY CHAIN	29
3.VII. AI ROLE UNDER DATA PRIVACY LAWS	29
Chapter 4. Implementing An AI Governance Program.....	32
4.I. OVERVIEW.....	33
4.I.A. European Union.....	33
4.I.B. United Kingdom	33
4.I.C. United States.....	34
4.I.D. Rest of World	34

4.I.E. AI Frameworks	34
4.II. ASSEMBLING AN AI GOVERNANCE TEAM	36
4.III. DATA GOVERNANCE.....	38
4.IV. LEGAL COMPLIANCE.....	41
4.V. RISK MANAGEMENT	42
4.V.A. Identify and Rank AI Risks	43
4.V.B. Likelihood and Severity of Harm	47
4.V.C. Document an AI Impact Assessment	49
4.VI. MITIGATION MEASURES.....	50
4.VI.A. Transparency and Explainability	51
4.VI.B. Fair and Unbiased	55
4.VI.C. Human-Centered and Beneficial for the Environment and Society	57
4.VI.D. Accuracy	58
4.VI.E. Robustness	60
4.VI.F. Safe and Secure	61
4.VI.G. Enhancing Privacy Protection	64
4.VI.H. Human Oversight.....	64
4.VI.I. Technical Documentation and Logs.....	66
4.VI.J. Post-Market Monitoring	67
4.VI.K. Communication Channels and Contestability	67
4.VI.L. Adopt Appropriate AI Contractual Provisions	67
4.VI.M. Decommissioning the AI System.....	69
4.VII. ACCOUNTABILITY	70
Chapter 5. Conclusion	72
5.I. Conclusion.....	72
Appendix A. Artificial Intelligence Leadership Policy	74
I. Purpose	74
II. Scope.....	74
III. Definitions	74
IV. Designation of the AI Oversight Committee Members.....	74
V. Authority of the AI Oversight Committee Members.....	74
VI. Responsibilities of the AI Oversight Committee	75
VII. Contact Information	75

VIII. Revision History	76
Appendix B. Artificial Intelligence Impact Assessment (AIIA)	78
Appendix C. Artificial Intelligence Developer Policy	100
I. Purpose	100
II. Scope.....	100
III. AI Governance.....	100
III.A. Involving the AI Oversight Committee	100
III.B. Data Governance.....	100
III.C. Legal Compliance	101
III.D. Risk Management.....	101
III.E. Mitigation Measures.....	103
III.F. Accountability.....	105
IV. Contact Information	105
V. Revision History	105
Appendix D. Artificial Intelligence Deployer Policy	106
I. Purpose	106
II. Scope.....	106
III. AI Governance.....	106
III.A. Involving the AI Oversight Committee	106
III.B. Data Governance.....	106
III.C. Legal Compliance	107
III.D. Risk Management	107
III.E. Mitigation Measures.....	108
III.F. Accountability.....	111
IV. Contact Information	111
V. Revision History	111
Appendix E. Employee Artificial Intelligence Use Policy.....	112
I. Purpose	112
II. Scope.....	112
III. AI Governance.....	112
IV. AI Principles.....	112
IV.A. Transparency	112
IV.B. Explainability	112
IV.C. Human Involvement.....	113

IV.D. Security and Safety	113
IV.E. Human Centered	113
IV.F. Fairness	113
IV.G. Privacy-Enhanced	113
IV.H. Accountability	113
V. Acceptable Use	113
VI. Prohibited Use of AI Systems	114
VII. Certain AI Use Cases That Create Extra Risks.....	115
VIII. Exercise of Rights.....	115
IX. AI Vendors	115
X. AI Systems Incident Reporting.....	115
XI. Training	116
XII. Contact Information	116
XIII. Revision History	116
Appendix F. Incorporating Artificial Intelligence Regulatory Requirements in Contracts	118

Preface

Current through September 23, 2025.

When modern generative artificial intelligence (AI) was released to the public in 2022, there was uncertainty among the business community regarding how to adopt this technology in a safe, secure, and trustworthy manner. At the time, companies considered: (A) the compliance programs they had to establish based on laws of general applicability, such as data privacy, cybersecurity, intellectual property (IP), employment, consumer protection, and product liability, (B) protecting their trade secrets, confidential information, and personal data, (C) AI principles that countries around the world were rapidly adopting based on the Organisation for Economic Co-operation and Development AI Principles, (D) allocation of risk and liability through commercial contracts, insurance, and representations and warranties in mergers and acquisitions agreements, (E) their existing corporate governance policies and procedures, and (F) global frameworks that government agencies and privacy organizations were releasing for the development and deployment of AI.

From 2023 to 2024, however, the AI regulatory landscape became a lot more clear to help companies tether their practices to legal obligations. During this time period, the European Union and Colorado passed comprehensive and horizontal AI laws, which will serve as the foundation for future legislation. Under these laws, it became evident that AI will be regulated based on two overarching criteria: (A) the organization's role in the AI value chain (i.e., providers and developers of AI systems, deployers, general-purpose AI model providers, product manufacturers, distributors, and importers), and (B) the AI's risk ranking (prohibited, high and low risk). Despite the clarity provided by these laws on how to govern AI, companies were still grappling with which of their departments had ownership of AI governance in their organizations and how to allocate responsibility for AI oversight.

Some companies took a fragmented approach to AI oversight by assigning responsibility for different aspects of AI governance to relevant departments, such as the privacy and security office, human resources, IP, and general compliance (i.e., a decentralized approach), while others governed AI by addressing all of the above obligations through a comprehensive AI governance program (i.e., a centralized approach). Both approaches are feasible from an AI-governance perspective, but companies that adopt a holistic governance framework will help ensure that there is a consistent process throughout the organization and visibility among departments regarding AI oversight.

The purpose of my book is to provide a program that companies may adopt to ensure that they have a baseline framework for tackling the evolving AI-regulatory environment and remain nimble as there are changes in laws, frameworks, guidelines, technology and threats in the future. Companies that expend significant person power and resources developing and deploying AI may find it useful to future-proof their programs by considering not only the current regulatory environment but also what lies ahead.

In publishing this book, I describe the component parts organizations may consider adopting to address AI oversight in their companies. However, this book is not intended to suggest that all of these measures are mandatory under applicable laws. Organizations may be subject to different laws depending on their jurisdiction and practices. Moreover, some of the frameworks cited in this book may not be practical for organizations depending on their size, sector, and/or industry. Please consult with appropriate professionals and legal counsel before developing an AI governance program that is right for your organization.

Introduction

ARSEN KOURINIAN

Mayer Brown

Current through September 23, 2025.

This book provides the components that an artificial intelligence (AI) governance program may include based on global guidelines, frameworks, and laws. When implementing an AI governance program, organizations should not view it as a traditional checklist of compliance tasks. Rather, AI professionals may need to be forward-looking and anticipate where the law, technology and threats are headed. Taking this approach will require buy-in from company leadership and stakeholders, who may naturally ask to see the specific law requiring them to expend time and resources to build an AI governance program.

In some instances, this may be straightforward, because there are laws of general applicability in different countries, industries, and sectors that require AI compliance or comprehensive AI governance laws that apply, such as the European Union’s Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (EU AI Act) and Colorado’s Concerning Consumer Protections in Interactions with Artificial Intelligence Systems (Colorado AI Law). For companies that are not hyper-regulated by such laws, AI professionals may communicate to internal stakeholders the policy and practical justifications for implementing an AI governance program, such as protecting the organization’s brand and customer trust, and avoiding dismantling AI systems if rapidly developing laws require specific requirements that the company did not consider when adopting AI, which can be costly.

Fortunately, the AI governance trends are progressing in a predictable manner, as countries on six continents and the G7 have subscribed to the Organisation for Economic Co-operation and Development’s (OECD) AI Principles and look at global standards when adopting laws suitable for their jurisdictions.¹

This book initially provides scoping topics that the AI governance team may consider before developing an AI governance program and background on AI. This book then describes the ingredients for an AI governance program that companies may consider, which is composed of the following high-level components: (A) forming an AI governance team, (B) data governance, (C) risk management, (D) legal compliance, (E) mitigation measures, and (F) accountability.

¹See EUR. COMM’N, HIROSHIMA PROCESS INTERNATIONAL GUIDING PRINCIPLES FOR ORGANIZATIONS DEVELOPING ADVANCED AI SYSTEM, (Oct. 30, 2023) (reflecting the G7’s adoption of a non-exhaustive list of guiding principles that “build on the existing OECD AI Principles”).



As described in this book, these high-level components address the issues below.

- **AI Governance Team.** The first component in developing an AI governance program is to assemble an AI oversight team. Developing an AI governance program is not a one-person task. Regardless of a company's size, different stakeholders and employees with varying skillsets are necessary to develop and operationalize an AI governance program, from individuals trained in law, data privacy, intellectual property (IP), technical skills (e.g., data scientists, engineers, and computer scientists), human resources, marketing, and procurement. An organization will need internal and external specialists in these and other areas to adopt and implement a safe, secure, and trustworthy AI governance program. The AI governance team should be trained in regular cadence regarding the rapidly evolving laws, technologies, and threats in this area to ensure that the AI governance program is up to date and remains effective.
- **Data Governance.** It is important to use high-quality data to train, test, and validate an AI model and representative data as the input to ensure unbiased and accurate results. The AI governance team should consider maintaining a data provenance record that traces the data lineage in the AI system to identify the source of problems when it is deployed in the market and to demonstrate accountability.
- **Legal Compliance.** The AI governance team should understand which laws apply to the organization's development and use of AI. There could be laws of general applicability that apply to the AI systems, such as data privacy, product liability, employment, IP, and antitrust, as well as AI-specific laws that the organization needs to comply with. While the AI governance program described here provides the structure for AI oversight, the AI governance team will need to supplement it as necessary depending on specific laws that apply in a given use case.
- **Risk Management.** The AI governance team should understand the risks present in the organization's development and/or use of AI and rank them as prohibited, high, limited, or minimal. If the AI practice is prohibited, the AI governance team must stop the practice. For high risks, the organization may only proceed with the AI practice if it implements appropriate mitigation measures to reduce the risks. For limited-to-minimal AI risks, the mitigation measures may be limited, such as providing a transparency notice for chatbots or AI-generated content. Ultimately, we recommend the AI governance team document in an AI impact assessment the risks, likelihood and severity of harm, mitigation measures, and benefits of AI to demonstrate accountability.

- **Mitigation Measures.** Depending on the risks identified, the AI governance team may need to implement mitigation measures to proceed with developing or using AI. These mitigation measures may include (A) providing transparency notices regarding the AI systems and explaining how the AI works, (B) measuring whether the AI systems are fair and without bias, (C) ensuring that the AI is beneficial for individuals, society, and the environment, (D) confirming that the AI is accurate, robust, safe, and secure, (E) enhancing privacy protection, (F) adopting appropriate human oversight (human-in-the-loop, over-the-loop, or out-of-the-loop), (G) keeping technical documentation and logs, (H) monitoring the AI systems after they have been placed in the market and making adjustments as necessary, (I) putting in place feedback and decision review channels, (J) adopting appropriate AI contractual provisions, and (K) having a process to decommission an AI system.
- **Accountability.** Finally, the AI governance team should implement policies and procedures to demonstrate accountability and conformity of its AI systems. The AI governance team should also keep records showing how these policies and procedures are applied in practice.

By adopting these components, multinational companies can help make their AI governance programs compatible with global standards and withstand rapidly evolving AI regulations, technologies, and threats.

Chapter 1. Initial Scoping Topics to Consider

[ARSEN KOURINIAN](#)

Mayer Brown

Current through September 23, 2025.

1.I. OVERVIEW

If you are just starting to develop an artificial intelligence (AI) governance program, you should consider some initial scoping topics before putting pen to paper in preparing your AI governance policies, procedures, and practices. Below are some issues you may consider at the outset before developing your AI governance program.

- **The company's AI objectives.** You should first understand what your company's objectives are with AI. Some organizations with the resources and infrastructure train their own AI models for public consumption, while others may fine-tune an existing foundation model to develop an AI system. Other organizations do not develop their own AI solutions. Rather, these organizations utilize AI internally to maximize productivity, enhance customer service, or assist human resources functions. Those responsible for AI within an organization should talk to internal stakeholders to understand these goals.
- **The AI technologies your company is already developing or using.** AI governance is a subset of IT asset management. Thus, AI systems should be reflected in your broader technology oversight plan. AI leaders should inventory the technologies in operation to determine if the company is already using or developing AI systems. AI professionals may be surprised to learn that the company likely already has some semblance of AI within the organization or is working with vendors who incorporate AI within their products and services. AI professionals should review applicable procurement contracts to understand the rights and obligations related to the input data and output of AI.
- **Your company's role vis-à-vis AI.** The AI governance components you may need to adopt could vary if your company is an AI developer or deployer. Depending on your role during the AI product lifecycle, different aspects of risk mitigation are in your control. For example, an AI developer (also referred to as a "provider") is responsible for properly training and developing the AI system, while the AI deployer is responsible for using the AI system based on the developer's instructions for use and addressing deficiencies once the AI system is placed in the market. That said, sometimes these obligations may overlap because of downstream and upstream reporting of new AI risks.
- **Data rights.** Before using data to train or fine-tune an AI model or as the input prompt, companies should consider if they have the legal right to use the data for this purpose. This touches on a number of legal issues. Under data privacy laws, AI professionals should assess, among other things, whether the company gave an appropriate privacy notice to data subjects describing its use of personal data in AI, if the company documented an adequate lawful basis for processing and, where required, if the company obtained consent to use personal data for AI. AI professionals should also determine if they have the IP right to use the data and if there are any contractual terms limiting the company's use of data for its own commercial benefit. These are critical

issues because an AI system developed on unlawfully sourced data could be subject to model disgorgement.¹

- **A workforce ready for AI.** Organizations may need a workforce with different skillsets to develop and use AI systems. Organizations making sophisticated uses of AI may need data scientists, engineers, computer scientists, attorneys, human resources personnel, and business professionals who understand the company’s objectives and goals in connection with AI. Organizations should consider training their existing workforce to address these needs and hiring external candidates or retaining contractors to supplement missing skillsets.
- **Your company’s risk tolerance level.** When the explosion of generative AI first started trending the news, companies had different perspectives on AI depending on their risk tolerance levels. On the one end of the risk spectrum, some companies took a hardline approach to AI and prepared policies and procedures to prevent their employees from developing and using AI within the company. These companies were not yet certain where the law and technology were headed and wanted to put a pause on AI until the horizon was clear. On the opposite side of the spectrum, some companies were apprehensive about falling behind in the AI arms race and were willing to move aggressively to develop or use AI. AI professionals should gauge their organization’s temperature and understand where it falls within this spectrum. Organizations with a low risk tolerance level will want to implement the full nuts and bolts of AI governance before developing or using AI, while companies with higher risk tolerance may establish their AI governance program in parallel with AI development and/or use. If companies take the latter approach, legal should confer with technical teams to ensure that the AI system has appropriate technical features to permit completing the AI governance program after the product is developed.
- **Understand your company’s management approach.** Companies have different approaches for compliance oversight. Some have centralized management, whereby a single group at the top oversees compliance and communicates the implementation plan to local offices and regions. Others take a decentralized approach, whereby each office and region are responsible for complying with their local laws. Another approach is a hybrid of the two, which has both top-to-bottom and bottom-up reporting for compliance. This book provides the component parts that an organization can use for each of these management styles.

¹See In the Matter of Everalbum, Inc., US FTC, Docket No. C-4743, Decision and Order (May 6, 2021) (requiring respondent to delete or destroy an AI model developed with data allegedly obtained without consumer consent).

Chapter 2. Background

ARSEN KOURINIAN

Mayer Brown

Current through September 23, 2025.

2.I. BACKGROUND.....	4
2.II. AI DEFINED	4
2.III. INCIDENTS, HAZARDS, AND HARMS	9
2.III.A. The OECD and EU AI Act.....	9
2.III.B. US AI Laws.....	10
2.III.C. The Harm Categories Under the National Institute of Standards and Technology (NIST)	11
2.III.D. The OECD Classification of AI Impacts	12
2.III.E. Examples of AI Incidents	13
2.IV. STRONG/BROAD V. WEAK/NARROW AI.....	14
2.V. MACHINE LEARNING, DEEP LEARNING, AND NEURAL NETWORKS.....	15
2.VI. METHODS OF AI TRAINING	15
2.VII. GENERATIVE V. DISCRIMINATIVE MODELS	16
2.VIII. MULTI-MODAL MODELS	17
2.IX. AGENTIC AI	17
2.X. AI LIFECYCLE.....	17

2.I. BACKGROUND

This section provides background to help readers understand the basics of artificial intelligence (AI) as they develop their governance programs.

2.II. AI DEFINED

The term “artificial intelligence” or “AI” was first coined at Dartmouth College by John McCarthy, a computer scientist, at the 1956 Dartmouth Summer Research Project on Artificial Intelligence.¹ Since then, as AI technology and law have evolved, global laws, guidelines, principles, and frameworks have adopted different definitions for AI. For this book, we will use the AI definition adopted by the world’s first comprehensive and horizontal AI governance law—the EU AI Act. The EU AI Act’s definition of AI is suitable because it is derived from the Organisation for

¹See John McCarthy, et al., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (Aug. 31, 1955), <https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>; DARTMOUTH COLL., *Artificial Intelligence Coined at Dartmouth*, <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth> (last visited Aug. 14, 2024).

Economic Co-operation and Development’s (OECD) AI Principles,² which have been adopted by countries on six continents, including the G7 (US, UK, Japan, Canada, France, Germany, and Italy). With some variations, the EU AI Act’s definition has also been adopted under the Colorado AI Law³ and South Korea’s Framework Act on AI Development and Establishment of a Foundation for Trustworthiness (the South Korea AI Act).⁴

The EU AI Act defines an AI system as:

a [1] machine-based system that is designed to [2] operate with varying levels of autonomy and that [3] may exhibit adaptiveness after deployment, and that, [4] for explicit or implicit objectives, [5] infers, from the input it receives, how to [6] generate outputs such as predictions, content, recommendations, or decisions that can influence [7] physical or virtual environments[.]⁵

Below is an overview of each of these elements.

- **Machine-based system.** Machine-based system means that AI systems run on machines.⁶ This includes systems using “techniques such as machine learning and knowledge-based approaches, and application areas such as computer vision, natural language processing, speech recognition, intelligent decision support systems, intelligent robotic systems, as well as the novel application of these tools to various domains.”⁷ Machines can include both hardware or software components.⁸ “The hardware components refer to the physical elements of the machine, such as processing units, memory, storage devices, networking units, and input/output interfaces, which provide the infrastructure for computation. The software components encompass

²See Org. for Econ. Co-operation and Dev. [OECD], *Recommendation of the Council on Artificial Intelligence*, Legal Instrument 449 (July 11, 2023), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

³See Colo. Rev. Stat. §6-1-1701(2) (defining an AI system as “any machine-based system that, for any explicit or implicit objective, infers from the inputs the system receives how to generate outputs, including content, decisions, predictions, or recommendations, that can influence physical or virtual environments”).

⁴See Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.2.2, Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. *available at* https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf (“‘AI system’ refers to an artificial intelligence-based system that deduces outputs such as predictions, recommendations, and decisions that affect real and virtual environments for a given goal with various levels of autonomy and adaptability.”).

⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

⁷See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.* Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&accname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

⁸See EUR. COMM’N, Approval of the Content of the Draft Communication from the Commission Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act), at 2 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>.

computer code, instructions, programs, operating systems, and applications that handle how the hardware processes data and performs tasks.”⁹

- **Levels of autonomy.** AI systems are designed to operate with some degree of independence and limited human intervention.¹⁰ AI systems can “learn or act without human involvement following the delegation of autonomy and process automation by humans.”¹¹ Thus, to qualify as an AI system, the technology cannot operate with full human involvement and intervention.¹² That said, AI systems require human oversight during their lifecycle, “such as during AI system design, data collection and processing, development, verification, validation, deployment, or operation and monitoring.”¹³
- **Adaptiveness.** AI systems may be capable of self-learning and changing while in use.¹⁴ AI systems, particularly those based on machine learning, can evolve after initial development and modify their behavior through direct interaction with input and data before or after deployment by inferring patterns and relationships in data.¹⁵ Through this training process, some AI systems can make new inferences that their programmers did not initially consider.¹⁶ An AI system “may, but does not necessarily have to, possess adaptiveness or self-learning capabilities after deployment to constitute an AI system. Accordingly, a system’s ability to automatically learn, discover new patterns, or identify relationships in the data beyond what it was initially trained on is a facultative and thus not a decisive condition for determining whether the system qualifies as an AI system.”¹⁷

⁹See EUR. COMM’N, *Approval of the Content of the Draft Communication from the Commission -Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, at 2 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>.

¹⁰See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹¹See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 6, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&acname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

¹²See EUR. COMM’N, *Approval of the Content of the Draft Communication from the Commission -Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, at 3 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>.

¹³See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 6, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&acname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

¹⁴See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://www.eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹⁵See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 6, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&acname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

¹⁶*Id.*

¹⁷See EUR. COMM’N, *Approval of the Content of the Draft Communication from the Commission -Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, at

- **Explicitly defined or implicit objectives.** AI systems operate based on explicitly defined or implicit objectives, which may be different from the AI system’s intended purpose in a specific context.¹⁸ Explicitly defined objectives means that the developer encodes the objective directly into the AI system, such as game-playing systems, reinforcement learning systems, combinatorial problem-solving systems, planning algorithms, simple classifiers and dynamic programming algorithms.¹⁹ Implicit objectives means that the rules dictate the action an AI system must take according to the current circumstances, such as applying a rule for a self-driving car to stop at a red traffic light.²⁰ The implicit objectives may also not be explicitly programmed to perform a task, but the objective is incorporated through training data and a system architecture that learns to emulate data.²¹
- **Inferences.** Inference means that the AI system generates an output from its input, usually after deployment.²² AI systems “learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved.”²³ AI systems are different than human-defined and rules-based traditional software systems or programming approaches, which automatically perform operations.²⁴
- **Generate output.** AI systems generate outputs, such as predictions, content, recommendations, or decisions, depending on the context of their operations.²⁵ These

4 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>.

¹⁸See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://www.eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹⁹See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 7, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&accname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

²⁰*Id.*

²¹*Id.*

²²See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 9, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&accname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

²³See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L202401689>.

²⁴See EUR. COMM’N, *Approval of the Content of the Draft Communication from the Commission -Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, at 5 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (internal quotation marks and citation omitted).

²⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L202401689>.

outputs correspond to the different levels of human involvement, with “decisions” being the most autonomous type of output and “predictions” the least autonomous.²⁶

- **Environments.** The AI system’s environment can be physical or virtual and include environments describing human activity, such as biological signals or human behavior.²⁷ The AI system observes the environment using data and sensor inputs and is influenced through actions (actuators). “Sensors and actuators are either humans or components of machines or devices.”²⁸

While the term “AI system” is broad in scope, it does not include traditional software systems or programming approaches and systems that are based on rules humans define to automatically execute operations.²⁹ Below is a list of systems that European Commission guidance considers outside the scope of the definition for AI systems.³⁰

- Linear or logistic regression methods, which have been used for many years and do not transcend basic data processing.
- Physics-based systems that approximate complex atmospheric processes and enable faster and efficient forecasts.
- Satellite telecommunication systems that optimize bandwidth allocation and resource management.
- Basic data processing systems that follow predefined, explicit instructions or operations and fixed human-programmed rules (e.g., database management systems that sort or filter data, spreadsheets, and calculation software).
- Systems that are intended for descriptive analysis, hypothesis testing and visualization (e.g., sales dashboard showing total sales or trends, charts and graphs, and other systems that simply present data in an informative way).
- Classical heuristics that involve rule-based approaches, pattern recognition, or trial-and-error strategies (e.g., a chess program).
- Systems whose performance can be achieved via a basic statistical learning rule (e.g., certain financial forecasting that predicts future stock prices, weather prediction systems that use last week’s average temperatures, static estimation systems, or demand forecasting for store products).

²⁶See Org. for Econ. Co-operation and Dev. [OECD], *Explanatory Memorandum on the Updated OECD Definition of an AI Sys.*, at 9, Artificial Intelligence Papers No. 8 (Mar. 5, 2024), <https://www.oecd-ilibrary.org/docserver/623da898-en.pdf?expires=1726780135&id=id&acname=guest&checksum=A143B27CE601FB8C7C801F979A3BF120>.

²⁷*Id.* at 7.

²⁸*Id.*

²⁹See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 12, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

³⁰See EUR. COMM’N, *Approval of the Content of the Draft Communication from the Commission -Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act)*, at 8-10 (Feb. 6, 2025), <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> (internal quotation marks and citation omitted).

2.III. INCIDENTS, HAZARDS, AND HARMS

One of the most critical functions of an AI governance program is to manage risks and increase safety. Below are some of the global definitions, guidelines, and issues considered AI incidents, hazards, and harms.

2.III.A. *The OECD and EU AI Act*

To help identify risks that may arise with AI system development and deployment, the OECD maintains the AI Incidents Monitor (AIM) to “show patterns and establish a collective understanding of AI incidents and their multifaceted nature and serve as an important tool for trustworthy AI.”³¹ For the past 10 years (2014-2024), the AIM has documented close to 10,000 AI incidents, which are tracked by AI principles, such as privacy and data governance, respect for human rights, robustness and digital security, transparency and explainability, accountability, reskill or upskill, performance, fairness, safety, and democracy and human autonomy, and harms, such as physical, psychological, economic/property, reputational, public interest, human rights, and other.³² These harms impact a number of industries, such as digital security, government, security and defense, mobility and autonomous vehicles, arts, entertainment and recreation, media, social platforms and marketing, health care, drugs and biotechnology, financial and insurance services, business processes and support services, education and training, and IT infrastructure and hosting.³³

In identifying these harms, the OECD applies a uniform definition for “AI incidents,” which is an event where the AI system development or use results in actual harm.³⁴ The OECD defines an AI incident as follows:³⁵

An AI incident is an event, circumstance or series of events where the development, use or malfunction of one or more AI systems directly or indirectly leads to any of the following harms:

- (a) injury or harm to the health of a person or groups of people;
- (b) disruption of the management and operation of critical infrastructure;
- (c) violations of human rights or a breach of obligations under the applicable law intended to protect fundamental, labour and IP rights;
- (d) damage to property, communities or the environment.

The EU AI Act has adopted a similar standard by defining the synonymous term “serious incident” as follows:³⁶

³¹See *OECD AI Incidents Monitor (AIM)*, [OECD.AI POLICY OBSERVATORY, https://oecd.ai/en/incidents?search_terms=%5B%5D&and_condition=false&from_date=2014-01-01&to_date=2024-04-15&properties_config=%7B%22principles%22:%5B%5D,%22industries%22:%5B%5D,%22harm_types%22:%5B%5D,%22harm_levels%22:%5B%5D,%22harmed_entities%22:%5B%5D%7D&only_threats=false&order_by=date&num_results=20](https://oecd.ai/en/incidents?search_terms=%5B%5D&and_condition=false&from_date=2014-01-01&to_date=2024-04-15&properties_config=%7B%22principles%22:%5B%5D,%22industries%22:%5B%5D,%22harm_types%22:%5B%5D,%22harm_levels%22:%5B%5D,%22harmed_entities%22:%5B%5D%7D&only_threats=false&order_by=date&num_results=20) (last visited Aug. 14, 2024).

³²*Id.*

³³*Id.*

³⁴See *OECD AI Incidents Monitor, Methodology and Disclosures*, OECD.AI POLICY OBSERVATORY, <https://oecd.ai/en/incidents-methodology> (last visited Aug. 14, 2024).

³⁵*Id.*

³⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(49), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

‘[S]erious incident’ means any incident or malfunctioning of an AI system that directly or indirectly leads to any of the following:

- (a) the death of a person, or serious damage to a person’s health;
- (b) a serious and irreversible disruption of the management and operation of critical infrastructure;
- (c) the infringement of obligations under Union law intended to protect fundamental rights;
- (d) serious harm to property or the environment.

Under the EU AI Act, serious incidents involving high-risk AI systems trigger reporting and investigation obligations.³⁷

The OECD has also adopted a definition for “AI hazard,” which is an event where the AI system’s development or use may potentially harm someone.³⁸ The OECD defines an “AI hazard” as follows:³⁹

An AI hazard is an event, circumstance or series of events where the development, use or malfunction of one or more AI systems could plausibly lead to an AI incident, i.e., any of the following harms:

- (a) injury or harm to the health of a person or groups of people;
- (b) disruption of the management and operation of critical infrastructure;
- (c) violations to human rights or a breach of obligations under the applicable law intended to protect fundamental, labour and IP rights;
- (d) harm to property, communities or the environment.

2.III.B. US AI Laws

In the United States, state AI laws are focused on protecting individuals from harms arising from discrimination and lack of transparency.

For example, under the Colorado AI Law, there is an obligation to use “reasonable care to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination. ...”⁴⁰ The Colorado AI Law defines “algorithmic discrimination” as:⁴¹

[A]ny condition in which the use of an artificial intelligence system results in an unlawful differential treatment or impact that disfavors an individual or group of individuals on the basis of their actual or perceived age, color, disability, ethnicity, genetic information, limited proficiency in the English language, national origin, race, religion, reproductive health, sex, veteran status, or other classification protected under the laws of this state or federal law.

³⁷See generally *id.* at Art. 73.

³⁸See *OECD AI Incidents Monitor, Methodology and Disclosures*, [OECD.AI](https://oecd.ai/en/incidents-methodology) POLICY OBSERVATORY, <https://oecd.ai/en/incidents-methodology> (last visited Aug. 14, 2024).

³⁹*Id.*

⁴⁰See COLO. REV. STAT. §6-1-1702(1).

⁴¹See COLO. REV. STAT. §6-1-1701(1)(a).

