

Responsible AI Enablement in Regulated Public Programs



A Governance Framework for Compliance & Program Integrity

Dr. Colin Beech, CEO

February 25, 2026

Table of Contents

Objective	3
The Institutional Risk Landscape	3
Generative AI Defined	4
Large Language Models.	4
The Transformer Architecture (“Attention Is All You Need,” 2017)	4
RLHF (Reinforcement Learning from Human Feedback) and Auto-Instruction	5
Implications	5
Current Service Offerings	6
Public Demonstration Interfaces.....	6
Individual Paid Accounts	6
Enterprise AI Platforms	7
API-Based Integration and Private Deployments	7
On-Premise/’Sovereign’ AI Deployments.....	8
Comparative Overview.....	9
Key Governance Insight.....	9
Enterprise AI Governance: A Practical Starting Model	9
Enterprise Platform Access Controls	10
Data Classification and Access Schema.....	10
Privileged Data	10
Confidential Data	10
Need-to-Know Data	10
Basic Access Data	10
AI Usage Policies and Guardrails	11
Breach Response and Incident Management	11
Auditability and Oversight.....	11
Alignment with Medicaid Program Integrity Requirements	12
Governance as an Enabler of Responsible Innovation.....	12
AI Use-Case Priority Matrix for Regulated Program Operations	12
Priority Matrix: AI Use Cases in Medicaid Program Operations	13
Fraud, Waste, and Abuse Analytics (Highest Priority).....	13

SOP and Compliance Documentation Generation	13
Staff Operational Support and Knowledge Assistance	14
Additional Organizational Affordances	14
Appendix A: Sample AI Acceptable Use Policy (Enterprise Environment)	15
Purpose	15
Scope	15
Approved AI Platforms	15
Data Protection Requirements.....	15
Prohibited Inputs.....	15
Permitted Uses	15
Prohibited Uses	16
Human Oversight Requirement	16
Incident Reporting.....	16
Policy Enforcement	16
References.....	17

Objective

The purpose of this white paper is to guide governmental organizations in establishing a controlled, secure, and auditable framework for AI enablement within their enterprise. Properly governed, AI can strengthen program integrity, operational consistency, and institutional accountability while maintaining rigorous security and risk controls.

This paper analyzes the technical foundations of contemporary generative artificial intelligence systems and examines their evolving modalities in the marketplace. It further considers the changing role of labor and professional practice in light of automation increasingly affecting traditionally white-collar work.

Finally, this paper offers a prescriptive governance framework and phased adoption model designed specifically for regulated public programs and mission-driven organizations, including NGOs and private firms serving government clients.

It is important here to carefully define precisely what ‘AI’ is from a contemporary, critical perspective. There is a rapidly burgeoning marketplace saturated with hype in the form of product features and rosy use cases; separating fact from fiction is vital to the practitioner looking to effectively implement AI within their organization. PCC brings long-standing experience in AI systems, enterprise IT, and regulated human services programs to this analysis.

The Institutional Risk Landscape

AI is already entering the workplace informally. The choice before leadership is not adoption versus avoidance — it is governed enablement versus unmanaged exposure.

This reality is driven by two structural conditions.

First, the internet has become fully embedded in daily organizational life. Smartphones, cloud services, and connected devices have normalized constant access to external platforms. While information security policies are designed to safeguard enterprise resources, most employees now operate with broad access to digital tools and relatively limited real-time oversight.

Second, generative AI technologies are not restricted or difficult to obtain. They are widely accessible, often free, and require no enterprise authorization to use. Market and cultural pressures increasingly incentivize employees to adopt these tools independently in order to remain competitive and efficient.

Generative AI is transformative. It materially alters how knowledge work is performed. Current capabilities represent only early-stage development, and organizations are still working to understand both its power and its limitations.

This creates a predictable pattern of uneven adoption. Some employees remain unaware. Some resist. The greatest risk, however, arises when employees embrace AI tools outside formal enterprise governance structures.

Unmanaged AI use introduces multiple exposure vectors:

- Potential leakage of PHI, PII, or confidential enterprise data
- Unvetted outputs presented as authoritative work product
- Intellectual property and data ownership ambiguity
- Regulatory and audit vulnerabilities
- Inconsistent or untrained application of tools

Leadership teams are therefore confronting a dual reality: AI offers measurable productivity gains, yet unmanaged adoption introduces material operational and compliance risk.

Enterprises have navigated technological risk before. The solution remains consistent: structured governance, defined guardrails, and deliberate enablement.

