Toolkit Overview

AI Governance Toolkit

*This toolkit is designed as an interlocking set of artifacts that work together to effectively **map**, **measure**, and **manage** AI risks across an organization. Tied together, these lightweight components provide a clear operating model for an effective AI governance program across an organization.*

Author: Taylor Galusha, Privacy & AI Counsel at Chime Financial [AIGP, CIPP/US, CIPP/E, CIPM, CIPT, FIP]

***Legal Disclosure:** This toolkit is provided for general informational and educational purposes only and does not constitute legal advice or a legal opinion. Use of these materials does not create an attorney-client relationship. Any views expressed are our own and do not necessarily reflect those of our employers or any affiliated parties. Sample materials are illustrative and must be tailored to your specific organizations and applicable law with advice from your own counsel. Nothing here is intended as a solicitation for legal services or as advice in any jurisdiction.*

Toolkit Components

1. 3 Core Components of AI Governance
2. AI Governance Policy
3. AI Governance Training
4. AI Do's and Don'ts
5. AI DPA Clauses
6. AI Assessment (internal)
7. AI Assessment (vendor)
8. AI Risk Management Tracker

3 Core Components of AI Governance

3 Core Components of AI Governance

*A practical and lightweight AI governance framework should be organized around three core functions: (1) **mapping** where and how AI is used across the company (including both vendor-provided and internally built use cases), (2) **measuring** the risks that flow from those specific use cases, and (3) **managing** those risks through clear decisions and controls—mitigation, elimination, or risk acceptance—based on the organization's risk tolerance.*

In that sense, the toolkit's components are meant to interlock rather than operate as standalone paperwork: the tracking and intake mechanisms help you map your AI footprint; the internal and vendor AI assessments and risk tracker help you measure likelihood and impact in a consistent way; and the policy, trainings, do's and don'ts, and contract clauses help you manage risk by translating assessment outcomes into enforceable requirements and day-to-day behaviors.

---

1. **Map the AI use:**
    - Purpose: Map AI Use Cases and risks to make informed decisions about what AI to build and deploy, how to use it, and what guardrails should be in place. Failure to adequately map AI usage will render it impossible to measure or manage AI risk.
    - Key Components:
        i. **Procurement and/or Third Party Risk AI Vendor Tracking:** Must appropriately identify and tag which vendors are providing AI services to the organization. Notably, this is not a responsibility of the AI governance team alone. Strong cross functional buy-in from the teams responsible for vendor onboarding is required.
        ii. **Product Led AI Product Tracking**: Must appropriately identify and tag which products and services are being developed and provided to customers and/or clients. Notably, this is not a responsibility of the AI governance team alone. Strong cross functional buy-in from the teams responsible for developing your organizations AI products and services is required.
2. **Measure the AI risks:**
    - Purpose: AI risks are measured so that the organization can effectively prioritize and manage those AI risks. AI risks are assessed on the likelihood of the risk occurring and the impact of the risk if it materializes. Some risks may be too high to accept while others can be mitigated and reduced to an acceptable level.
    - Key Components:
        i. **AI Risk Management Tracker**: Must measure the risk of all AI use cases in order to enable informed decisions about what AI to deploy, how to use it, and what guardrails should be in place.

ii. **AI Assessment (internal)**: Must measure the risk of all internally built AI tools and services through an impact assessment process customized to the organization's activities and intended to produce sufficient information to properly manage the AI risks.
iii. **AI Assessment (vendor)**: Must measure the risk of all vendor provided AI tools and services through an impact assessment process and intended to produce sufficient information to properly manage the AI risks.

3. **Manage the AI risks:**
   - Purpose: AI risks must be managed by an organization adequately through mitigation, elimination, or risk acceptance. The risk associated with different AI products and services can vary widely and its decisions about what AI risks to mitigate, eliminate, or accept should be based on the level of risk documented in the “measure” stage of the AI governance program.
   - Key Components:
     i. **AI Governance Policy**: Sets the governing principles, standards, and requirements for the use, development, and deployment of AI consistent with legal, compliance, risk, and reputation management. The policy should describe the organization's responsible AI mission statement, guiding governance principles, risk management framework, and the roles and responsibilities necessary to operationalize the framework.
     ii. **AI Governance Training**: Exists to translate the company’s AI Governance Policy into day-to-day practice. Its purpose is to ensure that employees understand how to responsibly design, use, procure, and oversee AI in a way that is consistent with the organization’s legal, compliance, risk, and reputational obligations.
     iii. **AI Do’s and Don’ts**: Simple and digestible tips for how to use and not use AI within the organization. Intended to serve as the front door to the AI Governance Program, translating dense policy requirements into practical, everyday guidance that employees can use.
     iv. **AI DPA Clauses**: A contractual control layer that governs how a vendor may process data in connection with AI systems. While a traditional DPA focuses on privacy, security, and regulatory compliance for data processing, an AI DPA (or simply AI provisions added to an existing DPA) extends those obligations to address AI-specific risks such as model training, output ownership, explainability, bias, autonomous behavior, and unacceptable AI uses.
     v. **AI Assessments Outcomes (internal & vendor)**: The outcomes, recommendations, and business decisions made during the AI Assessments (vendor and internal) are intended to help the company manage AI risk adequately through mitigation, elimination, or risk acceptance. Importantly, documenting mitigation measures and risk acceptances is more important than ever given the generative nature of AI. Whereas past governance programs avoided documenting risks out of fear of litigation, modern programs understand the generative and

always-present risks of AI being wrong and know how to document their mitigation strategies and risk acceptance rationales.

AI Governance Policy

AI Governance Policy

This document is a lightweight AI Governance Policy template designed for legal and privacy professionals. It requires adaptation for an organization's specific size, risk profile, and jurisdictional obligations (e.g., EU AI Act, US state laws). It sets the governing principles, standards, and requirements for the use, development, and deployment of AI consistent with legal, compliance, risk, and reputation management.

Version: 1.0

Effective Date:

Policy Owner: AI Governance Lead

Approver: Executive AI Steering Committee / CEO / Board-Delegated Risk Committee

Review Cadence: At least annually, and upon material legal or technology changes

Version Control

Version	Date	Owner	Summary of Changes
1.0		AI Governance Lead	Initial AI Governance Policy

---

1. Overview and Policy Purpose

[Company Name] ("Company") is committed to the responsible development, procurement, and deployment of Artificial Intelligence (AI) to enhance innovation while protecting fundamental rights, privacy, and safety.