If developers and deployers discover algorithmic discrimination arising from the AI system they must report it to the Colorado Attorney General within 90 days.⁴²

In addition, employment laws related to AI focus on harms arising from discrimination. For example, Illinois amended its Human Rights Act to prohibit an employer from using an AI system “that has the effect of subjecting employees to discrimination on the basis of protected classes ... or to use zip codes as a proxy for protected classes”⁴³ Similarly, New York City Local Law 144 is also focused on avoiding discrimination by requiring employers to conduct a bias audit before using AI to make employment decisions.⁴⁴

Lastly, another harm state legislatures are trying to protect consumers from is lack of transparency. Chatbot and generative AI laws in California,⁴⁵ New Jersey,⁴⁶ Maine⁴⁷ and Utah⁴⁸ require disclosing to individuals that they are interacting with AI and not a human, so that individuals are not deceived regarding communications with AI. California also passed AI laws that require developers of AI systems to: (a) make available disclosures in AI-generated content and provide users an AI detection tool so that individuals are not deceived by deepfakes;⁴⁹ and (b) post a high-level summary of the data they used to train AI systems so that consumers understand the output’s potential shortcomings.⁵⁰

2.III.C. The Harm Categories Under the National Institute of Standards and Technology (NIST)

The NIST organizes AI harms in three groups—harm to people, organization, and ecosystem.⁵¹ Harm to people can involve harm to (1) an individual’s civil liberties, rights, physical or psychological safety, or economic opportunity, (2) a group or community, such as discrimination against a population subgroup, and (3) society, such as to democratic participation or educational access.⁵² Harm to an organization can involve disruption to business operations, security breaches, monetary loss, and impact on reputation.⁵³ Harm to an ecosystem relates to the impact on interconnected and interdependent elements and resources, the global financial system, supply chain or interrelated systems, and natural resources, the environment and planet.⁵⁴

The NIST also identifies the following twelve risks unique to generative AI:⁵⁵

- The ability to use or develop AI that gives bad actors access to information related to chemical, biological, radiological, or nuclear weapons or other dangerous materials or agents;

⁴²See COLO. REV. STAT. §§6-1-1702(5); 6-1-1703(7).

⁴³See 775 ILCS 5/2-102(L).

⁴⁴See NEW YORK CITY, N.Y., CODE §20-871(a).

⁴⁵See CAL. BUS. & PROF. CODE §17941.

⁴⁶See N.J. STAT. ANN. §56:18-2.

⁴⁷See ME. REV. STAT. ANN. tit. 10, §1500-Y.

⁴⁸See U.C.A. 1953 §13-77-103.

⁴⁹See CAL. BUS. & PROF. CODE §22757, *et seq.*

⁵⁰See CAL. CIV. CODE §3110, *et seq.*

⁵¹See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

⁵²*Id.*

⁵³*Id.*

⁵⁴*Id.*

⁵⁵See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1, (July 2024), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf>.

- The ability of AI to confidently provide false information, referred to as confabulation, hallucination, or fabrication;
- AI making it easier for individuals to be exposed to dangerous, violent, or hateful content and providing individuals recommendations to carry out self-harm or illegal activities;
- Harm to individuals when AI discloses personal and sensitive information or gives individuals the ability to de-anonymize data so that it is identifiable;
- Harm to the environment caused by the energy and carbon emissions from training and operating AI models;
- Harmful bias or homogenization when non-representative training data is used and the AI amplifies historical, societal, and systemic biases based on historical data;
- Individuals attributing human characteristics to AI (i.e., anthropomorphism), which leads to automation bias, over-relying on AI, having algorithmic aversion, or becoming emotionally entangled with AI;
- AI watering down content integrity by failing to distinguish facts from opinions and being used for large-scale disinformation campaigns;
- AI lowering the barrier and making it easier for cybersecurity attacks, with AI also being vulnerable to unique attacks, such as prompt injection and data poisoning;
- AI making it easier to infringe the intellectual property rights of others, exposing trade secrets, and allowing others to plagiarize and illegally replicate content;
- AI making it easier to produce obscene, degrading, and/or abusive imagery, including related to children and nonconsenting adults; and
- The AI value chain can be non-transparent and difficult to trace, including the third-party components used and right to use data, and lack of transparency and accountability to the end users.

2.III.D. The OECD Classification of AI Impacts

The OECD Framework for the Classification of AI Systems describes the types of impact AI systems may have on human rights, democratic values, environment, well-being and society.⁵⁶ Examples of these impacts include:⁵⁷

- Liberty, safety and security;
- Physical, psychological and moral integrity;
- Freedom of thought, conscience and religion;
- Rule of law, absence of arbitrary sentencing;
- Equality and non-discrimination;
- Social and economic rights (e.g., health, education);
- Quality of democratic institutions (e.g., free elections);
- Right to property;
- Aggregate society-level risk;
- Physical and mental health;
- Housing;
- Income and wealth;
- Quality of job;

⁵⁶See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

⁵⁷*Id.*

- Quality of environment;
- Social connections;
- Civic engagement;
- Education, knowledge and skills;
- Work-life balance; and
- Duration of impact.

2.III.E. Examples of AI Incidents

The severity of AI incidents will vary and have different impact. In the most extreme situation, AI systems may cause death. For example, in February, 2024, a partially automated vehicle collided with a vehicle that was stopped on a highway, resulting in a fatality.⁵⁸ This incident underscores the need for drivers to be fully “in the loop,” particularly for partially automated driving vehicles.

An AI incident may also cause emotional distress and embarrassment. For example, it was reported that a mother was misidentified as a trespasser when a supermarket’s facial recognition technology scanned her face and incorrectly identified her on the store’s database of known offenders or suspects.⁵⁹ The supermarket staff told her to leave, even though she gave the store staff three forms of identification to show that she was not the culprit.⁶⁰ The mother said she felt discriminated because of her skin color and was embarrassed when the store staff accused her of theft.⁶¹ The New Zealand Privacy Commissioner launched an inquiry related to the supermarket’s alleged use of facial recognition technology based on concerns about the AI system’s bias and accuracy.⁶²

In the United States, the Federal Trade Commission (FTC) has also investigated a retail chain’s alleged use of facial recognition technology.⁶³ The FTC alleged that the retail chain used facial recognition technology “to identify customers who may have engaged in shoplifting or other problematic behavior.”⁶⁴ The FTC claimed that the facial recognition technology generated thousands of false-positive matches, which allegedly caused customers to be humiliated, embarrassed, and harassed by store staff.⁶⁵ The FTC also alleged that the facial recognition technology disproportionately impacted people of color.⁶⁶ These AI incidents demonstrate the need to use representative data for training AI and testing its accuracy and fairness before deploying it into the market.

⁵⁸See *Electric Ford SUV Driver Was Using Automated System Before Fatal Crash, Investigators Say*, CBS NEWS (Apr. 12, 2024), <https://www.cbsnews.com/detroit/news/electric-ford-suv-driver-was-using-automated-system-before-fatal-texas-crash-investigators-say/>.

⁵⁹See Sandra Conchie, *Supermarket Facial Recognition Trial: Rotorua Mother’s ‘Discrimination’ Ordeal*, ROTORUA DAILY POST (Apr. 12, 2024), <https://www.nzherald.co.nz/rotorua-daily-post/news/supermarket-facial-recognition-trial-rotorua-mothers-discrimination-ordeal/IK4ZEJHLOVFRMLMDE6LX4AR57PE/>.

⁶⁰*Id.*

⁶¹*Id.*

⁶²*Id.*

⁶³See Press Release, Fed. Trade Comm’n, *Rite Aid Banned from Using AI Facial Recognition After FTC Says Retailer Deployed Technology Without Reasonable Safeguards* (Dec. 19, 2023), <https://www.ftc.gov/news-events/news/press-releases/2023/12/rite-aid-banned-using-ai-facial-recognition-after-ftc-says-retailer-deployed-technology-without>.

⁶⁴*Id.*

⁶⁵*Id.*

⁶⁶*Id.*

Bad actors are also using AI systems to scam victims. In one incident, a scammer used a CEO’s deepfake audio to call an employee and obtain the employee’s contact details.⁶⁷ This incident demonstrates the need to be vigilant against deepfakes to protect personal information from similar security incidents.

Other AI manipulations may have a benign impact on individuals but harm public interest, such as a social media post that went viral about an AI-generated image depicting the sun covered by the moon’s shadow with swirling sunrays.⁶⁸ Upon investigation, the image was found to not be a genuine image of the April 8, 2024 solar eclipse.⁶⁹ While this situation did not lead to any physical or emotional harm, it could impact the veracity of photographs maintained by the scientific community.

2.IV. STRONG/BROAD V. WEAK/NARROW AI

There are different ways to define and categorize AI, but one common distinction is between weak/narrow and strong/broad AI. Weak/narrow AI is trained and designed to perform specific tasks.⁷⁰ Weak/narrow AI is a bit of a misnomer because it is anything but weak, as it includes the current most advanced AI systems.⁷¹ This includes virtual companions, real-time universal translation, next-gen cloud robotics, autonomous surgical robotics, robotic personal assistants, cognitive cybersecurity, neuromorphic computing, autonomous systems, spam filters, chatbots, and real-time emotion analytics.⁷²

Strong/broad AI refers to the hypothetical or aspirational goal of creating AI that can perform any intellectual task that a human can.⁷³ The AI, like a human, is self-aware, conscious, able to solve problems and constantly evolves over time.⁷⁴ Strong/broad AI is only a theoretical concept and has not been developed yet.⁷⁵ Some leading scientists even contend that the notion of strong/broad AI does not exist in the first place.⁷⁶ There is also the term “super AI,” which refers to AI that has cognitive abilities that surpass human intelligence.⁷⁷ However, this concept mostly exists in science fiction movies and does not exist.⁷⁸ Notably, despite significant advances, AI fails to exceed human abilities in “more complex cognitive tasks, such as visual commonsense reasoning and advanced-

⁶⁷See Michael Kan, *Scammers Target LastPass Employee With CEO Audio Deepfake*, PCMag (Apr. 12, 2024), <https://www.pcmag.com/news/scammers-target-lastpass-employee-with-ceo-audio-deepfake>.

⁶⁸See Srijanee Chakraborty, *AI Generated Image Viral As Photograph of Total Solar Eclipse 2024*, BOOM (Apr. 12, 2024), <https://www.boomlive.in/fact-check/fact-check-total-solar-eclipse-2024-artificial-intelligence-ai-image-viral-fake-news-24886>.

⁶⁹*Id.*

⁷⁰See Cole Stryker and Eda Kavlakoglu, *What is Artificial Intelligence (AI), Types of Artificial Intelligence: Weak AI vs. Strong AI*, IBM, <https://www.ibm.com/topics/artificial-intelligence> (Aug. 16, 2024).

⁷¹*Id.*

⁷²See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 14, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁷³*Id.* at 14–15.

⁷⁴See *What is Strong AI?*, IBM, <https://www.ibm.com/topics/strong-ai> (last visited Aug. 14, 2024); U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 14, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁷⁵See *What is Strong AI?*, IBM, <https://www.ibm.com/topics/strong-ai> (last visited Aug. 14, 2024); U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 14–15, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁷⁶See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 14, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁷⁷*Id.* at 15.

⁷⁸*Id.*

level mathematical problem-solving (competition-level math problems).”⁷⁹ This explains why we do not yet have strong/broad or super AI.

2.V. MACHINE LEARNING, DEEP LEARNING, AND NEURAL NETWORKS

Machine learning, deep learning, and neural networks are all sub-disciplines of AI.⁸⁰ Both machine learning and deep learning are trained to learn on data.⁸¹ However, machine learning requires more human involvement and structured data, while deep learning is a scalable type of machine learning that can learn from unstructured data in raw form.⁸² Once trained, a machine learning model can ingest new or unseen data and predict outcomes by detecting patterns contained in the new data.⁸³ Machine learning models also continue to improve their ability to make predictions through each iteration of training, testing and tuning.⁸⁴ Machine learning includes such applications as speech recognition, video recommendation models, spam filtering, and targeted advertisement.⁸⁵

Neural networks, also known as artificial neural networks, attempt to mimic the way a human brain works (but do not operate like the human brain).⁸⁶ Neural networks are made of node layers that contain input, hidden, and output layers, which are connected to each other and have an associated weight and threshold.⁸⁷ When any node’s output is higher than the threshold, that node is activated and sends data to the next node.⁸⁸ However, if the output node is not above the threshold, it will not pass data to the next node.⁸⁹ Deep learning involves greater depth in the neural network’s layers.⁹⁰ If a neural network has more than three layers, which includes inputs and the outputs, it is considered a deep-learning algorithm.⁹¹

2.VI. METHODS OF AI TRAINING

⁷⁹See NESTOR MASLEJ ET AL., INST. FOR HUMAN-CENTERED AI, STANFORD UNIV., THE AI INDEX 2024 ANNUAL REPORT 81 (Apr. 2024).

⁸⁰See Cole Stryker and Eda Kavlakoglu, *What is Artificial Intelligence (AI)*, IBM, <https://www.ibm.com/topics/artificial-intelligence> (Aug. 16, 2024).

⁸¹*Id.*

⁸²*Id.*

⁸³See DAVID LESLIE ET AL., THE ALAN TURING INST., AI ETHICS AND GOVERNANCE IN PRACTICE: AN INTRODUCTION 20 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-ai-ethics-an-intro_1.pdf.

⁸⁴*Id.*

⁸⁵See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 16, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024); David Leslie et al., THE ALAN TURING INST., AI ETHICS AND GOVERNANCE IN PRACTICE: AN INTRODUCTION 20 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-ai-ethics-an-intro_1.pdf.

⁸⁶See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 19, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁸⁷*Id.*; *What is a Neural Network?*, IBM, <https://www.ibm.com/topics/neural-networks> (last visited Aug. 14, 2024).

⁸⁸See How Neural Network Models in Machine Learning Work, TURING, <https://www.turing.com/kb/how-neural-network-models-in-machine-learning-work> (last visited Aug. 14, 2024).

⁸⁹*Id.*

⁹⁰See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 19, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁹¹See IBM Data and AI Team, *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What’s the Difference?*, IBM (July 6, 2023), <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>.

There are different methods to train an AI model—supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning involves training an AI model using labeled data.⁹² For example, if you want an AI system to distinguish between legitimate and spam email, you could label data as real email or spam as part of the training process. AI models trained using supervised learning perform (1) classification, which assign test data into categories through prediction (e.g., grouping images by cats or dogs), or (2) regression, which determines the relationship between features and a target variable (e.g., predicting household energy usage based on outside temperature).⁹³

Unsupervised learning is trained on unlabeled data and is able to look for patterns in the data without labels.⁹⁴ Unsupervised learning is commonly used for dimensionality reduction or clustering, which involves grouping data points based on a similar metric.⁹⁵ For example, an AI model developed through unsupervised learning that is fed sales data can identify different types of clients who commonly purchase products.⁹⁶ This AI model could help a company’s marketing team target specific customer groups based on consumption history and other characteristics.⁹⁷ Semi-supervised learning is trained on a mix of labeled and unlabeled data.⁹⁸ Reinforcement learning, on the other hand, learns through trial and error whereby correct actions are rewarded while incorrect actions receive negative feedback.⁹⁹ The AI is able to learn from its mistakes by continuously interacting with the environment, instead of existing data.¹⁰⁰

Reinforcement learning is used in games and autonomous driving with simulated environments.¹⁰¹

2.VII. GENERATIVE V. DISCRIMINATIVE MODELS

Generative AI refers to deep learning that focuses on creating new data or content that resembles the original data or content, such as images, text, and other content.¹⁰² It is used to generate art, music, and natural language processing tasks.¹⁰³ Examples of generative AI models include naïve Bayes models, hidden Markov models, linear discriminant analysis, and generative adversarial networks.¹⁰⁴ A discriminative AI model, however, is a type of machine learning method that learns

⁹²See Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. OF MGMT.: IDEAS MADE TO MATTER (Apr. 21, 2021), <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.

⁹³See DAVID LESLIE ET AL., THE ALAN TURING INST., AI ETHICS AND GOVERNANCE IN PRACTICE: AN INTRODUCTION 20 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-ai-ethics-an-intro_1.pdf.

⁹⁴See Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. OF MGMT.: IDEAS MADE TO MATTER (Apr. 21, 2021), <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.

⁹⁵See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 18, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

⁹⁶*Id.*

⁹⁷*Id.*

⁹⁸See Dave Bergman, *What is Semi-Supervised Learning?*, IBM (Dec. 12, 2023), <https://www.ibm.com/topics/semi-supervised-learning>.

⁹⁹See Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. OF MGMT.: IDEAS MADE TO MATTER (Apr. 21, 2021), <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>.

¹⁰⁰See U.A.E., NAT’L PROGRAM FOR ARTIFICIAL INTEL., AI GUIDE 18, https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_EN_v1-online.pdf (last visited Aug. 14, 2024).

¹⁰¹*Id.*

¹⁰²See Kim Martineau, *What is Generative AI?*, IBM (Apr. 20, 2023), <https://research.ibm.com/blog/what-is-generative-AI>.

¹⁰³Shradha Pujari, *Generative AI vs. Discriminative AI*, MEDIUM (Sept. 16, 2023), <https://pujarishradha.medium.com/generative-ai-vs-discriminative-ai-75415c8f7adf>.

¹⁰⁴See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

to discriminate between classes.¹⁰⁵ Discriminative AI models focus on predicting data labels by distinguishing between dataset classes, but are not capable of generating new data.¹⁰⁶ This type of AI is used to perform classification tasks, such as spam filtering and image classification.¹⁰⁷ Examples of discriminative AI models include regression analyses, support-vector machines, traditional neural networks, decision trees and random forests.¹⁰⁸

2.VIII. MULTI-MODAL MODELS

Multi-modal models refers to a subset of deep learning that can handle or integrate multiple types of data or modalities, such as images, text, speech, and video.¹⁰⁹ These models are trained on large amounts of data and learn patterns and the association between text descriptions and corresponding images, videos, or audio recordings.¹¹⁰

2.IX. AGENTIC AI

Agentic AI is a machine-based system that is capable of autonomously developing sub-goals, choosing actions, and adapting its behavior in real time to achieve a high-level objective.¹¹¹ Unlike simple chatbots, which merely generate outputs in response to specific prompts, agentic AI is designed to operate continuously, monitor the environment, and revise its own plans to reach an end goal.¹¹² As these systems scale, their capacity to act across different industries and functions (e.g., customer service, manufacturing, sales support, and health and social care) heightens both their commercial utility and the need for robust governance mechanisms.¹¹³

2.X. AI LIFECYCLE

An AI system's lifecycle is a cyclical process that can be split into four stages: (1) plan and design, (2) prepare input data, (3) build and validate, and (4) deploy and monitor.¹¹⁴

¹⁰⁵See Apostol Vassilev, et al., *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* 93, NIST (Jan. 2024) <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf>, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf>.

¹⁰⁶See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

¹⁰⁷Shradha Pujari, *Generative AI vs. Discriminative AI*, MEDIUM (Sept. 16, 2023), <https://pujarishradha.medium.com/generative-ai-vs-discriminative-ai-75415c8f7adf>.

¹⁰⁸See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

¹⁰⁹See *Multimodal Generative AI Systems*, META (Dec. 12, 2023), <https://ai.meta.com/tools/system-cards/multimodal-generative-ai-systems/>.

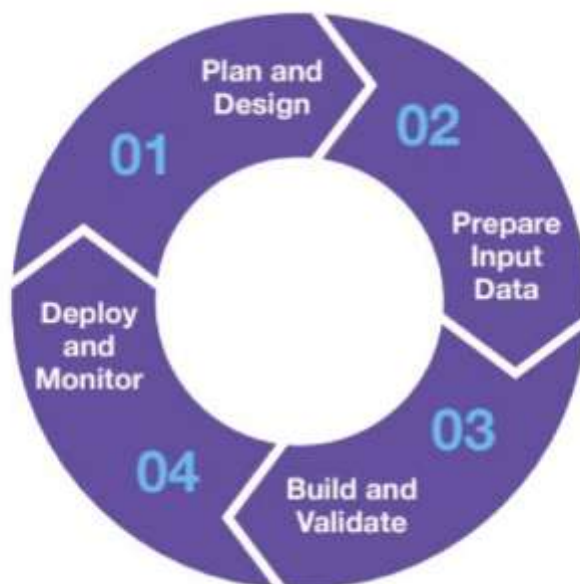
¹¹⁰*Id.*

¹¹¹Mark Purdy, *What is Agentic AI, and How Will it Change Work?*, HARVARD BUS. REV. (Dec. 12, 2024).

¹¹²*Id.*

¹¹³*Id.*

¹¹⁴See SAUDI DATA & AI AUTH., *AI ETHICS PRINCIPLES* 9 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>; see also Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, Digital Economy Papers No. 323, (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en> (“The AI system lifecycle can serve as a complementary structure for understanding the key technical characteristics of a system. The lifecycle encompasses the following phases that are not necessarily sequential: planning and design; collecting and processing data; building and using the model; verifying and validating; deployment; and operating and monitoring. ...”).



For the first stage, the developer will define the problem, find a data-driven approach to support the problem, select a framing approach on technology and system that govern AI, conduct a feasibility assessment for the selected approach, and define the key performance indicators.¹¹⁵ For the second stage, the developer will gather, discover, assess, cleanse, and validate the data and transform the data into AI model input features. For the third stage, the developer will train and test the model, tune the hyperparameters, validate model performance, and conduct a risk evaluation.¹¹⁶ And, for the final stage, the organization deploys the AI model to the system, creates versioning structures, periodically monitors the production model performance, and conducts an assessment to determine if there is a need to change the design according to the results of periodic reviews.¹¹⁷

¹¹⁵See SAUDI DATA & AI AUTH., AI ETHICS PRINCIPLES 9 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

¹¹⁶*Id.* at 10.

¹¹⁷*Id.*

Chapter 3. Players In the AI Supply Chain

ARSEN KOURINIAN

Mayer Brown

Current through September 23, 2025.

3.I. OVERVIEW.....	20
3.II. AI Architecture Providers.....	20
3.II.1. Semiconductor Chips.....	21
3.II.2. Accelerators.....	21
3.II.3. Cloud Computing.....	21
3.III. FOUNDATION MODELS.....	22
3.IV. AI SYSTEM PROVIDERS.....	28
3.V. AI SYSTEM DEPLOYERS.....	28
3.VI. OTHER ACTORS IN THE AI SUPPLY CHAIN.....	29
3.VII. AI ROLE UNDER DATA PRIVACY LAWS.....	29

3.I. OVERVIEW

As described in the scoping topics, it is important that the artificial intelligence (AI) governance team understand its role in connection with AI systems.¹ In the supply chain, there are a number of operators during the AI lifecycle—AI architecture providers, foundation model developers, AI system providers, AI system importers and distributors, product manufacturers, and AI system deployers. Moreover, because of the intersection of AI and data privacy laws, the organization needs to assess whether it is a controller or processor of the personal data it uses in connection with AI. Depending on its intended business plans, an organization may occupy several roles in the AI ecosystem and under data privacy laws.

While there are shared AI governance principles for all operators in the supply chain, there may be different risk mitigation obligations and considerations within the organization’s control depending on its role.² However, actors in the AI supply chain may need to work together to manage AI risks.³

3.II. AI Architecture Providers

The development of AI systems is a complex process that involves multiple stages and a variety of specialized components. Among these, AI architecture providers play a crucial role in ensuring that AI systems are both powerful and efficient. These providers include manufacturers of semiconductor chips, accelerators, and cloud computing services. Each of these components

¹See ISO/IEC 42001:2023(E), §4.1 (INT’L ORG. FOR STANDARDIZATION 2023) (“The organization shall consider the intended purpose of the AI systems that are developed, provided or used by the organization. The organization shall determine its roles with respect to these AI systems.”).

²See *id.* (“External and internal issues to be addressed under this clause can vary according to the organization’s roles and jurisdiction and their impact on its ability to achieve the intended outcome(s) of its AI management system.”).

³See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) 10, (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

contributes uniquely to the AI supply chain, facilitating the creation, training, and deployment of AI models.

3.II.1. Semiconductor Chips

Semiconductor chips are the foundational hardware for any AI system. These chips, often referred to as processors, are responsible for executing the complex calculations required for AI algorithms. The most common types of processors used in AI are Central Processing Units (CPUs), Graphics Processing Units (GPUs), and more recently, specialized AI chips like Tensor Processing Units (TPUs) and Neural Processing Units (NPU).⁴

- CPUs: Traditionally used for general-purpose computing, CPUs are versatile but may not be the most efficient for AI tasks that require parallel processing.⁵
- GPUs: Originally designed for rendering graphics, GPUs excel at parallel processing, making them ideal for training large AI models.⁶
- TPUs and NPUs: These are specialized chips designed specifically for AI workloads. They offer optimized performance for tasks such as deep learning and neural network inference.⁷

The development and manufacturing of these chips involve advanced semiconductor fabrication technologies, which are often concentrated in a few leading companies. This makes the supply chain for semiconductor chips highly specialized and critical for the performance of AI systems.

3.II.2. Accelerators

AI accelerators are hardware components designed to speed up AI computations.⁸ They can be integrated into existing systems or used as standalone devices. Accelerators are essential for handling the intensive computational demands of AI applications, particularly in real-time scenarios.⁹ Along with GPUs, addressed above, the following are examples of AI accelerators:

- Field-Programmable Gate Arrays (FPGAs): These are customizable chips that can be programmed to perform specific tasks efficiently. They offer flexibility and can be reconfigured for different AI applications.¹⁰
- Application-Specific Integrated Circuits (ASICs): Unlike FPGAs, ASICs are designed for a specific application. They offer high performance and efficiency but cannot be reprogrammed like FPGAs.¹¹

Accelerators reduce the time required for training AI models and improve the performance of AI inference, making them indispensable in the AI supply chain.

3.II.3. Cloud Computing

Cloud computing services provide the infrastructure and platforms necessary for developing, training, and deploying AI models. These services offer scalable resources that can be adjusted based on the computational needs of the AI system.

⁴See Stephen J. Bigelow, *GPUs vs. TPUs vs. NPUs: Comparing AI Hardware Options*, TECHTARGET (Aug. 27, 2024), <https://www.techtargget.com/whatis/feature/GPUs-vs-TPUs-vs-NPUs-Comparing-AI-hardware-options>.

⁵See *id.*

⁶See *id.*

⁷See *id.*

⁸See *What is an AI accelerator?*, IBM (June 12, 2024), <https://www.ibm.com/think/topics/ai-accelerator>.

⁹See *id.*

¹⁰See *id.*

¹¹See *id.*

- Infrastructure as a Service (IaaS): Provides virtualized computing resources over the internet. Companies can rent servers, storage, and networking capabilities, which are essential for large-scale AI training.¹²
- Platform as a Service (PaaS): Offers a platform allowing customers to develop, run, and manage applications without dealing with the underlying infrastructure. This is particularly useful for AI development environments.¹³
- Software as a Service (SaaS): Delivers software applications over the internet. AI-based SaaS solutions can include anything from machine learning frameworks to pre-trained models.¹⁴

Cloud computing enables organizations to access high-performance computing resources without the need for significant capital investment in physical hardware. This democratizes access to AI capabilities, allowing even smaller companies to develop sophisticated AI systems.

3.III. FOUNDATION MODELS

Foundation models—which are also referred to as dual-use foundation models,¹⁵ general-purpose AI (GPAI) models,¹⁶ and frontier models¹⁷—are pre-trained on high volume of data and often used to operate AI systems. Stanford University first coined this term in the publication, *On the Opportunities and Risks of Foundation Models*, which defines a foundation model as an AI “model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks. ...”¹⁸ Only a limited number of companies provide foundation models, because developing one requires massive amounts of data to train, expensive computational resources, and technical expertise.¹⁹ For example, two of the leading foundation model providers—OpenAI and Google—used an estimated \$78 and \$191 million worth of compute to train their AI models, respectively.²⁰ Since 2019, the United States has the greatest number of foundation models released, followed by China and the United Kingdom.²¹

¹²See *What is infrastructure as a service (IaaS)?*, IBM (Dec. 18, 2024), <https://www.ibm.com/think/topics/iaas>.

¹³See *What is platform as a service (PaaS)?*, IBM (Sept. 10, 2024), <https://www.ibm.com/think/topics/paas>.

¹⁴See *What is software as a service (SaaS)?*, IBM (Aug. 5, 2025), <https://www.ibm.com/think/topics/saas>.

¹⁵See, e.g., Exec. Order No. 14110 §3(k) (Oct. 30, 2023), <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.

¹⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(63), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹⁷See AUSTRALIAN GOV'T, DEPT. OF INDUS., SCI., AND RES., SAFE AND RESPONSIBLE AI IN AUSTRALIA CONSULTATION 4 (2024), https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-and-responsible-ai-in-australia-governments-interim-response.pdf.

¹⁸See RISHI BOMMASANI, ET AL., CTR. FOR RSCH. ON FOUNDATION MODELS, STANFORD INST. FOR HUMAN-CENTERED A.I., ON THE OPPORTUNITIES AND RISKS OF FOUNDATION MODELS 3 (July 12, 2023).

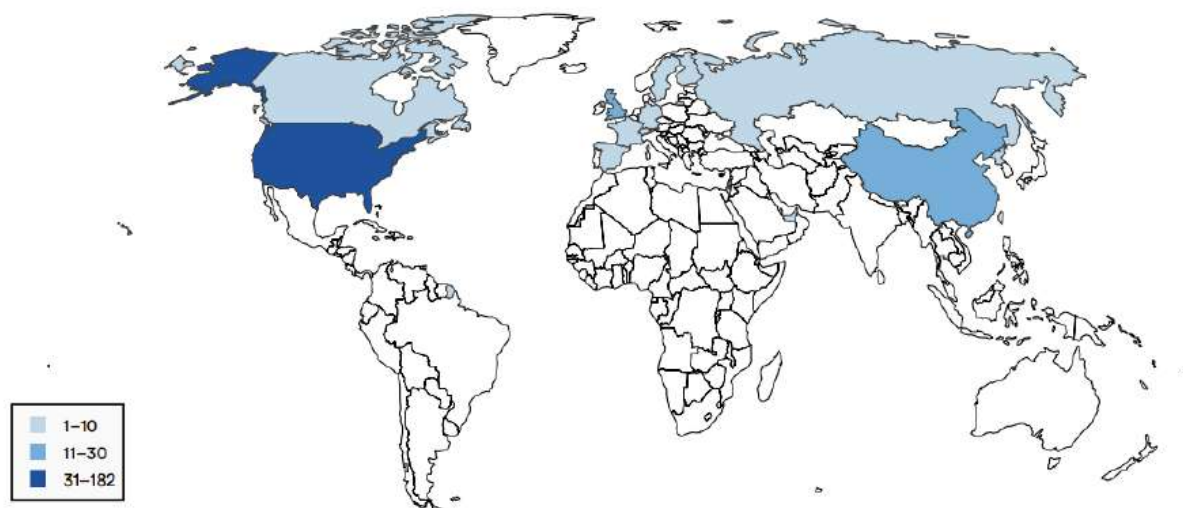
¹⁹See COMPETITION & MARKETS AUTH., AI FOUNDATION MODELS: INITIAL REPORT 27–39, (Sept. 18, 2023).

