eltropy AI

Safe AI

The AI Guide for Community Financial Institutions

Saahil Kamath, VP, Product
Rahul Prakash, Sr. Director, Engineering

April 2026

Please note: This content is provided for general informational purposes only and reflects Eltropy's current approach to AI safety and innovation as of March 2026. It does not constitute legal or financial advice. All institutions are encouraged to consult with their own legal counsel regarding the deployment of AI technologies.

INTRODUCTION

Building Trust Through Responsible AI Implementation

Executive Summary

Artificial Intelligence is transforming how Community Financial Institutions (CFIs) serve their consumers, offering unprecedented opportunities to enhance service quality, streamline operations, and strengthen consumer relationships. From intelligent automation to agentic AI that can take actions on consumers' behalf, these capabilities represent a new era in consumer service. However, with great power comes great responsibility.

This framework serves as your comprehensive guide to implementing AI safely and ethically. We've designed it to help community financial institution (CFI) leaders, staff, and consumers understand not just what AI can do, but how to confirm it works in everyone's best interest.

What You'll Find Here:

- ➔ Clear explanations of AI benefits and risks in plain language
- ➔ Practical frameworks for safe AI implementation
- ➔ Real-world use cases specifically for CFIs
- ➔ Step-by-step guidance for establishing AI governance
- ➔ Tools for detecting and preventing bias
- ➔ Transparent data privacy and security practices

01
Understanding Safe AI

Why This Matters Now

CFIs have always been built on trust, community, and putting consumers first. As AI becomes more prevalent in financial services, maintaining these core values while embracing innovation isn't just important, it's essential for your community financial institution's future.

The Bottom Line: AI should enhance human decision-making, not replace it. When implemented correctly, AI becomes a powerful tool that helps you serve consumers better while protecting their interests.

Importance of Safe AI in Financial Services

The integration of AI in financial services brings numerous advantages, including enhanced efficiency, personalized consumer services, and improved decision-making processes. However, these benefits come with inherent risks that must be carefully managed. The importance of safe AI in financial services can be summarized as follows:

1. **Building Trust with Consumers and Employees:** CFIs are built on trust and relationships. Ensuring AI systems are transparent, fair, and ethical helps maintain and strengthen this trust among both consumers and internal employees.
2. **Protecting consumer Privacy and Data Security:** With the increasing use of AI, large volumes of sensitive consumer data are processed. Implementing robust data privacy and security measures is essential to protect consumer information and maintain compliance with regulations.
3. **Mitigating Risks and Avoiding Bias:** AI systems can inadvertently introduce biases or make erroneous decisions that can harm consumers and employees. By adopting safe AI practices, CFIs can identify and mitigate these risks, facilitating fair and equitable outcomes for users.

4. **Supporting Regulatory Compliance:** The financial industry is heavily regulated, and non-compliance can result in severe penalties. Safe AI practices support the Community Financial Institution's compliance with relevant regulations and mitigating the risk of legal and financial repercussions.
5. **Enhancing Operational Efficiency:** While AI can significantly enhance operational efficiency, it is crucial that these systems operate reliably and safely. Safe AI practices create AI-driven processes that are efficient and trustworthy.
6. **Supporting Ethical AI Use:** As stewards of consumer trust, CFIs have a responsibility to use AI ethically. Adhering to ethical guidelines and best practices ensures that AI technologies are used in ways that align with the core values of CFIs and the broader community.

What Makes AI "Safe"?

Safe AI means technology that:

- ➔ Treats all consumers fairly without bias or discrimination
- ➔ Protects consumer privacy and keeps data secure
- ➔ Remains transparent about how decisions are made
- ➔ Keeps humans in control of important choices
- ➔ Follows applicable regulations and industry standards
- ➔ Continuously improves based on feedback and monitoring

The Eltropy Safe AI Framework: Four Layers of Protection

Think of AI safety like a building, you need a strong foundation and multiple layers of protection:

Layer 1: Model Foundation

What it means: The AI models we use are carefully selected and tested
What we do: We evaluate AI providers' training methods, data sources, and built-in safety measures before partnering with them

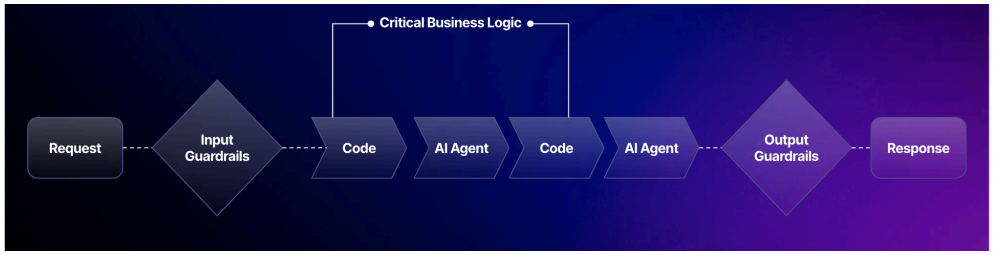
- Guardrails are enforced as a part of the training and alignment process. When Eltropy chooses a 3rd party vendor, we verify that the model vendor implements or enforces similar guardrails.
- It is critical to assess the steps taken by the model developer and the training data used to understand the risks the model implementation possesses.
- Implement a mitigation strategy at the Safety System or Application Layer.

Layer 2: Programmable Guardrails

What it means: Automated filters that catch problems before they reach you or your consumers. For agentic AI, systems that autonomously take actions on behalf of consumers, these guardrails are especially critical, ensuring agents operate only within approved boundaries.
What we do: We implement filters for harmful content, personal information protection, response quality checks, and action-level constraints that prevent agentic AI from exceeding its defined scope.

Constrained Agents (Programmable Guardrail)

Safe AI means our agents are built with defined constraints, keeping critical functions like authentication and escalations outside of the AI Agent scope.



- Guardrails can be enforced as a set of filters, rules, and tools that sit between inputs, the model, and outputs.
- These guardrails reduce the likelihood of erroneous/toxic outputs and unexpected formats.
- Improve conformity to expectations of values and correctness.

Layer 3: Application Design

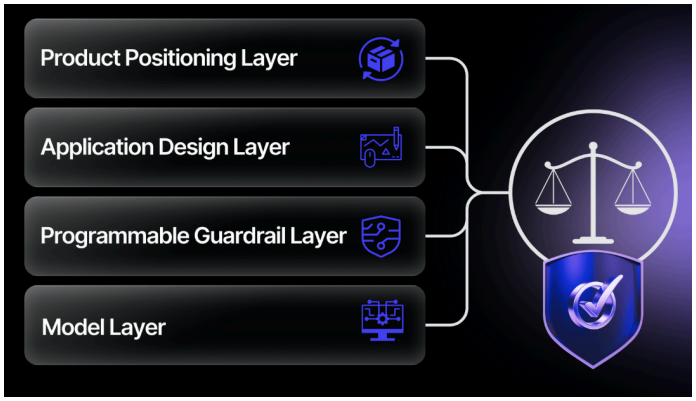
What it means: The way we build AI features includes multiple safety checks
What we do: We use techniques like cross-referencing answers, similarity checking, and confidence scoring

- Additional guardrails can be set at the application architecture level.
- Design mechanisms like Multi Prompting, Meta Prompting, Output Similarity Comparison, and Thresholding as a part of Application Development.