This Policy establishes:

- **Governance Structure:** Clearly defined roles, decision rights, and accountability frameworks.
- **Risk-Based AI Management:** A structured "Map, Measure, Manage" process to identify and mitigate AI-specific risks.
- **Compliance Alignment:** A baseline intended to meet or exceed global standards, including the NIST AI Risk Management Framework (RMF), ISO/IEC 42001 (AIMS), and the EU AI Act.

2. Scope, Applicability, and Definitions

2.1 Scope

This Policy applies to all employees, contractors, and third-party agents of the Company. It covers all AI Systems and AI Use Cases that are:

- Developed internally or fine-tuned on Company data.
- Procured or licensed from third-party vendors (SaaS, APIs, or open-source).
- Embedded in existing enterprise tools or customer-facing products.

2.2 Definitions

- **Artificial Intelligence (AI):** A machine-based system that, for a given set of human-defined objectives, makes predictions, recommendations, or decisions influencing real or virtual environments.
- **AI System:** The software, hardware, and data components (e.g., ML models, generative agents) used to produce outputs for a specific context.
- **AI Use Case:** A specific intended use of an AI System, defined by its user population, data inputs, and intended business outcome.
- **High-Risk AI:** Systems with the potential to materially impact human rights, health, safety, or significant economic interests (e.g., hiring, credit scoring, critical infrastructure).

---

3. Roles and Responsibilities

3.1 Governance Bodies

- **Executive Steering Committee:** Senior leadership (C-Suite) responsible for setting risk appetite, allocating resources, and approving high-risk AI use cases.
- **AI Governance Committee (AIGC):** A cross-functional group (Legal, Privacy, Security, Engineering, Product) that reviews AI Use Cases and manages the AI Use Case Inventory.
- **AI Governance Lead:** An IAPP AIGP-certified professional responsible for the day-to-day operation of the AI Governance Program.

3.2 Key Operational Roles

- **AI Use Case Owner (Business Lead):** Accountable for registering the use case, completing the AI Impact Assessment (AIA), and ensuring ongoing monitoring.
- **Technical Owner (Engineering/DS):** Responsible for technical validation, bias testing, and instrumentation for drift monitoring.
- **Data Legal/Privacy:** Responsible for AI Data Processing Addendums (AI DPA), fundamental rights impact assessments (FRIA), and ensuring lawful basis for training.

---

4. Policy Requirements

4.1 Responsible AI Mission Statement

The Company uses AI to empower users and drive progress while maintaining human-centricity, safety, and accountability in every AI tool and service we deploy.

4.2 Guiding Governance Principles

1. **Human Agency and Oversight:** AI should support, not replace, human judgment. Meaningful human-in-the-loop (HITL) controls are required for high-stakes decisions.
2. **Fairness and Non-Discrimination:** Systems must be tested for algorithmic bias and disparate impact on protected classes.
3. **Transparency and Explainability:** Users must be notified when interacting with AI, and significant decisions must be explainable.
4. **Security and Robustness:** AI must be resilient against adversarial attacks (e.g., prompt injection) and maintain accuracy throughout its lifecycle.

4.3 AI Risk Management Framework

The Company adopts the NIST AI RMF core functions:

- **GOVERN:** Establishing the culture of responsible AI use through this policy and other AI governance components, which requires the organization:
    - **MAP:** Maintaining a comprehensive AI Use Case Inventory and identifying context-specific risks.
    - **MEASURE:** Quantifying risks using a standardized likelihood vs. impact matrix.
    - **MANAGE:** Implementing controls, mitigations, or decommissioning systems that exceed risk tolerance.

---

5. Operational Procedures

5.1 Mandatory AI Inventory

Every AI Use Case must be registered in the centralized AI Use Case Inventory prior to pilot or production.

5.2 Risk Tiering Structure

Tier	Classification	Governance Requirement
Tier 0	**Prohibited**	**Banned.** Includes social scoring, untargeted facial scraping, and manipulative behavior distortion.

Tier 1	**High-Risk**	**AIGC Approval Required.** Full AIA, technical documentation, and mandatory human oversight.
Tier 2	**Moderate-Risk**	**Privacy/Legal Review.** Streamlined AIA and baseline transparency disclosures.
Tier 3	**Minimal-Risk**	**Inventory Registration.** Adherence to baseline security and code of conduct.

5.3 AI Impact Assessment (AIA)

For T1 and T2 systems, the Use Case Owner must complete an AIA covering:

- **Intended Purpose:** Context and population affected.
- **Data Governance:** Data source, quality, and minimization of sensitive data use risks.
- **Technical Safeguards:** Accuracy benchmarks, drift thresholds, and red-teaming results.
- **Human Oversight:** Specific procedures for human review and the ability to stop the line.

5.4 Procurement and Third-Party AI

The Company shall not procure AI services without an AI Data Processing Addendum (AI DPA) or equivalent clauses addressing:

- Restrictions on the vendor using Company data to train their foundation models.
- Warranties for bias testing and technical robustness.
- Right to audit the vendor's AI governance practices.

---

6. Monitoring, Testing, Audit, and Metrics

- **Post-Market Monitoring:** High-Risk systems must have active monitoring for model drift and performance degradation.
- **Independent Audit:** Periodic reviews by Internal Audit to ensure adherence to the AIMS and this Policy.
- **KPIs:** Percentage of registered use cases, AIA completion rates, and time-to-remediate AI incidents.

---

7. Training and Communication

- **AI Governance Training:** Mandatory for all staff, including dissemination of the Company's "AI Do's and Don'ts."

- **Specialized Training:** Role-based training for Engineers (TEVV standards) and Business Owners (AIA completion).

---

8. Exceptions

Exceptions to this Policy must be documented, include compensating controls, and receive AIGC approval for any T1 or T2 system.

---

9. Policy Administration

- **Sanctions:** Violations of this Policy may result in the suspension of the AI System and disciplinary action.
- **Retention:** All AIA and validation records must be retained for at least 10 years (aligned with EU AI Act requirements).

AI Governance Training

AI Governance Training Outline

Training is a core component of any AI governance program. Training should be tailored to the organization and/or delivered through a third-party provider. So treat the slide deck below as a sample reference, not a one-size-fits-all template.