²⁰See NESTOR MASLEJ ET AL., INST. FOR HUMAN-CENTERED AI, STANFORD UNIV., THE AI INDEX 2024 ANNUAL REPORT 5 (Apr. 2024).

²¹*Id.*

Number of foundation models by geographic area, 2019–23 (sum)

Source: Bommissani et al., 2023 | Chart: 2024 AI Index report



Foundation models are not developed to only perform a particular task. Rather, AI system providers can customize them through fine-tuning to perform specific functions. Fine-tuning is a process applied to the foundation model to add specific capabilities or improvements.²² Through fine-tuning, foundation model users can (1) specialize the pre-trained foundation model to a particular domain or task and (2) adjust the model so that it performs as intended (e.g., to avoid biased, false, or harmful results).²³ AI service providers can integrate foundation models into new and existing products and services, such as productivity software, search engines, social media, financial tools, healthcare, robotics, education, and legal tools.²⁴

In the EU, the regulation of foundation models was a contentious topic during the last phase of the EU AI Act’s legislative process in the fall of 2023.²⁵ France, Germany, and Italy wanted to avoid overregulation of foundation models and instead called for self-regulation through codes of conduct. Ultimately, a compromise was reached whereby unique obligations were included in the EU AI Act for GPAI models, depending on whether a model carries systemic or non-systemic risk.

Under the EU AI Act, a GPAI model is defined as “an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market.”²⁶

²²See COMPETITION & MARKETS AUTH., AI FOUNDATION MODELS: INITIAL REPORT 11, (Sept. 18, 2023).

²³*Id.* at 12.

²⁴*Id.* at 25–26.

²⁵See Luca Bertuzzi, *France, Germany, Italy Push for “Mandatory Self-Regulation” for Foundation Models in EU’s AI Law*, EURACTIV (Nov. 24, 2023), <https://www.euractiv.com/section/artificial-intelligence/news/france-germany-italy-push-for-mandatory-self-regulation-for-foundation-models-in-eus-ai-law/>.

²⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(63), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

The European Commission issued guidance to provide further clarification regarding whether an AI model meets this criteria under the EU AI Act. To determine whether an AI model is GPAI, the model's training compute needs to be greater than 10^{23} floating point operations (FLOPs) and it needs to be capable of generating language.²⁷ The European Commission settled on this standard for two reasons: (1) this threshold is based on the current standard for models trained on high volume of data and the approximate amount of compute commonly used to train a model with one billion parameters on large amounts of data; and (2) models that generate language are more often able to competently perform a wider range of tasks than other models.²⁸

To help illustrate the circumstances under which an AI model would be considered a GPAI model, the European Commission provides the following examples:²⁹

²⁷See EUR. COMM'N, *Approval of the content of the draft Communication from the Commission – Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act)*, at 7 (July 18, 2025), <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>.

²⁸See *id.*

²⁹See *id.* at 8–9.

Example of a model in scope

- A model is trained on a broad range of natural language data (i.e. text) curated and scraped from the internet and other sources (as is currently typical for language models) using 10^{24} FLOP.
 - The criterion from paragraph 17 indicates that the model should be a general-purpose AI model because it can generate text and its training compute is greater than 10^{23} FLOP. Training on a broad range of natural language further indicates that the model should display significant generality and should be capable of competently performing a wide range of distinct tasks. Therefore, the model likely is a general-purpose AI model.

Examples of models out of scope

- A model is trained on natural language data, using 10^{22} FLOP, and is not able to competently perform a wide range of distinct tasks.
 - While the model can generate text, its training compute is not greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should not be a general-purpose AI model. The fact that the model is not capable of performing a wide range of distinct tasks confirms that it is not a general-purpose AI model.
- A model is trained specifically for the task of transcribing speech to text, using 10^{24} FLOP.
 - The model can generate text and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (transcribing speech), it is not actually a general-purpose AI model.
- A model is trained specifically for the task of generating speech from text, using 10^{24} FLOP.
 - The model can generate speech and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (generating speech from text), it is not actually a general-purpose AI model.
- A model is trained specifically for the task of increasing the resolution of images, using 10^{24} FLOP.

- The model can generate images and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (upsampling images), it is not actually a general-purpose AI model.
- A model is trained specifically for the task of filling in damaged or missing parts of images, using 10^{24} FLOP.
 - The model can generate images and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (inpainting images), it is not actually a general-purpose AI model.
- A model is trained specifically for playing chess or video games, using 10^{24} FLOP.
 - While the model's training compute is greater than 10^{23} FLOP, if it cannot generate text, speech, text-to-image, and text-to-video, the criterion from paragraph 17 indicates that the model should not be a general-purpose AI model. If the model can only competently perform a narrow set of tasks (playing chess or video games), this fact confirms that it is not a general-purpose AI model.
- A model is trained specifically for modelling weather patterns or physical systems, using 10^{24} FLOP.
 - While the model's training compute is greater than 10^{23} FLOP, if it cannot generate text, speech, text-to-image, and text-to-video, the criterion from paragraph 17 indicates that the model should not be a general-purpose AI model. If the model can only competently perform a narrow set of tasks (modelling weather patterns or physical systems), this fact confirms that it is not a general-purpose AI model.
- A model is trained specifically for generating music, based on text prompts including lyrics, using 10^{24} FLOP.
- (21) The model can generate speech (in the form of song lyrics) and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (generating music, based on text prompts including lyrics) it is not actually a general-purpose AI model.
- A model is trained specifically for generating sound effects, using 10^{24} FLOP. It can generate language (speech), but the quality of the content of the speech generated is low (e.g. the content is not consistent over the course of a brief conversation).
 - The model can generate language and its training compute is greater than 10^{23} FLOP. Therefore, the criterion from paragraph 17 indicates that the model should be a general-purpose AI model. However, if the model can only competently perform a narrow set of tasks (sound effects) it is not actually a general-purpose AI model.

In addition, the EU AI Act distinguishes GPAI models that have “systemic risk,” which is defined as “a risk that is specific to the high-impact capabilities of general-purpose AI models, having

a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.”³⁰ A GPAI model is presumed to meet this criteria if it is trained using a total computing power of more than 10^{25} FLOPS,³¹ which reflects the most powerful foundation models.

Under the EU AI Act, non-systemic risk GPAI model providers are required to (1) prepare and maintain the model’s technical documentation, including its training and testing process and the results of its evaluation, (2) make available to providers information and documentation if they intend to integrate the GPAI model into their AI systems, (3) maintain a policy to respect EU copyright law, (4) make publicly available a summary of the content used to train the model, (5) cooperate with the European Commission and the national competent authorities, and (6) appoint an EU-authorized representative if the GPAI model provider is established outside the EU.³²

Systemic risk GPAI model providers are also required to comply with the aforementioned obligations, along with (1) performing model evaluation and adversarial testing to identify and mitigate systemic risks, (2) assessing other systemic risks that may stem from the development, placement in the market, or use of the model, (3) reporting serious incidents and taking corrective actions, and (4) ensuring that there is an adequate level of cybersecurity.³³ GPAI model providers may also rely on codes of practices to demonstrate compliance with the EU AI Act.³⁴

The EU AI Act limits the obligations of free and open-license GPAI model providers that do not have systemic risk, such as not requiring them to draw up and keep up-to-date technical documentation and making available information and documentation to providers of AI systems who intend to integrate the GPAI model into their AI systems.³⁵

As part of the AI value chain, downstream actors may modify or finetune GPAI models. The European Commission does not consider all downstream modifications of GPAI models as making the model modifier a GPAI provider under the EU AI Act.³⁶ Rather, a downstream modifier becomes a GPAI provider “only if the modification leads to a significant change to the model’s generality, capabilities, or systemic risk.”³⁷ The European Commission considers this criterion met when the training compute used for the modification is more than a third of the original model’s training compute.³⁸ If the downstream modifier does not know this value, then the threshold is a third of 10^{25} FLOP for systemic risk GPAIs and a third of 10^{23} FLOP for non-systemic risk GPAIs.³⁹ If the downstream modifier becomes a GPAI provider, its obligations are “limited to that

³⁰See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(65), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

³¹See *id.* at Art. 51.

³²See *id.* at Art. 53 & 54.

³³See *id.* at Art. 55.

³⁴See *id.* at Art. 53(4).

³⁵See *id.* at Art. 53(2).

³⁶See EUR. COMM’N, *Approval of the content of the draft Communication from the Commission – Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act)*, at 20 (July 18, 2025), <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>.

³⁷See *id.*

³⁸See *id.*

³⁹See *id.*

modification or fine-tuning, for example by complementing the already existing technical documentation with information on the modifications, including new training data sources. ...”⁴⁰

In the United States, California took the lead in regulating foundation models by enacting the Transparency in Frontier Artificial Intelligence Act (Frontier AI Act).⁴¹ Under the Frontier AI Act, developers of frontier models that were trained using a quantity of computing power greater than 10^{26} integer or FLOP are subject to a number of obligations (some of which get triggered if they meet an annual gross revenue of more than \$500 million), including: (1) writing, implementing, complying with, and publishing on their website a frontier AI framework that applies to the organization (including a national and international standard, such as NIST AI RMF and ISO/IEC 42001); (2) posting a transparency report on their website before making available a new frontier model to third parties; (3) transmitting to the California Office of Emergency Services (OES) catastrophic risk assessments; (4) reporting critical safety incidents to the OES; and (5) providing whistleblower protections to covered employees.

3.IV. AI SYSTEM PROVIDERS

A provider develops, or has developed, an AI system and puts it into service or makes it available in the market.⁴² A provider may also be the same organization that developed the foundation model, although it is not necessary for an AI system provider to train its own AI foundation model. Rather, a provider can integrate a foundation or other AI model into existing or new consumer-facing products or services.⁴³ Providers are expected to implement risk mitigation measures when developing AI, consider the potential AI uses and instruct users regarding the restrictions and limitations applicable to the AI system.⁴⁴ For the AI governance program described below, foundation model developers and AI system providers will generally be referred to as “developers” of AI systems.

3.V. AI SYSTEM DEPLOYERS

A deployer is the user of an AI system.⁴⁵ The competency level of deployers may vary from (1) an amateur, who has no training, (2) trained practitioner, who has some specific training on how to use an AI system, and (3) an AI expert, who has specific training and knowledge on how the AI system

⁴⁰Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Recital 109, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

⁴¹See S.B. 53, 2025-2026 Leg., Reg. Sess. (Cal. 2025).

⁴²See *id.* at Art. 3(3).

⁴³See COMPETITION & MARKETS AUTH., AI FOUNDATION MODELS: INITIAL REPORT 55–57, (Sept. 18, 2023).

⁴⁴See *The Artificial Intelligence and Data Act - Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>.

⁴⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(4), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, at 25, Digital Economy Papers No.323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en> (“Users of an AI system or application are the individuals or groups that utilize the system for a specific purpose.”).

works.⁴⁶ Deployers that use AI systems are expected to follow the instructions for use, assess and mitigate risks, and continuously monitor the AI system and the impact of its output.⁴⁷

3.VI. OTHER ACTORS IN THE AI SUPPLY CHAIN

There are other relevant actors in the AI supply chain, such as importers, distributors, product manufacturers, and impacted stakeholders.

The EU AI Act creates separate obligations for importers and distributors of AI systems. An importer is the entity that first makes a non-EU company’s AI system available in the EU market, while a distributor makes an AI system available in the EU market but is not otherwise a provider or importer.⁴⁸ AI importers and distributors are primarily obligated to verify whether an AI system is in conformity and not place the AI system in the market if it is not compliant.⁴⁹

While not defined, the EU AI Act also considers a “product manufacturer” as an operator in the AI value chain.⁵⁰ The manufacturer of certain regulated products may be considered a provider if it places a high-risk AI system on the market together with the product under its own name or trademark or puts the AI system into service under its name or trademark after the product is placed on the market.⁵¹ Finally, related to AI harms, the OECD indicates that “impacted stakeholders” are also parties in the AI value chain.⁵² These are individuals who “can be indirectly or directly affected by the deployment of an AI system or application but do not necessarily interact with the system.”⁵³

Examples of impact stakeholders include workers, consumers, business, government agencies and regulators, scientists and researchers, and children or other vulnerable or marginalized groups.⁵⁴

3.VII. AI ROLE UNDER DATA PRIVACY LAWS

Organizations should also assess whether they are a “controller” or “processor” of the personal data they use as developers or deployers of AI systems.

Under data privacy laws, the controller (also known as the “business”) decides the purposes and means of processing personal data (i.e., how and why personal data should be processed).⁵⁵ The processor (also known as a “service provider”), on the other hand, processes personal data on the

⁴⁶See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, at 25–26, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

⁴⁷See *The Artificial Intelligence and Data Act - Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>.

⁴⁸See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(6) & (7), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

⁴⁹*Id.* at Art. 23 & 24.

⁵⁰*Id.* at Art. 3(8).

⁵¹*Id.* at Art. 25(3).

⁵²See Org. for Econ. Co-operation and Dev. [OECD], *Framework for the Classification of AI Systems*, at 25, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

⁵³*Id.*

⁵⁴*Id.* at 26.

⁵⁵See, e.g., Regulation (EU) 2016/679, General Data Protection Regulation, art. 4(7), O.J. (L 119, 04.05.2016), <https://gdpr-info.eu/art-4-gdpr/>; CAL. CIV. CODE §1798.140(d)(1).

controller's behalf.⁵⁶ To help assess party roles, the UK Information Commissioner's Office (ICO) states that any of the following decisions could make the organization a controller in connection with AI processing activities:

- to collect personal data in the first place;
- what types of personal data to collect;
- the purpose or purposes the data are to be used for;
- which individuals to collect the data about;
- how long to retain the data; and
- how to respond to requests made in line with individuals' rights.⁵⁷

The ICO also advises that an organization is considered a processor if it does not have its own purpose for processing the data and is simply following instructions.⁵⁸ A processor, however, may still make technical decisions about the means of processing in connection with AI, such as:

- the IT systems and methods you use to process personal data;
- how you store the data;
- the security measures that will protect it; and
- how you retrieve, transfer, delete or dispose of that data.⁵⁹

Applying these guideposts, a developer would likely be a controller if it collects personal data to train its AI model because it decides to collect personal data, the type of personal data it needs to train its AI model, whose personal data it needs to collect (e.g., scrapping personal data from the internet), and how long it will keep the personal data in its model.⁶⁰ The developer, however, can also be a processor. For example, if a developer provides its AI system to customers, it could be a processor if it uses the personal data used as the input to only provide an output for the customer.⁶¹ In this scenario, the developer is providing the technical measures (i.e., the means) for the customer to make processing decisions, but is not deciding on the processing purpose.

A deployer would be the controller of personal data that it uses as input in an AI system, for AI processing activities that make decisions about data subjects, or if it used an AI service provider to process personal data or improve its own isolated AI model.⁶² A deployer, however, may also act as a processor. For example, if the deployer uses a developer's AI system to provide a service to its own customers, it may be a processor of the customers' personal data, and the AI system developer could be a subprocessor of that personal data.

⁵⁶See, e.g., Regulation (EU) 2016/679, General Data Protection Regulation, art. 4(7), O.J. (L 119, 04.05.2016), <https://gdpr-info.eu/art-4-gdpr/>; CAL. CIV. CODE §1798.140(ag)(1).

⁵⁷See *What are the Accountability and Governance Implications of AI?*, INFO. COMM'R.'S OFF. (Mar. 15, 2023), <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/what-are-the-accountability-and-governance-implications-of-ai/#howshouldweunderstand>.

⁵⁸*Id.*

⁵⁹*Id.*

⁶⁰*Id.* ("First, the prediction service provider decides how to create and train the model that powers its services, and processes data for these purposes. It is likely to be a controller for this element of the processing.")

⁶¹*Id.* ("An organisation provides live AI prediction and classification services to clients. It develops its own AI models, and allows clients to send queries via an API ('what objects are in this image?') to get responses (a classification of objects in the image). ... [T]he provider processes data to make predictions and classifications about particular examples for each client. The client is more likely to be the controller for this element of the processing, and the provider is likely to be a processor.")

⁶²*Id.* ("An AI service provider isolates different client-specific models. This enables each client to make overarching decisions about their model, including whether to further process personal data from their own context to improve their own model. As long as the isolation between different controllers is complete and auditable, the client will be the sole controller and the provider will be a processor.")

Ultimately, deciding whether the organization is a controller or processor is a fact-specific inquiry and may also depend on whether the organization can operate if it is restricted by contractual terms applicable to processors with respect to the personal data it is processing.

Chapter 4. Implementing An AI Governance Program

ARSEN KOURINIAN

Mayer Brown

Current through September 23, 2025.

4.I. OVERVIEW.....	33
4.I.A. European Union.....	33
4.I.B. United Kingdom.....	33
4.I.C. United States.....	34
4.I.D. Rest of World.....	34
4.I.E. AI Frameworks.....	34
4.II. ASSEMBLING AN AI GOVERNANCE TEAM.....	36
4.III. DATA GOVERNANCE.....	38
4.IV. LEGAL COMPLIANCE.....	41
4.V. RISK MANAGEMENT.....	42
4.V.A. Identify and Rank AI Risks.....	43
4.V.A.1. Prohibited AI Risks.....	44
4.V.A.2. High AI Risks.....	45
4.V.A.2.a. EU AI Act High Risks.....	46
4.V.A.2.b. Colorado AI Law.....	46
4.V.A.2.c. South Korea AI Act.....	46
4.V.A.2.d. Other High-Risk Considerations.....	46
4.V.A.3. Limited and Minimal AI Risks.....	47
4.V.B. Likelihood and Severity of Harm.....	47
4.V.C. Document an AI Impact Assessment.....	49
4.VI. MITIGATION MEASURES.....	50
4.VI.A. Transparency and Explainability.....	51
4.VI.B. Fair and Unbiased.....	55
4.VI.C. Human-Centered and Beneficial for the Environment and Society.....	57
4.VI.D. Accuracy.....	58
4.VI.E. Robustness.....	60
4.VI.F. Safe and Secure.....	61

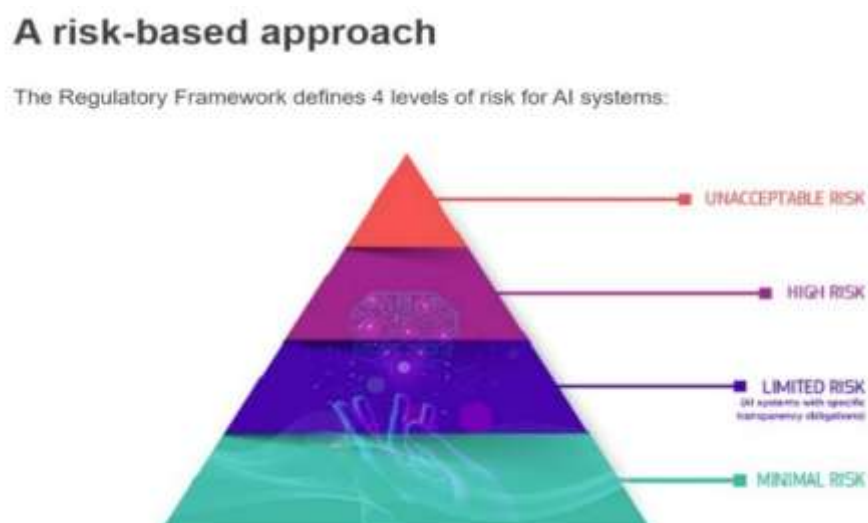
4.VI.G. Enhancing Privacy Protection	64
4.VI.H. Human Oversight.....	64
4.VI.I. Technical Documentation and Logs.....	66
4.VI.J. Post-Market Monitoring	67
4.VI.K. Communication Channels and Contestability	67
4.VI.L. Adopt Appropriate AI Contractual Provisions	67
4.VI.M. Decommissioning the AI System.....	69
4.VII. ACCOUNTABILITY	70

4.I. OVERVIEW

Globally, there are a number of guidelines, frameworks, and laws related to artificial intelligence (AI) governance. Adopting an AI governance program based on all of these global standards is possible, because they share common components but present the issues in a different way.

4.I.A. European Union

The EU AI Act takes a risk-based and role-specific approach to AI governance. Organizations need to identify and rank risks and understand whether they are a provider or deployer (among other party roles) in deciding which measures to employ for risk mitigation. The European Commission uses the triangle visual below to describe the risk-based approach under the EU AI Act.¹



4.I.B. United Kingdom

The United Kingdom takes a pro-innovation approach to AI regulation, which is principles-based and context-specific approach to AI governance. The United Kingdom's AI-principles are (A)

¹See EUR. COMM'N, *AI Act*, SHAPING EUROPE'S DIGITAL FUTURE (July 22, 2024), <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

safety, security, and robustness, (B) appropriate transparency and explainability, (C) fairness, (D) accountability and governance, and (E) contestability and redress.² The United Kingdom expects that regulators will implement these principles “by issuing guidance on best practice for adherence to these principles.”³

4.I.C. United States

Under the current administration, the United States federal government is taking a pro-innovation approach to AI regulation. Under the Removing Barriers to American Leadership in AI, which the White House issued on January 23, 2025, the United States federal government is taking steps to remove “existing AI policies and directives that act as barriers to American AI innovation, clearing a path for the United States to act decisively to retain global leadership in artificial intelligence.”⁴ To further this goal, the United States has also announced a \$500 billion private sector investment to fund infrastructure for AI, so that the United States stays competitive with other countries.⁵ Based on this new approach, it is not anticipated that the United States will pass comprehensive federal AI legislation, similar to the EU AI Act. However, states are taking the helm by passing legislation to protect individuals from potential AI harms.

4.I.D. Rest of World

South Korea appears to be following in the footsteps of the European Union by passing a comprehensive AI law, the South Korea AI Act, which goes into effect in January 2026.⁶ The South Korea AI Act requires AI businesses to develop a risk management plan, provide transparency and explainability notices, document the training data used to develop the AI, apply appropriate human oversight, conduct AI impact assessments, and maintain policies and procedures for the safe and trustworthy use and development of AI.⁷

Singapore, on the other hand, offers a unique framework, the Model Artificial Intelligence Governance Framework, Second Edition, which is neither risk- nor principles-based. Rather, Singapore’s framework provides practical guidance and use cases to help organizations implement measures to mitigate risks.⁸

4.I.E. AI Frameworks

The National Institute of Standards and Technology (NIST) AI RMF provides a three-prong governance structure that is focused on identifying and mitigating risks. This includes (A) map,

²DEP’T FOR SCI., INNOVATION, & TECH., A PRO-INNOVATION APPROACH TO AI REGULATION: GOVERNMENT RESPONSE, 2024, CP 1019, §3.2.3 (U.K.), <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>.

³*See id.*

⁴THE WHITE HOUSE, Removing Barriers to American Leadership in Artificial Intelligence (Jan. 23, 2025), <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.

⁵THE WHITE HOUSE, President Trump Positions U.S. as Global Superpower in Manufacturing (Mar. 20, 2025), <https://www.whitehouse.gov/articles/2025/03/president-trump-positions-u-s-as-global-superpower-in-manufacturing/>.

⁶Lee & Ko, *A New Era for AI, Republic of Korea Takes a Bold Step With AI Regulation* (Jan. 2025), <https://www.leeko.com/news/techai/202501/202501.pdf>.

⁷*See id.*

⁸*See* PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

which requires identifying risks based on context, (B) measure, which requires assessing, analyzing, and tracking risks, and (C) manage, which requires prioritizing risks and acting on them based on projected impact.⁹ The NIST depicts its AI governance framework in the three-prong circle below.¹⁰



The International Organization for Standardization's (ISO) Information Technology – Artificial Intelligence – Management System, ISO/IEC 42001, is another well-recognized AI governance framework. This standard provides a holistic approach to AI governance by addressing the organization's (A) leadership and oversight, (B) context, (C) planning, (D) support needs, (E) operational planning and control of AI risks, (F) performance evaluation, and (G) continuous improvement.¹¹ Like the EU AI Act and NIST, the ISO framework is also focused on assessing and managing AI risks through shared principles, along with broader considerations.

ISO/IEC 42001



⁹See NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 20 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹⁰*Id.*

¹¹See ISO/IEC 42001:2023(E) (INT'L ORG. FOR STANDARDIZATION 2023).

This book puts these and other major global guidelines, frameworks, and laws together to formulate a plan to ensure the safe, secure, and trustworthy development and use of AI. Ultimately, organizations should not view these global standards as mutually exclusive. This will hopefully alleviate any hesitation that AI professionals may have in implementing an AI governance program, because they are unable to settle on a preferred framework.

4.II. ASSEMBLING AN AI GOVERNANCE TEAM

Organizations that develop or use AI systems should consider forming an AI governance team that will oversee the technology.¹²

To form an AI governance team, organizations should identify relevant stakeholders and form a coordinating body.¹³ This may include an AI ethics board or committee that will oversee the AI governance process, provide independent advice, and share the guidelines, standards, and tools to help other team members in the organization develop or use AI systems responsibly.¹⁴

The oversight body should be composed of multidisciplinary advisors with different skillsets,¹⁵ backgrounds, and representation across the organization, such as the AI use case owner, engineers, data scientists, product developers, legal, human resources, customer support, marketing, and relevant leaders (e.g., Chief Operating Officer, Chief Data Officer, Chief Privacy Officer, Chief Information Security Officer, and/or Chief Technology Officer). The organization should clearly define each AI oversight team member's roles and responsibilities.

The AI governance team members should possess sufficient AI literacy and training to perform their duties.¹⁶ The AI governance team should receive training on, among other things, AI laws and

¹²See U.S. NAT'L SEC. AGENCY, A.I. SEC. CTR. ET AL., DEPLOYING AI SYSTEMS SECURELY: BEST PRACTICES FOR DEPLOYING SECURE AND RESILIENT AI SYSTEMS (Apr. 2024), <https://media.defense.gov/2024/Apr/15/2003439257/-1/-1/0/CSI-DEPLOYING-AI-SYSTEMS-SECURELY.PDF>.

¹³See PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 21-22 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

¹⁴See ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 18 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

¹⁵See Integrated Innovation Strategy Promotion Council, *Social Principles of Human-Centric AI 5*, CABINET SECRETARIAT OF JAPAN, <https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf> (last visited Aug. 14, 2024) ("It is important to have sufficient human resources with acquired application skills such as implementation and design of AI systems and a basic knowledge of data and AI. These skills would be acquired in a cross-disciplinary range of fields in a combined and integrated framework."); Smart Dubai, *AI Ethics Principles & Guidelines 7*, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) ("AI systems should be developed by diverse teams which include experts in the area in which the system will be deployed.").

¹⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 4, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 ("Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used."); PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 22 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf> ("Personnel and/or departments having internal AI governance functions should be fully aware of their roles and responsibilities, be properly trained,

regulations, how AI technology operates, AI use cases, AI risks, and sector- and geographic-specific issues applicable to the company. The AI training should not be stagnant. Rather, the AI oversight team should receive training in regular cadence (e.g., monthly, quarterly, semiannually, or annually) so that the organization is up to date on the latest developments in AI technologies and laws.¹⁷

The AI oversight team is responsible for the following tasks related to the organization's development and use of AI systems:

- Identifying the organization's AI objectives, which should be consistent with the AI policy, measurable, address applicable requirements, monitored, communicated, updated as necessary, and documented.¹⁸
- Ensuring that the organization's AI policies and objectives are established and compatible with the company's strategic direction.¹⁹
- Developing, documenting, and communicating within the organization an AI policy that is appropriate to the organization's purpose, provides a framework for setting AI objectives, and includes a commitment to meet applicable requirements and to continually improve.²⁰
- Securing the resources needed to establish, implement, maintain, and continually improve the AI management system.²¹
- Integrating the AI management system into the organization's business process.²²
- Communicating throughout the organization the importance of the AI governance program.²³
- Overseeing the AI governance program to ensure that it achieves its intended results.²⁴
- Providing direction and support to employees so that the AI governance program is effective.²⁵

Lastly, companies should consider the talent that they employ and their management style as they develop their AI governance programs (e.g., centralized versus decentralized).²⁶

and be provided with the resources and guidance needed for them to discharge their duties.”); Integrated Innovation Strategy Promotion Council, *Social Principles of Human-Centric AI* 8, CABINET SECRETARIAT OF JAPAN, <https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf> (last visited Aug. 14, 2024 (“AI users should have a general understanding of AI and should acquire sufficient education to use it properly, given that AI platforms are much more complicated than already developed conventional tools. Regarding developers of AI, meanwhile, it is of course necessary for them to master the basics of AI technology.”)); U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 23, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137>; ISO/IEC 42001:2023(E), §7.2 (INT’L ORG. FOR STANDARDIZATION 2023) (“The organization shall: ... where applicable, take actions to acquire the necessary competence, and evaluate the effectiveness of the actions taken.”).

¹⁷See Smart Dubai, *AI Ethics Principles & Guidelines* 11, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) (last visited Aug. 14, 2024) (stating that “[e]ducation should evolve and reflect the latest developments in AI, enabling to adapt to societal change”).