Generative AI Defined

Large Language Models.

Generative AI systems are built upon Large Language Models (LLMs).

Deviating sharply from more mathematical, game-theory oriented models, LLMs took this field in an entirely new direction. Unlike prior AI systems focused on narrow optimization tasks (e.g., chess engines), LLMs are trained on large corpora of text to model linguistic patterns across domains. Using neural-network models from machine learning, software agents, or “bots” were trained in parsing not only the text statements, but also their ordering and interrelationships. LLMs generate outputs by predicting statistically plausible token sequences based on learned probability distributions.

The Transformer Architecture (“Attention Is All You Need,” 2017)

The Transformer Architecture underpins contemporary generative AI systems. Introduced in 2017, enabled scalable parallel training of large language models through attention mechanisms. This breakthrough made it computationally feasible to train models on vast datasets efficiently.

This revolutionized training and processing algorithms that didn’t rely on long sequences of recurring symbols. In this way, “attention” emerges in machine learning as an exclusively forward-looking mechanism. Models of tokens and their vectors leading to “successes” could now be refined in a radically amplified way.

This allowed for *parallel* computation, small snippets of calculations that readily lent themselves to cheap, broadly available computer graphics processors, re-tooled from playing video games to neural-network training in LLMs. Faster, more effective training in real-time becomes possible with this methodology, enough so that it became a disruption to the industry.

RLHF (Reinforcement Learning from Human Feedback) and Auto-Instruction

As sophisticated as it is, the Attention model alone has not resulted in the current crop of GPTs in use today. Human feedback further refines model behavior through reward modeling. Human evaluators rank model outputs, reinforcing preferred behaviors through reward modeling. These rankings shape policy layers that guide model output toward preferred responses.

This methodology laid the groundwork for “instruction” based techniques, where human coding is captured more explicitly as the “right” way to perform a task, follow guidelines, or otherwise conform to scripted actions. InstructGPT, developed by OpenAI for example, explicitly focuses on parallel machine learning using detailed instruction sets as the locus of natural language processing.

This method becomes even more powerful with the advent of Auto-Instruction, where bots are allowed to generate their own meta-scripts based on simple instruction sets, with further modeling occurring as layers of these instructions are followed. Known as “fine-tuning”, this practice has now entered widespread adoption in the AI industry. It also has paved the way for multi-modality, where the same machine learning techniques are applied to visual arts, music, and other areas aside from natural language processing.

Implications

It cannot be stressed enough that Generative AI systems do not “understand” content in a human sense; they generate outputs by predicting statistically plausible sequences based on prior training data. Outputs may appear authoritative regardless of their underlying statistical uncertainty. They are statistically learned patterns layered atop a probabilistic core model, reflecting the choices, preferences, and verbal syntax of their developers.

It is for these very reasons that strong governance models are needed for generative pre-trained models:

- Errors can propagate through the system if they are in the original bank of model data, or instructions are applied ambiguously.
- Behavioral constraints are statistically reinforced rather than deterministically enforced. Guardrails are not hard-coded in the manner of traditional business rules engines.
- In some deployment models, user interactions may be incorporated into future model refinement unless explicitly restricted; entries and participation are added to the probabilistic domain, inadvertently capturing data provided by the human user into the larger domain for review and use by other users.
- The technology is still *interpretively flexible*, with a variety of unintended uses & consequences still emerging; risk exposure evolves as adoption stabilizes, often in ways not initially anticipated.

Current Service Offerings

Generative AI services are currently delivered through several distinct deployment and access models. Each model differs substantially in cost structure, method of technical integration, user accessibility, and - most importantly for regulated organizations - the level of governance control and exposure to sensitive data. Understanding these distinctions is essential before any responsible adoption strategy can be developed. Broadly speaking, contemporary offerings fall into four primary categories:

- Public Demonstration Interfaces
- Individual Subscription Accounts
- Enterprise Subscription Platforms
- API-Based Integration and Private Deployments

Each model represents a different balance of accessibility, capability, and governance risk.

Public Demonstration Interfaces

Public demonstration interfaces are freely accessible generative AI services available through standard web browsers. These systems typically operate through HTTPS-based web interfaces and require minimal user authentication. Examples include publicly available chatbot interfaces and demonstration models intended to showcase generative AI capabilities.

These models are usually free, or supported through advertising. Their connection method is usually through a web browser using conventional certificate-based security, though smartphone apps and access are probably the predominant form of access. They are open to any individual user with internet access.

