VERSATILITY

Our AI, Our Accountability: Shadow AI Discovery and Governance at Versa

Anusha Vaidyanathan, Sr. Director, Product Management

Jon Taylor

Shadow AI by the numbers

1 in 5

organizations have already had a breach tied to shadow AI

Adds an average of $670K to breach cost.

68% of employees use free-tier AI tools via personal accounts

57% input sensitive company data into public GenAI tools

63% of organizations have no AI governance policy

Sources: IBM Cost of a Data Breach Report 2025; Menlo Security 2025 State of Browser Security; TELUS Digital, Jan 2025 (n=1,000 enterprise employees).

VERSATILITY

Shadow AI is not Shadow IT

DATA FLOWS OUT

Employees paste source code, PII, and financial data into chatbots — bypassing DLP.

OUTPUT IS THE RISK

Generated code and uploaded files can become training data for third parties.

WEEKS, NOT YEARS

Teams adopt new GenAI tools daily; a browser is the only requirement.

AGENTS TAKE ACTION

Agents chain actions across systems with full user privileges — no audit trail.

VERSATILITY

What Shadow AI actually looks like

Consumer GenAI chatbots. ChatGPT, Claude, Gemini, Perplexity — accessed from corporate devices and BYOD.

Embedded AI in SaaS. Copilot, Notion AI, Salesforce Einstein, Slack AI — already live, often via auto-enable.

BYO models & APIs. Developers calling OpenAI, Anthropic, or HuggingFace APIs directly from code.

Agents and MCP servers. Autonomous AI that reads files, writes to systems, and chains tool calls — often with broad credentials.

AI Governance at Versa

VERSATILITY

Versa's Commitment to Responsible AI

EU AI Act

Regulation 2024/1689

All Versa AI systems mapped to risk classes. None classified as Unacceptable or High Risk.

NIST AI RMF 1.0

Risk Management Framework

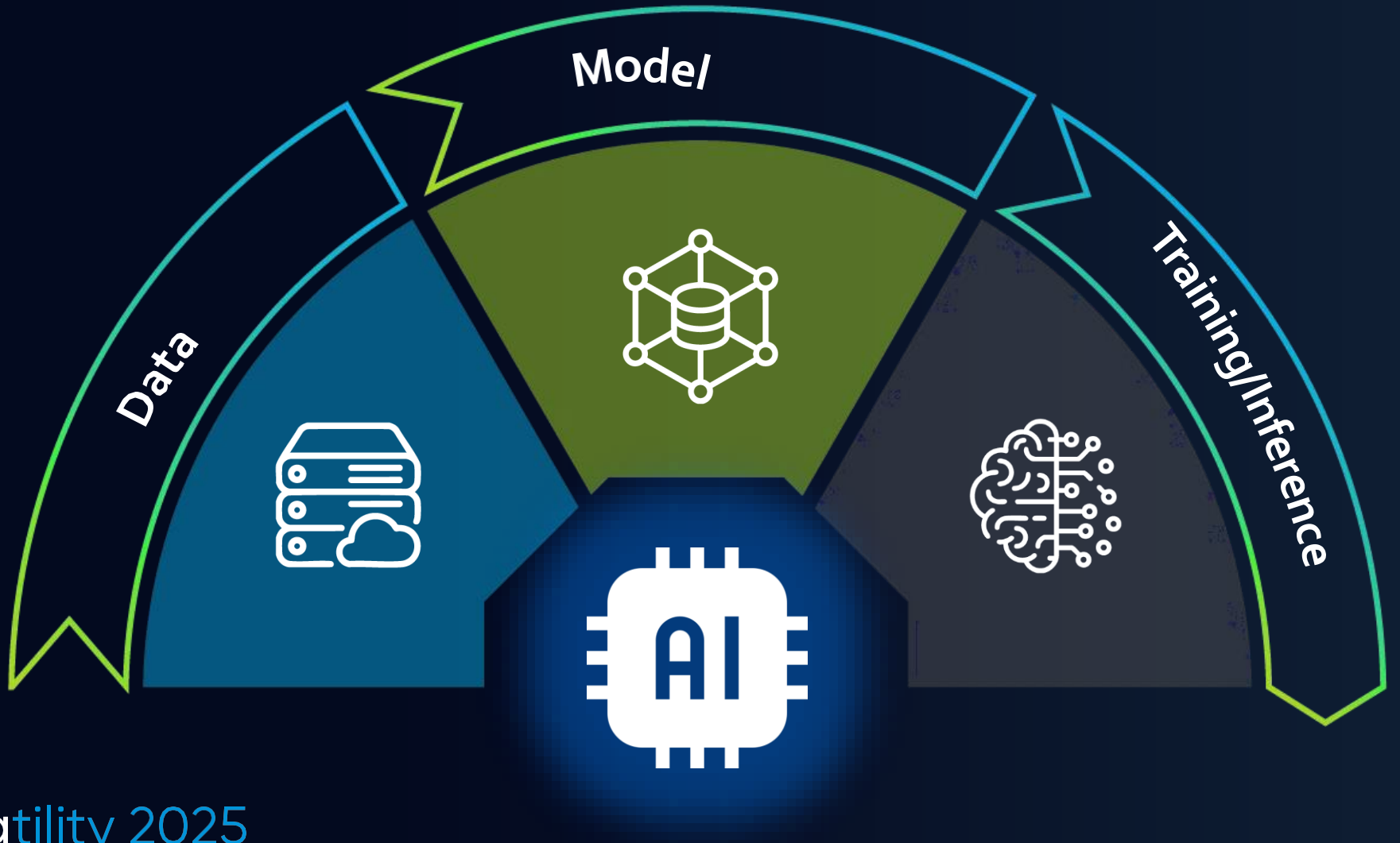
Govern, Map, Measure, Manage. Aligns AI risk practices to a US standard recognized across enterprises.