To make tailoring easy, we've included a slide-by-slide outline you can paste into your AI slide-generation tool to produce your own deck quickly. And yes: this is the future of slide creation—fewer late nights fiddling with slide layouts, and a lot fewer painfully boring decks.

What is AI Governance?

Defining AI Governance

A System of Rules & Practices
Ensuring AI is developed and deployed ethically, responsibly, and in compliance with regulations.

Guiding Principles
A framework to manage risks while maximizing business benefits.

From Policy to Practice
Translating high-level principles into actionable steps for developers and leaders.

Why AI Governance Matters

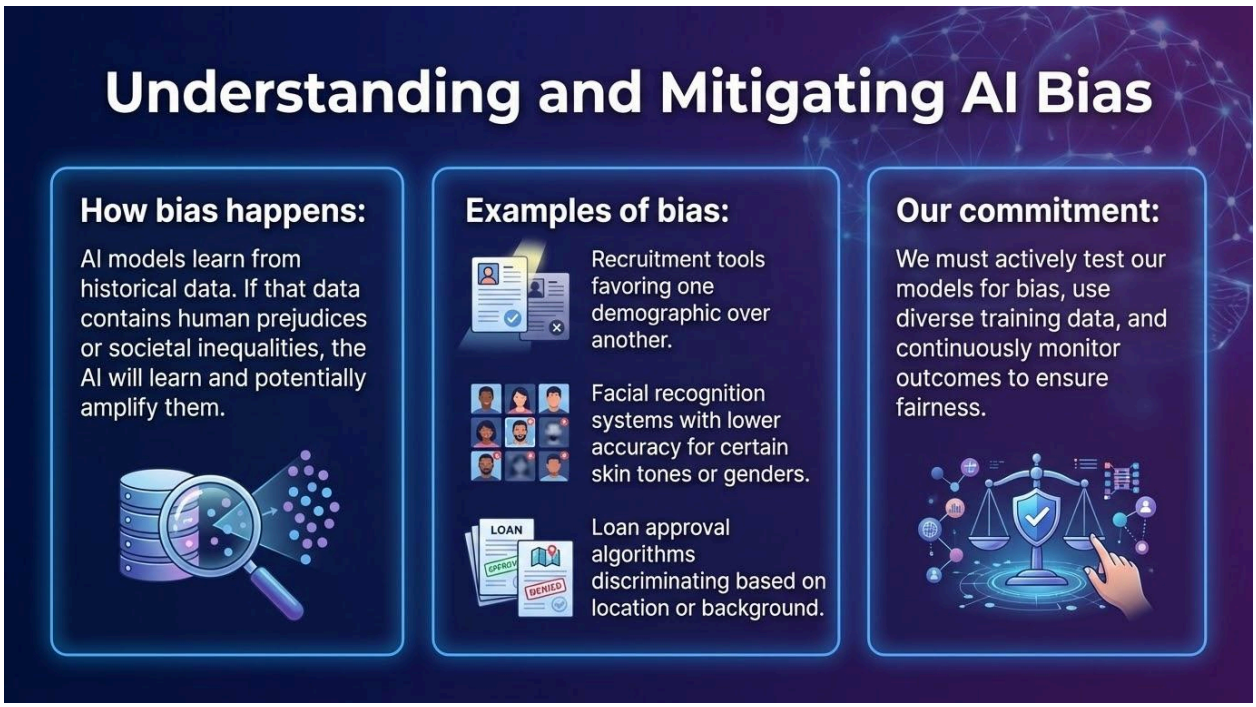
The Business Case for Governance

Builds Trust:
Demonstrating responsible AI use fosters trust with customers, employees, and stakeholders.

Mitigates Risk:
Proactive governance helps prevent legal fines, regulatory penalties, and costly lawsuits resulting from non-compliance or harmful outcomes.

Protects Reputation:
Prevents public relations disasters caused by biased, offensive, or unsafe AI behavior.

Ensures Ethical Alignment: Guarantees that our use of AI reflects our company's core values and mission.



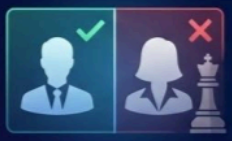
Real-World Failure Case Study: Hiring Bias

Case Study: When AI Recruitment Goes Wrong

The Scenario

A major tech company developed an AI tool to automate the review of job applicants' resumes.

The Failure



The model was trained on resumes submitted over a 10-year period, which came mostly from men. As a result, the AI taught itself to prefer male candidates. It penalized resumes that included the word "women's," such as references to a "women's chess club."

The Consequence

The project had to be abandoned after the bias was discovered, causing reputational damage and wasted resources.

The Lesson

Bias in training data leads to biased outcomes. Continuous fairness audits and human oversight are essential.

Data Privacy & Security in the Age of AI

Protecting Data in an AI World

The Fuel for AI: AI models require vast amounts of data to learn and operate.

Key Risks:

Data Breaches: Large datasets are attractive targets for cyberattacks.

Data Leakage: Using public AI tools can inadvertently expose sensitive information to third parties.

Unauthorized Use: Using personal data for AI training without proper consent violates privacy laws like GDPR.

The Golden Rule: Never input confidential company information, Personally Identifiable Information (PII), or trade secrets into unapproved, publicly available AI tools (e.g., standard ChatGPT, Midjourney).

Generative AI: Employee Do's and Don'ts

Practical Guide for Using Generative AI at Work

DO:

- Use company-approved AI tools to draft emails, summarize documents, and brainstorm ideas to improve productivity.
- Always verify and fact-check AI-generated output. AI can "hallucinate" incorrect information. You are responsible for your final work.
- Be transparent when significant content is created by AI.

DON'T:

- Share sensitive, confidential, or proprietary company data with public AI models.
- Rely solely on AI for critical decisions without human review ("human-in-the-loop").
- Use AI to create content that is discriminatory, offensive, or violates copyright.

Roles & Responsibilities

Who is Responsible for AI Governance?

Everyone:

Every employee is responsible for using AI tools according to company policy and ethical guidelines.

AI Governance Committee:

Sets overall policies, reviews high-risk use cases, and ensures strategic alignment.