It is important to note that public interfaces generally provide no contractual (or other) guarantees regarding data retention, logging, or downstream training usage. User inputs may be stored, analyzed, or used for model improvement depending on provider policies.

Public demonstration interfaces should **never be used with protected health information (PHI) or personally identifiable information (PII)**. Such use would likely violate HIPAA security expectations and internal information governance policies. These tools are appropriate only for non-sensitive exploratory or educational use.

Individual Paid Accounts

Many AI providers offer subscription-based access for individual users. These services remain primarily browser-based, but provide expanded model access, higher usage limits, and additional features. The cost is typically \$20 - \$50 per user, per month. Once again, they are usually accessed through web interfaces and/or smartphone apps via conventional security methods. They have limited plugin or tool integrations. Individual employees may subscribe independently to these services, but absent of any governance control or organizational oversight.

Despite their being a minimum set of data use agreements (user's content is considered their intellectual property), the use of individual subscription accounts with PHI or PII introduces *substantial* compliance risk. Data may pass through external processing environments without contractual assurances of isolation or retention control, or use exclusively within CONUS (continental United States). For regulated public programs, these accounts should generally be treated as **non-compliant environments for sensitive data analysis**.

Enterprise AI Platforms

Enterprise AI platforms provide organizationally managed access to generative AI systems through centrally administered accounts. These offerings typically include identity management integration, administrative controls, and contractual data governance provisions. The cost is typically per-seat licensing or enterprise agreements, with typically \$30–\$100+ per user per month depending on features. The connection method remains mostly web interface and conventional certificate-based security. Integration with enterprise identity providers (SSO) is a standard feature for such platforms.

Features often include administrative access controls, usage monitoring and logging, contractual commitments regarding data handling, and perhaps most importantly, isolation of enterprise data from broader model training outside the service engagement.

Enterprise AI platforms may be appropriate for certain internal productivity use cases if configured with appropriate data protection controls. However, they still rely on external vendor infrastructure, and organizations must carefully evaluate:

- Data retention policies
- Encryption practices
- Business Associate Agreements (BAAs) where applicable
- Regulatory compliance posture

Even in enterprise environments, direct processing of Medicaid PHI datasets may remain restricted depending on contractual terms.

API-Based Integration and Private Deployments

The most flexible (and governable) approach involves direct API-based integration of generative AI models into an organization's own application infrastructure. In this model, AI capabilities are accessed programmatically through secure API endpoints and embedded within enterprise systems under direct organizational control. These engagements rely on usage-based pricing (tokens, compute time) as well as infrastructure and development costs. Connection methods evolve into secure API calls over HTTPS channels, allowing for integration into internal applications and data pipelines.

With such engagements, organizations may implement input filtering and validation, logging and audit trails, data anonymization layers, and secure data processing pipelines within their own stack, making powerful tools available to internal development teams. In more advanced configurations, models may be deployed within **private cloud environments or virtual private networks**, further reducing external data exposure.

API-based or private deployments represent the most viable path for regulated environments seeking to use generative AI with sensitive datasets. When combined with appropriate governance controls (such as de-identification pipelines, audit logging, and secure processing environments) these architectures can support advanced analytics while maintaining regulatory compliance.

On-Premise/'Sovereign' AI Deployments

A final category of AI deployment involves hosting generative AI models entirely within infrastructure controlled by the organization itself. These are commonly referred to as **on-premise, sovereign, or self-hosted** AI deployments.

In this model, the AI system is deployed on servers owned or directly managed by the organization, typically within an internal data center or private cloud environment. The model weights, inference infrastructure, and associated data pipelines remain entirely under organizational control.

This architecture mirrors long-standing approaches used for sensitive analytics environments in finance, healthcare, and national security. Organizations operating sovereign deployments control:

- Model selection and versioning
- Data ingestion pipelines
- Logging and audit capabilities
- Security monitoring and network segmentation
- Retention and deletion policies

Because the organization maintains direct control over infrastructure, sensitive datasets can remain fully inside regulated environments. On-premise or sovereign AI deployments represent the **highest level of governance assurance** currently available for organizations handling regulated health data. When implemented within secure enterprise environments, this model allows for processing of PHI and PII under existing enterprise security controls, integration with existing compliance frameworks (HIPAA, CMS requirements, state policy), full auditability of system interactions, and clear data ownership and retention governance.