Layer 4: Product Positioning

What it means: Clear guidance on how to use AI tools effectively and responsibly
What we do: We provide training, usage guidelines, and clear information about AI limitations

- Guardrails can be enforced at the product positioning layer by educating the user on the limitations, best practices, and usage guidelines.
- It is essential to design and position AI systems in a way that user accountability is a part of the usage process.



02

Ethical AI Policy and Use Cases for AI Services

Introduction

Purpose and Scope: This policy outlines the ethical standards and use cases for Eltropy's AI-powered services (chatbots, voice assistants, and knowledge-based systems) provided to CFIs. Eltropy is committed to fair, transparent, and accountable AI usage that upholds consumer privacy and autonomy. This document establishes core ethical AI principles, defines approved use cases, lists prohibited practices, and describes risk mitigation measures, in alignment with industry best practices.

Core Ethical AI Principles

Our AI application development and deployments are guided by the following fundamental principles:

- **Fairness:** AI systems must treat all CFIs consumers and agents equitably and avoid unfair bias or discrimination. Eltropy designs and tests AI applications to focus on relevant financial criteria, providing equitable consumer treatment by prioritizing algorithmic integrity and data transparency. This provides CFIs with the technical framework necessary to evaluate and maintain their own fair lending and inclusivity standards. We also choose AI Models from third-party providers that adhere to the above policies. We provide a feedback mechanism in all our AI offerings to identify any bias or disparate impact, and it will be addressed immediately to uphold inclusivity and trust.

- **Accountability:** AI is viewed as a tool to assist human decision-makers, not replace them – in order to keep humans "in the loop." AI products are designed to assist the consumers' data or Agents using them to help in the decision-making process and will not decide on behalf of humans. AI products are required to share citations and sources for any generated content so that it can be reviewed by a human in the loop

before making a decision. Employees and stakeholders are trained on ethical AI use and understand how to use AI outputs responsibly. In practice, this means maintaining audit trails of AI decisions and being prepared to explain and justify outcomes to regulators and consumers. AI is viewed as a tool to assist human decision-makers, not replace them, keeping humans firmly "in the loop." Eltropy's AI products are designed to support CFI staff and consumers in the decision-making process, never to make decisions on their behalf. To enable meaningful human review, AI products are required to share citations and sources for any generated content before a decision is made. Employees and stakeholders are trained on ethical AI use and understand how to apply AI outputs responsibly. In practice, this means maintaining audit trails of AI-assisted decisions and being prepared to explain and justify outcomes to regulators and consumers.

- **Transparency:** Eltropy commits to AI transparency and explainability in both development and deployment. Consumers/Agents should always be informed when they are interacting with an AI (such as a chatbot) rather than a human, and how the AI is using their data. We strive to make AI decision-making processes understandable to non-technical stakeholders. For example, if our AI assistant provides citations to the answers generated. Eltropy commits to AI transparency and explainability in both development and deployment. Consumers/Agents should always be informed when they are interacting with an AI (such as a chatbot) rather than a human, and how the AI is using their data. We strive to make AI decision-making processes understandable to non-technical stakeholders, for example, by having our AI assistant provide citations for the answers it generates.

- **Privacy:** Protecting consumer privacy and data security is paramount. Eltropy follows a “privacy by design” approach for all AI systems, meaning data protection is built into each stage of development. We only collect and use consumer data that is necessary for the AI service and in accordance with the CFIs privacy policies and consumer consent. Data usage is transparent and limited to the stated purpose, for example, if an AI voicebot uses call audio to understand a consumer’s request, that audio will not be repurposed for any unrelated analysis or training. Protecting consumer privacy and data security is paramount. Eltropy follows a "privacy by design" approach for all AI systems, meaning data protection is built into each stage of development. We only collect and use consumer data that is necessary for the AI service and in accordance with the CFI's privacy policies and consumer consent.

Data usage is transparent and limited to its stated purpose — for example, if an AI voicebot uses call audio to understand a consumer's request, that audio will not be repurposed for any unrelated analysis or training.

- **User Autonomy:** Eltropy respects consumer autonomy and choice in all AI interactions. Our AI tools are designed to empower consumers with convenient self-service and personalized assistance, *not* to coerce or manipulate decisions. Consumers maintain control: they are free to accept or ignore AI-provided recommendations, and important decisions are never executed without the users (Consumer/Agents) informed consent. Consumers can easily request human assistance or review at any point. For instance, if a consumer is interacting with a chatbot and prefers to speak to a human representative, or if they want a human loan officer to review an AI-generated loan recommendation, those options are readily available. The AI will not override a consumer's decisions or hide the option to interact with a person. This principle aligns with emerging ethical standards that people should be able to opt out of automated systems in favor of human alternatives when appropriate.

Approved and Supported AI Use Cases

Eltropy supports a range of beneficial AI use cases that improve consumer service, operational efficiency, and decision support forCFIs. These use cases have been evaluated for ethical compliance and alignment with our principles. (As NCUA has noted, AI offers promise in areas like consumer communication and underwriting when used responsibly.) The following are approved AI applications that Eltropy provides:

What We Support: Ethical AI Applications

✅ Agentic AI Assistants

Purpose: Autonomously handle multi-step consumer requests
Why it's safe: Operates within strict predefined boundaries, maintains a human-in-the-loop escalation path, and logs all actions for audit and review

✅ Virtual Consumer Assistants

Purpose: Answer routine questions and handle simple transactions
Why it's safe: Always identifies as AI, maintains security protocols, transfers complex issues to humans

✅ Knowledge Support Systems

Purpose: Help staff find information quickly and accurately
Why it's safe: Uses only approved information sources, requires human verification for important decisions

✅ Service Quality Analysis

Purpose: Improve consumer experience through interaction analysis
Why it's safe: Anonymizes data, focuses on service improvement, maintains strict privacy controls

(The above list is not exhaustive, but covers the primary AI applications Eltropy enables. All approved use cases are subject to ongoing evaluation. New AI use cases will undergo an ethical review to ensure they meet these policy standards before deployment.)

What We Don't Allow: Prohibited Practices

❌ Automated Decision-Making for Important Matters

Never: Use AI alone to deny loans, close accounts, or make other adverse decisions
Why: Important financial decisions require human judgment and consumer interaction

❌ Behavioral Manipulation

Never: Use AI to pressure consumers into products or services they don't need
Why: This violates consumer trust and community financial institution values

❌ Discriminatory Practices

Never: Allow AI to make decisions based on race, gender, age, or other protected characteristics
Why: It's illegal, unethical, and against everything CFIs stand for

❌ Privacy Violations

Never: Use consumer data beyond its stated purpose or share it inappropriately
Why: Consumer privacy is fundamental to trust and regulatory compliance

(The above prohibited uses are representative. Eltropy will continually evaluate new AI applications and explicitly forbid any that present ethical, legal, or reputational risks. When in doubt, the default is to err on the side of consumer protection and not pursue questionable AI use.)