Data Scientists & Engineers:

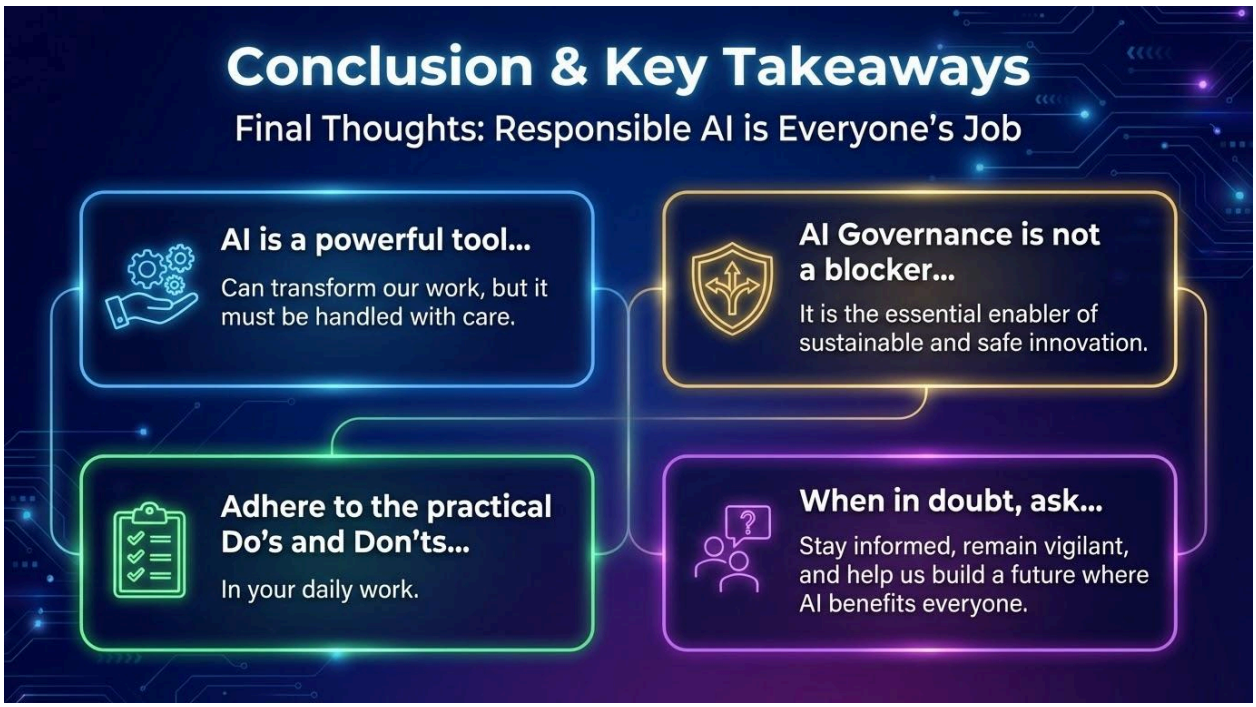
Responsible for building models that are fair, secure, transparent, and documented.

Legal & Compliance:

Ensures all AI initiatives comply with current laws and regulations.

Managers:

Oversee their team's AI usage, ensure proper training, and serve as a first point of contact for questions.



Below you will find the outline of the AI Governance Training posted above so that you can adapt and customize to your business.

Slide 1: Title Slide

Header: Corporate AI Governance Training

Subheader: Navigating the Future of AI Responsibly, Ethically, and Securely

Slide 2: What is AI Governance?

Header: Defining AI Governance

Content:

- **It's not just about technology.** AI Governance is a system of rules, practices, and processes used to ensure that artificial intelligence is developed and deployed ethically, responsibly, and in compliance with regulations.
- **Guiding Principles:** It provides the framework to manage the risks associated with AI while maximizing its business benefits.
- **From Policy to Practice:** It translates high-level principles like "fairness" into concrete actions, checks, and balances within our daily workflows.

Slide 3: Why AI Governance Matters

Header: The Business Case for Governance

Content:

- **Builds Trust:** Demonstrating responsible AI use fosters trust with customers, employees, and stakeholders.
- **Mitigates Risk:** Proactive governance helps prevent legal fines, regulatory penalties, and costly lawsuits resulting from non-compliance or harmful outcomes.
- **Protects Reputation:** Prevents public relations disasters caused by biased, offensive, or unsafe AI behavior.
- **Ensures Ethical Alignment:** Guarantees that our use of AI reflects our company's core values and mission.

Slide 4: Key Pillars of Our AI Governance Framework

Header: Our Foundation for Responsible AI

Content:

- **Accountability:** Clear ownership is defined for every AI system. There is always a "human-in-the-loop" responsible for its outcomes.
- **Fairness & Equity:** We actively work to identify and mitigate bias in our data and models to ensure fair outcomes for everyone.
- **Transparency & Explainability:** We strive to understand how our AI models make decisions and can explain them in plain language.
- **Data Privacy & Security:** Protecting sensitive data is paramount. We adhere to strict protocols for data handling used in AI.
- **Risk Management:** We proactively identify, assess, and mitigate AI-related risks throughout the entire product lifecycle.

Slide 5: Ethical Considerations: The Bias Problem

Header: Understanding and Mitigating AI Bias

Content:

- **How bias happens:** AI models learn from historical data. If that data contains human prejudices or societal inequalities, the AI will learn and potentially amplify them.
- **Examples of bias:**
    - Recruitment tools favoring one demographic over another.
    - Facial recognition systems with lower accuracy for certain skin tones or genders.
    - Loan approval algorithms discriminating based on location or background.
- **Our commitment:** We must actively test our models for bias, use diverse training data, and continuously monitor outcomes to ensure fairness.

Slide 6: Real-World Failure Case Study: Hiring Bias

Header: Case Study: When AI Recruitment Goes Wrong

Content:

- **The Scenario:** A major tech company developed an AI tool to automate the review of job applicants' resumes.
- **The Failure:** The model was trained on resumes submitted over a 10-year period, which came mostly from men. As a result, the AI taught itself to prefer male candidates. It penalized resumes that included the word "women's," such as references to a "women's chess club."
- **The Consequence:** The project had to be abandoned after the bias was discovered, causing reputational damage and wasted resources.
- **The Lesson:** Bias in training data leads to biased outcomes. Continuous fairness audits and human oversight are essential.