¹⁸See ISO/IEC 42001:2023(E), §6.2 (INT’L ORG. FOR STANDARDIZATION 2023).

¹⁹See *id.* at §5.1.

²⁰See *id.* at §5.2.

²¹See *id.* at §7.1.

²²See *id.* at §5.1.

²³See *id.*

²⁴See *id.*

²⁵See *id.*

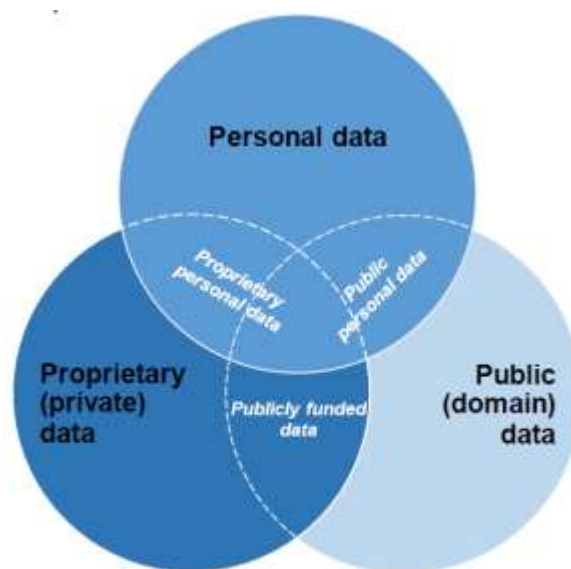
²⁶See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 21 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for->

Organizations may take different approaches for assembling their oversight bodies, depending on their size and resources. For example, smaller organizations may not have the resources or workforce to establish multidisciplinary, central governing bodies. Instead, smaller organizations may utilize existing staff to focus on mitigating high-risk areas in connection with their AI systems.

4.III. DATA GOVERNANCE

Organizations that develop or use AI systems should implement data governance to ensure that the AI systems are adequately trained, function as intended, and do not have harmful impacts on the public.

The data an organization may use generally fall within three categories: (1) proprietary data, which is private data a corporation has an economic interest to restrict access to and is protected under IP, trade secret, contract and/or cyber-criminal laws, (2) public data, which may in some situations not be protected by data privacy and IP laws and may be shared for access and reused through open data regimes, and (3) personal data, which relates to an identified or identifiable individual and is protected under data privacy laws.²⁷ As reflected in the OECD's Venn diagram below, these data sets may intersect:²⁸



[organisation/ai/sgmodelaigovframework2.pdf](https://www.oecd.org/organisation/ai/sgmodelaigovframework2.pdf) (“Organisations may also consider determining the appropriate features in their internal governance structures. For example, when relying completely on a centralised governance mechanism is not optimal, a de-centralised one could be considered to incorporate ethical considerations into day-to-day decisionmaking at the operational level, if necessary.”); ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 19 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics-beautified-201223_v2.pdf (“Notably, the degree of centralisation or decentralisation of the governance structure needs to be suitable for the organisational structure and culture. This entails identifying the appropriate balance between flexibility and rigidity to ensure optimal business and process execution. In the case where the business needs to be nimble and responsive to changes in operational requirements, it might be more effective to go with a more decentralised approach, where AI governance considerations and decisions are made on a more frequent basis at the operational level.”).

²⁷See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, at 37-38, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en>.

²⁸*Id.* at 38.

Organizations should initially assess if they have the legal right to use these types of data for training and/or as input in the AI system. This includes whether the organization has (1) given an appropriate privacy notice and/or obtained all necessary consents to use personal data for model training²⁹ or making decisions with AI systems, (2) the IP right to use data in the AI system,³⁰ and (3) the contractual right to use data, based on licensing rights and limitations in relevant commercial agreements pursuant to which the organization obtained the data.

Once an organization confirms that it has the legal right to use the data, it should then ensure that it adopts data governance best practices. While this AI governance component is particularly relevant to organizations developing an AI model or system, it is also important for deployers who are required by law to ensure that the input data is relevant and representative for their use case.³¹

When developing AI systems, organizations may need to rely on data scientists and other professionals to (1) prepare the data using annotation, labeling, cleaning, updating, enrichment, and aggregation, (2) formulate assumptions regarding the information the data is supposed to measure and represent, (3) assess the quantity, availability, and suitability of the datasets³² needed, (4) minimize or eliminate bias,³³ (5) seek to make the datasets representative of the environment,³⁴ of

²⁹See Complaint, Everalbum, Inc., Docket No. C-4743, FTC File No. 1923172 (May 6, 2021); *Interim Measures for Generative Artificial Intelligence Service Management*, CYBERSPACE ADMINISTRATION OF CHINA (Jul. 13, 2023), https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (stating generative AI service providers shall train AI models with data from legal sources and, if personal information is involved, obtain the individual's consent, before using it).

³⁰See W. Oremus, et al., *AI's Future Could Hinge on One Thorny Legal Question*, WASH. POST (Jan. 4, 2024), <https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/> (noting that the law is unclear in the US regarding whether there are IP violations based on using newspaper articles to train AI models); *Interim Measures for Generative Artificial Intelligence Service Management*, CYBERSPACE ADMINISTRATION OF CHINA (Jul. 13, 2023), https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (stating that the provision and use of generative AI services shall respect IP rights).

³¹See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 26(4), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (“[T]o the extent the deployer exercises control over the input data, that deployer shall ensure that input data is relevant and sufficiently representative in view of the intended purpose of the high-risk AI system.”).

³²See generally *id.* at Art. 10(2).

³³See Smart Dubai, *AI Ethics Principles & Guidelines 7*, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) (“Steps should be taken to mitigate and disclose the biases inherent in datasets”); Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM’N (Apr. 8, 2020), <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (“Focus on inputs, but also on outcomes. When we at the FTC evaluate an algorithm or other AI tool for illegal discrimination, we look at the inputs to the model – such as whether the model includes ethnically-based factors, or proxies for such factors, such as census tract. But, regardless of the inputs, we review the outcomes.”) (emphasis in original).

³⁴See Smart Dubai, *AI Ethics Principles & Guidelines 7*, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) (“Data ingested should, where possible, be representative of the affected population”).

good quality, complete,³⁵ free of error,³⁶ and low noise,³⁷ and (6) determine whether the data is appropriate for the purpose it will be used based on industry practice.³⁸ Developers should also use different datasets for training, testing, and validation.³⁹ The organization should train the AI model using training data, measure accuracy using test data, and validate the model using the validation data.⁴⁰

Both AI developers and deployers should understand the data lineage by tracking where the data came from, how it was collected, curated, and moved within the company, and how the data's accuracy is maintained over time.⁴¹ This involves looking at the data flows from its originating source, end use and backdating it to its source, and entire solution.⁴²

The organization should consider maintaining a data provenance record that documents (1) the date that the data was collected and last updated or modified, (2) the categories and types of data used, including for machine learning (e.g., training, validation, test, and production data), (3) the data quality (from origin to transformation), (4) the process for labeling data, (5) applicable data retention and disposal policies, (6) known or potential biases in the data, (7) the sources of error, (8) updates to the data, (9) intended use and purpose of the data; (10) how the data was prepared; (11) attributes data to their sources or owners; (12) the number of data points included in the datasets; (13) whether the datasets include IP-protected data; (14) whether the datasets were purchased or licensed; and (15) whether there was any cleaning, processing, or other modifications to the datasets.⁴³ AI developers should track the data lineage for training or fine-tuning the AI model, while deployers should monitor the input data and output for their use of the AI system.

In sum, data governance is an important process to ensure that high-quality data is used to develop an AI system and produce reliable output. Data governance is not a one-time project; the AI governance team should review the datasets periodically to ensure accuracy, quality, currency, relevance, and reliability and update them as necessary with new data⁴⁴ based on expert advice, including from data scientists. Developers should also be mindful that under applicable AI law,

³⁵See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, at 39, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en> (“Sample is complete, with minimal missing or partial values. Outliers must not affect the quality of data.”).

³⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 10(3), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689

³⁷See Org. for Econ. Co-operation and Dev. [OECD], *OECD Framework for the Classification of AI Systems*, at 39, Digital Economy Papers No. 323 (Feb. 2022), <https://doi.org/10.1787/cb6d9eca-en> (“Data is infrequently incorrect, corrupted or distorted (e.g. intentional or unintentional mistakes in survey data, data from defective sensors).”).

³⁸*Id.*

³⁹See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 40 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

⁴⁰*Id.*

⁴¹*Id.* at 37.

⁴²*Id.*

⁴³*Id.*; ISO/IEC 42001:2023(E), Annex B, §B.4.3 (INT’L ORG. FOR STANDARDIZATION 2023); CAL. CIV. CODE §3111(A).

⁴⁴See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 40 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

such as California’s Generative Artificial Intelligence: Training Data Transparency law, it may need to post a high-level summary of the training data used for its AI system on its website.⁴⁵

4.IV. LEGAL COMPLIANCE

AI is a hot topic among global policymakers. As mentioned in Stanford University’s Human-Centered AI Index Report for 2025, governments are stepping up on regulating AI.⁴⁶ According to the report, global “legislative mentions of AI rose 21.3% across 75 countries since 2023, marking a ninefold increase since 2016.”⁴⁷ In light of this dynamic regulatory environment, it is important for organizations to engage legal counsel to assess which laws and regulations apply to their development and/or use of AI systems.

While AI guidelines, frameworks, and principles are important sources for developing an AI governance plan, AI systems may trigger AI-specific laws (such as the EU AI Act, Colorado AI Law, and South Korea AI Act),⁴⁸ sector and industry-specific regulations,⁴⁹ and general laws, such as data privacy,⁵⁰ IP,⁵¹ antitrust,⁵² and employment.⁵³ Organizations should consider understanding

⁴⁵See CAL. CIV. CODE §3111(a).

⁴⁶See NESTOR MASLEJ ET AL., INST. FOR HUMAN-CENTERED AI, STANFORD UNIV., THE AI INDEX REPORT 2025 4 (Apr. 2025).

⁴⁷*Id.*

⁴⁸See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 9(1), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (comprehensive AI law); COLO. REV. STAT. §6-1-1701 et seq. (comprehensive AI law); U.C.A. 1953 §13-77-103 et seq. (AI transparency law); CAL. BUS. PROF. CODE §17941 (chatbot law); N.J. STAT. ANN. §56:18-2 (chatbot law); ME. REV. STAT. ANN. tit. 10, §1500-Y (chatbot law); CAL. CIV. CODE §3110 et seq. (AI training data transparency law); CAL. BUS. PROF. CODE §22757.1 et seq. (transparency regarding AI-generated content).

⁴⁹See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 17 (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf> (“[I]t should be noted that certain industry sectors (such as in the finance, healthcare, and legal sectors) may be regulated by existing sector-specific laws, regulations or guidelines relevant to the sector.”); DEP’T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 15 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf (“While AI is currently regulated through existing legal frameworks like financial services regulation, some AI risks arise across, or in the gaps between, existing regulatory remits.”); 3 Colo. Code Regs. §702-10:10-1-1 (AI law for life insurers).

⁵⁰See, e.g., VA. CODE ANN. §59.1-577(A)(5)(iii); COLO. REV. STAT. §6-1-1306(1)(a)(C); Regulation (EU) 2016/679, General Data Protection Regulation, Art. 22(1) O.J. (L 119, 04.05.2016), <https://gdpr-info.eu/art-4-gdpr/>.

⁵¹See Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (Mar. 16, 2023) (stating that copyright protection requires human authorship and providing guidance on when copyright may exist when there is both human-authored and AI generated material).

⁵²See Staff in the Bureau of Competition Off. of Tech., *Generative AI Raises Competition Concerns*, FED. TRADE COMM’N (June 29, 2023), <https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns> (“Generative AI may raise a variety of competition concerns. In particular, control over one or more of the key building blocks that generative AI relies on could affect competition in generative AI markets.”).

⁵³See EQUAL EMP. OPPORTUNITY COMM’N, OLC CONTROL NO. EEOC-NVTA-2023-2, SELECT ISSUES: ASSESSING ADVERSE IMPACT IN SOFTWARE, ALGORITHMS, AND ARTIFICIAL INTELLIGENCE USED IN EMPLOYMENT SELECTION PROCEDURES UNDER TITLE VII OF THE CIVIL RIGHTS ACT OF 1964 (May 18, 2023), <https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial>; 2021 N.Y.C. LOCAL LAW No. 144, N.Y.C. ADMIN. CODE. §20-870; Artificial Intelligence Video Interview Act, 820 ILCS 42/1 et seq.; ILLINOIS HUMAN RIGHTS ACT, 775 ILCS 5/2-101 et seq.

the legal landscape applicable to their practices and AI use cases to ensure legal compliance, particularly considering that enforcement authorities in major jurisdictions are motivated⁵⁴ to regulate AI under their general powers.

4.V. RISK MANAGEMENT

Organizations developing or using AI systems should define and establish an AI risk management system.⁵⁵ The risk management system helps the AI system achieve its objectives, with limited or no undesired effects, and continually improves to mitigate future risks.⁵⁶ The AI governance team needs to establish, implement, document, and maintain the risk management system throughout the AI lifecycle. Risk management requires monitoring and evaluating emerging risks⁵⁷ based on data collected after the AI system is deployed, because some of the risks may not become apparent until the AI is used in the real environment.⁵⁸ The AI governance team should ensure that any risk remaining in the AI system is acceptable⁵⁹ and communicated to the deployer and that the risk is

⁵⁴See Lina M. Khan et al., *Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems*, FED. TRADE COMM’N (Apr. 25, 2023), <https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/joint-statement-enforcement-efforts-against-discrimination-bias-automated-systems> (stating that US regulators will enforce improper use of AI under existing laws and regulatory authority); *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (stating that existing Canadian “consumer protection regulators are already moving to address some of the impacts of AI within their legislative authorities”).

⁵⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 9(1), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (requiring adopting a risk management system); COLO. REV. STAT. §6-1-1703(2)(a) (“[A] deployer of a high-risk artificial intelligence system shall implement a risk management policy and program to govern the deployer's deployment of the high-risk artificial intelligence system.”); Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.32, Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. *available at* https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf (stating that an AI business shall “[i]dentify, assess, and mitigate risks across the AI lifecycle” and “[e]stablish a risk management system to monitor and respond to AI-related safety incidents”); NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 4 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>. (“Addressing, documenting, and managing AI risks and potential negative impacts effectively can lead to more trustworthy AI systems.”); *Australia’s AI Ethics Principle*, AUSTL. GOV’T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024) (stating that problems identified with AI “should be addressed with ongoing risk management as appropriate”); *Asilomar AI Principles*, FUTURE OF LIFE INST. (Aug. 11, 2017), <https://futureoflife.org/open-letter/ai-principles/> (“Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.”); COLO. REV. STAT. §6-1-1703(2)(a).

⁵⁶See ISO/IEC 42001:2023(E), §6.1.1 (INT’L ORG. FOR STANDARDIZATION 2023).

⁵⁷See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 9(2)(c), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

⁵⁸See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 6 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“While measuring AI risks in a laboratory or a controlled environment may yield important insights pre-deployment, these measurements may differ from risks that emerge in operational, real-world settings.”).

⁵⁹*Id.* at 7 (“Risk tolerance and the level of risk that is acceptable to organizations or society are highly contextual and application and use-case specific.”).

proportionate to the organization's benefits and goals. Organizations may address this AI governance component by documenting and mapping to a major AI risk management framework, such as NIST AI RMF or ISO/IEC 42001.

Organizations should also use appropriate metrics and thresholds to test the AI systems before marketing or using them.⁶⁰ The organization should pay special attention to the possible impacts on vulnerable groups (e.g., elderly and children)⁶¹ and ensure that the risk management system is compatible with or integrated into any existing risk management procedures required by law.

Operationally, the organization should take the following steps for risk management: (1) identify and rank the risks applicable to the AI system, (2) assess the likelihood and severity of harm, and (3) document its risk assessment in an AI impact assessment, which contains a plan to mitigate risks.

4.V.A. Identify and Rank AI Risks

The initial step for a risk management system is to identify⁶² and rank the risks based on the AI system's domain and application, intended uses, and external and internal context.⁶³ The risks may include harms to individuals, groups, societies, organizations, and the environment.⁶⁴ The AI governance team should categorize the risks as unacceptable, high, limited, or minimal.⁶⁵ If the organization identifies the risk as unacceptable, it is prohibited from developing or using the AI system.⁶⁶ However, if the AI risks are high-to-minimal risk, the organization should implement mitigation measures that correspond to the level of risk, likelihood of harm, and severity of impact.

⁶⁰See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 9(8), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (“The testing of high-risk AI systems shall be performed, as appropriate, at any time throughout the development process, and, in any event, prior to their being placed on the market or put into service. Testing shall be carried out against prior defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system.”).

⁶¹See *id.*, Art. 9(9) (“When implementing the risk management system ... , providers shall give consideration to whether in view of its intended purpose the high-risk AI system is likely to have an adverse impact on persons under the age of 18 and, as appropriate, other vulnerable groups.”).

⁶²See Michael Atleson, *Keep Your AI Claims in Check*, FED. TRADE COMM’N: BUS. BLOG (Feb. 27, 2023), <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check> (“Are you aware of the risks? You need to know about the reasonably foreseeable risks and impact of your AI product before putting it on the market. If something goes wrong – maybe it fails or yields biased results – you can’t just blame a third-party developer of the technology. And you can’t say you’re not responsible because that technology is a ‘black box’ you can’t understand or didn’t know how to test.”).

⁶³See ISO/IEC 42001:2023(E), §6.1.1 (INT’L ORG. FOR STANDARDIZATION 2023).

⁶⁴See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 5 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>. (stating that there could be harm to people, organizations, or the ecosystem); ISO/IEC 42001:2023(E), Annex B, §§B.5.4 B.5.5 (INT’L ORG. FOR STANDARDIZATION 2023).

⁶⁵See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 1 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>. (“Like risks for other types of technology, AI risks can emerge in a variety of ways and can be characterized as long- or short-term, high or low-probability, systemic or localized, and high- or low-impact.”); SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 8 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>; (stating that “[t]he categories and levels of risks associated with the development and/or use of artificial intelligence are classified” as little or no risk, limited risk, high risk, or unacceptable risk).

⁶⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No

Practically speaking, it is not possible for organizations to eliminate negative risks entirely, and it may be counterproductive to do so.⁶⁷ It is important for organizations to rank risks so that they can prioritize mitigation resources on risks that present the greatest harm and regulatory impact.⁶⁸ Below are some guidelines to help an organization rank and prioritize risks.

4.V.A.1. Prohibited AI Risks

Applicable law may outright prohibit certain AI risks. Examples of prohibited AI risks include:⁶⁹

- Using AI for unfair or deceptive acts or practices;
- Use of an AI system that results in an unlawful differential treatment or impact that disfavors an individual or group of individuals based on protected classifications;
- Intentionally developing or deploying an AI system to incite or encourage a person to commit physical self-harm (including suicide), harm another person, or engage in criminal activity;
- Developing or deploying an AI system with the sole intent for the AI system to infringe, restrict, or otherwise impair an individual’s rights guaranteed under the United States Constitution;
- Developing or distributing an AI system in connection with certain sexually explicit content and child pornography;
- AI-based credit decisions that prevent creditors from accurately identifying the specific reasons for denying credit or taking other adverse actions;
- AI-based scoring systems used to screen rental applicants based on race;
- Social scoring for public and private purposes;
- Exploiting the vulnerabilities of persons and using subliminal techniques;
- Real-time remote biometric identification in publicly accessible spaces by law enforcement, subject to narrow exceptions;
- Biometric categorization of natural persons based on biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, or sexual orientation;
- Individual predictive policing;
- Emotion recognition in the workplace and education institutions, unless for medical or safety reasons (e.g., monitoring the tiredness levels of a pilot); and

168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 5, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 8 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>. (“In cases where an AI system presents unacceptable negative risk levels – such as where significant negative impacts are imminent, severe harms are actually occurring, or catastrophic risks are present – development and deployment should cease in a safe manner until risks can be sufficiently managed.”).

⁶⁷See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 7 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

⁶⁸*Id.*

⁶⁹See Arsen Kourinian, *Conducting an AI Risk Assessment*, BLOOMBERG LAW, <https://www.bloomberglaw.com/document/X3D03D2K000000> (last visited Aug. 14, 2024); see also COLO. REV. STAT. §6-1-1701(1); C.S.H.B. 149 2025 Gen. Assemb., Reg. Sess. (Tx. 2025); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 5, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

- Untargeted scraping of the internet or CCTV for facial images to build up or expand databases.

The AI governance team will need to stop developing or using the AI system if it falls within a prohibited category.

4.V.A.2. High AI Risks

If the AI use case is not prohibited, the AI governance team should next assess whether it is high risk.⁷⁰ Whether a risk is high may stem from specific legal authority that provides this designation, such as the EU AI Act, Colorado AI Law, or South Korea AI Act, be based on the context of use and development or be derived from applicable guidelines that indicate certain AI risks as presenting heightened concerns. Some examples of potential high-risk AI systems and processing activities are described below but, depending on applicable law, some of these high-risk AI systems may be prohibited, in which case the organization cannot develop or use the AI system.⁷¹ Further, some of the use cases below may fall outside the high-risk category under the EU AI Act and Colorado AI Law if they meet one of the following criteria:⁷²

- The AI system is intended to perform a narrow procedural task, such as an AI system that transforms unstructured data into structured data or categorizes or deduplicates documents (EU AI Act and Colorado AI Law);
- The AI system is intended to improve the results of a previously completed human activity (EU AI Act);
- The AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review (EU AI Act and Colorado AI Law);
- The AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the high-risk use cases (EU AI Act); or
- Certain simple technologies (e.g., anti-virus, calculators, spreadsheets, spell-checking, etc.), unless, when deployed, these technologies make, or are a substantial factor in making, a consequential decision.⁷³

⁷⁰See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 6, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (stating classification rules for high-risk AI systems); *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>.

⁷¹See Arsen Kourinian, *Conducting an AI Risk Assessment*, BLOOMBERG LAW, <https://www.bloomberglaw.com/document/X3D03D2K000000> (last visited Aug. 14, 2024) (citing legal sources for potential high AI risks); see also COLO. REV. STAT. §6-1-1701(3).

⁷²See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 6(3), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; COLO. REV. STAT. §6-1-1701(9)(b).

⁷³See COLO. REV. STAT. §6-1-1701(9)(b)(II) (“The following technologies, unless the technologies, when deployed, make, or are a substantial factor in making, a consequential decision: (A) Anti-fraud technology that does not use facial recognition technology; (B) Anti-malware; (C) Anti-virus; (D) Artificial intelligence-enabled video games; (E) Calculators; (F) Cybersecurity; (G) Databases; (H) Data storage; (I) Firewall; (J) Internet domain registration; (K) Internet website loading; (L) Networking; (M) Spam- and robocall-filtering; (N) Spell-checking; (O)

4.V.A.2.a. *EU AI Act High Risks*

- Critical infrastructure;
- Product safety component or certain regulated products;
- Biometric identification and surveillance;
- Education and vocational training;
- Employment and recruitment;
- Essential goods, services, and benefits;
- Law enforcement and administration of justice; and
- Immigration and border control.

4.V.A.2.b. *Colorado AI Law*

- Education enrollment or an education opportunity;
- Employment or an employment opportunity;
- A financial or lending service;
- An essential government service;
- Healthcare services;
- Housing;
- Insurance; and
- A legal service.

4.V.A.2.c. *South Korea AI Act*

- Supply of energy;
- Production processes of drinking water;
- Provision of healthcare;
- Development and use of medical devices;
- Nuclear materials;
- Analysis and use biometric information for criminal investigations or arrests;
- Judgments or evaluations that significantly affect the rights and obligations of individuals, such as hiring, loan screening, etc.;
- Operation of transportation;
- Government decisions that affect the public;
- Education; and
- Other areas designated by Presidential Decree as areas that have a significant impact on the protection of human life, physical safety, and fundamental rights.

4.V.A.2.d. *Other High-Risk Considerations*

- Consumer rights;
- Body scanners;
- Deepfakes;
- Sensitive personal data;
- Intrusion upon solitude, seclusion, or private affairs and other privacy violations;

Spreadsheets; (P) Web caching; (Q) Web hosting or any similar technology; or (R) Technology that communicates with consumers in natural language for the purpose of providing users with information, making referrals or recommendations, and answering questions and is subject to an accepted use policy that prohibits generating content that is discriminatory or harmful.”).

- Sale of personal data and data broker activities;
- Extensive profiling activities and behavioral advertising;
- Discrimination against population subgroups;
- Physical or psychological harm and safety;
- Civil liberties or rights and democratic participation;
- Harm to an organization’s business, reputation and security;
- Harm to the global financial system, supply chain, or interrelated systems;
- Harm to natural resources, the environment, and the planet; and
- Infringement of IP rights.

If the AI use case is high risk, the organization may need to implement the mitigation measures described below, as appropriate for the context, circumstances, and/or applicable law.

4.V.A.3. *Limited and Minimal AI Risks*

If the AI risk is neither prohibited nor high, then the risk may be limited or minimal. Limited AI risks include chatbots and AI-generated audio, image, video, or text content.⁷⁴ For limited-risk AI systems, developers and deployers may need to be transparent with individuals that they are interacting with an AI system or that content is AI-generated. All other risks, such as spam filters, are considered minimal risk. For these systems, organizations may voluntarily choose to apply mitigation measures.

4.V.B. *Likelihood and Severity of Harm*

In addition to risk ranking, organizations should assess the likelihood that the risks will materialize and the severity of their impact.⁷⁵ For this analysis, organizations may assign a risk score by using a risk matrix that multiplies the probability a risk will occur with the severity of harm. Below is an example of a simple three-by-three risk matrix.⁷⁶

⁷⁴See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 50, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; *Artificial Intelligence – Questions and Answers*, EUR. COMM’N (Aug. 1, 2024), https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683.

⁷⁵See ISO/IEC 42001:2023(E), §6.1.2 (INT’L ORG. FOR STANDARDIZATION 2023) (stating that an analysis of AI risks includes assessing “the realistic likelihood of the identified risks”); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 3(2), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (stating that “‘risk’ means the combination of the probability of an occurrence of harm and the severity of that harm”); NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 4 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“When considering the negative impact of a potential event, risk is a function of 1) the negative impact, or magnitude of harm, that would arise if the circumstance or event occurs and 2) the likelihood of occurrence. ...”).

⁷⁶Arsen Kourinian, *Conducting an AI Risk Assessment*, BLOOMBERG LAW, <https://www.bloomberglaw.com/document/X3D03D2K000000> (last visited Aug. 14, 2024).

Probability of Occurrence (Low to High)			
Severity of Harm (Low to High)	Low Risk / Low Likelihood (1)	Low Risk / Medium Likelihood (2)	Low Risk / High Likelihood (3)
	Medium Risk / Low Likelihood (2)	Medium Risk / Medium Likelihood (4)	Medium Risk / High Likelihood (6)
	High Risk / Low Likelihood (3)	High Risk / Medium Likelihood (6)	High Risk / High Likelihood (9)

Another sample comes from the New Zealand Government, which ranks the likelihood of impact from almost certain to rare, and impact severity from minimal to severe, with color-coded results.⁷⁷

Table 1 – Harm assessment matrix

		Impact				
		Minimal	Minor	Moderate	Significant	Severe
Likelihood	Almost Certain	Low	Medium	High	Severe	Critical
	Likely	Low	Medium	Medium	High	Critical
	Possible	Very Low	Medium	Medium	Medium	High
	Unlikely	Very Low	Low	Medium	Medium	Medium
	Rare	Very Low	Very Low	Very Low	Low	Low

Table 2 – Likelihood rating & criteria

Likelihood of occurrence	Criteria: probability in 12 month period
Almost Certain	> 80%
Likely	60 - 80%
Possible	40 - 60%
Unlikely	20 - 40%
Rare	< 20%

Table 3 – Impact rating & criteria

Impact of harm	Impact criteria (non-exhaustive & by way of illustration only)
Severe	Very significant or irreversible consequences: E.g. excluded from accessing services; significant breach of human rights; significant loss of rights or freedoms; significant negative impacts on specific groups; financial distress; significant loss of agency or autonomy; serious or long-term psychological or physical damage, death (including suicide); significant public health implications
Significant	Significant consequences and considerable damage: E.g. inability to access services; serious breach of human rights; loss of rights or freedoms; material negative impacts on specific groups; financial loss; loss of employment; loss of agency or autonomy; negative impacts on mental and/or physical health; public health implications
Moderate	Significant inconvenience: E.g. difficulty accessing services; breach of human rights; negative impacts on rights or freedoms; disadvantages to specific groups; extra costs; employment changes; anxiety, confusion, stress; minor physical ailments
Minor	Inconvenience: E.g. harder to access services; difficulty asserting rights or freedoms; significant inconvenience, frustration or irritation;
Minimal	Minor inconveniences or no impact

Organizations may also leverage existing processes and procedures for risk analysis, which may involve more complex risk matrices that account for other considerations, such as risk velocity, which reflects the time to impact, and risk contagion, which assesses the potential for risk in one area to impact the organization’s other areas.⁷⁸

⁷⁷New Zealand Gov’t, ALGORITHM IMPACT ASSESSMENT REPORT: ALGORITHM CHARTER FOR AOTEAROA NEW ZEALAND 9 (Dec. 2023).

⁷⁸See Brenda Boulwood, *How to Develop an Enterprise Risk-Rating Approach*, GLOBAL ASSOCIATION OF RISK PROFESSIONALS (Aug. 26, 2021), <https://www.garp.org/risk-intelligence/culture-governance/how-to-develop-an-enterprise-risk-rating-approach>.