However, sovereign deployments introduce **significant** operational complexity. Organizations must maintain sufficient technical expertise to manage model hosting, performance optimization, and ongoing security maintenance. As a result, this model is typically adopted by large enterprises (10,000+), government agencies with mature IT operations (DARPA, NASA, NSA, etc.) operating highly sensitive data environments.

For many regulated organizations, sovereign deployments may represent a **long-term maturity target**, like MITA 4.0+, while earlier adoption phases rely on enterprise platforms or API-based integrations.

Comparative Overview

<i>Model</i>	<i>Cost</i>	<i>Connection Method</i>	<i>User Access</i>	<i>Governance Control</i>	<i>PHI/PII Risk</i>
<i>Public Demo Interfaces</i>	Free	HTTPS web interface	Open public	None	Extremely High
<i>Individual Subscriptions</i>	Low monthly fee	HTTPS web interface	Individual users	Minimal	High
<i>Enterprise Platforms</i>	Moderate per-seat cost	HTTPS + SSO	Organizational users	Moderate	Moderate–High
<i>API/Private Deployment</i>	Usage-based + infrastructure	Secure APIs	Organizational systems	High	Lowest (with controls)
<i>On-Prem/Sovereign</i>	High initial infrastructure investment	Intranet-Based	Internal Users Exclusively	Very High	Very Low

Key Governance Insight

For regulated public programs such as Medicaid administration, the distinction between **consumer-facing AI tools and governed enterprise deployments** is critical.

The majority of widely publicized generative AI tools operate in environments unsuitable for sensitive data processing. Responsible adoption therefore requires deliberate architectural choices that prioritize governance, data protection, and regulatory compliance.

The appropriate AI deployment model for a regulated organization is therefore not determined by technological capability alone, but by governance requirements, data sensitivity, and institutional risk tolerance.

Enterprise AI Governance: A Practical Starting Model

The preceding analysis demonstrates that generative AI technologies can deliver significant operational benefits, but only when deployed within a structured governance environment. For regulated public programs such as Medicaid administration, AI adoption must occur within the same compliance, audit, and data protection frameworks that govern other enterprise information systems.

This governance framework assumes that the organization has implemented an **Enterprise AI platform** rather than relying on public or individually managed accounts. Enterprise deployments allow employees to access approved AI tools within a controlled environment, typically integrated with the organization's identity management system and existing security infrastructure.

The purpose of AI governance is therefore not to restrict technological capability, but to ensure that AI-enabled workflows operate within clearly defined boundaries that protect sensitive information,

maintain regulatory compliance, and preserve institutional accountability.

Enterprise Platform Access Controls

The foundation of any responsible AI governance program begins with controlled access to the technology itself. Enterprise AI platforms should be integrated with the organization's existing identity and access management systems, allowing user authentication through established mechanisms such as Single Sign-On (SSO) and multi-factor authentication. Administrative oversight should include:

- Role-based access control for AI platform users
- Centralized user provisioning and revocation
- Integration with existing cybersecurity monitoring tools
- Logging of user interactions for audit purposes

These controls ensure that AI tools function as managed enterprise applications rather than unmanaged external services.

Data Classification and Access Schema

Effective governance requires a clear framework for identifying and classifying data used within the organization. AI-enabled workflows must operate within the same data classification structure used for other enterprise systems.

A recommended classification schema includes four tiers:

Privileged Data

Privileged data represents the most sensitive category of information within the enterprise. In Medicaid program administration, this may include detailed beneficiary records, investigative materials related to fraud, waste, and abuse, or legal communications protected by attorney-client privilege.

Access to privileged data should be tightly restricted and typically excluded from AI processing environments unless explicitly authorized within secure analytics systems.

Confidential Data

Confidential data includes sensitive operational information that is not intended for broad distribution but may be required for internal program operations. Examples include internal investigative findings, provider performance analytics, and case management documentation. AI tools may interact with this data only under controlled conditions and within approved enterprise environments.

Need-to-Know Data

Need-to-know data includes operational information required by specific program teams in order to perform their responsibilities. Examples may include operational dashboards, provider utilization metrics, or program management reporting. Governance policies should ensure that access to this data is limited to appropriate organizational roles.

Basic Access Data

Basic access data includes information that may be safely shared across the organization without significant privacy or compliance risk. Examples include training materials, publicly available regulatory

guidance, and internal policy documentation. This category represents the safest initial domain for AI-assisted productivity tools such as document drafting and knowledge support.