03

Governance and Oversight Made Simple

Establishing AI Oversight Committees

Purpose and Objectives

AI Oversight Committees play a crucial role in overseeing the ethical development and deployment of AI systems within CFIs. These committees review the AI system's adherence to ethical principles, regulatory compliance, and the alignment of AI initiatives with organizational values and objectives.

Structure of the AI Oversight Committee

1. **Composition:** The committee should be composed of a diverse group of stakeholders, including senior management, AI experts, legal advisors, data scientists, and representatives from key business units.
2. **Leadership:** A designated AI Ethics Officer (CTO) or a senior executive should lead the committee to ensure accountability and effective decision-making.
3. **Meetings:** The committee should hold regular meetings to review AI projects, assess compliance with ethical standards, and address any emerging risks or concerns.

Key Responsibilities

1. **Policy Development:** Establish and maintain comprehensive AI ethics and governance policies that outline the organization's commitment to ethical AI practices.
2. **Review and Approval:** Review and approve AI projects to confirm they adhere to ethical guidelines, regulatory requirements, and organizational values.
3. **Monitoring and Evaluation:** Continuously monitor AI systems to detect and address any ethical or compliance issues, ensuring ongoing alignment with established policies.
4. **Risk Management:** Identify, assess, and mitigate risks associated with AI technologies, including biases, data privacy concerns, and security vulnerabilities.
5. **Training and Awareness:** Promote awareness and understanding of AI ethics and governance principles among employees through training programs and regular communications.

Roles and Responsibilities

AI Ethics Officer

1. **Leadership:** Lead the AI Oversight Committee and provide for the effective implementation of AI ethics and governance policies.
2. **Policy Enforcement:** Oversee the enforcement of ethical standards and regulatory compliance across all AI initiatives.
3. **Reporting:** Provide regular updates to senior management and the board of directors on the status of AI ethics and governance efforts.

Data Scientists and AI Developers

1. **Ethical Design:** Develop AI systems in accordance with ethical guidelines, ensuring fairness, transparency, and accountability.
2. **Bias Mitigation:** Implement techniques to detect and mitigate biases in AI models, continuously improving their fairness and accuracy.
3. **Documentation:** Maintain thorough documentation of AI development processes, including data sources, model architectures, and decision-making criteria.

Legal and Compliance Teams

1. **Regulatory Compliance:** Confirm that AI systems comply with relevant laws and regulations, such as GDPR, CCPA, and other industry-specific standards.

2. **Risk Assessment:** Conduct regular risk assessments to identify potential legal and ethical issues related to AI technologies.
3. **Advisory Role:** Provide legal guidance and support to the AI Oversight Committee and other stakeholders involved in AI projects.

Business Unit Representatives

1. **Alignment with Objectives:** Ensure that AI initiatives align with the strategic objectives and values of their respective business units.
2. **Stakeholder Engagement:** Engage with stakeholders to gather input and feedback on AI projects, ensuring they meet the needs and expectations of consumers and employees.
3. **Implementation Support:** Support the implementation and integration of AI systems within their business units, promoting ethical and responsible use.

IT and Security Teams

1. **Technical Safeguards:** Implement technical safeguards to protect the security and integrity of AI systems and data.
2. **Continuous Monitoring:** Continuously monitor AI systems for security vulnerabilities and compliance with ethical standards.
3. **Incident Response:** Develop and execute incident response plans to address any ethical or security breaches promptly and effectively.

04

Managing Risks Proactively

Understanding AI Risks in Simple Terms

The Risk Landscape: Where Each Risk Lives and who mitigates it

Risk Category	Specific Risk	Where It Lives	Who Mitigates It
Fairness & Bias	Bias in model training data	LLM model layer	Azure / Google + Eltropy application controls
Fairness & Bias	Disparate quality of responses across consumer groups	Output layer	Eltropy (audits, RAG grounding, human oversight)
Privacy	PII / PCI leakage in inputs	Input layer	Eltropy (redaction pipeline)
Privacy	PII reconstructed in outputs	Output layer	Eltropy (output scrubbing)
Privacy	Consumer data is not used for model training	LLM provider infrast...	Azure / Google contractual commitments
Security	Prompt injection attacks	Input layer	Eltropy (input filtering + system prompts)
Security	Harmful or out-of-scope content	Input / output layer	Eltropy (profanity & content filtering)
Transparency	Hallucinations / factual errors	LLM output layer	Eltropy (RAG architecture + output validation)
Transparency	consumers unaware they are interacting with AI	Application / UX layer	Eltropy (disclosure design + user education)
Transparency	Inability to explain AI decisions to regulators	Application layer	Eltropy (audit trails, citations, human-in-the-loop)

Eltropy's Risk Mitigation Strategy: Risk mitigation at different layers.

Layer 1: Model Foundation

Primary risks addressed: Fairness & Bias Risk · Privacy Risk · Security Risk

Before any data reaches Eltropy's application, our LLM providers, Microsoft Azure OpenAI and Google Gemini, enforce their own foundational safeguards. When Eltropy selects a third-party LLM provider, we evaluate their training methods, data sources, and built-in safety measures before partnering with them. This is our first line of defense against bias baked into the model itself, privacy violations at the infrastructure level, and security gaps in how the model handles inputs.

The model-level protections we verify and contractually require include:

- **No data retention:** Queries and responses are not stored beyond the immediate request, protecting consumer privacy at the infrastructure level
- **No training use:** Your data and your consumers' data is never used to train, fine-tune, or improve the underlying foundation models, a contractual commitment that aligns with privacy standards
- **Geographic controls:** All LLM processing occurs within US-based data centers, in compliance with data residency requirements
- **Built-in content safety:** Both Azure and Google enforce baseline content moderation and safety alignment at the model level, providing a foundational layer against Security Risk
- **Model bias controls:** We assess the steps taken by each model developer during training to understand bias exposure and use this assessment to inform the additional Fairness & Bias controls we apply at higher layers

Our AI architecture leverages the foundational security of enterprise providers who adhere to rigorous safety and privacy standards. Further, Eltropy conducts due diligence on third-party LLM providers to verify their alignment with our core principles. However, Eltropy does not rely on provider-level controls alone. The layers that follow describe the additional, Eltropy-specific protections we apply on top.

Layer 2: Programmable Guardrails (Input Controls)

Primary risks addressed: Privacy Risk · Security Risk · Fairness & Bias Risk

Everything a user types or submits passes through Eltropy's input pipeline before it is ever sent to the LLM. These are automated filters and rules that sit between the user and the model, reducing the likelihood of erroneous, harmful, or privacy-violating inputs reaching the LLM. This layer is our primary defense against Security Risk at the point of entry and our first enforcement point for Privacy Risk.

PII and PCI Redaction *(Privacy Risk)* Personally identifiable information (PII), such as Social Security numbers, account numbers, dates of birth, and contact details, as well as payment card information (PCI) is automatically detected and redacted from user inputs before they are transmitted to the LLM. This means sensitive consumer data never reaches the model in raw form, eliminating a critical Privacy Risk at source.