Figure 7: Risk-Scoring Examples

Risk	Probability	Impact	Velocity	Contagion	Risk Score
A	5	6	0	0	30
B	5	6	0	.5	33
C	2	4	.75	0	11
D	2	4	0	.2	8.8
E	2	4	.75	.5	13

Developing these risk matrices may be difficult for organizations, because there is a lack of consensus on appropriate measurement methods.⁷⁹ This may lead to pitfalls in the risk measurement, such as oversimplified analysis, gamed outcomes, lack of critical nuance, unexpected reliance, and failure to account for differences in affected groups and context.⁸⁰ Organizations should consider having their risk matrices vetted by third-party auditors to ensure they are objective and adopt a methodology that recognizes context and harms to different groups, subgroups, and communities other than direct AI users.⁸¹

4.V.C. Document an AI Impact Assessment

The AI governance team should document the risk analysis above in an AI impact assessment.⁸² In an AI impact assessment, an organization should address data privacy, AI legal requirements, and other suggested topics in AI guidelines and frameworks.⁸³ The AI impact assessment should document the risks, benefits, steps taken to mitigate risks, and whether, on balance, the organization should proceed with the AI use case. Developers may need to provide deployers documentation and information that they will need to complete an AI impact assessment,⁸⁴ which is

⁷⁹See NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 6 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

⁸⁰*Id.*

⁸¹*Id.*

⁸²See Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.35, Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. available at https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf.

⁸³See, e.g., ISO/IEC 42001:2023, §6.1.4 Annex B, §B.5.3 (INT'L ORG. FOR STANDARDIZATION 2023); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 26(9) 27, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; 4 COLO. CODE REGS. §904-3-9.06 (listing the requirements under the Colorado Privacy Act that need to be included in a data protection impact assessment if profiling is used for automated decision-making); Regulation (EU) 2016/679, General Data Protection Regulation, Art. 35(3)(a) O.J. (L 119, 04.05.2016), <https://gdpr-info.eu/art-4-gdpr/>; PRIV. COMM'R FOR PERS. DATA, H.K. THE INFO. ACCOUNTABILITY FOUND., ETHICAL ACCOUNTABILITY FRAMEWORK FOR HONG KONG, CHINA 29 (Oct. 2018); U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 26, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (stating that ethical impact assessments should “identify and assess benefits, concerns and risks of AI systems, as well as appropriate risk prevention, mitigation and monitoring measures, among other assurance mechanisms”); COLO. REV. STAT. §6-1-1703(3).

⁸⁴See COLO. REV. STAT. §6-1-1702.

an issue the parties should negotiate in their commercial agreement. A nonexclusive list of issues to cover in an AI impact assessment include:

- A summary of the AI processing activity;
- Relevant internal and external parties contributing to the impact assessment;
- Any internal or external audits conducted in relation to the impact assessment;
- The dates the impact assessment was reviewed and approved and the names, positions, and signatures of the individuals responsible for the review and approval;
- The categories and sources of data processed by the AI;
- The applicable transparency notice, including a privacy notice provided to individuals if their personal data will be processed and the lawful basis for processing;
- Any contractual or licensing rights related to using data in the AI processing activity;
- An explanation of the training data and logic used to create the AI system, including if the AI system is sourced from another party;
- A description of the AI system's output and how it will be used;
- The period of time and frequency the AI system will be used;
- The AI processing activity context, including the relationship between the organization and the categories of individuals impacted;
- If any decisions will be made about individuals using the AI system;
- The level of human involvement in the AI processing activity;
- Whether the AI system has been evaluated for fairness and disparate impact and the results of such evaluation, including metrics used;
- The technology or processors used and details regarding their involvement;
- Names or categories of data recipients as part of the AI processing activity;
- The number of individuals impacted and the volume of data processed using AI and applicable retention periods;
- The purpose and benefits of the AI processing activity, including the steps taken to ensure that it is used for its intended purpose;
- The sources and nature of risks, foreseeable misuse, and predictable failures;
- The mitigation measures and safeguards the organization will employ to reduce or eliminate the risks, misuse, and potential failures, including post-deployment monitoring;
- The measures to take if the identified risks materialize, including the organization's arrangements for internal governance and complaint mechanisms; and
- Whether the AI benefits outweigh the risks.

In preparing the AI impact assessment, the AI governance team should be mindful that the organization may need to produce this assessment if requested by a regulator or as part of an investigation. As such, the AI governance team should carefully assess the impact assessment's accuracy and whether it can implement the mitigation measures justifying the organization proceeding with the AI use case. A sample AI impact assessment template is attached to this book for reference.

4.VI. MITIGATION MEASURES

Organizations should consider implementing mitigation measures to eliminate and/or reduce the risks identified during the risk management stage. As described below, there are a number of risk mitigation measures available under AI laws, guidelines, and frameworks that an organization may consider during this stage.

4.VI.A. Transparency and Explainability

Organizations should be transparent that they are using AI systems and explain how it works to foster trustworthiness with the public.⁸⁵ Transparency addresses the question “what happened” in the AI system, while explainability provides information regarding “how” the AI system made the decision.⁸⁶ The purpose of transparency and explainability is to allow individuals to know that they are interacting with an AI system, how it works, the steps taken to manage risks, the type of decisions it makes, the impact an AI system can have on them, their ability to contest decisions, and that they have the option to request human review for significant decisions.⁸⁷ A transparency and explainability notice can also provide the organization an opportunity to solicit external feedback regarding how its AI systems function in the real environment, detect model drift, and establish bug bounty programs to help improve the AI systems.⁸⁸

To address this requirement, the AI governance team should prepare and publicly post a transparency and explainability notice regarding its AI use case. The notice should be written in

⁸⁵See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 53-54, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; DUTCH DATA PROT. AUTH., DEP’T FOR THE COORDINATION OF ALGORITHMIC OVERSIGHT, ALGORITHMIC RISKS REPORT NETHERLANDS 6 (July 2023), https://autoriteitpersoonsgegevens.nl/uploads/2023-08/Algorithmic%20Risks%20Report%20Netherlands%20-%20July%202023_0.pdf (stating that “transparency and the complimentary understandable explanations can increase public trust and help improve the quality of these tools”); ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 11 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf (“In order to build public trust in AI, it is important to ensure that users are aware of the use of AI technology and understand how information from their interaction is used and how the AI system makes its decisions using the information provided.”).

⁸⁶See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 17 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

⁸⁷See *Artificial Intelligence – Questions and Answers*, EUR. COMM’N (Aug. 1, 2024), https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683; THE WHITE HOUSE OFF. OF SCI. AND TECH. POL’Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 40 (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf> (“Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible.”); *Canadian Guardrails for Generative AI – Code of Practice*, INNOVATION, SCI. AND ECON. DEV. CAN. (Aug. 16, 2023), <https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice> (stating that “[i]t is important to ensure that individuals realize when they are interacting with an AI system or with AI-generated content”); Smart Dubai, *AI Ethics Principles Guidelines* 7, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) (“People should be told when significant decisions about them are being made by AI” and “[d]ecisions and methodologies of AI systems which have a significant effect on individuals should be explainable to them, to the extent permitted by available technology”); Org. for Econ. Co-operation and Dev. [OECD], Recommendation of the Council on Artificial Intelligence, Legal Instrument 449 (July 11, 2023), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 44, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM’N (Apr. 8, 2020), <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (stating that consumers “must know what data is used in your model and how that data is used to arrive at a decision.”) (emphasis in original); INFO. COMM’R’S OFF. THE ALAN TURING INST., EXPLAINING DECISIONS MADE WITH AI 17 (Oct. 17, 2022), <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>.

⁸⁸See generally, NAT’L INST. OF STANDARDS AND TECH., AI RMF PLAYBOOK 40 (2023), https://airc.nist.gov/AI_RM_F_Knowledge_Base/Playbook.

plain language and easy to understand,⁸⁹ which organizations may validate using readability tools, such as the Fry readability graph, the Gunning Fog Index, and the Flesch-Kincaid readability test.⁹⁰ Organizations may also use visualization tools, graphical representations, and/or summary tables in their AI notices to enhance readability.⁹¹

In drafting a transparency and explainability notice, the AI governance team may consider incorporating the following elements:

- The name of the organization accountable for the AI system and its outcomes, and the type of AI systems it uses or makes available.⁹²
- A statement that individuals are interacting with AI, the nature and purpose of the AI, and what decisions are made using an AI system.⁹³
- The type of data (including personal and sensitive data) that were or will be processed as part of the AI decision, along with the data used to train the AI.⁹⁴
- An explanation of the logic used in the AI decision, including the key parameters that affect the AI system's output.⁹⁵
- The AI system's intended output (e.g., numerical score).
- An explanation of how the AI is used in the decision-making process, including the role of human involvement.⁹⁶
- How risks are managed and whether the AI system has been evaluated for accuracy, validity, reliability, fairness, or bias and the outcome of any such evaluation.⁹⁷
- The benefits and potential consequences of the decision made using AI.⁹⁸

⁸⁹See THE WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 40 (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>.

⁹⁰See PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 57, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

⁹¹*Id.*

⁹²See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 75 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; S.B. 2 §2(d), 2024 Leg., Feb. Sess., (Conn. 2024).

⁹³See *Artificial Intelligence – Questions and Answers*, EUR. COMM'N (Aug. 1, 2024), https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683; CAL. BUS. PROF. CODE §17941; DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 75 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; 4 Colo. Code Regs. §904-3-9.03; COLO. REV. STAT. §6-1-1703(4)(a); Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.31, Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. *available at* https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf.

⁹⁴See 4 Colo. Code Regs. §904-3-9.03; DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 75 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; COLO. REV. STAT. §6-1-1703(5)(a)(III).

⁹⁵See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 75 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; 4 Colo. Code Regs. §904-3-9.03.

⁹⁶See 4 Colo. Code Regs. §904-3-9.03.

⁹⁷See *id.*; COLO. REV. STAT. §6-1-1703(5)(a)(II).

⁹⁸See 4 Colo. Code Regs. §904-3-9.03.

- Information about how individuals may exercise rights in connection with the AI, such as access to further information about the AI system and opting-out of or appealing AI decisions.⁹⁹
- Contact methods if the public has any questions or feedback regarding the AI system.¹⁰⁰

In addition, deployers may address transparency and explainability by focusing on the AI system's quality and maintaining documents that will help build user confidence in the AI's outcome.¹⁰¹

This includes documents reflecting (1) that the AI system's results are repeatable, (2) the traceability of the decision-making process, (3) the data provenance record and a digitally centralized process log, and (4) AI model cards, which accompany AI systems and indicate their intended use, performance evaluations, and other information relevant to the AI system.¹⁰²

Further, AI system developers should be transparent with deployers by providing them instructions for use and information regarding the AI system's limitations. The instructions for use and other documentation should contain the following information:¹⁰³

- The developer's identity and contact details.
- The AI system's intended purpose, benefits, known limitations, and unacceptable uses.
- The AI system's level of accuracy, including its metrics, robustness, and cybersecurity, that can be expected and any known and foreseeable circumstances that may impact that level of accuracy, robustness, and cybersecurity.
- Any known or foreseeable situations that may lead to risks based on the AI system's intended purpose or under conditions of reasonably foreseeable misuse, including risk of algorithmic discrimination.
- Information to explain the AI system's output, such as its technical capabilities and characteristics.

⁹⁹See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 28 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf ("Parties directly affected by the use of an AI system should also be able to access sufficient information about AI systems to be able to enforce their rights."); 4 Colo. Code Regs. §904-3-9.03; PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 57, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; COLO. REV. STAT. §6-1-1703(4)(a)(b).

¹⁰⁰See PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 56–57, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; COLO. REV. STAT. §6-1-1703(4)(a)(II).

¹⁰¹See ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 12 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf; ISO/IEC 42001:2023, Annex B, B.7.2 (INT'L ORG. FOR STANDARDIZATION 2023) (noting that transparency and explainability include "data provenance and the ability to provide an explanation of how data are used for determining an AI system's output if the system requires transparency and explainability"); NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 16 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> ("Maintaining the provenance of training data and supporting attribution of the AI system's decisions to subsets of training data can assist with both transparency and accountability. Training data may also be subject to copyright and should follow applicable intellectual property rights laws.").

¹⁰²See ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 12 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf

¹⁰³See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 13(3), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; COLO. REV. STAT. §6-1-1702.

- How the AI system will perform with respect to specific persons or groups on which it is intended to be used.
- Information regarding the input, training, validation, and testing data, taking into account the AI system's intended purpose.
- Information to enable deployers to interpret the AI system's output and use it appropriately.
- Any updates or changes to the AI system and its performance from the initial instruction for use the developer provided to the deployer.
- The human oversight measures the developer has implemented, including the technical measures put in place to facilitate the deployer's interpretation of the AI system's output.
- The computational and hardware resources needed, the AI system's expected lifetime, and any necessary maintenance and case measures, including their frequency, to ensure the AI system's proper functioning and necessary software updates.
- A description of the mechanisms included in the AI system that would allow users to properly collect, store, and interpret the logs.
- How the AI system has been evaluated for performance and mitigation of risks, including the mitigation measures implemented.

Developers may also need to disclose on their website a high-level summary of the datasets used to develop an AI system, including:¹⁰⁴

- The sources or owners of the datasets;
- A description of how the datasets further the intended purpose of the AI system or service;
- The number of data points included in the datasets, which may be in general ranges, and with estimated figures for dynamic datasets;
- A description of the types of data points within the datasets (e.g., types of labels used or general characteristics);
- Whether the datasets include any data protected by copyright, trademark, or patent or whether the datasets are entirely in the public domain;
- Whether the developer purchased or licensed the datasets;
- Whether the datasets include personal information or aggregate consumer information;
- Whether the developer cleaned, processed, or modified the datasets and the intended purpose of those efforts in relation to the AI system or service;
- The time period during which the data in the datasets was collected, including a notice if the data collection is ongoing;
- The dates the datasets were first used during the development of the AI system or service; and
- Whether the generative AI system or service used or continuously uses synthetic data generation in its development.

Lastly, organizations developing or using AI systems that generate or manipulate image, audio, or video content should consider placing appropriate watermarks for individuals to know that the content has been artificially generated or manipulated.¹⁰⁵ This will help mitigate the risk of

¹⁰⁴See CAL. CIV. CODE §3111(a).

¹⁰⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU)

individuals being deceived by deepfakes.¹⁰⁶ In fact, certain developers of AI systems that have over 1 million monthly visitors or users may need to make available an AI detection tool, offer AI users the option to include a manifest disclosure that content is AI generated, and include a latent disclosure in AI-generated content.¹⁰⁷

4.VI.B. *Fair and Unbiased*

The AI governance team needs to ensure that AI systems are developed and used in a manner that is fair for all human beings and not biased against individuals or groups based on protected categories.¹⁰⁸

Fairness means that the organization takes necessary steps in the AI system's design, data, development, deployment, and use to eliminate bias, discrimination, or stigmatization of individuals, communities, or groups.¹⁰⁹ Bias may occur in an AI system because of data, representation, or algorithms and lead to discrimination.¹¹⁰

Organizations may need to identify and mitigate biases that can lead to different forms of discrimination, including (1) direct discrimination, whereby individuals are treated adversely due to their membership in a protected class, (2) indirect discrimination, whereby individuals in a protected class are disparately harmed or unfairly disadvantaged by the AI system, even though neutral policies, criteria, rules, and practices are applied, and (3) discriminatory harassment, whereby unwanted or abusive behavior linked to a protected characteristic violates someone's

2020/1828 (Artificial Intelligence Act), Art. 50, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹⁰⁶See *id.* at Recital 134.

¹⁰⁷See CAL. BUS. PROF. CODE §22757.1 et seq.

¹⁰⁸See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 29 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 21, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (“AI actors should promote social justice and safeguard fairness and non-discrimination of any kind in compliance with international law. This implies an inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all, taking into consideration the specific needs of different age groups, cultural systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized and vulnerable people or people in vulnerable situations.”); *Canadian Guardrails for Generative AI – Code of Practice*, INNOVATION, SCI. AND ECON. DEV. CAN. (Aug. 16, 2023), <https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>; Org. for Econ. Co-operation and Dev. [OECD], *Recommendation of the Council on Artificial Intelligence*, Legal Instrument 449, at 1.2, (July 11, 2023), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., *ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0)*, at 17-18 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>; *Australia's AI Ethics Principle*, AUSTL. GOV'T, DEP'T OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024); THE WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, *BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 23* (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>; *Interim Measures for Generative Artificial Intelligence Service Management*, CYBERSPACE ADMIN. OF CHINA Art. 4(2) (Jul. 13, 2023), https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm; Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020), <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (“Don't discriminate based on protected classes. Cavalier use of AI could result in discrimination against a protected class.”) (emphasis in original).

¹⁰⁹See SAUDI DATA AI AUTH., *AI ETHICS PRINCIPLES 12–14* (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

¹¹⁰*Id.* at 12.

dignity, degrades their identity, or creates an offensive environment for them.¹¹¹ Examples of discrimination in connection with AI systems include:

- AI chatbots that learn hate speech and biases when trained on data from the internet. When an organization places the AI chatbot in the market, it may generate hateful or discriminatory content.¹¹²
- An AI system trained on historic human resources data for scoring applicants may incorrectly infer that, because companies have previously hired white male applicants, this is an appropriate factor to consider to assign a higher score. As a result, the AI system may assign a lower score and filter out qualified female and minority candidates.¹¹³
- A medical diagnosis AI system that is available on a cell phone application could predominantly be used by younger, digitally literate, and affluent groups. This may discriminate against elderly, less digitally literate, and economically disadvantaged groups, who may not have access to the application.¹¹⁴

The NIST also identifies three categories of bias an organization should mitigate against: (1) systemic bias, which is present in AI datasets, the organization, and society, (2) computational and statistical biases, which are present in AI datasets and algorithmic processes and stem from failing to use representative data, and (3) human cognitive bias, which relates to how an individual or group perceives AI systems.¹¹⁵ These categories of bias can occur even if there is no prejudice, partiality, or discriminatory intent.¹¹⁶

Fairness and bias are difficult to define, because cultures and groups may perceive these issues differently and standards may shift depending on application.¹¹⁷ However, the organization's risk mitigation measures will be more effective if they account for these cultural and group differences.¹¹⁸ Some steps an organization may consider taking to mitigate unfair and biased impacts are as follows:¹¹⁹

- Ensure that the AI system is trained, tested, and validated using representative data that is fit for the purpose, relevant, accurate, and applicable to broader groups;
- Ensure that the AI system's policies and objectives are non-discriminatory, acceptable, and within the expectations of impacted individuals;

¹¹¹See DAVID LESLIE, ET AL., THE ALAN TURING INST., AI FAIRNESS IN PRACTICE 13–14 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-fairness_1.pdf.

¹¹²*Id.* at 15.

¹¹³*Id.*

¹¹⁴See *id.* at 16.

¹¹⁵See NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 18 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹¹⁶*Id.*

¹¹⁷*Id.*

¹¹⁸*Id.* at 17.

¹¹⁹See DAVID LESLIE, ET AL., THE ALAN TURING INST., AI FAIRNESS IN PRACTICE 23 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-fairness_1.pdf; THE WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 23 (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>; *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (“Appropriate actions must be taken to mitigate discriminatory outcomes for individuals and groups.”).

- Ensure that the AI system’s model architecture does not include discriminatory target variables, features, processes, or analytical structures and that it avoids encoded social and historical patterns of discrimination;
- Develop clear metrics for determining whether the AI system is fair and make them available to relevant stakeholders and impacted individuals;
- Ensure that AI system users are sufficiently trained on the limitations and strengths of the AI system and are aware of potential biases to affected individuals;
- Be aware of the economic, legal, cultural, and political structures or institutions in which the AI system operates and ensure that they do not impact AI research in ways that amplify asymmetrical and discriminatory power dynamics or generate inequitable outcomes for protected groups;
- Ensure that the AI systems are accessible for disadvantaged groups;
- Conduct disparity testing and migration; and
- Implement active oversight of the AI system.

The organization should monitor the biases and potential discriminatory outcomes throughout the AI system’s development and operation and correct issues as they come up. The organization should also prepare and make publicly available a fairness position statement in plain and nontechnical language, explaining the metrics used to ensure fairness.¹²⁰ This fairness position statement will help the organization explain the rationale and logic behind the AI system’s output by sharing the reasons and underlying fairness values used in the AI’s decision-making process.¹²¹

4.VI.C. Human-Centered and Beneficial for the Environment and Society

Organizations should ensure that their AI systems are beneficial to individuals, society, and the environment throughout the AI’s lifecycle, including during design, development, and deployment.¹²² The human-centered principle focuses on building AI systems that are just and ethical and incorporate human rights and cultural values to generate a short- and long-term beneficial impact.¹²³

To address this issue,¹²⁴ organizations (1) can test their AI systems with small groups of internal users from different backgrounds and demographics and incorporate their feedback in the AI

¹²⁰See DAVID LESLIE, ET AL., THE ALAN TURING INST., AI FAIRNESS IN PRACTICE 54 (2023), https://www.turing.ac.uk/sites/default/files/2023-12/aieg-ati-fairness_1.pdf.

¹²¹See SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 34 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

¹²²See *Australia’s AI Ethics Principle*, AUSTL. GOV’T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024) (stating that “AI systems should be used for beneficial outcomes for individuals, society and the environment”); Org. for Econ. Co-operation and Dev. [OECD], Recommendation of the Council on Artificial Intelligence, Legal Instrument 449, at 1.1, 1.2, (July 11, 2023), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 18, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (“No human being or human community should be harmed or subordinated, whether physically, economically, socially, politically, culturally or mentally during any phase of the life cycle of AI systems.”); ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 14 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

¹²³See SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 17 (Sept. 2023), .

¹²⁴See ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 14 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf; SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 17–19 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

systems, (2) should not use AI systems for malicious purposes, use dark patterns, or attempt to sway or deceive users into making decisions that are not beneficial for them or society, (3) should ensure that adopting AI systems does not significantly disrupt labor and job prospects until a proper assessment is conducted to determine if the AI systems can replace workers, (4) should periodically assess their AI systems after they have been deployed to ensure that the results align with human rights and cultural values, (5) should ensure that appropriate resources and energy levels are consumed in connection with their AI systems, and (6) should communicate their AI systems' benefits in their ESG messaging to reduce the public's concern¹²⁵ regarding the proliferation of AI in their daily lives.

In short, organizations should take steps to ensure that their AI systems are beneficial to humans, society, and the environment, document the steps they have taken, and properly message this information to the public.

4.VI.D. Accuracy

The AI governance team should test and document the AI system's accuracy by benchmarking how close initial observations, computations, or estimates are to true values.¹²⁶ The organization should also document the relevant metrics that it used to evaluate the accuracy level.¹²⁷

The AI governance team can measure accuracy by assessing whether the AI system is underfit or overfit by comparing how it performs on training data to test and holdout data. This will help the organization determine if the AI system functions as intended in the real environment.

Underfitting occurs when an AI model is unable to accurately draw a relationship between the input and output variables, which leads to a high error rate and poor performance.¹²⁸ If there is high bias and low variance, the AI model is likely underfit.¹²⁹ To correct underfitting, the organization should use more input features and training time so that the AI model is capable of generalization, which will allow it to make predictions and classify data.¹³⁰ However, this is a balancing act, because, if the AI model is overtrained, it may also lead to high error rates because of overfitting.¹³¹

¹²⁵See LEE RAINE, ET AL., PEW RESEARCH CTR., HOW AMERICANS THINK ABOUT ARTIFICIAL INTELLIGENCE, (Mar. 17, 2022), <https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence/> (“In broad strokes, a larger share of Americans say they are ‘more concerned than excited’ by the increased use of AI in daily life than say the opposite. Nearly half of US adults (45%) say they are equally concerned and excited.”); Sabrina Ortiz, *Most Americans Want Federal Regulation of AI, Poll Shows*, ZDNET (Aug. 14, 2023), <https://www.zdnet.com/article/most-americans-want-federal-regulation-of-ai-poll-shows/> (“Of the participants polled, 62% reported being somewhat or mostly ‘concerned’ about AI, with 86% believing AI could accidentally cause a catastrophic event.”).

¹²⁶See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 14 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>; Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM’N (Apr. 8, 2020), <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (stating that, to avoid consumer harm, algorithm operators should ask “[h]ow accurate are your predictions based on big data”).

¹²⁷See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 15, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹²⁸See *What is Underfitting?*, IBM, <https://www.ibm.com/topics/underfitting> (last visited Aug. 14, 2024).

¹²⁹*Id.*

¹³⁰*Id.*

¹³¹*Id.*

An overfit model is fit too closely to its training data and is unable to generalize.¹³² To remedy overfitting, an organization can diversify and scale its training data and use data science techniques, such as early stopping (stopping the model in time to avoid learning the noise in data), pruning (eliminating irrelevant features in the training set and focusing on the important ones), regularization (eliminating the factors that do not impact the prediction outcomes), ensembling (using multiple models), and data augmentation (periodically changing the input data in small ways).¹³³ An organization should work with data scientists to ensure that the AI model is neither under nor overfit.

The AI governance team can also measure accuracy using traditional metrics and methodologies, such as by calculating the AI system's overall accuracy, precision, recall, and/or F1 score, as described below:¹³⁴

- **Overall accuracy.** Overall accuracy is measured by calculating the AI system's number of correct predictions divided by the total number of predictions.
- **Precision.** Another method to calculate accuracy is by evaluating the precision, which assesses the portion of positive indications that was actually correct. The formula for precision is true positive divided by the total of true positive and false positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

- **Recall.** Recall measures what portion of actual positives was correctly identified. The formula for recall is true positive divided by the total of true positive and false negative.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- **F1 Score.** The F1 score accounts for both precision and recall in a single metric. The formula for the F1 score is $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$. F1 score helps balance these two metrics.

There are trade-offs between precision and recall.¹³⁵ For example, if you place more importance on recall to find as many positive cases as possible, this could result in some false positives, which lowers precision.¹³⁶ Notably, AI systems cannot achieve 100% in both recall and precision, which means that they can make wrong predictions at times.¹³⁷

¹³²*Id.*

¹³³See *What is Overfitting?*, AWS, <https://aws.amazon.com/what-is/overfitting/> (last visited Aug. 14, 2024).

¹³⁴See Koo Ping Shung, *Accuracy, Precision, Recall or F1?*, Medium: Towards Data Sci. (Mar. 15, 2018), <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>.

¹³⁵See *What Do We Need To Know About Accuracy And Statistical Accuracy?*, INFO. COMM'R'S OFF. (Mar. 15, 2023), <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/what-do-we-need-to-know-about-accuracy-and-statistical-accuracy>.

¹³⁶*Id.*

¹³⁷See Eur. Union Agency for Cybersecurity, *Cybersecurity of AI and Standardisation*, at 10 (Mar. 14, 2023), <https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation>.

Finally, the organization should test the AI system's accuracy continuously and after it has been placed in the market.¹³⁸ The organization may conduct the initial accuracy testing on static test data that is held back from the training data. However, the AI system's accuracy level may change when it is applied in the real environment, which has new and changing populations.¹³⁹ This phenomenon is called concept or model drift.¹⁴⁰ Organizations can mitigate this risk by measuring the distance between classification errors over time.¹⁴¹ If the measurements show that there is an increasing error frequency, this may suggest model drift.¹⁴²

4.VI.E. Robustness

The organization should ensure that the AI system is robust, which reflects its capability to perform comparably on new data as it did on its training data or the data used in typical operations.¹⁴³ Robustness requires the system to perform as intended under expected uses and to minimize potential harms to people if it is operating in unexpected settings.¹⁴⁴

To ensure robustness, the AI governance team can take a number of steps, including (1) through technical redundancy solutions, which may include backup or fail-safe plans, (2) conducting rigorous testing before deployment to ensure consistent results across different situations and environments, (3) maintaining proper documentation of data sources, tracking data processing steps and data lineage, which help troubleshoot the AI system if issues arise, (4) adopting continuous learning practices, where the AI system's learned parameters are not fixed and can continue to change when it is deployed in the real world, (5) conducting continuous monitoring to ensure that the AI system does not learn unintended behavior in the process, and (6) putting in place back-up systems, protocols, or procedures in case the AI system produces unacceptable or inaccurate results or fails to function.¹⁴⁵

¹³⁸See *What Do We Need To Know About Accuracy And Statistical Accuracy?*, INFO. COMM'R'S OFF. (Mar. 15, 2023), <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/what-do-we-need-to-know-about-accuracy-and-statistical-accuracy>.

¹³⁹*Id.*

¹⁴⁰*Id.*

¹⁴¹*Id.*

¹⁴²*Id.*

¹⁴³See ISO/IEC 42001:2023(E), Annex C, §C.2.8 (INT'L ORG. FOR STANDARDIZATION 2023).

¹⁴⁴See NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 14 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹⁴⁵See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 15(4), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 16, 37, 69 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

4.VI.F. *Safe and Secure*

The AI governance team should ensure that the AI system is safe and secure throughout its lifecycle¹⁴⁶ by conducting continuous monitoring and managing risks.¹⁴⁷ For AI safety, the AI governance team should implement processes to prevent danger to human life, health, property, and the environment.¹⁴⁸ For example, an autonomous vehicle can pose a risk to people’s lives if it does not recognize a pedestrian on the road or if it malfunctions.¹⁴⁹ To help ensure that the AI system is safe, the governance team can (1) employ responsible design, development, and deployment practices, (2) provide clear instructions to deployers regarding how to use the AI system, (3) if it is a deployer, make responsible decisions using AI systems, and (4) document and explain AI risks based on empirical evidence of incidents.¹⁵⁰ The AI governance team can also consider adopting safety guidelines,¹⁵¹ as relevant to the organization’s industry or sector.

The organization’s risk-mitigation measures should be tailored to the AI system’s safety risk, with risk of serious injury or death requiring the most urgent priority and risk management.¹⁵² Critically, the organization should have a process in place to intervene and safely turn off an AI system that poses a high safety risk.¹⁵³

Security refers to the organization’s ability to protect the AI system from malicious attacks.¹⁵⁴ These attacks can include data poisoning, model inversion, exfiltration of models, training data or IP, tampering of datasets, byzantine attacks in federated learning, and other reverse engineering

¹⁴⁶See *Asilomar AI Principles*, FUTURE OF LIFE INST. (Aug. 11, 2017), <https://futureoflife.org/open-letter/ai-principles/> (“AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.”); U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 20, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137>; *FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments From Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, THE WHITE HOUSE, (July 21, 2023) <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/> (stating that “[c]ompanies have a duty to make sure their products are safe before introducing them to the public ... build systems that put security first”).

