Carve Generative AI Governance Framework

Valency's strategy for responsible, secure, and transparent use of AI in project assurance

wwww.valencyinc.com

Table of Contents

Introduction 2

Design Principles 2

- Human-in-the-loop 2
- Privacy First Architecture 2
- Professional and Ethical Standards 2

Risk Management 3

- Human Oversight to Prevent Misuse 3
- Quality Standards and Ethical Safeguards 3
- Prompt Injection Protection 3
- Data Privacy and Model Containment 3

Compliance Commitments 3

- Canada's Voluntary Code of Conduct for Generative AI 3
- EU Artificial Intelligence Act 4
- U.S. Framework Alignment for Generative AI 4

Future Outlook 5

Introduction

Valency has introduced generative AI capabilities within the Carve project assurance solution to enhance facilitator insights and drastically reduce the time required to write concise assessment insights. This Governance Framework outlines our responsible approach to deploying Generative AI while safeguarding intellectual property, promoting transparency, and meeting the expectations of global organizations.

Design Principles

Valency's approach to generative AI is grounded in principles that reflect the needs of our clients – global organizations that demand transparency, trust, and control. Our design framework ensures that AI integration supports users without compromising data security, compliance, or professional integrity.

Human-in-the-loop oversight
All AI-generated content within Carve is designed to assist, not replace, human judgment. Users remain fully in control of whether and how they engage with AI features. AI-generated summaries are clearly labeled, and every summary must be reviewed and approved by a human before it is included in an Executive Summary. This model ensures that assessment insights reflect the knowledge and discretion of experienced professionals.

Privacy-first architecture
Carve is built to meet the data protection needs of globally distributed organizations. All data interactions with the large language model (LLM) occur within Valency's secured infrastructure, hosted on AWS Bedrock. The LLM operates within the same regional instance selected by each customer, guaranteeing data residency and sovereignty. Customer data is never transferred across borders, shared with external providers, or used to train or fine-tune public AI models. Prompts and responses are not stored or logged, preserving confidentiality.

Professional & ethical standards
We evaluated and selected our foundational LLM models based on their alignment with workplace norms, professional tone, and ethical integrity. Our AI features are developed through a disciplined, repeatable process that includes prompt engineering, behavioral testing, and continuous improvement based on user feedback.

The Director of Software Engineering is accountable for the oversight, refinement, and performance of AI models, ensuring they meet Valency's internal quality and integrity standards.

Risk Management

Human Oversight to Prevent Misuse
All AI-generated summaries are reviewed and approved by a Carve user (human-in-the-loop) before finalization, mitigating the risk of hallucinated or misleading insights reaching project stakeholders.

Quality Standards and Ethical Safeguards
Valency addresses the inherent risks of generative AI with a robust QA process. Each insight is evaluated using a rubric that includes Professionalism & Ethical Compliance as a core criterion. Outputs are expected to be culturally sensitive and free from offensive, biased, or inappropriate language.

Prompt Injection Protection
Valency reduces the risk of prompt injection using two key principles:

- Least Privilege: The LLM accesses only the minimum data required to generate an insight.
- Controlled Prompt Engineering: Prompts are professionally designed to prevent adversarial manipulation. End users cannot write or modify prompts.

Data Privacy and Model Containment
Valency ensures strict data privacy and model containment:

- No customer data is ever used to train or fine-tune AI models.
- All LLM requests remain fully within Valency-controlled infrastructure—no data is transmitted externally.

Compliance Commitments

Valency aligns its generative AI practices with evolving international frameworks for responsible AI deployment. This includes both voluntary and regulatory codes of conduct

Canada's Voluntary Code of Conduct for Generative AI
Although Canada's Code of Conduct for Generative AI is voluntary, Valency treats its principles as mandatory guardrails in our application of AI. We support and implement the following key principles:

- **Transparency:** AI-generated content is clearly labeled, and users are informed when they are interacting with AI features.
- **Accountability:** Internal oversight is led by Valency's Director of Software Engineering, ensuring responsible deployment and ongoing model evaluation.

- **Safety:** The underlying LLM includes safeguards to minimize harmful, biased, or inappropriate outputs.
- **Fairness:** The LLMs upholds workplace norms and supports equitable use across industries and geographies.
- **Human Oversight:** AI outputs are never final without user review, reinforcing human judgment in all project assurance workflows.

These commitments guide our ongoing application of AI features in Carve and reflect our dedication to responsible innovation.

EU Artificial Intelligence Act

Based on Carve's current functionality and the definitions outlined in **Article 52 of the EU AI Act**, Valency has determined that its AI usage qualifies as 'limited risk.' We will reassess this classification regularly in line with regulatory updates and functional changes. We meet the required obligations through the following measures:

- **Clear Disclosure:** AI-generated content is always labeled so users understand when they are interacting with AI.
- **Human Oversight:** All AI-generated summaries are reviewed and approved by a Carve user before they are included in deliverables.
- **User Awareness:** Carve provides context and transparency about what the AI feature does—and does not do—ensuring users understand its role and limitations.
- **Professional Output:** The AI model is tuned to maintain a neutral, workplace-appropriate tone, avoiding manipulation or misleading suggestions.

US Framework Alignment for Generative AI

While the United States does not yet have a unified federal law governing AI, Valency proactively aligns with key federal guidance and emerging standards to ensure responsible implementation. Our approach reflects the following frameworks:

- **NIST AI Risk Management Framework (RMF 1.0).** We apply the core principles of trustworthy AI to our internal practices — governance, data quality, transparency, and risk mitigation—aligned with the National Institute of Standards and Technology's voluntary guidelines.
- Recent White House Executive Orders on AI. We monitor and respond to Executive Orders (e.g. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Oct 2023) that promote safe, secure, and trustworthy AI development, including expectations for transparency, risk assessments, and content provenance.

These efforts ensure Valency stays ahead of regulatory trends while meeting the expectations of our international client base for ethical and transparent use of AI.

Future Outlook

Valency's application of generative AI in Carve is just beginning. While our initial focus is generation of insights for executive summaries, we are actively exploring expanded use cases. Future applications will follow the same rigorous governance framework.

References

Canadian Voluntary Code of Conduct for the Responsible Development and Management of Advanced Generative AI Systems. Innovation, Science and Economic Development Canada (ISED).

EU Artificial Intelligence Act (2024). European Parliament and Council Regulation laying down harmonised rules on Artificial Intelligence (AI Act).

NIST AI Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, U.S. Department of Commerce. January 2023.

White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Executive Order 14110, issued October 30, 2023. Office of the President of the United States.

151 Frobisher Drive, Unit B105
Waterloo, ON, CANADA N2V 2C9
+1 (519) 883-7136
Email: info@valencyinc.com
www.valencyinc.com