ISO/IEC 42001

AI Management System

End-to-end controls for AI lifecycle: design, data, deployment, monitoring, and continual improvement.

Components of AI



VERSA

EU AI Act Highlights and Recommendations

Risk-Based Classification

- Unacceptable risk (banned)
- High risk (regulated)
- Limited risk (transparency)
- Minimal risk (unregulated)

Prohibited AI Practices

- Social scoring systems
- Real-time biometric ID in public
- Emotion recognition in workplaces/schools
- Systems exploiting vulnerabilitie

Transparency Requirements

- Disclosure for AI-generated content
- Clear labeling of chatbots/deepfakes

High-Risk AI Requirements

- Risk assessment and mitigation
- High-quality data governance
- Technical documentation
- Human oversight
- Robustness and security

Governance & Enforcement

- National competent authorities
- AI Office within Commission
- Scientific Panel of experts
- Penalties: up to €35M or 7% of turnover

Implementation

- Regulatory sandboxes for innovation
- Special provisions for SMEs
- Gradual timeline (24 months after adoption)

AI Systems & EU AI Act Risk Classification

Product	Models	EU AI Act Risk	Model Source
Verbo (with Zero Trust MCP server)	4	Minimal	3rd-party LLMs (opt-in)
VANI (analytics)	4 hybrid	Minimal	100% proprietary
UEBA	2 hybrid	Limited	100% proprietary
ATP	40+	Minimal	Open-source
AI DLP	5	Limited	100% proprietary

No Versa AI system is High Risk or Prohibited — Conformity Assessment, FRIA, and DPIA under EU AI Act Articles 27/43 are not required.

Data Handling & Tenant Isolation

Multi-Tenant Isolation

Per-tenant isolation across control, data, and management planes.

No Generic Training

Customer data never used for generic training. Fine-tuning is opt-in only.

PII at Inference Only

PII scanned at inference for detection — never stored or trained on.

Encryption & Access

Encrypted at rest and in transit. Role-based access with full audit trails.

