APPENDIX D

Sample AI Acceptable Use Policy

Introduction

The advancement of Artificial Intelligence (AI) technologies presents unprecedented opportunities for innovation, automation, and productivity enhancement. However, the adoption and integration of AI into enterprise environments bring with it a range of security, privacy, ethical, legal, and operational risks. The purpose of this policy is to define acceptable and responsible use of AI systems across [Your Company Name] and its affiliated entities.

This policy governs the development, deployment, and use of AI technologies, including generative AI, machine learning (ML) models, and AI-driven services (internal or third-party). It is designed to:

- Prevent data leakage or misuse
- Ensure ethical, secure, and fair use of AI technologies
- Safeguard confidential and sensitive company or customer information
- Maintain trust in AI-powered decision-making systems
- Meet legal and compliance obligations

This policy applies to all employees, contractors, consultants, partners, and vendors who create, manage, or use AI technologies on behalf of [Your Company Name].

Definitions

- **AI (Artificial Intelligence):** Computational systems capable of performing tasks that typically require human intelligence, including natural language processing (NLP), image recognition, and predictive analytics

- **Generative AI:** AI systems that generate new content (text, image, code, etc.), such as ChatGPT, Google Gemini, and image generators like DALL·E
- **Public AI models:** Models hosted by external providers (e.g., OpenAI, Google, Anthropic) and available over the internet
- **Self-hosted models:** AI models deployed within [Your Company Name]'s controlled infrastructure (on-premise or private cloud) with strict access and security controls
- **Confidential data:** Information not publicly available, including but not limited to PII, PHI, financial records, source code, strategy documents, trade secrets, and customer data

Core Principles

All AI use within [Your Company Name] must comply with the following core principles:

- **Privacy and confidentiality:** No confidential or sensitive data shall be used in ways that could expose it to public, unvetted AI systems.
- **Security by design:** AI use must be proactively risk assessed and hardened against common AI threats.
- **Transparency and accountability:** AI use must be auditable, explainable, and documented.
- **Permission-based data use:** Any data used for training or fine-tuning must be legally acquired, authorized, and documented.
- **Governance and oversight:** AI usage must be monitored by approved teams and audited on a regular basis.
- **Documentation:** All AI systems, processes, and decision flows must be documented in IBM AI Factsheets format to ensure traceability, facilitate reviews, and support compliance obligations.

Authorized AI Usage

The following use cases are considered acceptable, provided they meet all security, privacy, and governance criteria:

- Internal process automation and summarization using internal data (external exposure not allowed)

- Code generation and productivity tools (if non-sensitive)
- Internal analytics, dashboards, and forecasting models
- AI-enhanced customer support (e.g., chatbot assistants using sanitized datasets)
- Responsible experimentation in sandboxed environments, governed by this policy

All use cases involving AI must be submitted to appropriate data governance or security review teams for evaluation and written approval before deployment.

Prohibited Activities

The following uses of AI technologies are strictly prohibited:

- Inputting any confidential, customer, PII, PHI, or regulated data into public AI models, such as ChatGPT, Google Gemini, or Claude, unless explicitly authorized and protected by contractual and technical safeguards
- Using AI models to generate or infer personally identifiable information or conduct profiling based on sensitive attributes (e.g., race, health, political views)
- Training or fine-tuning any AI system on datasets without explicit, documented permission or license rights
- Bypassing security or data governance review processes
- Using AI to support decisions with significant legal, financial, or employment implications without human review
- Allowing AI models to autonomously deploy actions in production environments without appropriate human-in-the-loop controls

Prohibited Use (Expanded)

In addition to general restrictions, the following categories are explicitly prohibited:

Unauthorized AI Tools

AI tools that lack contractual agreements or vetted data-handling controls are not approved for use with Controlled or Confidential company data. This includes free or consumer-grade versions of AI tools such as ChatGPT, GitHub Copilot, or other unmanaged SaaS-based AI offerings.

Sensitive Information

No data classified as Confidential or Controlled, including PII, PHI, customer records, source code, financials, or proprietary company materials, may be used in unauthorized or unvetted AI tools under any circumstances.

Non-Public Output

AI tools must not be used to generate or manipulate non-public information, including:

- Proprietary or unpublished research
- Legal analysis or legal advice
- Hiring decisions or performance evaluations
- Academic evaluations or intellectual property generation
- Creation of confidential training materials or business documentation not intended for public release

Fraudulent or Illegal Activities

AI tools must never be used to:

- Conduct or support fraudulent schemes
- Engage in plagiarism or impersonation
- Facilitate phishing, spam, or misinformation
- Violate federal, state, local, or international laws
- Breach internal company policies or contractual obligations

Data Protection and Confidentiality

Public AI Systems

- Confidential or regulated data (e.g., PII, PHI, PCI) must never be entered into publicly available AI models unless:
  - Explicit contractual protections are in place (e.g., private tenancy, model isolation, non-training clauses)
  - Use is approved by legal and security review

- Usage of tools like ChatGPT, Google Gemini, or other cloud-based models should be restricted to non-sensitive content unless behind enterprise-managed controls.

Self-Hosted AI for Sensitive Data

- AI systems that process or generate outputs using confidential data must be self-hosted or deployed in a secured, enterprise-controlled environment.
- Examples of acceptable platforms include: custom fine-tuned LLMs deployed in a private cloud (e.g., Amazon Bedrock with dedicated endpoints), open-source models hosted on internal infrastructure.

Responsible AI Training Practices

Training Data Permissions

- No dataset may be used to train, fine-tune, or reinforce an AI model unless the following conditions are met:
  - The data is either public domain, open source (under proper license), or internally owned and authorized for this use.
  - Legal, compliance, and data privacy teams have approved the use.
  - Consent has been obtained from data subjects where required (especially for customer or employee data).

Data Minimization and Anonymization

- Before training or fine-tuning:
  - Data must be de-identified or anonymized wherever feasible.
  - Avoid training on raw logs, chat messages, or unstructured user content unless scrubbed of sensitive identifiers.

AI Inventory and Documentation

- All AI systems and tools in use must be cataloged in an AI Inventory, managed by the team responsible for the model or use case.
- This inventory must contain:

- Description of the AI system or tool
- Model type and version
- Data sources used for training/inference
- Purpose and business function
- Risks identified and mitigated
- Approval history and owners
- Dates of deployment and review schedule

- The inventory will be regularly audited by the Security and Data Governance teams.

Collaboration with Security Teams

Teams deploying or using AI systems must partner with Information Security to:

- Conduct AI-specific threat modeling
- Implement defenses against:
  - **Data poisoning** (malicious training data corrupting model behavior)
  - **Hallucinations** (fabricated or false outputs from generative AI)
  - **Evasion attacks** (inputs crafted to bypass AI detection)
  - **Model tampering** (unauthorized modifications or adversarial perturbations)
  - **Confused deputy attacks** (misuse of AI services via indirect invocation)
- Ensure:
  - Logging and monitoring of AI activity
  - Role-based access controls
  - Encryption of data at rest and in transit
  - Incident response playbooks specific to AI risks