AI Usage Policies and Guardrails

Governance programs must establish clear usage policies defining how AI systems may be used within the organization. These policies should explicitly identify both permitted and prohibited uses. Examples of permitted activities may include:

- Drafting internal documentation
- Summarizing regulatory guidance
- Assisting staff with operational procedures
- Generating training materials

Examples of restricted or prohibited uses may include:

- Submission of PHI or PII into unapproved AI environments
- Automated decision-making affecting beneficiary eligibility
- Unsupervised AI analysis of sensitive program datasets

Clear policy guidance helps prevent misuse while allowing staff to safely benefit from AI-enabled productivity tools.

Breach Response and Incident Management

AI governance programs must include clear procedures for responding to potential data exposure or misuse of AI systems. These procedures should align with existing enterprise cybersecurity and compliance response frameworks. Key components include:

- Defined reporting channels for suspected AI-related incidents
- Immediate containment procedures for data exposure events
- Investigation protocols coordinated with cybersecurity and compliance teams
- Documentation of incidents for regulatory reporting and audit review

Organizations should treat AI-related data exposures with the same seriousness as other cybersecurity incidents involving sensitive information.

Auditability and Oversight

Maintaining an auditable record of AI system usage is essential for both internal governance and regulatory compliance. Enterprise AI platforms should support:

- Logging of user prompts and outputs where appropriate
- Administrative review capabilities
- Retention policies aligned with enterprise data governance rules

These records provide important visibility into how AI tools are being used and allow organizations to identify emerging risks or misuse.

Alignment with Medicaid Program Integrity Requirements

For organizations operating within Medicaid programs, AI governance must align with broader program integrity expectations established by federal and state authorities. Governance frameworks should therefore demonstrate alignment with principles commonly expected by oversight bodies such as the Centers for Medicare & Medicaid Services (CMS), including:

- Protection of beneficiary data
- Transparency of analytical methods used in program integrity investigations
- Preservation of audit trails supporting investigative actions
- Adherence to HIPAA and related privacy regulations

AI tools used in program integrity activities should function as **analytical support systems**, not as autonomous decision-making authorities; the abilities of individual team members are augmented, rather than a sub-set of team members replaced. Human oversight must remain central to any investigative or enforcement activity.

Governance as an Enabler of Responsible Innovation

When implemented effectively, AI governance programs allow organizations to adopt new technologies with confidence rather than hesitation. By establishing clear policies, data classification frameworks, and oversight mechanisms, institutions can safely explore AI-enabled capabilities while maintaining the high standards of accountability required in public service.

In this sense, governance should be understood not as a constraint on technological innovation, but as the institutional architecture that allows innovation to occur responsibly.

AI Use-Case Priority Matrix for Regulated Program Operations

Once an organization has implemented a governed enterprise AI platform, the next step is to identify operational areas where generative AI can provide immediate value while maintaining acceptable risk exposure.

For regulated programs such as Medicaid administration, not all AI applications are equally appropriate. The most effective governance models begin by prioritizing use cases that offer high operational benefit while maintaining clear oversight and compliance controls.

The following matrix illustrates several priority areas for AI-assisted operations within regulated program environments:

Priority Matrix: AI Use Cases in Medicaid Program Operations

Use Case	Operational Value	Data Sensitivity	Governance Complexity	Recommended Priority
<i>Fraud, Waste, and Abuse (FWA) Analytics</i>	Very High	High	Moderate	Priority 1
<i>SOP and Compliance Documentation Generation</i>	High	Low	Low	Priority 2
<i>Staff Operational Support / Knowledge Assistant</i>	High	Low–Moderate	Low	Priority 3
<i>Policy Analysis and Legislative Tracking</i>	High	Low	Low	Priority 4
<i>Case Narrative Drafting and Documentation Assistance</i>	Moderate–High	Moderate	Moderate	Priority 5
<i>Workforce Training and Knowledge Management</i>	Moderate	Low	Low	Priority 6
<i>Data Quality Monitoring and Anomaly Detection</i>	High	Moderate–High	Moderate	Priority 7

Fraud, Waste, and Abuse Analytics (Highest Priority)

Generative AI systems demonstrate strong capability in analyzing complex patterns within large transaction datasets; it is their primary recommended Use Case in this white paper. In regulated programs such as Medicaid administration, this capability can be applied to identify indicators of fraud, waste, and abuse across large populations of program activity. Potential applications include:

- Transaction pattern analysis
- Anomaly detection within provider billing behavior
- Identification of irregular service utilization patterns
- Investigation support for program integrity teams

When implemented through governed enterprise platforms or API-based integrations, AI systems can assist analysts in identifying patterns that may otherwise remain difficult to detect through traditional query-based analytics alone. Because these applications involve PHI and PII, strong governance controls—including de-identification pipelines and secure analytics environments—are essential.