Data Residency & Deployment Modes

Verbo deploys flexibly with any head-end variant

- **On-prem** — full data sovereignty
- **Cloud-hosted** — GCP, US region today
- **Expanding** — Europe, APAC tied to public cloud regions

Explainability, Audit & Oversight

Explainability

- **Verbo, VANI, UEBA —** dedicated explainability models
- **AI DLP —** task-specific explainable outputs
- **ML lifecycle —** full audit of experiments, versions, metrics
- **QA signals —** Verbo chat feedback, ATP drift detection

Audit & Oversight

- **Cross-functional review —** Product, Engineering, Privacy, Legal
- **Admin override —** for block, quarantine, UEBA risk scoring
- **Third-party models —** inventory in Versa Trust Center
- **Contact —** compliance@versa-networks.com / PSIRT

Explainability & Quality Assurance

MLOps Pipeline

- Model training vs Inference

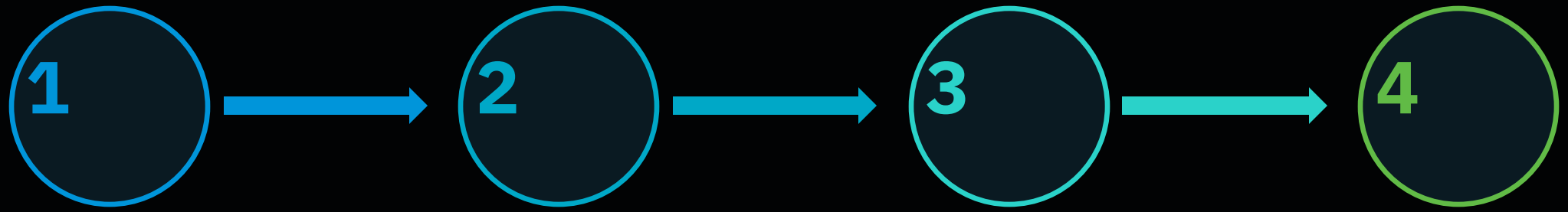
Versa AI

- Models & Systems
- Components & Algorithms
- Tech Stacks
- Explainability & Testing

FAQs

The MLOps Pipeline

Four pillars — Versa does all of them



Data
Collect, clean & prepare

Training
Build foundational model

Testing & Fine-Tuning
Validate, specialize, refine

Inference
Deploy to production

Model Training

Data: Millions–billions
Compute: GPU clusters · weeks
Cost: $$$

VERSA EXAMPLES
- Malware Detection (40+ models)
- Mistral-7B + ModernBERT

Fine-Tuning

Data: Thousands
Compute: Single GPU · hours–days
Cost: $$

VERSA EXAMPLES
- AI-Powered DLP (5 models)
- UEBA + VANI (6 models)

Inference

Data: Single samples
Compute: CPU/GPU/edge · ms
Cost: $

VERSA EXAMPLES
- Verbo (Reranker, Qwen3)
- VersaGPT (GPT-4o + RAG)

VERSA

Differentiator: Versa does all three

VersaAI™ - Models & Systems

Threat Protection

AI/ML Malware Detection

40+ models, 3-4 per file type
ELF, PE, JS, PDF, OLE
Mistral-7B + ModernBERT

Data Protection

AI-Powered DLP

5 models, 1 per task
Source code & PII detection
Image classification & OCR

AI Ops

Versa UEBA + VANI

6 models total, none proprietary
1 model for explainability
Anomaly + behavior analytics

Co-Pilots

Verbo

Reranker (BGE)
Embedder (Qwen3)
Versa GPT (GPT-4o)

VERSATILITY

VersaAI™ - Components & Algorithms

Threat Protection

Malware Detection

Tree-based classifiers
Deep Neural Networks
Embedding + Transformer LLMs

Data Protection

GenAI Firewall + DLP

Built on URL filtering + DLP
CNN-based classifiers
Named Entity Recognition (NER)

AI Ops

UEBA + VANI

Graph DB + unified query
Search, retrieval, Graph ML
Time-series + Isolation Forest

Co-Pilots

Verbo + Versa GPT

OpenAI API on LLMs
Retrieval-augmented (RAG)
Tool call / MCP debugger

Explainability/Testing

- **MLflow / ClearML**— for MLOps tracking
- **Experiment tracking**— params, metrics, artifacts
- **Model registry**— versioning, staging, production
- **Deployment**— Docker, Kubernetes, Cloud
- **CI/CD integration**

Example: Verbo Explainability

- **Synthetic + human-verified data**— for testing
- **Tool-call validation**— right tool, right context
- **Example**— VersaGPT vs. MCP vs. debugger
- **User feedback loop**— thumbs up / down on answers

Secure AI. Enable Innovation.