Slide 7: Data Privacy & Security in the Age of AI

Header: Protecting Data in an AI World

Content:

- **The Fuel for AI:** AI models require vast amounts of data to learn and operate.
- **Key Risks:**
    - **Data Breaches:** Large datasets are attractive targets for cyberattacks.
    - **Data Leakage:** Using public AI tools can inadvertently expose sensitive information to third parties.
    - **Unauthorized Use:** Using personal data for AI training without proper consent violates privacy laws like GDPR.
- **The Golden Rule:** Never input confidential company information, Personally Identifiable Information (PII), or trade secrets into unapproved, publicly available AI tools (e.g., standard ChatGPT, Midjourney).

Slide 8: Generative AI: Employee Do's and Don'ts

Header: Practical Guide for Using Generative AI at Work

Content:

- **DO:**
    - Use company-approved AI tools to draft emails, summarize documents, and brainstorm ideas to improve productivity.
    - Always verify and fact-check AI-generated output. AI can "hallucinate" incorrect information. You are responsible for your final work.
    - Be transparent when significant content is created by AI.
- **DON'T:**
    - Share sensitive, confidential, or proprietary company data with public AI models.
    - Rely solely on AI for critical decisions without human review ("human-in-the-loop").
    - Use AI to create content that is discriminatory, offensive, or violates copyright.

Slide 9: Roles & Responsibilities

Header: Who is Responsible for AI Governance?

Content:

- **Everyone:** Every employee is responsible for using AI tools according to company policy and ethical guidelines.
- **AI Governance Committee:** Sets overall policies, reviews high-risk use cases, and ensures strategic alignment.
- **Data Scientists & Engineers:** Responsible for building models that are fair, secure, transparent, and documented.

- **Legal & Compliance:** Ensures all AI initiatives comply with current laws and regulations.
- **Managers:** Oversee their team's AI usage, ensure proper training, and serve as a first point of contact for questions.

Slide 10: How to Spot and Report AI Risks

Header: Be Our Eyes and Ears: Reporting AI Risks

Content:

- **What to Spot:**
    - **Biased Outcomes:** AI decisions that seem unfair to a particular group.
    - **Inaccurate Information:** An AI tool consistently providing wrong or misleading answers.
    - **Unexpected Behavior:** An AI system acting in an erratic or unexplainable way.
    - **Data Concerns:** Potential misuse of sensitive data or privacy violations.
- **How to Report:** Don't ignore it. If you see something, say something. Use our designated reporting channel \[Insert Link/Email Here\] to flag potential issues.
- **Our Culture:** We encourage a culture of open communication. Reporting risks is a positive action that helps us improve and protect the company.

Slide 11: Conclusion & Key Takeaways

Header: Final Thoughts: Responsible AI is Everyone's Job

Content:

- AI is a powerful tool that can transform our work, but it must be handled with care.
- AI Governance is not a blocker; it is the essential enabler of sustainable and safe innovation.
- Adhere to the practical Do's and Don'ts in your daily work.
- When in doubt, ask. Stay informed, remain vigilant, and help us build a future where AI benefits everyone.

AI Do’s and Don’ts

🤖 AI Do’s and Don’ts

This guide outlines the "Rules of the Road" for using Artificial Intelligence at our organization. Whether you are using a public LLM to summarize a meeting or building a custom machine learning model, these guardrails apply.

✅ The "Do’s": How to Innovate Safely

- **Use Sanctioned Tools Only:** Only use AI tools and platforms that have been vetted and approved by IT, Security, and Legal. If it’s not on the Approved AI Registry, don't put company data into it.
- **Register Your Use Case:** Before deploying a new AI-backed feature or workflow, submit it through the AI Intake Process. We need to track *how* we use AI to comply with emerging global regulations.
- **Maintain a "Human-in-the-Loop":** Treat AI output as a "first draft" only. You are ultimately responsible for the accuracy of your work. Always review, verify, and fact-check AI-generated content before it is finalized or shared.
- **Practice Data Minimization:** Only provide the AI with the minimum amount of information necessary to complete the task. Use synthetic or anonymized data whenever possible.
- **Disclose AI Interaction:** Be transparent with our customers and partners. If they are interacting with a chatbot or viewing AI-generated content, ensure it is clearly labeled as such.
- **Be Mindful of Bias:** Proactively check AI outputs for discriminatory language or biased results. If the AI seems to be "hallucinating" or showing prejudice, stop use and report it to the Governance team.

❌ The "Don’ts": What to Avoid

- **No PII or Sensitive Data:** Never upload Personally Identifiable Information (PII), Protected Health Information (PHI), or customer passwords into public or "unmanaged" AI tools (like the free version of ChatGPT).
- **Protect our Intellectual Property:** Do not input unreleased source code, trade secrets, or confidential product roadmaps into public AI models. These inputs may be used to train future versions of the model, effectively "leaking" our IP.

- **Don't "Set it and Forget it":** AI is not a static tool. Don't assume a model that worked yesterday will be accurate today. Continuous monitoring is required for high-risk applications.
- **No Legal or Medical Advice:** Do not use AI to generate definitive legal, medical, or financial advice for the company or our users without expert human oversight.
- **Avoid "Shadow AI":** Do not sign up for new AI "browser extensions" or "productivity plugins" using your corporate email without a formal security review. These are common vectors for data exfiltration.
- **Don't Ignore Copyright Warnings:** Be cautious when using AI to generate images or code for commercial use. Ensure we have the necessary rights to the output to avoid potential copyright infringement claims.

🛑 When in Doubt: The "Pause & Consult" Rule

If you are unsure if your use case is high-risk, ask yourself these three questions:

1. Does this AI make a significant decision about a human being (e.g., hiring, credit, access to services)?
2. Does this involve sensitive categories of data (e.g., race, religion, health)?
3. Is this output customer-facing without a human reviewer?

If the answer to any of these is "Yes," contact the AI Governance Team before proceeding.

AI DPA Clauses

AI DPA Clauses

The following AI DPA model clauses are designed for inclusion in a Master Services Agreement (MSA) or Data Processing Addendum (DPA) when engaging with AI vendors. These clauses should be adapted to your organization and adjusted based on your unique risk tolerance and commercial contracting process.

1. Values-Based Principles & Compliance

The Provider shall develop and implement AI Systems that adhere to broadly recognized ethical frameworks, such as the OECD AI Principles. At a minimum, these systems must prioritize:

- **Inclusive Growth & Well-being:** Benefits for people and the planet.
- **Human Rights & Fairness:** Mitigating bias and protecting democratic values.
- **Robustness & Safety:** Ensuring systems are secure and function reliably throughout their lifecycle.