¹⁴⁷See DEP’T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 27 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf.

¹⁴⁸See Nat’l Inst. Of Standards And Tech., U.S. Dep’t of Com., Artificial Intelligence Risk Management Framework (AI RMF 1.0), At 14 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> *Canadian Guardrails for Generative AI – Code of Practice*, INNOVATION, SCI. AND ECON. DEV. CAN. (Aug. 16, 2023), <https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>.

¹⁴⁹See SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 21 (Sept. 2023), <https://sdaia.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

¹⁵⁰See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 14 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹⁵¹*Id.* at 15.

¹⁵²*Id.* at 14.

¹⁵³See ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 13 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

¹⁵⁴See *id.* at 14; NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 15 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“While resilience is the ability to return to normal function after an unexpected adverse event, security includes resilience but also encompasses protocols to avoid, protect against, respond to, or recover from attacks.”).

attacks.¹⁵⁵ Related to security is resilience, which reflects the AI system’s ability to withstand such adverse events and return to normal function.¹⁵⁶

Organizations may apply the traditional cybersecurity paradigm of confidentiality, integrity, and availability (CIA) to safeguard AI systems.¹⁵⁷ Attackers may try to compromise the “confidentiality” of the AI system by using carefully crafted inputs and observing the outputs.¹⁵⁸ By making the model talk, bad actors may obtain information about the AI model or its training data.¹⁵⁹ The AI model’s parameter values may also leak because of human error or inadequate security.¹⁶⁰ The AI system’s “integrity” may also be compromised when attackers work on the AI model’s inputs to find small perturbations, which lead to modified outputs.¹⁶¹ In addition, bad actors may modify the AI model’s behavior by altering the data or models.¹⁶² An AI system’s “availability” can be jeopardized in a denial of service attack if an attacker uses an inappropriate input data format to increase the model’s computation time.¹⁶³

To address these security threats, organizations can apply traditional cybersecurity techniques to AI systems, like they do with other IT assets.¹⁶⁴ The organization’s security team can monitor and regularly check log files to detect anomalies in the AI system, apply technical protection measures, such as firewalls, encryption, multi-factor authentication and security patches, limit access to the AI system during development and inference time, require the AI system’s core developers to undergo background checks, protect proprietary data sources used to train or fine-tune the AI model, apply security-by-design and zero-trust frameworks, monitor abnormal behavior and potential security threats, cryptographically protect important information for the AI system’s entire lifecycle, and conduct audits and penetration testing.¹⁶⁵ Cybersecurity defenses should be

¹⁵⁵See ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 13 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf; NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 15 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹⁵⁶See NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 15 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“AI systems, as well as the ecosystems in which they are deployed, may be said to be *resilient* if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal. ...”) (emphasis in original).

¹⁵⁷See Eur. Union Agency for Cybersecurity, *Cybersecurity of AI and Standardisation*, at 16 (Mar. 14, 2023), <https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation>; NAT’L INST. OF STANDARDS AND TECH., U.S. DEP’T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 15 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“AI systems that can maintain confidentiality, integrity, and availability through protection mechanisms that prevent unauthorized access and use may be said to be *secure*.”) (emphasis in original).

¹⁵⁸See Eur. Union Agency for Cybersecurity, *Cybersecurity of AI and Standardisation*, at 16 (Mar. 14, 2023), <https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation>.

¹⁵⁹See *id.*

¹⁶⁰See *id.*

¹⁶¹See *id.*

¹⁶²See *id.*

¹⁶³See *id.*

¹⁶⁴See *id.* at 17; FED. OFF. FOR INFO. SEC., AI SECURITY CONCERNS IN A NUTSHELL – PRACTICAL AI SECURITY GUIDE 5 (Sept. 3, 2023) (Ger.), https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Practical_AI-Security_Guide_2023.pdf? (“AI systems are IT systems, meaning classical measures can be applied to increase IT security. Moreover, AI systems, in practice, do not operate in isolation but are embedded in a more extensive IT system consisting of various components.”).

¹⁶⁵See Fed. Off. for Info. Sec., AI SECURITY CONCERNS IN A NUTSHELL – PRACTICAL AI SECURITY GUIDE 5 (Sept. 3, 2023) (Ger.), https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Practical_AI-Security_Guide_2023.pdf?; U.S. NAT’L SEC. AGENCY, A.I. SEC. CTR. ET AL., DEPLOYING AI SYSTEMS SECURELY:

diverse and comprehensive because there are many attack vectors in connection with AI systems.¹⁶⁶

The AI governance team should also ensure that its existing cybersecurity policies and procedures incorporate the organization's AI use cases.¹⁶⁷ For example, the organization can update its incident response¹⁶⁸ plan to address third-party bad actors¹⁶⁹ potentially misusing¹⁷⁰ the AI system and exploiting vulnerabilities.¹⁷¹ The AI governance team can also regularly test and conduct diligence on its AI system throughout its lifecycle¹⁷² by conducting red-teaming of models or systems in high-risk areas and updating technical standards for safety and security.¹⁷³ Further, the organization should include AI use cases in their business continuity and disaster recovery plans.¹⁷⁴

BEST PRACTICES FOR DEPLOYING SECURE AND RESILIENT AI SYSTEMS 4–8 (Apr. 2024),

<https://media.defense.gov/2024/Apr/15/2003439257/-1/-1/0/CSI-DEPLOYING-AI-SYSTEMS-SECURELY.PDF>.

¹⁶⁶See U.S. NAT'L SEC. AGENCY, A.I. SEC. CTR. ET AL., DEPLOYING AI SYSTEMS SECURELY: BEST PRACTICES FOR DEPLOYING SECURE AND RESILIENT AI SYSTEMS 2 (Apr. 2024),

<https://media.defense.gov/2024/Apr/15/2003439257/-1/-1/0/CSI-DEPLOYING-AI-SYSTEMS-SECURELY.PDF>.

¹⁶⁷*Id.* at 1 (“As agencies, industry, and academia discover potential weaknesses in AI technology and techniques to exploit them, organizations will need to update their AI systems to address the changing risks, in addition to applying traditional IT best practices to AI systems.”).

¹⁶⁸See ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 13 (2024),

https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf (“Deployers should also develop incident response plans to safeguard AI systems from ... attacks.”).

¹⁶⁹See Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020),

<https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (“If you're in the business of developing AI to sell to other businesses, think about how these tools could be abused and whether access controls and other technologies can prevent the abuse.”).

¹⁷⁰See *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>.

¹⁷¹See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 15(5), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 15 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (“Common security concerns relate to adversarial examples, data poisoning, and the exfiltration of models, training data, or other intellectual property through AI system endpoints.”); *Australia's AI Ethics Principle*, AUSTL. GOV'T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024) (stating that AI system security measures include identifying “potential security vulnerabilities, and assurance of resilience to adversarial attacks”).

¹⁷²See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 27 (U.K.),

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf (“AI systems should function in a robust, secure and safe way throughout the AI life cycle, and risks should be continually identified, assessed and managed.”).

¹⁷³See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 27 (U.K.),

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf (“System developers should be aware of the specific security threats that could apply at different stages of the AI life cycle and embed resilience to these threats into their systems. Other actors should remain vigilant of security issues when they interact with an AI system.”).

¹⁷⁴See ASS'N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 14 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

While adopting existing cybersecurity techniques are helpful to address an AI system's security, it is worth noting that proper AI governance is not just about protecting the AI asset.¹⁷⁵ Instead, it also involves correctly implementing trustworthiness in cybersecurity¹⁷⁶ through other mitigation measures described above, such as data governance, risk assessment, transparency, etc.

4.VI.G. *Enhancing Privacy Protection*

There are laws of general applicability, such as data privacy laws, that apply to AI systems. Some of the core requirements to address data privacy compliance when developing and/or using AI systems include (1) being transparent that the organization is using personal data to train or make decisions with AI systems, (2) documenting a lawful basis for the AI system processing personal data (e.g., legitimate interest or obtaining consent, based on applicable privacy laws and practices), (3) employing privacy-by-design techniques that, among other things, allow the organization to honor data subject rights requests for personal data used in AI systems (e.g., right to delete, access, correct, and opt-out/objection to processing), (4) including appropriate data privacy terms in agreements with AI service providers, and (5) employing privacy enhancing techniques (PETs) to minimize the use of personal data in AI systems, such as anonymization, de-identification, aggregation, pseudonymization, masking, encryption, and tokenization.¹⁷⁷

The AI governance team's privacy specialists should be involved in this process to ensure that the above-mentioned practices comport with privacy laws and apply appropriate PET techniques to help prevent the AI system revealing personal data as output or during a malicious attack.

4.VI.H. *Human Oversight*

When organizations develop or deploy AI systems, it is critical that automation does not completely override human agency, particularly in sensitive domains like critical infrastructure.¹⁷⁸ The AI governance team should ensure that there is appropriate human involvement¹⁷⁹ when

¹⁷⁵See Eur. Union Agency for Cybersecurity, *Cybersecurity of AI and Standardisation*, at 19 (Mar. 14, 2023), <https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation>.

¹⁷⁶See *id.* at 19 24..

¹⁷⁷See THE WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 6 (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>; NAT'L INST. OF STANDARDS AND TECH., U.S. DEP'T OF COM., ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0), at 17 (Jan. 2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>; SAUDI DATA AI AUTH., AI ETHICS PRINCIPLES 39 (Sept. 2023), <https://sdaja.gov.sa/en/SDAIA/about/Documents/ai-principles.pdf>.

¹⁷⁸See SANJEEV SANYAL, ET. AL., ECON. ADVISORY COUNCIL TO THE PM, A COMPLEX ADAPTIVE SYSTEM FRAMEWORK TO REGULATE ARTIFICIAL INTELLIGENCE 22 (Jan. 2024) (India), https://eacpm.gov.in/wp-content/uploads/2024/01/EACPM_AI_WP-1.pdf.

¹⁷⁹See PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 30–31, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art.14, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (“[H]igh-impact AI systems must be designed and developed in such a way as to enable people managing the operations of the system to exercise meaningful oversight. This includes a level of interpretability appropriate to the context.”); *Asilomar AI Principles*, FUTURE OF LIFE INST. (Aug. 11, 2017), <https://futureoflife.org/open-letter/ai-principles/> (“Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.”); Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.34.4,

developing or using AI that is commensurate with the risk level and the severity and probability of harm.¹⁸⁰

Generally speaking, there are three levels of human involvement to consider: (1) human-out-of-the-loop, whereby the AI system operates with no human oversight, (2) human-over-the-loop, whereby a human can control and override unexpected and undesirable events while monitoring or supervising AI decisions, and (3) human-in-the-loop, whereby a human is in full control over the AI system and treats its output as a recommendation.¹⁸¹ If the risk score identified during the risk assessment is high (e.g., in the medical context),¹⁸² the AI governance team should utilize human-in-the-loop to mitigate risks.¹⁸³ On the other hand, if the risk score is low (e.g., spam filters and product recommendations on websites), human-out-of-the-loop may be adequate to a certain degree.¹⁸⁴ In some situations, a moderate risk score could justify using human-over-the-loop (e.g., GPS recommending driving routes).¹⁸⁵

Lastly, the AI governance team may consider other variables as part of proper human oversight, such as (1) incorporating a kill switch¹⁸⁶ if the AI system goes awry and poses a danger, (2) ensuring that the organization understands that AI systems are not always accurate by avoiding

Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. available at https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf (stating that an operator of a high-impact AI system shall implement “[h]uman management and oversight of high-impact AI”).

¹⁸⁰See generally, PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 30–31, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (stating that human oversight requires “considering the scale of deployment, the manner in which the system is being made available for use, and its user base.”).

¹⁸¹See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 30, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>; ASS’N OF SE. ASIAN NATIONS, ASEAN GUIDE ON AI GOVERNANCE AND ETHICS 25–26 (2024), https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf.

¹⁸²See U.N. Educ., Sci. and Cultural Org. [UNESCO], *Recommendation on the Ethics of Artificial Intelligence*, 22, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (“As a rule, life and death decisions should not be ceded to AI systems.”).

¹⁸³See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 30, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

¹⁸⁴*Id.*

¹⁸⁵*Id.*

¹⁸⁶See PERSONAL DATA PROT. COMM’N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 31, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf> (“For safety-critical systems, it would be prudent for organisations to ensure that a person be allowed to assume control, with the AI system providing sufficient information for that person to make meaningful decisions or to safely shut down the system where human control is not possible.”); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 14(4)(e), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689 (stating that humans should be able “to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.”).

automation bias¹⁸⁷ (e.g., an attorney relying on a brief written by an AI system without checking the case law and analysis), (3) understanding the AI system’s capacities and limitations, (4) receiving training on how to properly interpret the AI system’s output, and (5) knowing when to override, ignore, or reverse the AI system’s decisions, which may be due to inaccuracy, hallucination, or drift.¹⁸⁸

4.VI.I. *Technical Documentation and Logs*

The organization may need to maintain technical documentation and logs throughout the AI system’s lifecycle to demonstrate accountability and trace the AI system’s sources of error if it drifts or hallucinates.

The technical documentation should include, among other things described under the EU AI Act, a description of (1) the AI system, (2) the AI system’s elements and development process, (3) the AI system’s monitoring, functioning, and control, particularly with regard to its capabilities and limitations in performance (including accuracy), (4) the appropriateness of the AI system’s performance metrics, (5) the risk management system the organization adopted, (6) any relevant changes made to the AI system through its lifecycle, and (7) the process in place to evaluate the AI system’s performance when deployed in the market.¹⁸⁹

The AI governance team should also ensure that there is a process to automatically record events (i.e., logs) during the AI system’s lifetime, such as (1) the period of each AI use (start and end date and time of each use), (2) the reference database against which the AI system has checked the input data, (3) the input data that led to a match for a search, and (4) the individuals who verified the AI system’s results.¹⁹⁰

These documents will allow the AI governance team to identify and mitigate risks that may arise when the AI system is operating in the real environment by ensuring that there is a sufficient level of traceability. The documents will also verify the measures taken to ensure the AI system’s safety and reliability.¹⁹¹

¹⁸⁷See Integrated Innovation Strategy Promotion Council, *Social Principles of Human-Centric AI* 4, CABINET SECRETARIAT OF JAPAN, <https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf> (last visited Aug. 14, 2024) (noting that society should not be “overly dependent on AI” or let AI “control human behavior”).

¹⁸⁸See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 14(4)(d) (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; Smart Dubai, *AI Ethics Principles Guidelines* 9, DIGITAL DUBAI, <https://www.digitaldubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf> (last visited Aug. 14, 2024) (stating that AI “systems should be able to be overridden or their decisions reversed by designated people”).

¹⁸⁹See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 11, Annex IV (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

¹⁹⁰See *id.* at Art. 12.

¹⁹¹See Basic Act on the Development of Artificial Intelligence and Establishment of Trust, Art.34.5, Enforced January 22, 2026, Law No. 20676, Enacted January 21, 2025, translated by Etcetera Language Group, Inc. available at https://cset.georgetown.edu/wp-content/uploads/t0625_south_korea_ai_law_EN.pdf.

4.VI.J. Post-Market Monitoring

AI governance is not a one-time task. Rather, after an organization develops and/or deploys an AI system into the market, it needs to monitor it throughout its lifecycle, because the AI system may drift or hallucinate.¹⁹² For post-market monitoring, the AI governance team needs to have procedures to collect, document, and analyze how the AI system is performing and interacting with other systems or environments after deployment.¹⁹³ While monitoring the AI system, the AI governance team should identify and address any risks, defects, or non-conformities arising from the AI system. This will then allow the AI governance team to correct any deficiencies and make necessary adjustments.

4.VI.K. Communication Channels and Contestability

While AI systems may perform well on test and validation data, they may function differently when interacting with users. Thus, in addition to having an automated process for post-market monitoring, the organization should have open communication channels for the public to provide feedback regarding their observed concerns.¹⁹⁴

The AI governance team should also have internal procedures in place to give individuals an opportunity to opt out and request human review of AI decisions, especially if the organization is subject to emerging privacy laws giving consumers the right to opt out of profiling from automated decisions that have legal or similarly significant effects.¹⁹⁵ Individuals should have an opportunity to contest decisions¹⁹⁶ and access to a person who can quickly consider and remedy problems they may have experienced. The AI governance team should designate individuals in the organization to monitor the public's feedback and to ensure that communication channels are accessible, equitable, effective, and maintained and do not impose an unreasonable burden on the public.¹⁹⁷

4.VI.L. Adopt Appropriate AI Contractual Provisions

Whether your organization is a developer or deployer of an AI system, it may need to adopt appropriate contractual provisions to mitigate risks.¹⁹⁸

Generally speaking, unlike data privacy laws that require specific terms between controllers and processors, comprehensive AI laws, like the Colorado AI Law and EU AI Act, do not have a long

¹⁹²See generally, *id.* at Art. 72.

¹⁹³See *id.*

¹⁹⁴See PERSONAL DATA PROT. COMM'N SINGAPORE, MODEL ARTIFICIAL INTELLIGENCE GOVERNANCE FRAMEWORK 57, (2nd ed. Jan. 21, 2020), <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf>.

¹⁹⁵See, e.g., VA. CODE ANN. §59.1-577(A)(5)(iii); COLO. REV. STAT. §6-1-1306(1)(a)(C); Regulation (EU) 2016/679, General Data Protection Regulation, Art. 22(1), (O.J. L 119, 04.05.2016), <https://gdpr-info.eu/art-4-gdpr/>; *Australia's AI Ethics Principle*, AUSTL. GOV'T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024) (“In the case of decisions significantly affecting rights, there should be an effective system of oversight, which makes appropriate use of human judgment.”).

¹⁹⁶See DEP'T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 31-32 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf.

¹⁹⁷See THE WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 7 (Oct. 2022), <https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>

¹⁹⁸See ISO/IEC 42001:2023(E), Annex A, §§A.10.3 A.10.4 (INT'L ORG. FOR STANDARDIZATION 2023).

list of specific terms that developers and deployers must include in their agreements. That said, there is a narrow situation where the EU AI Act requires an agreement between a high-risk AI system provider “and the third party that supplies an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI system. ...”¹⁹⁹ The EU AI Act requires the agreement to specify “the necessary information, capabilities, technical access and other assistance based on the generally acknowledged state of the art, in order to enable the provider of the high-risk AI system to fully comply with the obligations set out in [the EU AI Act].”²⁰⁰

Moreover, if the parties are subject to the EU AI Act, they should include a provision addressing shifting of party roles. Under Article 25 of the EU AI Act, a distributor, importer, deployer or other third-party could transform into an AI system provider (New Provider) if one of the following conditions are met:²⁰¹

- (a) they put their name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are otherwise allocated;
- (b) they make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in such a way that it remains a high-risk AI system pursuant to Article 6;
- (c) they modify the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such a way that the AI system concerned becomes a high-risk AI system in accordance with Article 6.

If the New Provider meets one of the above conditions, the initial provider (Initial Provider) will no longer be the provider of that specific AI system under the EU AI Act.²⁰² Instead, the Initial Provider will only be obligated to “closely cooperate with new providers and ... make available the necessary information and provide the reasonably expected technical access and other assistance that are required for the fulfilment of the obligations set out in [the EU AI Act], in particular regarding the compliance with the conformity assessment of high-risk AI systems.”²⁰³

To address this scenario, the parties should include a provision in the agreement stating whether the AI system may be changed. If the AI system provider includes a provision prohibiting the counterparty from changing the AI system, the Initial Provider is not required to assist the New Provider under Article 25.²⁰⁴

Further, the parties may consider utilizing voluntary model terms under the EU AI Act. For example, on October 5, 2023, the European Commission published EU model contractual clauses

¹⁹⁹See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 25(4), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689.

²⁰⁰See *id.*

²⁰¹See *id.* at 25(1).

²⁰²See *id.* at Art. 25(2).

²⁰³See *id.*

²⁰⁴See *id.*

for public organizations procuring AI systems.²⁰⁵ Further, under the EU AI Act, “[t]he AI Office may develop and recommend voluntary model terms for contracts between providers of high-risk AI systems and third parties that supply tools, services, components or processes that are used for or integrated into high-risk AI systems.”²⁰⁶ Model contractual terms for AI agreements may ultimately become commercial practice, similar to the EU standard contractual clauses for cross-border data transfers.

Organizations may also consider including a mutual provision in the agreement requiring both developers and deployers to comply with their respective obligations under AI laws and not engage in prohibited AI practices. The parties may either draft this provision broadly or they can detail the specific obligations each of them must comply with based on their role. For example, an AI system developer may want provisions requiring the deployer to (1) provide the transparency and explainability notice to end users, (2) use the AI system pursuant to the developer’s documentation and instructions of use, (3) represent that it has the legal right to use the input data and that the input data is appropriate for the AI system’s intended use, and (4) be primarily responsible for honoring individual rights requests under AI laws (with the developer merely providing assistance to the deployer).

On the other hand, the deployer of the AI system may want provisions requiring the developer to represent that it (1) developed the AI system using appropriate data governance techniques and had the legal right to the data used to train, test, validate and/or fine-tune the AI system, (2) has provided the deployer with accurate and appropriate documentation and instructions of use for the AI system, (3) tested the AI system for accuracy, robustness, and security, (4) designed the AI system with the capability to automatically record events (i.e., logs), and (5) developed the AI system so that when end users prompt the AI, it communicates that it is an AI system and not a human.

The parties may also consider expanding the compliance with AI laws provision to require adopting industry best practices in connection with their development or use of AI systems, such as NIST AI RMF or ISO/IEC 42001. Including such a requirement will depend on the parties’ negotiation posture, leverage, size and sophistication. For example, if the AI system developer is a small startup, it may not have adopted a mature AI governance program to comfortably provide such a representation to its customers. Likewise, a small company contracting for AI services may not have the resources to comply with a major risk management framework or the leverage needed to require a large AI provider to agree to such a term.

Lastly, organizations should consider including a provision in the agreement requiring the parties to cooperate together with respect to certain AI obligations, such as investigating, correcting and reporting serious incidents or instances of algorithmic discrimination, preparing AI impact assessments, handling individual rights, post-market monitoring of AI systems and human oversight, and responding to regulatory inquiries and investigations.

4.VI.M. Decommissioning the AI System

²⁰⁵See *EU Model Contractual AI Clauses to Pilot in Procurements of AI*, EUR. COMM’N, (Sept. 29, 2023), <https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/eu-model-contractual-ai-clauses-pilot-procurements-ai>.

²⁰⁶See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 25(4), (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:202401689>.

Organizations should have processes and procedures in place to decommission the AI system in a safe manner so that it does not increase the risks or decrease the organization’s trustworthiness.²⁰⁷ Before decommissioning or deleting AI systems, organizations may need to consider whether the AI systems are subject to regulatory or other legal requirements that require archiving in a model inventory for a certain period of time.²⁰⁸ Organizations should also assess whether the AI system is dependent or linked to other systems or currently utilized by individuals who rely on it before decommissioning or deleting.²⁰⁹

4.VII. ACCOUNTABILITY

A mature AI government program should ensure that the organization is accountable for its AI systems.²¹⁰ The AI governance team should consider implementing measures to oversee the AI system, with accountability throughout its lifecycle.²¹¹

The AI governance team should consider documenting the above steps through auditable policies and procedures,²¹² which may include:

- An AI leadership policy describing the roles and responsibilities of individuals responsible for AI oversight in the organization;
- An IT asset inventory that includes the AI systems the organization develops and/or uses and a data provenance record reflecting the AI systems’ data lineage;
- A deployer’s AI use policy;
- A developer’s governance policy for AI development;
- An employee AI-use policy

²⁰⁷See NAT’L INST. OF STANDARDS AND TECH., AI RMF PLAYBOOK 14 (2023), https://airc.nist.gov/AI_RM_F_Knowledge_Base/Playbook.

²⁰⁸See *id.*

²⁰⁹See *id.*

²¹⁰See *Australia’s AI Ethics Principle*, AUSTL. GOV’T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024) (stating that the accountability principle acknowledges “organisations’ and individuals’ responsibility for the outcomes of the AI systems that they design, develop, deploy and operate”); Org. for Econ. Co-operation and Dev. [OECD], Recommendation of the Council on Artificial Intelligence, Legal Instrument 449, at 1.5, (July 11, 2023), <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (“AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.”).

²¹¹DEP’T FOR SCI., INNOVATION TECH., A PRO-INNOVATION APPROACH TO AI REGULATION, 2023, CP 815, at 30 (U.K.), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf; *Australia’s AI Ethics Principle*, AUSTL. GOV’T, DEPARTMENT OF INDUS., SCI. AND RES., <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principle> (last visited Aug. 14, 2024); *Canadian Guardrails for Generative AI – Code of Practice*, INNOVATION, SCI. AND ECON. DEV. CAN. (Aug. 16, 2023), <https://ised-isde.canada.ca/site/ised/en/consultation-development-canadian-code-practice-generative-artificial-intelligence-systems/canadian-guardrails-generative-ai-code-practice>.

²¹²See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Art. 17, (OJ L, 2024/1689, 12.7.2024) (July 12, 2024), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689; *The Artificial Intelligence and Data Act (AIDA) – Companion Document*, INNOVATION, SCI. AND ECON. DEV. CAN. (Mar. 13, 2023), <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document> (stating that accountability “includes the proactive documentation of policies, processes, and measures implemented”).

- An AI impact assessment that identifies the AI risks and mitigation measures implemented to reduce the risks, which incorporates by reference the technical documentation and logs maintained for the AI system and the risk management system adopted by the organization (e.g., NIST AI RMF or ISO/IEC 42001);
- A legal assessment that identifies the AI laws applicable to the organization and the steps taken (or to be taken) to address compliance, such as procedural requirements under the EU AI Act (e.g., conformity assessments, EU declaration of conformity, the CE marking affixed to the AI system, registration requirements, and appointment of an EU representative);
- An AI trust center for transparency and explainability with the public that also includes mechanisms for internal and external stakeholders to provide feedback regarding the organization's AI systems and contest decisions;
- An AI bias and fairness report that analyzes whether the AI system presents risks of algorithmic discrimination or other unfair outcomes based on the circumstances; and
- Contractual agreements with licensees and licensors of AI systems.

Attached to this book are sample high-level and generic template developer and deployer policies, an employee AI-use policy, a template AI impact assessment, and a checklist for incorporating AI regulatory requirements in agreements. Organizations will ultimately need to draft custom policies, assessments and contractual provisions as relevant to their practices and applicable laws.

The AI governance team may also review other policies within the organization and integrate them with the company's AI use and development, such as incorporating AI in a security incident response plan, data privacy policies and procedures, or human resources and marketing practices. The AI governance team may also consider conducting third-party audits to objectively test the organization's AI systems to validate its governance program.²¹³

²¹³See Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020), <https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms> (“Consider how you hold yourself accountable, and whether it would make sense to use independent standards or independent expertise to step back and take stock of your AT. ... Such outside tools and services are increasingly available as AT is used more frequently, and companies may want to consider using them.”).

Chapter 5. Conclusion

[ARSEN KOURINIAN](#)

Mayer Brown

Current through September 23, 2025.

5.I. Conclusion

An organization developing or deploying an AI system should consider being proactive and developing the internal framework necessary to address rapidly evolving AI laws, technology and threats. To address these requirements, multi-national corporations may consider governing their AI systems using an approach that harmonizes global standards, frameworks and laws, which are composed of common components. This will give organizations the ability to develop and deploy AI systems at scale, without having siloed and fragmented governance programs for each country they operate in. Adopting this strategy may also allow organizations to gain the public's trust by having a consistent approach across jurisdictions, instead of extending additional protections in certain countries (e.g., the EU), but not others.

Appendix A. Artificial Intelligence Leadership Policy

I. Purpose

The purpose of this Artificial Intelligence Leadership Policy (“Policy”) is to set forth the policy on how our organization and its affiliates (the “Company”) shall designate leaders for its artificial intelligence oversight committee (“AI Oversight Committee”) and the tasks they will need to perform. This Policy may be updated as there are developments in AI Requirements.

II. Scope

This Policy applies to members of the AI Oversight Committee. The Company’s current list of AI Oversight Committee members can be found in the Company’s directory.

III. Definitions

- A. “*AI Governance Policies*” means policies, procedures, guidelines, frameworks, and/or principles adopted by the Company for AI governance.
- B. “*AI Laws*” means laws, legislation, regulations, and regulatory guidance applicable to the Company’s use and/or development of AI Systems, each as updated or replaced from time to time.
- C. “*AI Requirements*” means AI Laws and AI Governance Policies.
- D. “*AI Systems*” means any machine-based system that, for any explicit or implicit objectives, infers from the inputs the system receives how to generate outputs, including content, decisions, predictions, decisions or recommendations, that can influence physical or virtual environments they interact with.

IV. Designation of the AI Oversight Committee Members

The Company shall designate AI Oversight Committee Members based on the following criteria:

- A. The AI Oversight Committee members shall be designated on the basis of professional qualities, knowledge of the AI Requirements, and the ability to fulfill the tasks described in this Policy.
- B. The AI Oversight Committee members shall be composed of individuals with different skillsets, backgrounds, and representation across the Company, such as those with legal, data privacy, cybersecurity, engineering, technical, data science, management, human resources, and other relevant skillsets and perspectives.
- C. The AI Oversight Committee members shall be designated according to specific geographic regions within the Company.
- D. The AI Oversight Committee members may be Company employees or, if necessary, external professional services providers.

V. Authority of the AI Oversight Committee Members

- A. The AI Oversight Committee members will report directly to senior leadership in the Company.

- B. The AI Oversight Committee members will have an appropriate degree of independence to fulfill the responsibilities described in this Policy.
- C. The AI Oversight Committee members will be trained in regular cadence regarding AI Requirements.
- D. The AI Oversight Committee members will be informed and involved, properly and in a timely manner, in all issues that are related to the Company's oversight, development and use of AI Systems.
- E. Individuals may contact the AI Oversight Committee members in their geographic region with respect to issues pertaining to AI Requirements.