Versa GenAI Firewall delivers multi-layered defense across thousands of AI applications, integrated natively within the Versa SASE platform. No separate agent. No separate product. One architecture.

4 Layers
of defense

1000s
of GenAI apps

39 AppIDs
with DPI

9 CASB Apps
activity control

VERSATILITY

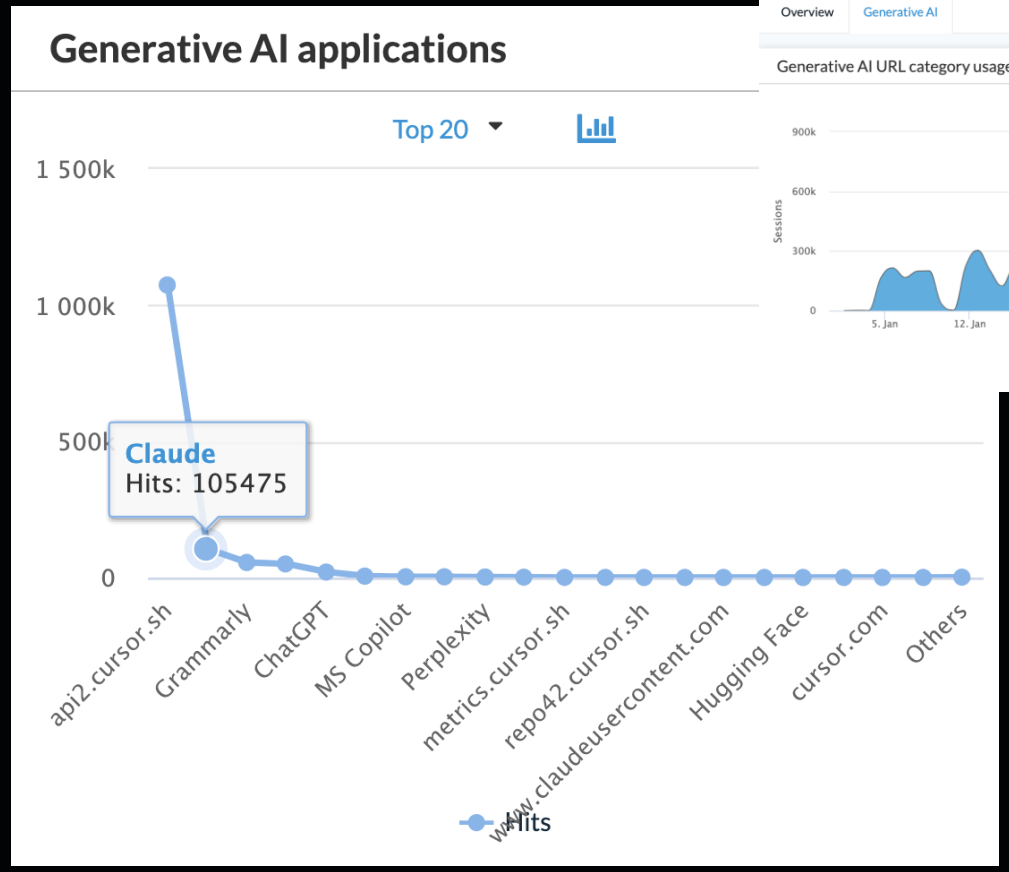
Solution – Versa Generative AI Firewall (dog food n/w)