2. Prohibited "Unacceptable Risk" Activities

The Provider represents and warrants that the AI System is not designed for, and shall not be used for, activities deemed to pose an "Unacceptable Risk," including:

- **Deceptive Techniques:** Using subliminal or manipulative tactics to distort behavior.
- **Social Scoring:** Evaluating or classifying individuals based on social behavior or predicted personality traits.
- **Biometric Scraping:** Untargeted scraping of facial images from the internet or CCTV.
- **Workplace Emotion Inference:** Using AI to infer emotions in a workplace or educational setting.

3. Transparency & "Explainability" Rights

Upon request, the Provider must be able to provide a meaningful explanation of the AI System's logic. This includes:

- **Development Methodology:** Descriptions of the data selection process, assumptions made during training, and categories of data used.
- **Output Rationale:** Explanations of the key factors that led to a specific output and how variables can be adjusted to arrive at different results.
- **Risk Documentation:** Providing a record of identified risks and the measures implemented to manage them.

4. Ownership & Model Training Restrictions

- **Input/Output Ownership:** The Company retains all right, title, and interest in all "Inputs" provided to the AI and all "Outputs" generated by the AI based on those Inputs.
- **No Training:** The Provider is strictly prohibited from using Company data, Inputs, or Outputs to train, fine-tune, or improve its AI models for any third-party use or for the Provider's general commercial gain.
- **IP Non-Infringement:** The Provider warrants that the training of the AI model and the subsequent use of the system do not infringe upon or misappropriate the intellectual property rights of any third party.

5. AI-Specific Data Quality & Security

- **Bias Mitigation:** Provider shall structure and analyze training data to minimize inaccuracies, errors, and socially constructed bias.
- **Logging & Retention:** AI-specific logs (including prompts and responses) shall not be stored for longer than necessary to perform the Services.
- **Feature Notification:** Provider must notify the Company in writing prior to introducing new AI features or capacities that significantly impact the functionality of the system

AI Assessment (internal)

AI Assessment (Internal AIA)

The AI Assessment (Internal) is the organization's standardized impact assessment for internally built or internally configured AI use cases. Its purpose is to ensure every internal AI system is mapped, risk-tiered, and reviewed before pilot or production (and re-reviewed after material changes), so the company can make an informed launch decision and implement the right safeguards (e.g., human oversight, privacy and data minimization, security controls, bias testing, monitoring, and "stop-the-line" controls). The assessment produces the documentation and decisions needed to comply with the AI Governance Policy's approach and to operationalize risk management through clear ownership, mitigations/conditions, and escalation paths.

Who completes this / when it's used

- **Owner:** Business Owner (with support from the Technical Owner).
- **When:** Before pilot or production, and again after any material change (new model, new data, new user population, new decisioning, new automation, etc.).

---

1) Intake

Process / Use Case Name:
Requester Name + Team:
Business Owner:
Technical Owner (Eng/DS):
Product / System / Workflow:
Current Stage: ☐ Idea ☐ Pilot/POC ☐ Production
Intended users: ☐ Internal ☐ Customer-facing ☐ Partner-facing
Brief description (2–4 sentences):

---

2) Model + system details

1. **Model name/provider/version** enabling this use case (or explain if not applicable):
2. **How the model is used:** ☐ Classification/Scoring ☐ Ranking ☐ Summarization ☐ Generation ☐ Agent/Automation ☐ Other:
3. **Where it runs / is hosted:** ☐ SaaS ☐ Cloud-hosted ☐ On-prem ☐ Other:

4. **Key inputs + outputs:** What goes in, what comes out, and how outputs are consumed downstream.

---

3) High-stakes decisioning screen

1. **Confirm this AI will not be used to take adverse action or make (or materially influence) consequential decisions** about an individual's access to financial services, credit, employment, housing, or government benefits.
   ☐ Confirm ☐ Not true
   - If not true: describe the decision, impacted population, safeguards, and escalation path.
2. **Human-in-the-loop:** Confirm outputs are reviewed by a human before any high-impact action is taken.
   ☐ Confirm ☐ Not true
   - If not true: describe safeguards (e.g., thresholds, failsafes, dual controls, audit logging) and why human review isn't feasible.

---

4) Accountability + remediation

1. **Single-threaded owner:** Who is responsible for the AI output and remediation if something goes wrong (name/role/team):
2. **Incident response:** Confirm there is a process to detect issues, triage, remediate harm, and document actions taken.
   ☐ Confirm ☐ Not true
   - If not true: explain the gap and proposed plan.
3. **Kill switch / rollback:** Confirm there is an operational "stop-the-line" control to disable the feature/model quickly.
   ☐ Confirm ☐ Not true
   - If not true: explain how the team will stop harm in production.

---

5) Bias, fairness, and hallucinations

1. **Confirm the AI will not be biased, violate anti-discrimination laws, or produce hallucinations that may harm users**, especially members of protected classes.
   ☐ Confirm ☐ Not true / Unknown
   - If not true/unknown: describe likely failure modes, planned testing (metrics + slices), and mitigation plan.

2. **Testing plan**
    - What you'll test (quality + safety + bias):
    - What data you'll use (representativeness):
    - Pass/fail thresholds:
    - Who signs off:

---

6) User recourse: correction / appeal

1. **Confirm there are mechanisms to correct, challenge, or appeal material errors** in the AI output (customer-facing or internal operator workflows, as applicable).
   ☐ Confirm ☐ Not true
    - If not true: explain why recourse is not applicable and what alternative control exists (e.g., monitoring + proactive remediation).

---

7) Safety risks + robustness

1. **Potential harms:** Describe realistic harms (financial harm, privacy harm, discrimination, security abuse, misinformation, operational risk).
2. **Mitigations:** What controls prevent/detect/respond to those harms.
3. **Adversarial robustness:** Confirm the system can handle noisy/corrupted inputs and adversarial attacks relevant to the use case (e.g., prompt injection, data exfiltration attempts).
   ☐ Confirm ☐ Not true
    - If not true: describe compensating controls and launch conditions.