VI. Responsibilities of the AI Oversight Committee

The AI Oversight Committee shall perform the following tasks within their designated geographic region of administrative responsibility:

- A. Identify the Company's objectives with AI Systems and ensure that they are established and compatible with the Company's strategic direction.
- B. Document and update the Company's AI Governance Policies at planned intervals or additionally as needed based on changes in AI Laws, industry best practices, and technological developments to ensure their continuing suitability, adequacy and effectiveness.
- C. Assess whether other policies in the Company may be impacted by or apply to the Company's objectives with respect to AI Systems.
- D. Oversee and monitor compliance with the AI Requirements.
- E. Define and put in place a process for Company employees and external parties to report concerns about the Company's use and/or development of AI Systems.
- F. Identify, document and secure the resources needed to establish, implement, maintain, and continually improve the Company's development and/or use of AI Systems and integrate them into the Company's business process.
- G. Identify and document the human resources and their competencies utilized for the development and/or use of AI Systems.
- H. Communicate the importance of the AI Requirements throughout the Company.
- I. Provide direction, support and training to employees commensurate with their roles and experience so that the AI Requirements can be effectively implemented.
- J. Develop and impose corrective measures for violations of the AI Requirements.
- K. Manage and oversee compliance with AI Requirements vis-à-vis third-party vendors, suppliers, service providers, contractors, processors and third parties.

VII. Contact Information

Please consult other members of the AI Oversight Committee if you have any questions or need to request an exception to the Policy requirements.

VIII. Revision History

Version	Date	Author	Description	Sections Affected

Appendix B. Artificial Intelligence Impact Assessment (AIIA)

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 1	

APPENDIX B: Artificial Intelligence Impact Assessment (AIIA)

AIIA DETAILS		
Contact details of project owner.	Click or tap here to enter text.	
Target launch date, period of time within which the AI system will be used and frequency of usage.	Click or tap here to enter text.	
Has this AIIA been reviewed and approved at the company?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Provide the approval date and names and positions of the individuals responsible for the review and approval. Click or tap here to enter text.
Has this AIIA been presented to company leadership, including its data protection officer (if applicable)?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Has an internal or external audit been conducted in relation to this AIIA?	<input type="checkbox"/> Yes <input type="checkbox"/> No	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 2	

	Name and title	Click or tap here to enter text.
	Company	Click or tap here to enter text.
	Date	Click or tap here to enter text.
	Results of audit (include report as attachment)	Click or tap here to enter text.
Provide a list of internal employees or teams and external parties that contributed to this AIIA.	Click or tap here to enter text.	
Who is responsible for ensuring that this project is maintained as stated in this AIIA? Describe how that person monitors compliance and the process for reporting non-compliance.	Click or tap here to enter text.	
Which of the following does this project involve?	<input type="checkbox"/> Commencement of a new project <input type="checkbox"/> Changes to an existing project	
Has an AIIA or data protection impact assessment been previously completed for this project?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
	If yes, please describe the changes or developments triggering an update to the AIIA.	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 3	
	Click or tap here to enter text.	
GENERAL DESCRIPTION OF THE AI SYSTEM		
Provide a description of the AI system (e.g., purpose, intended use cases, deployment context and general benefits).	Click or tap here to enter text.	
Is the Company developing or deploying an AI system?	<input type="checkbox"/> Developing <input type="checkbox"/> Deploying	
	<p>If the company is deploying a third-party's AI system, provide the name of the third party's AI system and attach copies of any internal or external evaluations sufficient to show the accuracy and reliability of the AI system.</p> <p>Click or tap here to enter text.</p>	
Will individuals directly interact with the AI system ?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
What output is produced by the AI system and how will it be used?	Click or tap here to enter text.	
If applicable, describe the decisions that will be made using or	Decisions	Benefits Over Manual Processing
	Click or tap here to enter text.	Click or tap here to enter text.

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 4	

supported by the AI system and its benefits over manual processing.		
	Click or tap here to enter text.	Click or tap here to enter text.
Describe the logic used by the AI system, and any assumptions of the logic.	Click or tap here to enter text.	
What metrics are used to evaluate the performance and known limitations of the AI system?	Click or tap here to enter text.	
Describe how the AI system's use directly and reasonably relates to the company's goods or services.	Click or tap here to enter text.	
Describe the technology that will be used in connection with the AI system?	Click or tap here to enter text.	
AI OVERSIGHT		
Has an AI oversight team been assembled?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Does the AI oversight team have training and AI literacy?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Describe the frequency of AI training received by the oversight team.	Click or tap here to enter text.	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 5	

DATA GOVERNANCE									
Was a foundation model utilized for the AI system?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes, please describe: Click or tap here to enter text.								
If the company is training an AI model and/or developing an AI system, describe the data governance technics used by data scientists or other specialists in the development.	Describe or attach a copy of a report regarding the data governance technics used, which address, among other things, the below. <ul style="list-style-type: none"> • <i>Relevant design choices.</i> • <i>Data collection processes and the origin of data, and in the case of personal information, the original purpose of the data collection.</i> • <i>Relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation.</i> • <i>The formulation of assumptions, in particular related to the information that the data are supposed to measure and represent.</i> • <i>Whether an assessment been done regarding the availability, quantity and suitability of the data sets that are needed.</i> • <i>Whether an assessment has been done for biases in the data, and mitigation steps necessary for same.</i> • <i>Whether there are any gaps or shortcomings that would prevent proper data governance.</i> • <i>Whether the datasets are sufficiently representative, free of errors, and complete for the intended purpose.</i> 								
What data will be used for the AI system?	<table border="1"> <thead> <tr> <th>Data Type</th> <th>Source</th> <th>Purpose</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/> Name</td> <td> <input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider </td> <td> <input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI system <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system </td> </tr> </tbody> </table>	Data Type	Source	Purpose	<input type="checkbox"/> Name	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI system <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system		
Data Type	Source	Purpose							
<input type="checkbox"/> Name	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI system <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system							

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 6	

	<input type="checkbox"/> Other		
	<input type="checkbox"/> Phone number	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Address	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Email address	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
<input type="checkbox"/> Job title	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 7	

	<input type="checkbox"/> Username	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Password	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Internet protocol address	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Persistent identifiers	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 8	

	<input type="checkbox"/> Internet or other electronic network activity information	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Audio, electronic, visual, thermal, olfactory, or similar information	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Professional or employment-related information	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Financial or business information	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Government identification and	<input type="checkbox"/> Directly from individual	<input type="checkbox"/> Train AI model

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 9	

	numbers (Social Security, driver's license, passport, or other government-issued numbers)	<input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Geolocation data	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Biometric data	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Health and medical data	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Genetic data	<input type="checkbox"/> Directly from individual	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 10	

		<input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Demographic information (family, race, ethnicity, tribal origin, religious or philosophical beliefs, sex life or sexual orientation, or immigration or citizenship status)	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Intellectual or political opinions	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Trade union membership	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Children's data	<input type="checkbox"/> Directly from individual	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)		DOCUMENT NUMBER	DATE
		PAGE 11	
		<input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
	<input type="checkbox"/> Other (<i>describe</i>)	<input type="checkbox"/> Directly from individual <input type="checkbox"/> Web scrapping or crawling <input type="checkbox"/> Third-party data provider <input type="checkbox"/> Other	<input type="checkbox"/> Train AI model <input type="checkbox"/> Fine-tune AI model <input type="checkbox"/> Test or validate AI system <input type="checkbox"/> Input prompt for AI system
Attach or provide a link to the privacy policy applicable to the personal information processed by the AI system or in connection with the project.	Click or tap here to enter text.		
If applicable, describe the lawful bases for processing the data, if personal information is involved.	<input type="checkbox"/> Consent <input type="checkbox"/> Contract with the individual <input type="checkbox"/> Public interest <input type="checkbox"/> Legitimate interests of our company <input type="checkbox"/> Privacy notice at or before the point of collection <input type="checkbox"/> Other: Click or tap here to enter text.		
	If consent is a legal basis for the processing of personal information, is such consent being obtained in compliance with privacy laws, particularly in relation to the following requirements? <input type="checkbox"/> Consent is being obtained in accordance with legally prescribed formalities		

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 12	
	<input type="checkbox"/> Separate consent is being obtained for each category of processing (e.g., collection/use, provision to third parties, and cross-border transfers) <input type="checkbox"/> Separate consent is being obtained for sensitive data and unique identification data <input type="checkbox"/> N/A consent is not required for this processing activity	
Does the company have procedures in place to honor individuals' privacy rights?	<input type="checkbox"/> Yes <input type="checkbox"/> No	
Describe the nature and approximate volume of personal information to be collected, including the number of individuals' personal information involved.	Click or tap here to enter text.	
Describe (a) from which category or categories of individuals the personal information will be collected; (b) the relationship of the individuals to the company; and (c) whether it is within the reasonable expectation of the individuals that their personal information will be processed in connection with the AI system.	Click or tap here to enter text.	
Which of the following individuals and entities	<input type="checkbox"/> Our employees <input type="checkbox"/> Our contractors	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)		DOCUMENT NUMBER	DATE
		PAGE 13	
associated with the company will have access to the personal information?	<input type="checkbox"/> Our customers <input type="checkbox"/> Our suppliers <input type="checkbox"/> Other (please specify): Click or tap here to enter text.		
Can the purpose of processing be achieved by collecting less personal information?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes, specify and explain which personal information elements could be excluded: Click or tap here to enter text.		
Is the personal information subject to a retention schedule?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes, please attach or provide a link to the retention policy Click or tap here to enter text.		
Where and how will personal information be stored?	Click or tap here to enter text.		
Is the personal information going to be returned, deleted, or anonymized/de-identified/aggregated at the conclusion of the project?	<input type="checkbox"/> Yes <input type="checkbox"/> No If no, please explain whether the personal information will be pseudonymized instead. Click or tap here to enter text.		
List the countries in which individuals' whose personal	Click or tap here to enter text.		

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 14	

information are collected are located.		
Cross-border data transfers.	In which country and systems will the personal information be stored? Click or tap here to enter text.	
	List of countries from which the personal information will be transferred and/or accessed (including remote access through login). Click or tap here to enter text.	List of countries to which the personal information will be transferred to and/or accessed (including remote access through login). Click or tap here to enter text.
	List the bases or mechanisms for any cross-border transfers of personal data (e.g., binding corporate rules, data privacy agreements, standard contractual clauses, derogations, consent, disclosure of details in the privacy policy, or other). Click or tap here to enter text.	
For what purpose will data be processed for the AI system?	<input type="checkbox"/> As a product offering that is sold or licensed to our customers <input type="checkbox"/> To provide other products or services to our customers that are not the AI system itself <input type="checkbox"/> To carry out internal business functions <input type="checkbox"/> To make decisions about individuals (if selected complete the below) <input type="checkbox"/> Other (describe)	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 15	

	<p>If applicable, select the decision(s) that will be made using the AI system.</p> <ul style="list-style-type: none"> <input type="checkbox"/> Provision or denial of financial or lending services <input type="checkbox"/> Housing <input type="checkbox"/> Insurance <input type="checkbox"/> Education enrollment or opportunity or vocational training <input type="checkbox"/> Criminal justice or law enforcement <input type="checkbox"/> Employment or recruitment <input type="checkbox"/> Healthcare services <input type="checkbox"/> Access to essential goods or services <input type="checkbox"/> Safety component of a product or AI system <input type="checkbox"/> Biometric and biometric-based systems <input type="checkbox"/> Management and operation of critical infrastructure <input type="checkbox"/> Administration of justice and democratic processes <input type="checkbox"/> Immigration and border control <input type="checkbox"/> Other: Click or tap here to enter text.
Is any of the data explicitly restricted or prohibited from being collected or processed by applicable laws or court decisions?	<ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> No <p>If yes, please describe:</p> <p>Click or tap here to enter text.</p>
Will external service providers or third parties have access to personal information	<ul style="list-style-type: none"> <input type="checkbox"/> Yes <input type="checkbox"/> No

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 16	

or other data processed in connection with the AI system?	If yes, describe the steps taken to ensure that disclosing or making available personal information to service providers or third parties is in compliance with privacy laws and complete the details below. Click or tap here to enter text.				
	Name	Has the service provider or third party signed a data protection agreement?	Is the service provider or third party in compliance with an approved code of conduct or any certifications, seals, or marks? If yes, please list.	Purpose of processing	Countries the personal information will be processed in.
	Click or tap here to enter text.	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No Click or tap here to enter text.	Click or tap here to enter text.	Click or tap here to enter text.
	Click or tap here to enter text.	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No Click or tap here to enter text.	Click or tap here to enter text.	Click or tap here to enter text.
Please fill out the following table reflecting the result	Processing activity	Purpose and benefit	Necessity	Proportionality	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 17	

of the necessity and proportionality assessment performed, taking into consideration each processing activity and its respective purpose. This assessment is based on the premise that the data retention periods indicated above are implemented.				
	Based on the above, are the processing activities carried out necessary and proportionate? <input type="checkbox"/> Yes <input type="checkbox"/> No			
RISK MANAGEMENT				
Has legal conducted a review to determine if the AI system falls within a prohibited category?	<input type="checkbox"/> Yes – the AI system is not prohibited <input type="checkbox"/> No			
Risk analysis <u>without</u> mitigation measures in place.	Description of risk and ranking (high, limited, minimal).	Likelihood (high, medium, low).	Severity (high, medium, low).	
Risk analysis <u>with</u> mitigation measures described below implement.	Description of risk and ranking (high, limited, minimal).	Likelihood (high, medium, low).	Severity (high, medium, low).	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 18	

MITIGATION MEASURES		
Have measures been taken to address transparency and explainability?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Website statement regarding AI systems <input type="checkbox"/> AI pre-use notice(s), including for adverse decisions (if applicable) <input type="checkbox"/> Instructions for use and documentation to deployer <input type="checkbox"/> Disclosure in privacy policy <input type="checkbox"/> The AI system discloses to individuals with whom it interacts, if asked or prompted by the individual, that the individual is interacting with an AI system and not a human <input type="checkbox"/> Other (describe)
Watermarking	Does the AI system generate synthetic audio, image, video or text content? <input type="checkbox"/> Yes <input type="checkbox"/> No	If yes, are the AI system's outputs marked in a machine-readable format and detectable as artificially generated or manipulated? <input type="checkbox"/> Yes <input type="checkbox"/> No
Has a risk management system (RMS) been adopted?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> RMS has been adopted per a legal standard (<i>specify</i>) <input type="checkbox"/> NIST AI RMF <input type="checkbox"/> ISO/IEC 42001 <input type="checkbox"/> Other (<i>specify</i>)
Do you maintain technical documentation	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> The company maintains documentation provided by the developer of the AI system

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)		DOCUMENT NUMBER	DATE
		PAGE 19	
regarding the AI system?		<input type="checkbox"/> The company has prepared technical documentation regarding the AI system that it has developed	
Record-keeping	Does the AI system technically allow for the automatic recording of events (logs) over the lifetime of the system?	<input type="checkbox"/> Yes <input type="checkbox"/> No If yes, do the logging capabilities provide: <ul style="list-style-type: none"> <input type="checkbox"/> Recording of the period of each use of the system (start date and time and end date and time of each use) <input type="checkbox"/> The reference database against which input data has been checked by the system <input type="checkbox"/> The input data for which the search has led to a match <input type="checkbox"/> The identification of the natural persons involved in the verification of the results for remote biometric identification systems. Check this box if not applicable <input type="checkbox"/>. 	
Human oversight	Level of human oversight <ul style="list-style-type: none"> <input type="checkbox"/> Human in the loop (a human is in full control over the AI system and treats its output as a recommendation) <input type="checkbox"/> Human over the loop (a human can control and override unexpected and undesirable events while 	Describe: Click or tap here to enter text.	

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)	DOCUMENT NUMBER	DATE
	PAGE 20	

	<p>monitoring or supervising AI decisions)</p> <p><input type="checkbox"/> Human out of the loop (the AI system operates with no human oversight)</p>	
Accuracy	<p>Has the AI system been tested for accuracy?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Description of accuracy testing, including as applicable, precision, recall, F1 score, and overfitting/underfitting:</p> <p>Click or tap here to enter text.</p>
Robustness	<p>Has the AI system been tested for robustness to assess its resiliency to errors, faults or inconsistencies that may occur within the system or the environment it operates in?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Describe:</p> <p>Click or tap here to enter text.</p>
Bias and fairness	<p>Has the AI system been tested for bias, fairness, and algorithmic discrimination?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Describe:</p> <p>Click or tap here to enter text.</p>
Human-centered and beneficial for society and the environment	<p>Has the AI system been assessed to determine its benefit for humans, society and the environment?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>Describe:</p> <p>Click or tap here to enter text.</p>
Safety	<p>Has the AI system been tested for safety?</p>	<p>Describe:</p>

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)		DOCUMENT NUMBER	DATE
		PAGE 21	
	<input type="checkbox"/> Yes <input type="checkbox"/> No	Click or tap here to enter text.	
Cybersecurity	Have appropriate cybersecurity measures been adopted to safeguard the AI system? <input type="checkbox"/> Yes <input type="checkbox"/> No	Description of cybersecurity measures, including procedures for reporting serious incidents: Click or tap here to enter text.	
Post-deployment monitoring	Has a post-deployment monitoring system been established and documented? <input type="checkbox"/> Yes <input type="checkbox"/> No	Describe: Click or tap here to enter text.	
Privacy-enhanced	Have steps been taken to minimize use of personal information and adopt other privacy-protective measures? <input type="checkbox"/> Yes <input type="checkbox"/> No	Describe: Click or tap here to enter text.	
Communication channels	Have communication channels been established for external feedback regarding the AI system? <input type="checkbox"/> Yes <input type="checkbox"/> No	Describe: Click or tap here to enter text.	
Does the company provide opt-out, access to decision, appeal and other rights?	<input type="checkbox"/> Right to opt-out of decisions using the AI system <input type="checkbox"/> Right to access information about decisions made using the AI system <input type="checkbox"/> Right to appeal the decision made by the AI system <input type="checkbox"/> Not applicable – the company does not make regulated decisions about individuals using the AI system <input type="checkbox"/> Rights provided under applicable data privacy laws		

ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (AIIA)		DOCUMENT NUMBER	DATE
		PAGE 22	
Decommissioning	Has a plan been established for decommissioning the AI system?	Describe: Click or tap here to enter text.	
ACCOUNTABILITY			
Does the company maintain policies and procedures related to the development or deployment of the AI system?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Describe: Click or tap here to enter text.	
CONCLUSION			
Based on the risks, mitigation measures implemented, and the benefits of the AI system, should the company proceed with developing or deploying the AI system?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Describe: Click or tap here to enter text.	

Appendix C. Artificial Intelligence Developer Policy

I. Purpose

Our organization (the “Company”) is committed to the safe, secure and responsible development of artificial intelligence (“AI”) systems (“AI Systems”) as part of our business operations. The purpose of this AI Developer Policy (“Policy”) is to establish the governance principles and practices the Company should abide by when developing AI Systems as part of our business operations. While the definition of AI Systems will vary across countries, frameworks, guidelines and laws, we define AI Systems to mean any machine-based system that, for any explicit or implicit objectives, infers from the inputs the system receives how to generate outputs, including content, decisions, predictions, decisions or recommendations, that can influence physical or virtual environments they interact with.

II. Scope

This Policy applies to officers, directors and employees in our organization that are involved in and oversee the development of AI Systems as part of our business operations (“AI Members”).

III. AI Governance

The Company is committed to the safe, secure and trustworthy development of AI Systems within our organization. For this reason, the Company requires AI Members to follow the AI governance procedures described below before developing AI Systems as part of our business operations. The AI governance procedures include: (A) involving the AI Oversight Committee; (B) implementing proper data governance; (C) conducting a legal compliance analysis; (D) conducting a risk assessment; (E) initiating measures to mitigate risks; and (F) demonstrating accountability.

III.A. Involving the AI Oversight Committee

The Company has formed an AI Oversight Committee to oversee the use and development of AI Systems in our organization. The AI Oversight Committee is subject to the Company’s AI Leadership Policy. Before developing an AI System, AI Members shall inform and consult the AI Oversight Committee and follow any instructions and directions they provide.

III.B. Data Governance

AI Members should involve individuals with appropriate training in data science, engineering and other relevant skillsets to ensure that the Company’s development of AI Systems comports with proper data governance practices. If the Company is fine-tuning AI Systems developed by another organization, AI Members should conduct due diligence on the AI System or model developer to ensure that proper data governance practices were involved in developing the AI System or model, and that the data used for finetuning the AI Systems comports with appropriate data governance practices.

For data governance, the Company needs to adopt appropriate data collection practices; prepare the data (e.g., through annotation, labeling, cleaning, updating enrichment, and aggregation); formulate assumptions for the information the data is supposed to measure and represent; assess the availability, quantity, and suitability of the datasets needed; minimize or eliminate bias; and ensure that the datasets are representative of the environment, free of errors, complete, of good quality, and respectful of the intellectual property rights of others. The Company should also use different

datasets for training, testing, and validation. The AI System needs to be trained using training data, assessed for accuracy using test data, and validated using the validation dataset.

AI Members also should understand the data lineage by tracking where the data came from; how it was collected, curated, and moved within the Company; and how the data's accuracy is maintained over time. AI Members should maintain a data provenance record that documents the data quality (from origin to transformation), traces sources of error, updates the data, and attributes data to their sources.

III.C. Legal Compliance

Before developing AI Systems, AI Members should identify which laws and regulations apply, assess whether additional steps should be taken to ensure compliance, and implement compliance steps as necessary. AI laws may require the Company to take additional procedural steps as part of developing AI Systems, such as indicating the name, registered trade name or mark and address on the AI Systems, packaging or accompanying documentation; maintaining a quality management system that ensures compliance with AI laws; ensuring that the AI Systems undergo a conformity assessment procedure prior to placing it on the market or putting it into service; preparing a declaration of conformity; affixing mandatory markings on the AI Systems; registering the AI Systems and the Company in a database; cooperating with regulatory authorities; and appointing a representative in certain regions. AI Members should involve legal counsel to help with this assessment.

III.D. Risk Management

AI Members should follow the Company's risk management system, which is used to identify, mitigate, and manage risk to ensure that the development of the AI Systems is safe, secure and trustworthy. As part of the risk management program, AI Members should rank the AI risks, identify the likelihood and severity of harm, and document an AI impact assessment.

III.D.1. Risk Ranking

III.D.1.a. Prohibited AI Systems

AI Members should evaluate the AI Systems and determine whether the AI practice is prohibited under applicable laws. Examples of prohibited AI practices may include the following:

- AI Systems using subliminal techniques to distort a person's behavior in a harmful manner;
- AI Systems used to exploit vulnerabilities of disadvantaged groups (e.g., due to age, disability, social or economic situation, and other protected classifications);
- AI Systems used for social scoring based on social behavior or personal characteristics;
- AI Systems used to create or expand facial recognition databases through the untargeted scrapping of facial images from the internet or CCTV footage;
- AI Systems used for emotion recognition in the workplace;
- AI Systems used for biometric categorization to infer sensitive data;
- AI Systems used for predictive policing or discipline;
- AI Systems that discriminate against an applicant or employee due to protected classifications;
- AI-based credit decisions that prevent creditors from accurately identifying the specific reasons for denying credit or taking other adverse actions;

- Unfair and deceptive practices;
- AI Systems that cause algorithmic discrimination;
- AI Systems that incite or encourage a person to commit physical self-harm (including suicide), harm another person, or engage in criminal activity;
- AI Systems that infringe, restrict, or otherwise impair an individual's rights guaranteed under the United States Constitution; and
- AI Systems developed or used in connection with certain sexually explicit content and child pornography.

If the AI processing activity is prohibited, AI Members must cease such practices and reconsider developing the AI Systems.

III.D.1.b. High-Risk AI Systems

If the AI Systems are not prohibited, AI Members should assess whether the AI Systems are high risk, which may vary depending on applicable laws and jurisdictions. Examples of potential high-risk AI Systems are provided below:

- Critical infrastructure;
- Product safety component or certain regulated products;
- Biometric identification and surveillance;
- Education and vocational training;
- Employment and recruitment;
- Essential goods, services, and benefits;
- Law enforcement and administration of justice;
- Immigration and border control;
- Financial or lending services;
- Essential government services;
- Health-care services;
- Housing;
- Insurance; and
- Legal services.

If the AI Systems are identified as high risk, AI Members should take additional mitigation measures to reduce the risk.

III.D.1.c. Medium or Low Risk AI

Certain AI Systems that are not high risk may fall within the medium- or low-risk category, depending on how they are used. Examples include GPS navigation, chatbots or AI in video games. AI Members may still need to implement some mitigation measures in connection with low-to-medium risk AI Systems (e.g., transparency and explainability).

III.D.2. Likelihood and Severity of Harm

AI Members should conduct a risk analysis to determine the likelihood that the AI risks will materialize and severity of harm. Depending on the likelihood and severity of harm, AI Members may need to implement additional mitigation measures to reduce the risk.

III.D.3. AI Impact Assessment

AI Members should document the above risk analysis in an AI impact assessment. The AI impact assessment should address requirements under applicable data privacy and AI laws, frameworks, and guidelines. The Company maintains a template AI impact assessment that AI Members should utilize following consultation with the Company's AI Oversight Committee.

III.E. Mitigation Measures

Depending on the severity and likelihood of harm and applicable laws and best practices, AI Members should consider the following mitigation measures in connection with developing AI Systems. AI Members should discuss with individuals with appropriate technical and legal training regarding the mitigation measures necessary based on the risk score.

III.E.1. Transparency and Explainability

AI Members should prepare transparency and explainability notices providing information and instructions of use regarding the Company's AI Systems. The AI transparency and explainability notice and information should also be included in our public-facing trust center.

III.E.2. Fairness, Avoidance of Bias, Algorithmic Discrimination, and Accessibility

AI Members should implement appropriate measures (including during the data governance stage) to ensure that the AI Systems are fair, unbiased and avoid algorithmic discrimination. The Company should engage internal or external auditors to evaluate fairness, bias and algorithmic discrimination as needed depending on the intended use of the AI Systems. In addition, the AI Members should take steps to make the AI Systems accessible for those with disabilities under applicable standards.

III.E.3. Human-Centered and Beneficial for the Environment and Society

AI Members should communicate to the public the benefits of the AI Systems to individuals, the environment, and society. AI Members should also incorporate the benefits of the AI Systems to humans, society, and the environment as part of the Company's Environmental, Social, and Governance messaging.

III.E.4. Accuracy

AI Members should test and document the accuracy of the AI Systems by benchmarking how close initial observations, computations, or estimates are to true values. The Company should engage individuals with relevant skillsets to assess the AI Systems for underfitting or overfitting, and accuracy levels (e.g., precision, recall, and/or F1 score).

III.E.5. Robustness

AI Members should test the AI Systems to ensure that they are robust so that they can cope with erroneous inputs or errors during execution and function correctly in non-ideal circumstances. This also includes ensuring that the AI Systems perform in a manner that will not harm people or operate in unexpected settings. AI Members should assess the AI Systems' robustness by conducting ongoing testing or monitoring to confirm that the AI Systems perform as intended, including through adversarial "red team" testing on the AI Systems to ensure that they are able to handle a broader range of unexpected input variables.

III.E.6. Safe and Secure

AI Members should test the AI Systems to ensure that they are safe and secure throughout the AI lifecycle by continually identifying, assessing, and managing risks. AI Systems should be safe so that they do not endanger human life, health, property, or the environment. For security, AI Members should implement appropriate cybersecurity (i.e., confidentiality, integrity, and availability) and ensure that AI security incident use cases are implemented within the Company's incident response plan to address potential misuse of the AI Systems and respond to third-party bad actors who may, for example, try to exploit vulnerabilities and manipulate the training dataset (data poisoning) or pre-trained components used in training (model poisoning), which may lead to harmful decision-making. AI Members should regularly test and conduct diligence on the AI Systems throughout their lifecycle and update technical standards addressing safety and security. This also includes, for example, conducting internal and external red-teaming of the AI Systems.

III.E.7. Privacy-Enhanced

AI Members should test the AI Systems to ensure that they are compliant with applicable privacy laws (e.g., providing privacy notices, honoring privacy rights, and implementing internal business obligations), comport with the Company's data privacy policies and procedures, and adopt privacy-enhancing technologies ("PETs"), as necessary. PETs include data minimization methods, such as anonymization, de-identification, or aggregation so that, where possible, personal data is not used in the AI Systems.

III.E.8. Human Oversight

AI Members should determine the level of human oversight necessary for the AI Systems to prevent or minimize risks. To decide on the level of human oversight, AI Members should consider the severity and probability of harm based on the above-described risk assessment. The levels of human oversight may include: (a) human-out-of-the-loop, which has no human oversight over AI decisions; (b) human-over-the-loop, which has a human involved in monitoring or supervising AI decisions with the ability to take over control when the AI Systems encounter unexpected or undesirable events; and (c) human-in-the-loop, which has active and involved human oversight, with a human retaining full control and the AI Systems only providing recommendations or input. If the severity and probability of harm are high, human-in-the-loop is the appropriate level of oversight. However, if the severity and probability of harm are low, human-out-of-the-loop is sufficient. Further, if there is moderate severity and likelihood of harm, AI Members may consider human-over-the-loop as an option.

AI Members should also consider the following factors as part of the human oversight: (a) developing a kill switch if the AI Systems pose a danger; (b) avoiding automation bias by understanding that AI Systems are not always right; (c) understanding the capacities and limitations of the AI Systems; (d) learning how to correctly interpret the AI Systems' output; and (e) knowing when to disregard, override, or reverse AI decisions.