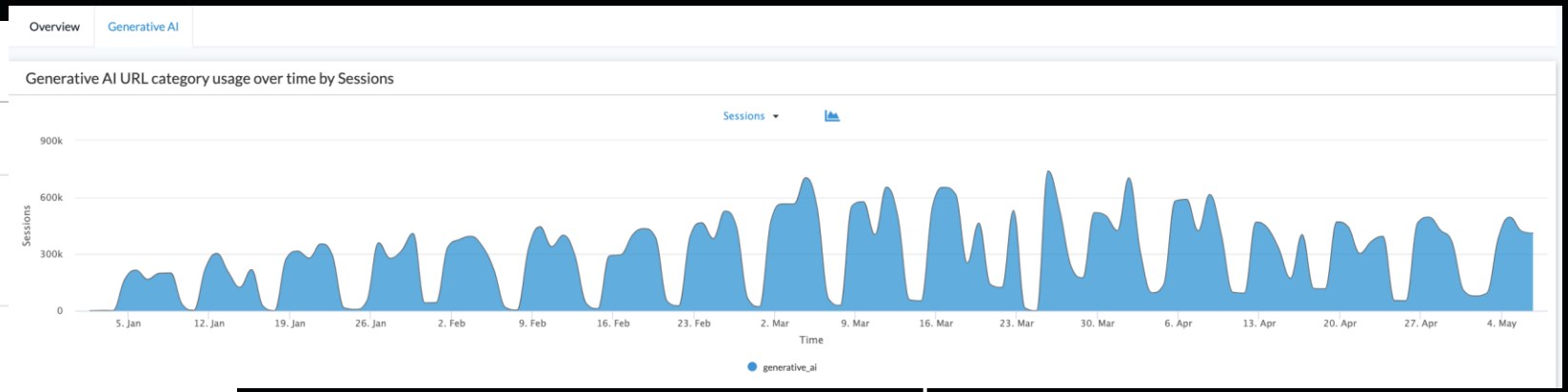
Visibility: Get the insight into who is using generative AI tools and how

Security: Prevent access to risky Gen AI sites and protect sensitive data exfiltration

Compliance: Risk based analysis of Gen AI applications

Versa Shadow AI Discovery





Monday – Thursday productivity blip

Top Gen AI apps
Cursor – enterprise subscription 80%
Claude 7%
Grammarly4%
ChatGPT, Perplexity, Copilot, Huggingface

VERSATILITY

Drops

Generative AI SASE Web Monitoring logs (drop-session) (ALS powered)

Show Domain Names

Click to set a filter Apply Clear

Application	User	App Category	URL Category	URL Reputation	SSL Decrypted	SSL Version	Policy Action	Policy Module	Policy Rule	Server Name Indication	Traffic Scope	Flow Duration (ms)
chatgpt	akini@versa-networks.com	web	generative_ai	trustworthy	no		drop-session	policy	Drop_Quic_App	chatgpt.com		11s
chatgpt	akini@versa-networks.com	web	generative_ai	trustworthy	no		drop-session	policy	Drop_Quic_App	chatgpt.com		17s
ssl	Unknown	encrypted	generative_ai	trustworthy	no		drop-session	idp	Allow_From_Trust	api.anthropic.com		4h 25m 50s
ssl	Unknown	encrypted	generative_ai	trustworthy	no		drop-session	idp	Allow_From_Trust	api.anthropic.com		4h 23m 44s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		6h 35m 41s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		6h 25m 23s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		6h 29m 1s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		6h 17m 28s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		7h 32m 48s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		15h 47m 48s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api3.cursor.sh		4h 24m 30s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api3.cursor.sh		8h 36m 47s
ssl	Unknown	encrypted	generative_ai	low_risk	no		drop-session	idp	Allow_From_Trust	api2.cursor.sh		9h 13m 28s