Note: *PII/PCI redaction is active across most Eltropy products. Where it is not yet fully deployed, this is tracked as a remediation item under our ongoing security roadmap.*

Profanity and Harmful Content Filtering *(Security Risk · Fairness & Bias Risk)* Inputs are screened for profanity, abusive language, and content that falls outside acceptable use boundaries before being processed. This prevents harmful content from influencing model behavior and reduces the risk of biased or inappropriate responses being triggered by manipulative inputs.

Prompt Injection Detection *(Security Risk)* Malicious actors may attempt to manipulate AI systems by embedding hidden instructions within their inputs, a technique called prompt injection. Left undetected, these attacks can cause the AI to behave outside its intended boundaries, expose data, or produce harmful outputs. Eltropy's input layer includes detection mechanisms to identify and neutralize these attempts before they can influence the LLM's behavior.

System Prompt Guardrails *(Security Risk · Fairness & Bias Risk · Transparency Risk)* Every Eltropy AI product is configured with a carefully designed system prompt that defines the model's role, scope, and behavioral boundaries. These prompts restrict the LLM to topics relevant to the community financial institution's services, prevent the model from acting outside its defined purpose, enforce consistent and equitable behavior across all consumer interactions, and establish the tone, constraints, and escalation behaviors the model must follow. For agentic AI, systems that can take actions on behalf of consumers, these

guardrails also include action-level constraints that prevent the agent from operating outside its approved boundaries.

Layer 3: Programmable Guardrails (Output Controls)

Primary risks addressed: Privacy Risk · Security Risk · Transparency Risk · Fairness & Bias Risk

LLM responses pass through Eltropy's output pipeline before being displayed to consumers or staff. Just as with inputs, automated filters sit between the model and the user to catch problems before they surface. This layer closes the loop on risks that the model may have introduced in its response, regardless of how clean the input was.

Profanity and Harmful Content Filtering *(Security Risk · Fairness & Bias Risk)* Outputs are screened for profanity, inappropriate content, and responses that fall outside the defined scope of the product, mirroring the input-side filtering to create a complete content safety boundary around the LLM. This also serves as a secondary check against biased or discriminatory language appearing in consumer-facing responses.

PII Scrubbing *(Privacy Risk)* Output responses are checked to ensure no PII has been inadvertently included or reconstructed by the model before being surfaced to the end user. This is a critical Privacy Risk control, even when inputs are redacted, models can occasionally reconstruct sensitive patterns from context, and this filter catches that before it reaches the consumer.

Response Validation *(Transparency Risk · Fairness & Bias Risk)* Responses are checked for format, relevance, and coherence against the query and the retrieved knowledge base content before delivery. This improves conformity to expectations of correctness, reduces the risk of unexpected output formats, and helps ensure that responses are consistent and equitable across different consumer queries, directly supporting both Transparency Risk and Fairness & Bias Risk management.

Layer 4: Application Design

Primary risks addressed: Transparency Risk · Fairness & Bias Risk · Privacy Risk

The most significant risk mitigations Eltropy applies are built into the architecture of our applications themselves. These are design-level mechanisms that go beyond filtering, they shape how the AI thinks, what it draws from, and how confident it needs to be before responding. This layer is our primary defense against Transparency Risk and our strongest architectural answer to Fairness & Bias Risk, because it constrains the model to verified, approved content rather than allowing it to reason freely from training data.

Retrieval-Augmented Generation (RAG) *(Transparency Risk · Fairness & Bias Risk)* RAG is central to how Eltropy's AI products work and is one of our most important tools for managing both Transparency Risk and Fairness & Bias Risk. Rather than asking an LLM to generate answers purely from its training data, which can lead to hallucinations, outdated information, or responses influenced by model-level bias, our system first retrieves relevant, verified content from the community financial institution's own approved knowledge base, then provides that content to the LLM as the basis for its response.

This means:

- Answers are grounded in documents the community financial institution has explicitly approved, not in the model's unconstrained interpretation
- The LLM is constrained to what is in the knowledge base, significantly reducing the influence of model-level bias on outputs
- Citations and sources are provided with every response, enabling human review before any decision is made, a direct Transparency Risk control
- The risk of fabricated or factually incorrect answers (hallucinations) is significantly reduced, supporting both Transparency Risk and regulatory explainability requirements

Confidence Thresholding *(Transparency Risk)* Eltropy's application logic includes confidence scoring. If the AI's certainty about an answer falls below a defined threshold, the system escalates to a human agent rather than delivering a potentially unreliable response to the consumer. This directly mitigates Transparency Risk by ensuring the AI does not present low-confidence outputs as authoritative answers.

Output Similarity Comparison *(Transparency Risk · Fairness & Bias Risk)* Responses are compared against expected outputs and knowledge base content to detect answers that deviate significantly from what the system should be producing, an additional check against hallucination, model drift, and inconsistent treatment of similar queries across different consumer groups.

Human-in-the-Loop Design *(Transparency Risk · Fairness & Bias Risk)* No Eltropy AI product is designed to make final decisions autonomously. AI outputs are presented as recommendations, summaries, or suggested responses, always subject to human review and approval before consequential action is taken. This is especially important for any output that could affect a consumer's financial standing, access to credit, or account status. Audit trails of all AI-assisted decisions are maintained to support regulatory review, directly addressing Transparency Risk and regulatory explainability obligations under ECOA and fair lending laws.

Scope Restriction *(Security Risk · Fairness & Bias Risk)* AI agents and assistants deployed through Eltropy are scoped to specific, predefined tasks. They cannot take actions outside their defined boundaries, and any attempt to operate outside scope triggers an escalation or a graceful handoff to a human representative, preventing both Security Risk from out-of-bounds behavior and Fairness & Bias Risk from the model operating in areas it has not been tested or validated for.

Layer 5: User Education

Primary risks addressed: Transparency Risk · Fairness & Bias Risk

Guardrails are only as effective as the people using the system. It is essential to design and position AI systems so that user accountability is part of the usage process. Transparency Risk in particular cannot be fully addressed through technology alone, consumers and staff must understand what AI is doing, what it cannot do, and how to apply their own judgment. Eltropy enforces guardrails at this layer by:

- **Disclosure *(Transparency Risk):*** Clearly informing consumers that they are interacting with an AI, not a human, before and during every AI interaction, ensuring no consumer is misled about the nature of the system they are using
- **Human escalation *(Transparency Risk · Fairness & Bias Risk):*** Providing easy, always-available pathways to escalate to a human agent at any point, ensuring consumers are never trapped in an AI interaction and always have access to human judgment
- **Communicating limitations *(Transparency Risk):*** Clearly explaining AI limitations to consumers and staff so that appropriate judgment is applied to AI outputs, especially for financially significant decisions

- **Staff training *(Fairness & Bias Risk · Transparency Risk):*** Providing community financial institution staff with training materials, usage guidelines, and best practices for responsible use of AI recommendations, including how to recognize and escalate potentially biased or incorrect outputs
- **Accountability by design *(Transparency Risk):*** Designing AI interfaces so that reviewing, confirming, or overriding AI suggestions is a natural and expected part of the workflow, not an optional extra step

---

Active Protections: What Is In Place Today

Protection	Layer	Risk Category
LLM provider safety alignment & content moderation	Model Foundation	Security · Bias
No data retention / no training use (contractual)	Model Foundation	Privacy
US-only data residency	Model Foundation	Privacy
Model bias assessment prior to vendor selection	Model Foundation	Fairness & Bias
PII / PCI redaction on inputs	Programmable Guardr...	Privacy
Profanity & harmful content filtering on inputs	Programmable Guardr...	Security · Bias
Prompt injection detection	Programmable Guardr...	Security
System prompt guardrails & scope restriction	Programmable Guardr...	Security · Bias · Trans...