SOP and Compliance Documentation Generation

Generative AI is particularly well suited to structured document generation. Organizations operating in regulated environments must maintain extensive libraries of Standard Operating Procedures, compliance policies, and regulatory documentation. AI systems can significantly accelerate:

- Drafting new policies
- Summarizing CMS guidance
- Aligning SOPs with federal and state regulations

- Maintaining compliance documentation

Because these use cases typically rely on regulatory text rather than sensitive operational data, they represent **low-risk, high-value applications** of AI technology.

Staff Operational Support and Knowledge Assistance

Enterprise AI platforms can also function as internal knowledge assistants, helping staff quickly access information related to policies, procedures, and operational guidelines. Typical use cases include:

- Answering procedural questions
- Assisting staff in interpreting policy documentation
- Generating summaries of internal guidelines
- Providing structured checklists for operational tasks

These tools can improve staff productivity while reducing reliance on informal institutional knowledge networks.

Additional Organizational Affordances

Beyond these initial priority areas, generative AI systems offer several additional operational opportunities that organizations should be aware of as governance models mature.

Policy Analysis and Legislative Tracking

AI systems can assist policy teams in tracking evolving federal and state regulations, summarizing new CMS guidance, and identifying potential operational impacts.

Case Narrative Drafting and Documentation Assistance

Generative AI tools can assist investigators by summarizing case documentation, generating structured reports, and organizing investigative findings.

Workforce Training and Knowledge Management

AI platforms can help develop training materials, generate learning modules, and assist in onboarding new staff members.

Data Quality Monitoring and Anomaly Detection

AI models can assist analysts in identifying anomalies or inconsistencies within large operational datasets, improving data quality across program systems.

Appendix A: Sample AI Acceptable Use Policy (Enterprise Environment)

The following policy template illustrates how organizations may define acceptable use of enterprise AI systems.

Purpose

This policy establishes guidelines for the responsible use of generative artificial intelligence (AI) tools within the organization. The goal is to enable productivity and innovation while ensuring the protection of sensitive data, regulatory compliance, and program integrity.

Scope

This policy applies to all employees, contractors, and authorized users who access enterprise AI platforms or AI-enabled applications through organizational systems.

Approved AI Platforms

Employees may only access AI tools that have been approved and provisioned through the organization's enterprise technology environment.

Use of external or public AI services for organizational work is prohibited unless explicitly authorized.

Data Protection Requirements

Users must comply with the organization's data classification and security policies when interacting with AI systems. The following restrictions apply:

Prohibited Inputs

- Protected Health Information (PHI)
- Personally Identifiable Information (PII)
- Confidential investigative materials
- Internal legal communications
- Sensitive operational datasets

Unless the AI platform has been explicitly authorized for such data under controlled governance conditions.

Permitted Uses

Approved uses of enterprise AI systems may include:

- Drafting internal documentation
- Summarizing policy or regulatory guidance
- Generating training materials
- Assisting with internal research or analysis
- Improving productivity in routine administrative tasks

All AI-generated content should be reviewed by a human user before being used in official communications or operational decisions.

Prohibited Uses

Employees may not use AI systems to:

- Make automated eligibility or benefits determinations
- Generate investigative findings without human review
- Submit sensitive data to unapproved ai services
- Represent AI-generated content as official organizational policy without validation

Human Oversight Requirement

AI systems are intended to support human decision-making, not replace it. All outputs produced by AI tools must be reviewed by qualified personnel before being used in operational, investigative, or policy-related contexts.

Incident Reporting

Any suspected misuse of AI tools or potential data exposure must be reported immediately through established cybersecurity or compliance reporting channels.

AI-related incidents will be handled under the organization's existing security and breach response procedures.

Policy Enforcement

Failure to comply with this policy may result in disciplinary action and potential revocation of system access.

References

Pinch, T. J., & Bijker, W. E. (1987). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. J. Pinch (Eds.), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (pp. 17–50). MIT Press.

Billings, D., Davidson, A., Schaeffer, J., & Szafron, D. (2002). *The Challenge of Poker*. Artificial Intelligence, 134(1–2), 201–240.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. Advances in Neural Information Processing Systems (NeurIPS 30)