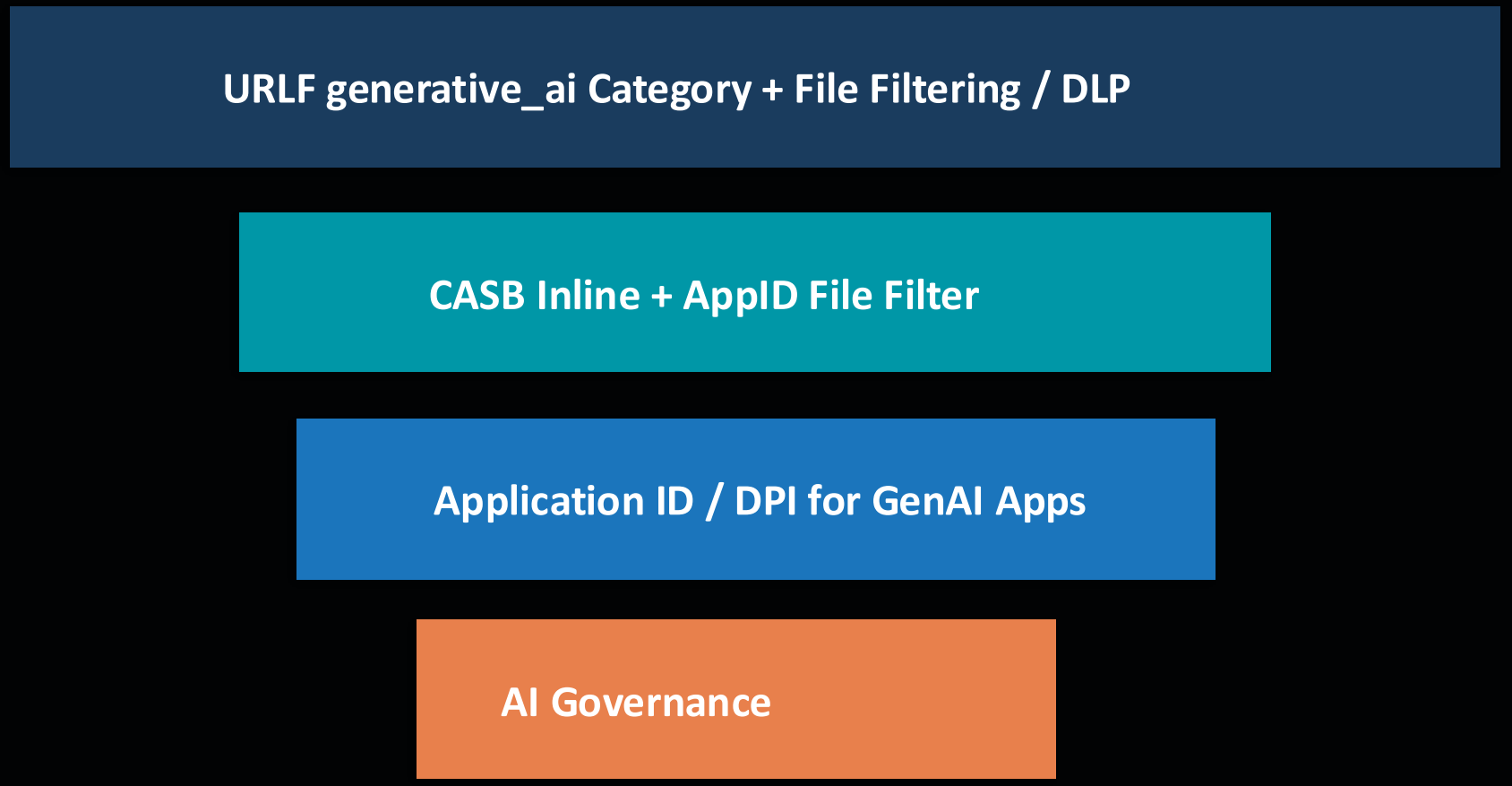
QUIC is dropped, renegotiated

Encrypted long lived session dropped

VERSATILITY

GenAI Firewall: Defense in Depth

Layered security controls from broadest coverage to granular content inspection



~1,000s of GenAI apps, domain based
File Filter blocks binary uploads; DLP inspects content

Blanket Block on all Uploads with 39 AppIDs
CASB Inline - 9 apps, 12 activity types
71 granular app-activity pairs
Requires TLS Decryption