Profanity & harmful content filtering on outputs	Programmable Guardr...	Security · Bias
PII scrubbing on outputs	Programmable Guardr...	Privacy
Response validation & output similarity comparison	Programmable Guardr...	Transparency · Bias
RAG-grounded responses with citations	Application Design	Transparency · Bias
Confidence thresholding	Application Design	Transparency
Human-in-the-loop escalation & audit trails	Application Design	Transparency · Bias
AI disclosure & human escalation pathways	User Education	Transparency
Staff training & usage guidelines	User Education	Transparency · Bias

---

Ongoing Risk Management

Risk management is not a one-time activity. Eltropy maintains the following ongoing practices across all risk categories:

1. **Continuous Monitoring** *(Security Risk · Transparency Risk)* AI system performance, error rates, escalation rates, and flagged interactions are monitored continuously. Anomalies trigger review by Eltropy's AI and security teams, with Security Risk events prioritized for immediate response.
2. **Regular Bias and Fairness Audits** *(Fairness & Bias Risk)* AI outputs are periodically reviewed for patterns that may indicate bias or disparate impact on consumer groups. Findings feed directly into product, prompt, and knowledge base improvements.

3. **LLM Provider Evaluation** *(Fairness & Bias Risk · Privacy Risk · Security Risk)* As the AI provider landscape evolves, Eltropy continuously evaluates both Azure OpenAI and Google Gemini against our security, privacy, safety, and fairness standards. New models are assessed before deployment. New providers undergo the full vendor due diligence process described in Section 10.
4. **Incident Response** *(All Risk Categories)* Eltropy maintains documented incident response procedures for AI-specific failures, including hallucination events, data handling anomalies, bias incidents, and security breaches. Any confirmed issue triggers immediate containment, root cause analysis, and remediation with communication to affected CFIs.
5. **Consumer and Staff Feedback Integration** *(Fairness & Bias Risk · Transparency Risk)* Feedback mechanisms built into Eltropy's products allow consumers and staff to flag AI responses that seem incorrect, biased, or inappropriate. This feedback is reviewed and used to improve system prompts, knowledge base content, and filtering rules on an ongoing basis, keeping human experience at the center of continuous improvement.

05

Ensuring Fairness and Preventing Bias

Introduction

This section aims to provide an overview of AI guardrails implementation strategy to mitigate the risks related to bias and fairness in AI. It covers the implementation strategy for technical safeguards and incorporating the right feedback tools and processes for continuous improvement.

Ensuring Fairness in AI Algorithms

Overview

Fairness in AI algorithms is crucial to ensure that AI systems do not perpetuate or exacerbate existing biases and inequalities. Ensuring fairness involves designing AI models that treat all individuals equitably, regardless of their demographic or personal characteristics. This requires a proactive approach to identify, prevent, and mitigate biases throughout the AI development lifecycle.

Principles of Fairness in AI

1. **Equity:** AI systems should provide equitable outcomes for all individuals, avoiding favoritism or discrimination based on characteristics such as race, gender, age, or socioeconomic status.
2. **Inclusivity:** AI models should be trained on diverse and representative data sets that reflect the varied experiences and characteristics of the population they serve.
3. **Accountability:** Developers and organizations must take responsibility for the fairness of their AI systems, implementing protections against producing biased outcomes.
4. **Transparency:** The decision-making processes of AI systems should be transparent, allowing stakeholders to understand and challenge the fairness of AI-driven outcomes.

Bias Detection and Mitigation Techniques

Identifying Bias in AI Models

1. **Data Analysis:** Conduct thorough analyses of training data to identify potential biases, such as underrepresentation or overrepresentation of certain groups.
2. **Fairness Metrics:** Utilize fairness metrics to evaluate AI models, including disparate impact, equal opportunity, and demographic parity, to assess whether the model treats different groups equitably.
3. **Bias Audits:** Perform regular bias audits during the AI development and deployment phases to detect and address biases that may arise.

Mitigating Bias in AI Models

1. **Pre-processing Techniques:** Modify training data to reduce biases before training AI models. This can include re-sampling, re-weighting, or anonymizing data to ensure a more balanced representation.

2. **In-processing Techniques:** Adjust the learning algorithms to incorporate fairness constraints directly into the model training process. Techniques such as adversarial debiasing and fairness constraints can help reduce bias during model development.
3. **Post-processing Techniques:** Alter the outputs of AI models to provide fairness after the model has been trained. This can include re-ranking, threshold adjustment, or other methods to correct biased outcomes.

Regular Audits for Bias and Discrimination

Importance of Regular Audits

Regular audits are essential to maintain the fairness and integrity of AI systems over time. They help AI models continue to operate equitably and prevent the development of biases as they encounter new data or evolving environments.

Audit Processes

1. **Frequency:** Conduct bias audits at regular intervals, such as quarterly or bi-annually, and after significant updates or changes to the AI system.
2. **Scope:** Define the scope of each audit to cover all aspects of the AI system, including data sources, model performance, and decision-making processes.
3. **Stakeholder Involvement:** Involve diverse stakeholders, including data scientists, ethicists, legal advisors, and representatives from affected communities, to provide comprehensive perspectives during audits.

Key Audit Activities

1. **Data Review:** Analyze the data used for training, validation, and real-time operations to identify potential sources of bias or discrimination.
2. **Model Evaluation:** Assess AI models using fairness metrics and bias detection techniques to evaluate their performance and identify any biased behaviors.
3. **Outcome Analysis:** Examine the outcomes produced by AI systems to confirm they are fair and non-discriminatory, considering the impact on different demographic groups.
4. **Compliance Checks:** Verify that AI systems comply with relevant laws, regulations, and organizational policies related to fairness and non-discrimination.

5. **Corrective Actions:** Implement corrective actions to address any identified biases, including retraining models, adjusting decision-making processes, and updating policies or procedures.

Documentation and Reporting

1. **Audit Reports:** Document the findings of each bias audit, including identified biases, mitigation actions taken, and recommendations for improvement.
2. **Transparency:** Share audit reports with relevant stakeholders, including internal teams, regulatory bodies, and, where appropriate, the public, to demonstrate commitment to fairness and accountability.
3. **Continuous Improvement:** Use the insights gained from bias audits to continuously improve AI systems, updating training data, algorithms, and processes to enhance fairness and reduce discrimination over time.

06

AI Transparency and Explainability

Introduction

This section provides a comprehensive framework for ensuring transparency and explainability in AI operations. It covers strategies and techniques for making AI models more understandable and their decisions more transparent to stakeholders, particularly CFIs. The goal is to build trust and accountability by providing clear insights into how AI systems operate and make decisions.

Ensuring Transparency in AI Operations

Overview

Transparency in AI operations involves making the inner workings of AI systems accessible and understandable to stakeholders. This includes clarifying how AI models are developed, how they make decisions, and how they are monitored for compliance and performance.