III.E.9. Technical Documentation, Logs, and AI Inventory

AI Members should maintain technical documentation related to the AI Systems and automatically recorded events. For the technical documentation, AI Members should include, among other things, a description of: (a) the AI Systems; (b) the elements of the AI Systems and their development process; (c) the monitoring, functioning, and control of the AI Systems, particularly with regard to their capabilities and limitations in performance; (d) the appropriateness of the performance

metrics for the AI Systems; (e) the risk management system adopted; (f) any change made to the AI Systems through their lifecycle; and (g) the system in place to evaluate the AI Systems' performance in the post-market phase. AI Members should also ensure that the AI Systems are capable of automatically recording events (i.e., logs) while they are in operation, including: (a) the period of each AI System use (start and end date and time of each use); (b) the reference database against which the AI Systems have checked the input data; (c) the input data for which the search has led to a match; and (d) the persons involved in verifying the results. AI Members should also ensure that the AI Systems developed are documented in the Company's IT asset inventory.

III.E.10. Post-Market Monitoring System and Communication Channels

AI Members should implement a post-market monitoring system to ensure that the AI Systems are in compliance throughout their life cycle. This involves collecting, documenting, and analyzing data about how the AI Systems perform and interact with other systems or environments, and identifying and addressing any risks, defects, or non-conformities that may arise from the AI Systems. This also includes establishing communication channels for members of the public to provide feedback and report issues related to the AI Systems.

III.E.11. Adopt Appropriate AI Contractual Provisions

Before making the Company's AI Systems available to customers, the AI Members should enter into appropriate contracts that describe the roles, responsibilities and allocation of liability and risks of the Company and customers. The Company maintains a template AI contract that should be utilized following consultation with legal counsel.

III.E.12. Decommissioning the AI System

If AI Members decide to decommission the AI System because they will no longer be used by the Company, they should consider the risks posed to linked systems, legal or regulatory concerns, and impact to customers and the public in doing so.

III.F. Accountability

AI Members should ensure that measures are in place to oversee the AI Systems, with accountability across the AI lifecycle. AI Members should confirm that the above steps are properly documented through auditable policies, procedures, and practices. This may also require third-party audits to objectively test the AI Systems to validate that the Company has implemented appropriate governance.

IV. Contact Information

AI Members should contact the AI Oversight Committee if they have any questions regarding this Policy or need to request an exception to the Policy requirements.

V. Revision History

Version	Date	Author	Description	Sections Affected

Appendix D. Artificial Intelligence Deployer Policy

I. Purpose

Our organization (the “Company”) is committed to the safe, secure, and responsible use of artificial intelligence (“AI”) systems (“AI Systems”) as part of our business operations. The purpose of this AI Deployer Policy (“Policy”) is to establish the governance principles and practices AI Members should abide by when using AI Systems as part of the Company’s business operations. While the definition of AI Systems varies across countries, frameworks, guidelines, and laws, we define AI Systems to mean any machine-based system that, for any explicit or implicit objectives, infers from the inputs the system receives how to generate outputs, including content, decisions, predictions, decisions or recommendations, that can influence physical or virtual environments they interact with.

II. Scope

This Policy applies to officers, directors, and employees in the Company that are involved in and oversee the use of AI Systems as part of our business operations (“AI Members”).

III. AI Governance

The Company is committed to the safe, secure, and trustworthy use of AI Systems within our organization. For this reason, the Company requires AI Members to follow the AI governance procedures described below before using AI Systems as part of our business operations. The AI governance procedures include: (A) involving the AI Oversight Committee; (B) implementing proper data governance; (C) conducting a legal compliance analysis; (D) conducting a risk assessment; (E) initiating measures to mitigate risks; and (F) demonstrating accountability.

III.A. Involving the AI Oversight Committee

The Company has formed an AI Oversight Committee to oversee the use and development of AI Systems in our organization. The AI Oversight Committee is subject to the Company’s AI Leadership Policy. Before using an AI System, AI Members shall inform and consult the AI Oversight Committee and follow any instructions and directions they provide.

III.B. Data Governance

AI Members should conduct diligence on the AI System providers to ensure that they developed the AI Systems using appropriate data governance techniques and that the AI Systems are developed with data that is sufficiently representative and relevant to the Company’s use of the AI Systems. Before using data as the input prompt for the AI Systems, AI Members should ensure that there are no confidentiality, data privacy, commercial, or legal restrictions precluding them from using the data for the AI processing activity. This involves consulting with legal and other stakeholders in the Company to ensure that the Company has the legal right to use the data, which may require obtaining licensing rights, entering into commercial agreements, providing a privacy notice, documenting a lawful basis for processing personal data under data privacy laws, or obtaining consent from individuals.

AI Members also should understand the data lineage by tracking where the data came from; how it was collected, curated, and moved within the Company; and how the data’s accuracy is maintained over time. AI Members should maintain a data provenance record that documents the data quality

(from origin to transformation), traces sources of error, updates the data, and attributes data to their sources. AI Members may need to coordinate with the AI System providers to obtain documents relevant to data governance.

III.C. Legal Compliance

Before using AI Systems, AI Members should work with legal counsel to identify which laws and regulations apply, assess whether additional steps should be taken to ensure compliance, and implement compliance steps as necessary. AI laws may require the Company to take additional procedural steps as part of using AI Systems, such as providing notice to individuals regarding the use of AI Systems, honoring individual rights under AI and data privacy laws, registering the AI System use in a database, and cooperating with regulatory authorities.

III.D. Risk Management

AI Members should follow the Company's risk management system, which is used to identify, mitigate, and manage risk to ensure that the AI System use is safe, secure, and trustworthy. As part of the risk management program, AI Members should rank the AI risks, identify the likelihood and severity of harm, and document an AI impact assessment.

III.D.1. Risk Ranking

III.D.1.a. Prohibited AI Systems

AI Members should evaluate the AI Systems and determine whether the AI use is prohibited under applicable laws. Examples of prohibited AI practices may include the following:

- AI Systems using subliminal techniques to distort a person's behavior in a harmful manner;
- AI Systems used to exploit vulnerabilities of disadvantaged groups (e.g., due to age, disability, social or economic situation, and other protected classifications);
- AI Systems used for social scoring based on social behavior or personal characteristics;
- AI Systems used to create or expand facial recognition databases through the untargeted scrapping of facial images from the internet or CCTV footage;
- AI Systems used for emotion recognition in the workplace;
- AI Systems used for biometric categorization to infer sensitive data;
- AI Systems used for predictive policing or discipline;
- AI Systems that discriminate against an applicant or employee due to protected classifications;
- AI-based credit decisions that prevent creditors from accurately identifying the specific reasons for denying credit or taking other adverse actions;
- Unfair and deceptive practices;
- AI Systems that cause algorithmic discrimination;
- AI Systems that incite or encourage a person to commit physical self-harm (including suicide), harm another person, or engage in criminal activity;
- AI Systems that infringe, restrict, or otherwise impair an individual's rights guaranteed under the United States Constitution; and
- AI Systems developed or used in connection with certain sexually explicit content and child pornography.

If the AI processing activity is prohibited, AI Members must cease such practices and reconsider using the AI Systems.

III.D.1.b. High-Risk AI Systems

If the AI Systems are not prohibited, AI Members should assess whether the AI Systems are high risk, which may vary depending on applicable laws and jurisdictions. Examples of potential high-risk AI Systems are provided below:

- Critical infrastructure;
- Product safety component or certain regulated products;
- Biometric identification and surveillance;
- Education and vocational training;
- Employment and recruitment;
- Essential goods, services, and benefits;
- Law enforcement and administration of justice;
- Immigration and border control;
- Financial or lending services;
- Essential government services;
- Health-care services;
- Housing;
- Insurance; and
- Legal services.

If the AI Systems are identified as high risk, AI Members should take additional mitigation measures to reduce the risk.

III.D.1.c. Medium- or Low-Risk AI

Certain AI Systems that are not high risk may fall within the medium- or low-risk categories, depending on how they are used. Examples include GPS navigation, chatbots, or AI in video games. AI Members may still need to implement some mitigation measures in connection with low-to-medium-risk AI Systems (e.g., transparency and explainability).

III.D.2. Likelihood and Severity of Harm

AI Members should conduct a risk analysis to determine the likelihood that the AI risks will materialize and severity of harm. Depending on the likelihood and severity of harm, AI Members may need to implement additional mitigation measures to reduce the risk.

III.D.3. AI Impact Assessment

AI Members should document the above risk analysis in an AI impact assessment. The AI impact assessment should address requirements under applicable data privacy and AI laws, frameworks, and guidelines. The Company maintains a template AI impact assessment that AI Members should utilize following consultation with the Company's AI Oversight Committee.

III.E. Mitigation Measures

Depending on the severity and likelihood of harm and applicable laws and best practices, AI Members should consider the following mitigation measures in connection with using AI Systems.

AI Members should discuss with individuals with appropriate technical and legal training regarding the mitigation measures necessary based on the risk score.

III.E.1. Transparency and Explainability

The Company may need to prepare a transparency and explainability notice before using AI Systems, which informs individuals that they are interacting with AI Systems or the AI Systems are making consequential decisions about them. If required, AI Members shall ensure that the Company's transparency and explainability notice is provided to individuals before using AI Systems. The transparency and explainability notice may need to be included in the Company's public-facing trust center that contains the required information. If the AI Members use AI Systems to make certain adverse decisions about individuals, they may also need to provide additional notices and instructions on how to appeal the decision.

III.E.2. Fairness, Avoidance of Bias, Algorithmic Discrimination, and Accessibility

AI Members should implement appropriate measures (including during the data governance stage) to ensure that the AI Systems are fair, unbiased, and avoid algorithmic discrimination. The Company should engage internal or external auditors to evaluate fairness, bias, and algorithmic discrimination as needed depending on the intended AI use case. In addition, AI Members should take steps to make the AI Systems accessible for those with disabilities under applicable standards.

III.E.3. Human-Centered and Beneficial for the Environment and Society

AI Members should communicate to the public the benefits of the AI Systems to individuals, the environment, and society. AI Members should also incorporate the benefits of the AI Systems to humans, society, and the environment as part of the Company's Environmental, Social, and Governance messaging.

III.E.4. Accuracy

AI Members should test and document the accuracy of the AI Systems by benchmarking how close initial observations, computations, or estimates are to true values. If the AI Systems are providing inaccurate results or drifting, AI Members may need to coordinate with the AI System developers to take corrective action.

III.E.5. Robustness

AI Members should test the AI Systems to ensure that they are robust so that they can cope with erroneous inputs or errors during execution and function correctly in non-ideal circumstances. This also includes ensuring that the AI Systems perform in a manner that will not harm people or operate in unexpected settings. AI Members should assess the AI Systems' robustness by conducting ongoing testing or monitoring to confirm that the AI Systems perform as intended, including through adversarial "red team" testing on the AI Systems to ensure that they are able to handle a broader range of unexpected input variables.

III.E.6. Safe and Secure

AI Members should test the AI Systems for safety and security before and after deployment. AI Systems should be safe so that they do not endanger human life, health, property, or the environment. For security, AI Members should implement appropriate cybersecurity (i.e., confidentiality, integrity, and availability) and ensure that AI security incident use cases are

implemented within the Company's incident response plan to address potential misuse of the AI Systems and respond to third-party bad actors who may, for example, try to exploit vulnerabilities and manipulate the training dataset (data poisoning) or pre-trained components used in training (model poisoning), which may lead to harmful decision-making. AI Members should regularly test and conduct diligence on the AI Systems throughout their lifecycle and update technical standards addressing safety and security. This also includes, for example, conducting internal and external red-teaming on the AI Systems.

III.E.7. Privacy-Enhanced

AI Members should test the AI Systems to ensure that they are compliant with applicable data privacy laws (e.g., providing privacy notices, honoring privacy rights, and implementing internal business obligations), comport with the Company's data privacy policies and procedures, and adopt privacy-enhancing technologies ("PETs"), as necessary. PETs include data minimization methods, such as anonymization, de-identification, or aggregation so that, where possible, personal data is not used in the AI Systems.

III.E.8. Human Oversight

AI Members should determine the level of human oversight necessary for the AI Systems to prevent or minimize risks. To decide on the level of human oversight, AI Members should consider the severity and probability of harm based on the above-described risk assessment. The levels of human oversight may include: (a) human-out-of-the-loop, which has no human oversight over AI decisions; (b) human-over-the-loop, which has a human involved in monitoring or supervising AI decisions with the ability to take over control when the AI Systems encounter unexpected or undesirable events; and (c) human-in-the-loop, which has active and involved human oversight, with a human retaining full control and the AI Systems only providing recommendations or input. If the severity and probability of harm are high, human-in-the-loop is the appropriate level of oversight. However, if the severity and probability of harm are low, human-out-of-the-loop is sufficient. Further, if there is moderate severity and likelihood of harm, AI Members may consider human-over-the-loop as an option.

AI Members should also consider the following factors as part of the human oversight: (a) confirming with the AI developers whether there is a kill switch if the AI Systems pose a danger; (b) avoiding automation bias by understanding that AI Systems are not always right; (c) understanding the capacities and limitations of the AI Systems; (d) learning how to correctly interpret the AI Systems' output; and (e) knowing when to disregard, override, or reverse AI decisions.

III.E.9. Instructions for Use, Technical Documentation, Logs, and AI Inventory

AI Members should maintain any technical documentation provided by the AI developer and ensure that they use the AI Systems pursuant to the instructions for use. AI Members should also keep the logs automatically generated by the AI Systems to the extent such logs are under their control. Further, AI Members should ensure that the AI Systems they use are documented in the Company's IT asset inventory.

III.E.10. Post-Market Monitoring System and Communication Channels

AI Members should implement a post-market monitoring system, including pursuant to the developers' instructions for use, to ensure that the AI Systems operate as intended throughout their

lifecycle. This involves collecting, documenting, and analyzing data about how the AI Systems perform and interact with other systems or environments, and identifying and addressing any risks, defects, or non-conformities that may arise from the AI Systems. This also includes establishing communication channels for members of the public to provide feedback and report issues related to the AI Systems.

III.E.11. Adopt Appropriate AI Contractual Provisions

Before using an AI System, AI Members should ensure that the Company has an appropriate contract in place with the AI provider, containing provisions that describe the roles, responsibilities and allocation of liability and risks between the parties. The Company maintains a template AI contract that should be utilized following consultation with legal counsel.

III.E.12. Decommissioning the AI Systems

If AI Members decide to decommission the AI Systems because they will no longer be used by the Company, they should consider the risks posed to linked systems, legal, or regulatory concerns, and impact to customers and the public in doing so.

III.F. Accountability

AI Members should ensure that measures are in place to oversee the AI Systems, with accountability across the AI lifecycle. AI Members should confirm that the above steps are properly documented through auditable policies, procedures, and practices. This may also require third-party audits to objectively test the AI Systems to validate that the Company has implemented appropriate governance.

IV. Contact Information

AI Members should contact the AI Oversight Committee if they have any questions regarding this Policy or need to request an exception to the Policy requirements.

V. Revision History

Version	Date	Author	Description	Sections Affected

Appendix E. Employee Artificial Intelligence Use Policy

I. Purpose

Our organization (the “Company”) is committed to the safe and responsible use of artificial intelligence (“AI”) in relevant business operations. The purpose of this employee AI use policy (“Policy”) is to establish guidelines on how to use AI systems safely and lawfully at the Company.

AI has the potential to significantly improve our productivity, efficiency and profitability. However, AI also presents risks and challenges that must be carefully managed to ensure that we use AI systems and associated data in a secure and responsible manner. “AI” is a tool that allows you to give a command or instruction (the input) to get content, decisions, predictions or recommendations (the output). The AI’s response could be in the form of text, images, audio or video. AI is trained on a high volume of data from different sources so that it can infer what would be a helpful response to your command or instruction. To be clear, the response you get from AI is not from a human being; rather, it is a machine that could be wrong. Always double check any information or content you get from AI as an output and avoid deferring to what AI says. The Company has a list of company-approved AI systems, which are the only AI systems employees may use in our organization while carrying out the Company’s work functions.

II. Scope

This Policy applies to all employees. Nothing in this Policy is intended to nor should modify or supersede any of our other policies and procedures that employees are subject to, nor does this Policy limit our efforts to comply with any applicable law or regulation and other governance documents. Nothing in this Policy is intended to violate any local, state, or federal law.

III. AI Governance

We have designated individuals in the Company who are responsible for overseeing the organization’s responsible and lawful use of AI systems. If employees have any questions regarding their use of AI systems and this Policy, particularly AI practices that may be prohibited or high-risk, they should contact our AI oversight committee.

IV. AI Principles

Employees need to use AI systems aligned with the Company’s principles, which are intended to mitigate AI risks. Employees must comply with the following principles when they use AI.

IV.A. Transparency

The Company is committed to using AI systems in a transparent manner. Employees using AI should not hide that they used AI as part of their job functions or that materials are AI-generated.

IV.B. Explainability

The Company is committed to making sure that its use of AI systems is explainable. AI systems work in a complex manner, which makes it difficult to explain why it generated a particular output in response to a command. For this reason, it is important that employees assess the output and explain based on their own analysis, research and judgment support the reason for any decisions made using the AI’s output.

IV.C. Human Involvement

The Company believes that while AI carries benefits, it cannot replace human involvement in our business operations. AI systems must be used as a tool and resource to assist employees, not as a replacement for human judgment and expertise. Employees should not use the AI output without human review to ensure accuracy, completeness, fairness, and applicability to a given task or function. Employees are ultimately responsible for all content produced with the assistance of AI, as if employees were the original creators.

IV.D. Security and Safety

The Company is committed to the secure and safe use of AI. AI systems may be vulnerable to unique cybersecurity attacks and generate content that is not safe, such as harmful recommendations or hateful content. If employees learn that their use of an AI system may have compromised the Company's security or the safety of others, they must follow our incident response procedures.

IV.E. Human Centered

The Company is committed to using AI in a manner that is beneficial to individuals, society and the environment. Employees should be mindful of the Company's Environmental, Social, and Governance goals and how they can contribute to them by using AI for good.

IV.F. Fairness

The Company is committed to using AI in a manner that is fair and avoids bias. Employees should consider whether the command they provide to the AI system could lead to biased outcomes, and carefully review the output of the AI system to check for any biased and/or unfair responses.

IV.G. Privacy-Enhanced

The Company is committed to protecting the privacy rights of individuals. For this reason, we prohibit employees from using personal information in AI systems. Instead, employees may work with appropriate teams at the Company to anonymize, deidentify or aggregate information before using it in AI.

IV.H. Accountability

The Company believes that our employees are essential to helping the organization remain accountable for the lawful and trustworthy use of AI. The Company is providing employees this Policy so that our organization can use AI systems in a manner that is compliant with AI regulations. It is important that employees follow this Policy because failure to do so could impact the Company's ability to comply with applicable laws. Employees are also ultimately accountable for the work product that they produce using AI. For this reason, employees will need to always confirm that the output is reliable and trustworthy.

V. Acceptable Use

Employees are permitted to use AI systems based on the below acceptable practices.

- If you use AI for authorized, company-related activities, you may use the Company-approved accounts created with work email addresses and credentials.
- Your use of AI must comply with this Policy and other related company policies.

- You may use AI systems for inspiration or generic output. The tools are less likely to produce infringing content when used as inspiration or to create generic output that can easily be replaced.
- You may only use supplier integrations or products featuring AI that have been approved by the Company.

VI. Prohibited Use of AI Systems

Employees are not permitted to use AI systems that are within prohibited categories. If you are uncertain whether your use of AI is prohibited, please contact our legal department before moving forward. These prohibited uses include, but are not limited to, the below.

- Using an AI system to subliminally change someone's behavior in a harmful manner;
- Using an AI system to take advantage of vulnerable groups like the elderly, disabled, or economically disadvantaged;
- Using an AI system to incite or encourage a person to commit physical self-harm (including suicide), harm another person, or engage in criminal activity;
- Using an AI system to infringe, restrict, or otherwise impair an individual's rights guaranteed under the United States Constitution;
- Using an AI system in connection with certain sexually explicit content and child pornography;
- Using an AI system to score people based on social behavior or personal characteristics;
- Using an AI system to build or expand facial recognition databases by collecting images from the internet or CCTV footage;
- Using an AI system for emotion recognition in the workplace;
- Using an AI system that employs biometric categorization to guess sensitive information about people, such as race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation;
- Using AI systems to predict crimes or discipline people;
- Using AI systems in a manner that may discriminate against jobs applicants, employees, and consumers based on protected classifications;
- Using AI systems to make credit decisions without clearly explaining why credit was denied or other negative actions were taken;
- Using AI systems to engage in unfair or deceptive practices, such as making misleading claims about costs, prices, products or services, tricking consumers to buy products or services, and omitting or hiding important limitations or conditions from an offer;
- Using confidential or personal information in an AI system;
- Using the output of an AI system without thoroughly reviewing it for accuracy and completeness;
- Using personal accounts with AI systems for Company-related purposes;
- Using AI systems to gain access to information that could be the IP of third parties;
- Using AI systems to create content or advertising that will be consumer-facing without approval from the Company; and
- Using AI systems without disabling settings that would permit the AI system provider to use the input data to train or improve their AI models.

VII. Certain AI Use Cases That Create Extra Risks

Certain AI use cases are permitted but come with additional risks. If you plan to use AI in the following ways, our legal team may need to take extra steps to ensure compliance. Please contact our legal department before proceeding.

These high-risk use cases include, but are not limited to, AI systems used in connection with:

- Biometric identification and surveillance (e.g., facial recognition technology);
- Providing individuals essential goods, services and benefits;
- Critical infrastructure;
- Education and vocational training;
- Employment and recruitment;
- Law enforcement and administration of justice;
- Immigration and border control;
- Healthcare services;
- Financial or lending services;
- Housing;
- Insurance;
- Legal services;
- Product safety component of certain regulated products; and
- Activities that may cause other legal or significant impact on individuals (i.e., important decisions).

VIII. Exercise of Rights

Under data privacy and AI laws, individuals, such as consumers, employees, and job applicants, have certain rights, such as the right to access, delete, correct, opt out of certain processing activities and decisions made with AI systems, and appeal decisions. If employees receive a rights request from individuals, employees will not respond to them directly. Rather, employees will direct individuals to our legal department so that they can submit the request through our designated methods, and appropriate employees responsible for handling privacy and AI rights requests can process the request.

IX. AI Vendors

The Company has approved only certain AI tools for employees to use in our organization. This is because the Company has to ensure that appropriate contractual provisions are in place with the AI vendors so that our organization remains protected and compliant with laws and regulations. Employees are not permitted to use other AI vendors who have not gone through our procurement process.

X. AI Systems Incident Reporting

If you have reason to believe that your or another employee's use of an AI system may have been subject to a data security or safety incident, please contact our information security team. The Company has incident response procedures related to such incidents that we must adhere to under such circumstances.

XI. Training

The Company may hold training sessions so that employees can learn how to use AI systems in compliance with applicable laws and regulations and in accordance with the Company’s policies, procedures and guidelines. All employees must attend these training sessions. Those responsible for coordinating the training sessions should maintain documentation reflecting who attended, date of the training, content of training, and materials provided to employees.

XII. Contact Information

If you have any questions about this Policy, contact our AI oversight committee.

XIII. Revision History

Version	Date	Author	Description	Sections Affected

Appendix F. Incorporating Artificial Intelligence Regulatory Requirements in Contracts

Under comprehensive artificial intelligence (AI) regulations, like the European Union AI Act (EU AI Act) and Colorado’s Concerning Consumer Protections in Interactions With AI Systems (Colorado AI Law), the provider (also known as the developer) and deployer of an AI system are subject to different obligations. The below chart provides an issues list for AI providers and deployers to consider when drafting AI agreements, which is mapped to each party’s compliance obligations under the EU AI Act and Colorado AI Law. The list below also includes a description of the mandatory contractual provision providers must include under the California AI Transparency Act when they license AI systems that generate content containing latent disclosures.

Issue	Provider	Deployer
Legal compliance	The provider would want the deployer to use the AI system in compliance with applicable laws, including avoiding prohibited AI practices.	The deployer would want the provider to have developed the AI system in compliance with applicable laws, including avoiding prohibited AI practices.
High-risk practices	The provider should assess whether its AI system is intended to be used for high-risk uses. If not, the provider should specify in the agreement that the AI system should not be used for high-risk practices. If yes, the provider should ensure that the agreement includes provisions that require the deployer to meet AI regulatory requirements for high-risk areas.	The deployer should assess whether it plans on using the AI system for high-risk uses. If yes, the deployer should ensure that the agreement includes provisions that require the provider to meet AI regulatory requirements for high-risk areas.
Data* <i>*Outside of the AI regulatory environment, the parties should consider their positioning under data privacy laws in connection with personal data. For example, if the provider is operating as a service provider or processor under data privacy laws, it may not be able to use the deployer’s personal data to train its AI model and service other customers. A data processing agreement between the parties will clarify party-role obligations vis-à-vis personal data. Further, the parties should consider ownership of data processed by the AI system and the output.</i>	The provider would want the deployer to ensure that it has the right to use the input data and the data is appropriate and representative for the intended use and audience.	The deployer would want the provider to ensure that it had the legal right to the data used to train, test and validate the AI system.
Procedural requirements and documentation	The provider would want to ensure that the deployer maintains logs automatically generated by the AI system.	The deployer would want to ensure that the provider complied with certain procedural requirements under the law, such as: (1) indicating the provider’s name and contact information on the AI system, the packaging or accompanying documentation; (2) having a quality management system; (3) maintaining technical and other documentation required by law; (4) keeping auto-generated logs; (5) undergoing a conformity assessment; (6) preparing a declaration of conformity; (7) affixing a CE marking on the AI system; and (8) complying with registration obligations.

Corrective actions	The provider would want to ensure that the deployer provides notice to the provider if corrective actions are necessary for the AI system.	The deployer would want to ensure that the provider has a process in place to take necessary corrective actions, or withdraw, disable, or recall the AI system.
Accessibility	The provider would want to ensure that the deployer takes measures to ensure that disabled end users have sufficient accessibility to the AI system.	The deployer would want to ensure that the provider's AI system complies with accessibility requirements under the law.
Risk management	The provider would want to ensure that the deployer has adopted an industry-standard risk management policy, such as NIST AI RMF or ISO/IEC 42001.	The deployer would want to ensure that the provider has adopted an industry-standard risk management policy, such as NIST AI RMF or ISO/IEC 42001.
Transparency, explainability and instructions of use	The provider would want to ensure that the deployer follows its instructions for use. The provider would also want to ensure that the deployer provides individuals an appropriate transparency and explainability pre-use notice and any notices required if there is an adverse decision regarding an individual. If required by applicable law, the provider would want the deployer to include appropriate disclosures on its website regarding its use of the AI system.	The deployer would want to ensure that the provider gave it appropriate instructions of use for the AI system and the information the deployer needs to comply with its own obligations. The deployer would want to ensure that the provider developed the AI system so that, when prompted, it indicates to end users that they are interacting with AI and not a human being. If required by applicable law, the deployer would want the provider to include appropriate disclosures on its website regarding the AI system it offers.
Human oversight and monitoring	The provider would want to ensure that the deployer has adopted appropriate human oversight to prevent or minimize AI risks and monitor the AI system after it has been deployed.	The deployer would want to ensure that the provider has adopted appropriate human oversight to prevent or minimize AI risks and monitor the AI system after it has been deployed.
Accuracy, robustness, and cybersecurity	The provider would want to ensure that the deployer provides it notice if the AI system is providing inaccurate outputs, is not functioning as intended or if there are security incidents or concerns.	The deployer would want to ensure that the provider developed the AI system with suitable accuracy, robustness and cybersecurity measures.
Algorithmic discrimination and bias	The provider would want to ensure that the deployer has appropriate policies, procedures and practices to avoid algorithmic discrimination and bias.	The deployer would want to ensure that the provider developed the AI system to avoid algorithmic discrimination and bias.
AI impact assessment	The provider would want to ensure that the deployer prepares an AI impact assessment before using the AI system.	The deployer would want to ensure that the provider prepared an AI impact assessment for its AI system and gave the deployer the information the deployer needs to prepare an AI impact assessment.
AI rights	The provider would want to ensure that the deployer is responsible for honoring AI rights that may be available to individuals, such as opting out of automated decisions, appealing adverse decisions, and other rights they may have under AI and related data privacy laws (e.g., right to correct, delete, and access personal data used in connection with AI).	The deployer would want to ensure that the provider can assist the deployer so that it is capable of honoring AI and related data privacy rights.
AI incidents	The provider would want to ensure that the deployer has policies, procedures and practices in place to provide notice and/or report AI security, safety, algorithmic discrimination and other incidents.	The deployer would want to ensure that the provider has policies, procedures and practices in place to provide notice and/or report AI security, safety, algorithmic discrimination and other incidents.

Cooperation with regulators	The provider would want to ensure that the deployer will cooperate with the provider if there are regulatory inquiries regarding the AI system.	The deployer would want to ensure that the provider will cooperate with the deployer if there are regulatory inquiries regarding the AI system.
Shifting of party roles	The provider should assess whether it wants to permit the deployer to take actions (e.g., by placing the deployer's trademark or name on the AI system, making substantial modifications to the AI system, or modifying the intended purpose of the AI system) that would shift the deployer's role to that of a provider. If party-role shifting is permissible, the provider would want to clarify the type of cooperation, information, and assistance it will provide to the counterparty for the new provider to be able to comply with provider obligations.	The deployer should assess whether it may take steps that would make it a provider under AI regulations (e.g., by placing its trademark or name on the AI system, making substantial modifications to the AI system, or modifying the intended purpose of the AI system). If deployer contemplates this occurring, it would want to specify in the contract the type of cooperation, information and assistance the initial provider will offer it to be able to comply with provider obligations.
Prohibition against tampering with latent disclosures	If the provider offers a generative AI system that contains a latent disclosure in AI-generated image, video, or audio content, the provider should require the deployer to maintain the AI system's capability to include the disclosure. The provider should include a contract provision requiring the deployer to provide notice within 96 hours to provider if the deployer modified the AI system such that it is not capable of including the latent disclosure. The provision should permit the provider to revoke the license it provided to the deployer if this occurs.	The deployer would want a provision in the contract requiring the provider to include appropriate disclosures in AI-generated content and provide an AI-detection tool, if required by law.