---

8) Privacy + data protection

1. **Data categories used:** Personal data? Sensitive data? Internal confidential data? Customer support content? Transactional metadata?
2. **Data minimization:** Confirm the use case uses only the minimum necessary data.
   ☐ Confirm ☐ Not true
    - If not true: explain what's additional and why it's needed.
3. **Security controls:** Confirm appropriate safeguards exist (access controls, logging, encryption in transit/at rest, environment segregation).
   ☐ Confirm ☐ Not true
    - If not true: explain.

4. **Data retention + deletion:** How long are inputs/outputs/logs retained, and how are they deleted?
5. **Data separation (if applicable):** Confirm company data is not exposed or reused in a way that creates leakage or training risk.
   ☐ Confirm ☐ Not true / Unknown
   - If not true/unknown: explain what you know and what controls mitigate the risk.

---

9) IP and third-party rights

1. **Confirm the use case is designed to avoid infringing third-party rights**, and that appropriate guardrails exist (especially for generated content/code).
   ☐ Confirm ☐ Not true / Unknown
   - If not true/unknown: describe the risk, guardrails (e.g., restrictions on use, provenance controls, human review), and any approvals needed.

---

10) Transparency

1. **Confirm users will be made aware when they are interacting with AI** (or receiving AI-generated content) when applicable.
   ☐ Confirm ☐ Not true / Not applicable
   - If not true: explain why and document alternate transparency measures.

---

11) Monitoring + change management

1. **Monitoring:** What metrics will be monitored (quality, safety incidents, drift, bias indicators) and who receives alerts.
2. **Revalidation cadence:** When do you re-test (e.g., quarterly, after model/prompt/data changes)?
3. **Material changes:** What changes require re-review and re-approval?

---

12) Launch decision

Risk tier: ☐ High ☐ Moderate ☐ Minimal
Decision: ☐ Approve ☐ Approve w/ conditions ☐ Block pending changes ☐ Reject

Conditions / required mitigations before launch:
Approvers (Privacy/Legal/Security/Product/Eng as needed):

AI Assessment (vendor)

AI Assessment (Vendor AIA)

The AI Assessment (Vendor) is the company's standardized diligence and risk assessment for third-party AI tools and services. Its purpose is to gather use-case-specific, evidence-backed information from vendors about how their AI works and how they manage key risks so the company can approve, approve with conditions, or block the procurement and deployment consistent with the AI Governance Policy. The assessment also ensures the company can translate vendor claims into enforceable requirements (e.g., contracting controls + operational guardrails) and accurately reflect the resulting risk tier and mitigations in the AI Risk Management Tracker.

This list of questions is rather long and is intended to provide a robust set of questions that your organization can tailor down and configure for your particular industry and risk tolerance.

Instructions to vendor

- Answer all questions for the specific use case in scope. If something is not applicable, say so and explain why.
- Where you answer "Confirm/Yes," briefly describe the control behind it (policy, technical measure, process) and provide evidence where available.
- If any answer is "No," describe the gap, mitigation plan, and timeline.
- Attach the evidence listed at the end.

---

1) Scope and use case

1. Describe the AI tool's intended use case in detail (what it does, who uses it, how outputs are used downstream).
2. Identify the end users (roles) and the affected individuals/populations (if any).
3. List all integrations supported or planned (SSO, email, ticketing, chat, data warehouse, APIs, connectors).
4. Provide a high-level data flow: inputs → processing → outputs → storage/logging → deletion.

---

2) Model and system architecture

5. Identify the AI capability type: ☐ classification/scoring ☐ ranking ☐ summarization ☐ generation ☐ agentic automation ☐ other.

6. Provide the model(s) used (name/provider/version) and how they are accessed (API, hosted, embedded).
7. Describe whether the tool uses: prompting, retrieval (RAG), embeddings/vector DB, fine-tuning, custom models, third-party models.
8. Describe any guardrails (filters, policies, constrained prompting, allowlists/denylists, tool/function restrictions).
9. Confirm you will notify the customer before material changes (model change, new subprocessor, new training practice, new retention, new agentic capability).
   ☐ Confirm ☐ Not true (explain)

---

3) Data processing, training, and minimization

10. List the categories of customer data you expect to process (e.g., end-user data, employee/admin data, support content, transactional data, identifiers, device data).
11. Confirm the tool is designed to process only the minimum necessary data to provide the service.
    ☐ Confirm ☐ Not true (explain)
12. If the tool supports data redaction/masking, describe what can be redacted and where (client-side, server-side, both).
13. Confirm whether any customer data (inputs, outputs, logs, embeddings, metadata) is used to train, improve, or fine-tune any model(s).
    ☐ Confirm ☐ Not true (explain)
    - If true: specify what data, whether opt-in/out exists, purpose, retention, and whether training benefits other customers.
14. Confirm whether you support zero-retention (or equivalent) modes for prompts/outputs/logs; if yes, describe limitations.
15. Provide your data retention schedule for: raw inputs, outputs, embeddings, logs, backups, analytics/telemetry.
16. Describe your deletion process (timing, scope, backups) and how customers can request deletion (including upon termination).

---

4) Security controls

17. Confirm you maintain a formal information security program aligned to industry standards (e.g., SOC 2 Type II / ISO 27001 or equivalent).
    ☐ Confirm ☐ Not true (explain)
18. Confirm customer data is protected with encryption in transit and at rest; describe key management (KMS/HSM, rotation, access).
    ☐ Confirm ☐ Not true (explain)
19. Describe tenant isolation and whether customer data is logically/cryptographically separated from other customers.

20. Describe identity and access controls: SSO/SAML, MFA, RBAC, SCIM, least privilege, admin roles.
21. Confirm you maintain audit logs for admin actions and user activity; describe what's logged and retention.
    ☐ Confirm ☐ Not true (explain)
22. Confirm you have a vulnerability management program (SAST/DAST, patch SLAs, pen tests, bug bounty if applicable).
    ☐ Confirm ☐ Not true (explain)
23. Confirm you have an incident response program; provide notification timelines for security incidents affecting customer data.
    ☐ Confirm ☐ Not true (explain)

---

5) Subprocessors and onward transfers

24. List all subprocessors involved in delivering the AI tool (including model providers, hosting, logging, analytics).
25. Confirm you enter into written agreements with subprocessors that impose equivalent privacy/security obligations.
    ☐ Confirm ☐ Not true (explain)
26. Confirm you will notify customers of subprocessor changes in advance and provide a reasonable objection mechanism.
    ☐ Confirm ☐ Not true (explain)

---

6) Data residency and hosting

27. Where is customer data hosted (regions/cloud providers)?
28. Confirm whether customers can select a data residency region (if available).
    ☐ Confirm ☐ Not true (explain)
29. Describe cross-border transfers (if any) and how you manage them.