Key Components of Transparency

1. **Documentation:** Comprehensive documentation of AI models, including data sources, model architectures, training processes, and performance metrics.
2. **Governance Policies:** Clear policies and procedures governing the development, deployment, and monitoring of AI systems.
3. **Stakeholder Involvement:** Engaging stakeholders throughout the AI lifecycle to ensure their perspectives and concerns are considered.
4. **Performance Metrics:** Regular reporting of AI system performance, including accuracy, fairness, and reliability metrics.
5. **Audit Trails:** Maintaining detailed logs of AI system operations, decisions, and updates to facilitate audits and reviews.

Implementation Strategies

1. **Developing Detailed Documentation:** Ensure all aspects of AI systems are well-documented, including data preprocessing steps, feature selection, model training procedures, and decision-making processes.
2. **Establishing Clear Governance Policies:** Create and enforce governance policies that define roles, responsibilities, and procedures for AI development and oversight.
3. **Engaging Stakeholders:** Regularly consult with stakeholders, including consumers, employees, and regulatory bodies, to gather feedback and ensure transparency.
4. **Reporting Performance Metrics:** Implement a reporting mechanism to regularly share AI system performance metrics with stakeholders.
5. **Maintaining Audit Trails:** Ensure that all AI system operations and decisions are logged and can be reviewed during audits.

Explainability Techniques for AI Models

Overview

Explainability involves making AI models and their decisions understandable to humans. It helps stakeholders comprehend how AI systems work and why they make specific decisions, fostering trust and accountability.

Key Explainability Techniques

1. **Feature Importance:** Identifying and ranking the features that most influence the AI model's decisions.
2. **Rule-Based Approaches:** Converting complex model decisions into simpler, rule-based explanations.
3. **Counterfactual Explanations:** Providing examples of how different inputs could lead to different outputs, helping to illustrate decision boundaries.

Implementation Strategies

1. **Applying Feature Importance Methods:** Use techniques like permutation feature importance or tree-based feature importance to rank and explain the influence of different features on model decisions.
2. **Utilizing Model-Agnostic Methods:** Implement LIME and SHAP to generate local explanations for individual predictions and global explanations for overall model behavior.

3. **Leveraging Visualization Tools:** Develop dashboards and visual tools that illustrate how AI models process data and make decisions.
4. **Implementing Rule-Based Approaches:** Where possible, translate complex AI model decisions into simple rules that can be easily understood by stakeholders.
5. **Generating Counterfactual Explanations:** Use counterfactual techniques to show stakeholders how different inputs would affect model outcomes, helping them understand decision boundaries and model sensitivity.

Communicating AI Decisions to CFIs

Overview

Effective communication of AI decisions is essential for building trust and ensuring that stakeholders understand and accept AI-driven outcomes. This involves providing clear, accessible explanations of how decisions are made and their implications.

Key Communication Strategies

1. **Simplifying Technical Jargon:** Translate complex technical concepts into simple, layman's terms that can be easily understood by non-experts.
2. **Using Case Studies and Examples:** Provide real-world examples and case studies to illustrate how AI decisions are made and their impacts.
3. **Interactive Demonstrations:** Use interactive tools and demonstrations to show how AI models work and respond to different inputs.
4. **Regular Updates and Reports:** Provide regular updates and detailed reports on AI system performance, decisions, and any changes or improvements made.
5. **Feedback Mechanisms:** Implement channels for stakeholders to provide feedback and ask questions about AI decisions.

Implementation Strategies

1. **Creating Accessible Documentation:** Develop user-friendly documentation that explains AI models and their decisions in clear, non-technical language.
2. **Developing Case Studies:** Compile case studies that demonstrate how AI decisions are made and their real-world applications and impacts.
3. **Using Interactive Tools:** Create interactive tools that allow stakeholders to experiment with AI models and see how different inputs affect outcomes.

4. **Providing Regular Updates:** Establish a schedule for regular updates and reports on AI system performance and decision-making processes.
5. **Implementing Feedback Channels:** Set up mechanisms for stakeholders to provide feedback, ask questions, and receive answers about AI decisions.

Ensuring transparency and explainability in AI operations is critical for building trust and accountability in AI systems. By implementing detailed documentation, clear governance policies, and effective communication strategies, CFIs can make AI systems more transparent and their decisions more understandable to stakeholders. This framework provides a roadmap for achieving transparency and explainability in AI, fostering trust, and ensuring responsible and ethical AI use.

07

Data Privacy and Security in Plain Sight

Where Your Data Lives

Storage Location: Secure data centers in the United States Encryption Standard: Bank-level AES-256 encryption at rest (the same protection used by major financial institutions)
Access Controls: Only authorized personnel can access data, with full audit trails

How We Protect Your Information

Before Processing

- **PII Redaction:** Personal information is automatically masked before storing
- **Access Controls:** Only approved users can access specific information
- **Authentication:** Multiple verification steps before any data access

During Processing

- **Secure Transmission:** All data travels through encrypted connections
- **Minimal Data Use:** We only process information necessary for the requested service
- **No Training Use:** Your data never trains AI models or improves external systems

After Processing

- **Secure Storage:** Results are encrypted and stored with restricted access
- **Audit Trails:** Complete records of who accessed what information when
- **Retention Policies:** Data is kept only as long as necessary and then securely deleted

Your AI Provider Partners

Primary LLM Provider: Microsoft Azure OpenAI , Google Gemini

Please note: This content is provided for general informational purposes only and reflects Eltropy's current approach to AI safety and innovation as of March 2026. It does not constitute legal or financial advice. All institutions are encouraged to consult with their own legal counsel regarding the deployment of AI technologies.

Why This Matters:

- Your data stays within Microsoft's secure environment / Google Secure environment
- Information is never shared with OpenAI directly (for OpenAI models)
- Data isn't used to train external models or the foundation models from these partners
- Full compliance with financial industry regulations

08

Communicating with Consumers About AI

When We Use AI

- We tell consumers upfront when they're talking to an AI system
- We explain what the AI can and cannot do
- Consumers must agree to interact with AI before continuing
- We always provide an easy way to switch to a human agent

Consumer Rights

- Consumers can see transcripts of AI conversations
- Consumers can request corrections to their data
- We don't discriminate against any consumer groups
- Consumers can easily opt out and talk to humans instead

Who's Responsible

- AI Product Owner: Makes sure AI disclosures work properly
- Developers: Build the consent and opt-out features
- Legal/Compliance: Reviews all AI-related policies
- Support Team: Handles consumer concerns and escalations

- Communications: Trains staff and informs consumers

09

Managing AI Vendors

Third-Party Vendor Management of AI Service Providers

Eltropy's SafeAI Framework requires that our selected third-party LLM service providers meet the highest standards for security, regulatory compliance, and ethical practices. The following principles govern our vendor selection, oversight, and due diligence.

---

1. Commitment to Ethical and Regulatory Excellence

Data Privacy & Protection: We utilize the cloud vendors (AWS and GCP) state-of-the-art security measures, including robust encryption, granular access controls, and strict data residency practices, aligned with regulatory frameworks (e.g., GDPR, GLBA, HIPAA). All data processing and storage must occur only in approved data center regions. Vendors are contractually obligated to process consumer data in accordance with these geographical requirements and are never permitted to use our data or our customers' data to train, fine-tune, or otherwise enhance their AI systems.

Protecting consumer Privacy and Data Security: Vendors must have robust privacy and data security measures such as end-to-end encryption, data minimization, and strict access controls. All consumer data must be handled solely for the purposes of providing our service and must not be stored beyond immediate processing requirements.

Mitigating Risks and Avoiding Bias: Vendors must use safe AI practices, including regular algorithm audits, transparency in decision-making processes, and built-in human oversight. These measures facilitate AI outcomes that are fair, equitable, and free from bias.

---

2. Rigorous Vendor Selection & Enhanced Due Diligence

Security & Privacy Assessments

- **Encryption & Access Controls:** All data must be encrypted using industry-standard methods (e.g., AES-256) with role-based access controls rigorously enforced.
- **Data Isolation & No Retention:** Data must be processed in isolated environments. Vendor systems must be designed to prevent retention of LLM query/response data.
- **No Data Retention Policy:** Vendors must provide written assurances that LLM inputs/outputs are processed only on a per-request basis and discarded immediately after processing.

Regulatory & Compliance Verification

- **Certifications & Audits:** Up-to-date certifications (e.g., SOC 2) and independent audit reports are required.
- **Contractual Non-Usage Clause:** Our data and our customers' data must never be used for training or any secondary purpose.

Financial & Operational Stability

- Vendors are evaluated on financial health, support track record, and long-term service continuity plans.

Model Development & Guardrails

- **Training Data Integrity:** Proprietary or sensitive data must never be integrated into broader AI training pipelines.
- **Risk Mitigation:** Comprehensive model-level guardrails must be in place to mitigate risks such as bias and data leakage, with transparent documentation of safety and alignment processes.

Availability, SLAs & Data Center Region Requirements

- **High Availability**: Vendors must maintain high uptime and have redundancy, failover mechanisms, and robust disaster recovery plans.

- **Detailed SLAs:** Service Level Agreements must specify uptime guarantees, incident response times, and remediation measures with clear penalties for service disruptions.
- **Data Center Regions:** All data processing and storage must occur exclusively in approved regions. Vendors must document data center locations, comply with local data residency laws, and notify us immediately of any changes.

3. Robust Contractual Agreements

Our contracts with LLM vendors include data protection and performance standard clauses:

- **Data Ownership & Confidentiality:** Vendors must not use or store our data beyond immediate processing and are expressly forbidden from using it for AI training or secondary purposes.
- **Performance Standards & SLAs:** Contracts include detailed SLAs specifying uptime (e.g., 99.9% availability), incident response, and resolution timelines, with clear penalties for breaches.
- **Data Center Region Obligations:** Vendors must agree that all data is processed and stored exclusively within pre-approved regions, in accordance with applicable data residency requirements.
- **Long-Term Support & Maintenance:** Vendors must provide ongoing updates, proactive security enhancements, and continuous technical support.

4. Continuous Performance Monitoring & Risk Management

- **Ongoing Oversight:** Continuous monitoring frameworks include security audits, real-time alerting, and regular performance monitoring to verify adherence to contractual obligations.
- **Risk Management:** Risk assessment is performed for every vendor, covering security, compliance, operational controls.
- **Incident Response & Remediation:** Robust, documented incident response plans address any issues related to service availability, or regulatory breaches immediately.
- **Availability & Data Center Assessments:** Uptime metrics, SLA compliance, and data center location adherence are reviewed regularly. Any deviations trigger immediate corrective action.

5. Enforcing Model-Level Guardrails

- **Training & Alignment Controls:** Vendors must integrate model-level guardrails throughout the model's development lifecycle, ensuring no sensitive data is used improperly and that risks of bias or error are minimized. Examples include profanity detection and mitigation, and steps taken to ensure reduced or no bias in model outputs.
- **Transparency & Reporting:** Vendors must provide regular, detailed reports (excluding sensitive information) on model performance, risk mitigation efforts, and any incidents related to data handling or bias.

10
Best Practices

Eltropy is committed to responsibly harnessing AI to serve CFIs. This policy outlines our "Safe AI" best practices for implementing AI solutions – including third-party large language models (LLMs), retrieval-augmented generation (RAG) systems, and agentic AI (chatbots, virtual assistants) – in a way that maximizes value to our customers while minimizing risks. It aligns with Eltropy's Safe AI principles and guardrails, ensuring AI is used *responsibly, safely, ethically, and in line with human values* . By adhering to these guidelines, Eltropy's engineering and implementation teams, as well as our community financial institution partners, can deploy AI solutions that are reliable, compliant, and effective without disrupting operations.

Guiding Principles: Eltropy's AI initiatives are grounded in the following core principles, which reflect our Safe AI and industry best practices:

- **Reliability & Accuracy:** Prioritize factual correctness and consistency in AI outputs. Implement measures to reduce hallucinations (fabricated answers) and ensure the AI provides accurate, up-to-date information . AI systems should be thoroughly tested so they perform reliably under expected conditions (e.g. high consumer inquiry volumes) without failures or wild deviations.

- **Security & Privacy:** Protect consumer data and sensitive information at all stages. This includes secure integration with third-party LLMs (e.g. using encryption, access control) and preventing any unintended data leakage. Strict filters and context-aware checks must prevent AI from revealing confidential data . All data handling must comply with privacy laws and Eltropy's data security policies.

- **Compliance & Accountability:** Ensure AI solutions operate within the stringent regulations governing financial services. Systems are built with transparency so that their decisions or recommendations can be explained and audited. *Regulatory compliance is a top priority* – AI tools must respect consumer protection laws (e.g. fair lending rules) and produce outputs that adhere to industry regulations . Eltropy takes accountability for AI outcomes by maintaining audit logs and enabling oversight of AI-driven interactions.

- **Fairness & Ethical Use:** Proactively mitigate bias in AI models. The AI should treat all users fairly and be blind to sensitive characteristics like race or gender in decision-making . We employ rigorous testing and ongoing monitoring to ensure the AI does not perpetuate unfair biases or unethical behavior. Ethical guidelines (aligned with our values and the CFI's mission of inclusivity) are integrated from design through deployment.

- **Transparency & Trust:** Build AI systems that are transparent in operation and limitations. Consumers and staff interacting with AI should be informed they are engaging with an AI assistant and understand its role. Where applicable, AI-driven recommendations or decisions should be explainable in plain language to maintain trust . We foster trust by educating users on AI capabilities and limitations, and by providing clear fallback options to human support whenever needed.

- **Human-Centric Design:** Use AI to augment and empower humans, not to replace the human touch that is central to CFI. Our AI chatbots and assistants are designed to enhance consumer service and employee productivity, while seamlessly handing off to human agents for complex issues or on request . Human oversight is built into the loop – critical decisions or exceptions are reviewed by staff, ensuring AI remains a tool under human control rather than an autonomous authority .

By following these principles, we aim to deploy trustworthy AI tools that deliver tangible benefits (faster service, improved insights, efficiency gains) *in line with the values and regulatory standards of CFIs*.