---

7) Output quality, testing, and monitoring

30. Describe how you evaluate output quality for the relevant use case (accuracy, precision/recall, hallucination rate, toxicity, etc.).
31. Confirm you have controls to reduce hallucinations and unsafe outputs; describe the controls (retrieval constraints, citations, refusals, filters).
    ☐ Confirm ☐ Not true (explain)
32. Describe your approach to regression testing after model upgrades and product releases.

33. Describe monitoring for: drift/performance degradation, unsafe content, abuse/misuse, anomalous access patterns.
34. Confirm you provide operational controls for customers to set/tune safety thresholds (where applicable).
    ☐ Confirm ☐ Not true (explain)

---

8) Bias, fairness, and non-discrimination

35. Confirm the tool is designed to avoid biased outcomes and align with anti-discrimination expectations.
    ☐ Confirm ☐ Not true / Unknown (explain)
36. Describe what bias testing you perform (pre-release and ongoing), including any subgroup performance analysis.
37. If the tool is used for ranking, scoring, or decision support, describe how you detect and mitigate disparate impact risks.

---

9) Human oversight and “stop-the-line” controls

38. Confirm the tool supports meaningful human oversight (review workflows, confidence indicators, explanations, override).
    ☐ Confirm ☐ Not true (explain)
39. Describe how customers can disable or constrain the system quickly (admin controls, kill switch, feature flags, API controls).
40. If the tool supports agentic behaviors, describe guardrails (tool permissions, approval gates, sandboxing, rate limits).

---

10) Prompt injection, data leakage, and abuse resistance

41. Describe how your system mitigates prompt injection and data exfiltration attempts (especially where RAG, tools, or connectors exist).
42. Describe how you prevent leakage of sensitive/confidential data through outputs (training leakage, retrieval leaks, memorization risks).
43. Confirm you have controls to detect and respond to abuse/misuse (policy enforcement, blocking, alerting, escalation).
    ☐ Confirm ☐ Not true (explain)

---

11) Transparency and user-facing disclosures

44. If user-facing, confirm users can be informed that they are interacting with AI or receiving AI-generated content.
    ☐ Confirm ☐ Not true / Not applicable (explain)
45. Describe what transparency/explainability you can provide for outputs (why a result was produced; key factors; reproducibility/traceability).

---

12) IP, ownership, and legal risk allocation

46. Confirm the customer retains ownership of its **inputs** and **outputs** generated specifically from customer inputs.
    ☐ Confirm ☐ Not true (explain)
47. Confirm you do not assert ownership over customer prompts, embeddings, or outputs (except limited rights to provide the service).
    ☐ Confirm ☐ Not true (explain)
48. Describe how you address IP infringement risk for generated content (training data hygiene, filters, claims process, indemnities if offered).
49. Confirm you will not use customer data to train or improve models for general commercial use without explicit agreement.
    ☐ Confirm ☐ Not true (explain)

---

13) Documentation and auditability

50. Confirm you can provide documentation on request: system description, risk controls, testing summaries, monitoring approach.
    ☐ Confirm ☐ Not true (explain)
51. Confirm you support reasonable audit rights (or provide independent audit reports) to validate security and AI governance controls.
    ☐ Confirm ☐ Not true (explain)
52. Provide a point of contact for AI governance/security escalations and incident response.

AI Risk Management Tracker

AI Risk Management Tracker

The AI Risk Management Tracker is a centralized directory designed to map all AI use cases across the company and measure their associated risks in a single, accessible location. Its primary function is to serve as the "single source of truth" for documenting what AI tools are in use, who is responsible for them, and exactly what controls or guardrails the AI Governance Team has recommended to mitigate potential harm. By maintaining this inventory, the organization ensures it can make informed decisions about which AI to deploy while ensuring every system is handled with appropriate care and accountability. Oftentimes, product teams will have their own trackers for AI tools, in which case you should attempt to integrate into those systems rather than recreate the wheel.

Lastly, given the extensive proliferation of AI tools and the rate at which current tools add new AI features, the process for AI risk management and tracking remains an active and evolving process—adjust accordingly.

AI Use & Risk Log

This simplified table focuses on the essentials: identifying the "who" and "what," and documenting the specific safety measures required for a responsible launch.

Use Case	Owner	Business Function	Description	Risk Rating	Controls	Status
Example: *Customer Support Chatbot*	*Head of CX*	*Customer Experience*	*GPT-4 based agent for Tier 1 support. (Pilot Stage)*	***Moderate***	*1. Human-in-the-loop review for all refunds. 2. Mandatory transparency disclosure to users. 3. PII masking on all inputs.*	*Live*

Example: *AI Resume Screener*	*HR Director*	*Human Resources*	*Automated ranking of inbound job applicants. (Production)*	***High***	*1. Quarterly bias audits for protected classes. 2. "Stop-the-line" kill switch if drift exceeds 5%. 3. Full AI Impact Assessment (AIA) completed.*	*In Development*

Key Definitions

To keep this tracker easy for your team to fill out, here is how to handle the "Risk" and "Controls" columns:

- **Overall Risk Rating:** Categorize the risk based on the company AI Governance Policy risk tiers [LINK]. More information can be found in the Policy.
    - **Tier 0 (Prohibited):** Banned (e.g., social scoring).
    - **Tier 1 (High-Risk):** Significant impact on safety or rights (e.g., hiring, credit scoring).
    - **Tier 2 (Moderate-Risk):** Typical enterprise uses requiring privacy or legal review.
    - **Tier 3 (Minimal-Risk):** Standard productivity tools with baseline security.
- **Controls & Mitigations:** This is where you paste the specific technical requirements suggested by the AI Governance Team. This might include things like:
    - **Human-in-the-loop:** Ensuring a person reviews an output before it affects a customer.
    - **Data Minimization:** Only feeding the AI the minimum info needed to finish the task.
    - **Bias Testing:** Checking for discriminatory language or results.
    - **DPA Contracting Clauses**: Commercial-led DPA clauses to ensure vendor adherence to company's AI Governance principles.