11
Building Tomorrow's AI

Your Ongoing Commitment to Excellence

Developing safe and effective AI chatbots isn't a destination, it's a continuous journey of learning, refining, and innovating. This framework has equipped you with the essential principles and practical strategies needed to build AI assistants that users can trust and rely on within the Eltropy platform.

The Path Forward

The AI landscape moves fast. New breakthroughs, regulatory frameworks, and best practices emerge regularly, reshaping how we approach responsible AI development. That's why we're committed to keeping our guidelines current and our training materials fresh. But staying ahead requires partnership, your curiosity, vigilance, and real-world insights are invaluable. The AI landscape moves fast. New breakthroughs, regulatory frameworks, and best practices emerge regularly, including the rise of agentic AI, where systems don't just respond but autonomously act on behalf of consumers and staff. That's why we're committed to keeping our guidelines current and our training materials fresh as we responsibly expand into these more autonomous capabilities.

Your Role as an AI Steward

- Learn continuously: Extract lessons from every deployment
- Share actively: Your discoveries strengthen the entire team
- Adapt thoughtfully: Embrace change while maintaining our core principles
- Question boldly: Challenge assumptions and push for better solutions

Our North Star

No matter how sophisticated our technology becomes, one principle remains unchanging: we deliver exceptional value through AI while never compromising safety, ethics, or user trust. This isn't just a guideline, it's our promise to every user who interacts with our systems.

A Living Document, A Growing Community

This framework will evolve as rapidly as the technology it governs. Your experiences, insights, and feedback don't just inform our future updates, they shape the very culture of responsible AI development we're building together.

Thank you for embracing these principles and helping us create AI that truly serves people. Together, we're not just building better chatbots, we're pioneering a more trustworthy digital future.

12

The Evolving AI Regulatory Landscape

In 2026, AI service providers and CFIs operate in a "two-track reality". Because there is no single comprehensive federal AI law, we must simultaneously navigate a fast-moving patchwork of state mandates while aligning with existing financial and other regulations that are already being applied to AI. Compliance is not a static checkbox but a continuous process of monitoring and adaptation.

Our Four Pillars of Regulatory Alignment

- **Consumer Protection (UDAAP):** Regulators have made it clear that "existing laws apply to AI". We monitor for Unfair, Deceptive, or Abusive Acts or Practices to promote AI interactions, like chatbot responses, that remain truthful and fair.
- **Privacy & Data Integrity:** We align with GLBA and evolving state privacy laws to validate sensitive consumer data that is not used for unauthorized AI training.
- **State-Level Management:** State-level activity is accelerating, with major laws in California (SB 53) and Texas (RAIGA) taking effect in January 2026, followed by Colorado in June. We monitor these state-by-state requirements to maintain national service standards.
- **Industry Best Practices:** At the federal level, the AI landscape is primarily shaped by Executive Orders and agency-led enforcement rather than a single, comprehensive “AI Act” from Congress.

13

Frequently Asked Questions

Where is the Data Stored?

Documents (Knowledge) uploaded by the Community Financial Institution are securely stored in Eltropy's document store, with logical separation to ensure bot access is restricted only to authorized documents. These documents are encrypted using server-side encryption with AES-256 standard. Metadata and extracted text data is stored in the database encrypted at rest with the same encryption standard.

What Region is the Data Stored in?

The data is stored in Amazon Web Services (AWS) - US regions.

What is the Encryption Standard Used in Storage?

- Document store: Encrypted using AES-256 encryption.
- Metadata and Extracted Text Data: Encrypted at rest using AES-256 encryption.

Who is the LLM Provider?

Microsoft Azure hosts the OpenAI Models used. In future releases, we will be integrating other large language models such as Google Gemini, Anthropic Claude, etc., maintaining all security and safe AI measures in place to adhere to the same standards. Eltropy commits to implementing industry-standard security and safe AI practices, recognizing that while no system can incorporate every possible safeguard, we follow established recommendations and best practices to protect consumer data and maintain responsible AI deployment.

What Region is the LLM Hosted in?

The LLM is hosted in the US

Difference Between Data Handled by GenAI and Knowledge Assist:

- GenAI Knowledge Assistant (Employee Facing): Private, with access restricted to specific users on the Eltropy platform based on user groups.
- GenAI Agent Chat (Consumer Facing): Public, accessible to any user (Consumer/Lead) via the conversation panel. Ensure only publicly accessible knowledge is uploaded.

What Data is Sent to the LLM?

User queries and extracted text data from documents are sent through a PII and PCI redaction pipeline to the LLM to answer user queries.

AI Guardrails (Available in GA, limited in BETA)

Both user queries and extracted text data are processed through a PII and PCI redaction pipeline before being sent to the LLM.

Data Privacy and Security Policy of Azure and Google for LLM Models

The Azure OpenAI Service is fully controlled by Microsoft and does not interact with any OpenAI-operated services. Additionally we also partner with Google Cloud for Gemini Models. The data sent to the LLM is:

- Not available to other consumers.
- Not available to OpenAI, Google or any other Foundation model provider for training.
- Not used to improve OpenAI, Google or any other foundation model providers offerings..
- Not used to improve any Microsoft, Google or other third-party products or services.
- Not used for automatically improving Azure OpenAI models or any other foundation model for your use in your resource.

For more information on Azure AI Privacy policies, please refer to the documentation: Azure OpenAI Data Privacy.

For more information on Google Cloud Gemini , please refer to the documentation here: Google Paid API Services Data Policies

For Community Financial Institution Leadership

Q: How do we know AI is making fair decisions?
A: We use multiple fairness metrics, conduct regular bias audits, and maintain human oversight for all important decisions. Every AI recommendation includes transparency features showing how conclusions were reached.

Q: What happens if something goes wrong?
A: We have incident response procedures that immediately flag problems, alert human staff, and implement corrective measures. All AI decisions can be reviewed and overridden by qualified staff.

Q: How much will this cost our community financial institution?
A: AI implementation costs vary based on your needs, but most CFIs see positive ROI within 6-12 months through improved efficiency and consumer satisfaction.

For Community Financial Institution Staff

Q: Will AI replace my job?
A: No. AI is designed to handle routine tasks so you can focus on complex consumer needs, relationship building, and strategic work that requires human judgment and empathy.

Q: How do I know when to trust AI recommendations?
A: Always verify AI suggestions, especially for important decisions. Look for confidence scores, check cited sources, and use your professional judgment to evaluate recommendations.

Q: What if I don't understand how the AI reached a conclusion?
A: Ask questions! AI systems should be able to explain their reasoning. If you can't understand the logic, escalate to your supervisor or the AI oversight committee.

For Consumers

Q: How do I know when I'm interacting with AI?
A: AI systems will always identify themselves clearly. You'll see labels like "AI Assistant" or receive explicit notifications when AI is involved in your interaction.

Q: Can I opt out of AI interactions?
A: Absolutely. You can always request to speak with a human representative. AI is designed to enhance service, not replace human interaction when you prefer it.

Q: Is my personal information safe with AI?
A: Yes. We use bank-level security, encrypt all data, and never share your information inappropriately. AI systems process only the minimum information needed to help you.