39 GenAI apps
Requires TLS Decryption

Model-level detection, AI safety scoring, prompt injection, Shadow AI
End-2026

VERSATILITY

Layer 1: URLF Category + File Filtering / DLP

Broadest Coverage: Thousands of GenAI Applications

Option A: File Filtering (Upload Block)

- Blocks binary file uploads (DOCX, PDF, XLSX, images, archives) to any GenAI app in the generative_ai URL category
- Covers thousands of apps automatically via security package updates
- Cannot block TXT/JSON files (prompts use the same content types)
- **Best for: blanket file upload prevention across all GenAI**

Option B: DLP Content Inspection

- DLP profiles attached to all URLs matching generative_ai category, inspecting both prompts and file content
- Detects PII, source code patterns, financial data, credentials, and custom dictionaries
- Inspects text prompts (not just files), catching copy-paste data leakage
- **Best for: content-aware protection with selective blocking**

Both options can be combined: File Filter blocks uploads broadly; DLP inspects allowed text prompts for sensitive content.

GenAI Granular Activity Controls

12 activity types across 9 GenAI applications (71 app-activity pairs)

Authentication

login, login_failed, login_successful, logout

Control user access and session management

Data Transfer

upload_file, download_file

Control data movement to/from GenAI apps

Collaboration

share, post

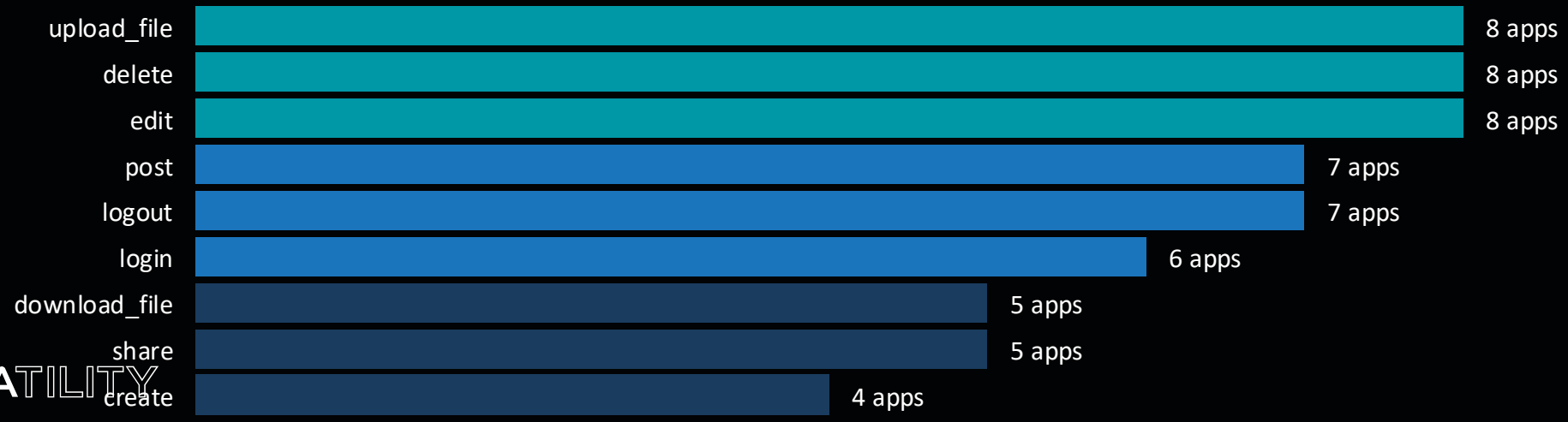
Control content sharing and prompt submissions

Content Mgmt

create, edit, delete, preview

Control content lifecycle within GenAI apps

TOP ACTIVITIES BY GENAI APP COVERAGE



VERSATILITY

Layer 3: AppID + DLP (Content Inspection)

When uploads are allowed, DLP inspects content for sensitive data

Upload Allowed in Layer 2

CASB policy permits file upload for this user/app combination

AppID Identifies GenAI App

DPI-based identification for GenAI apps provides app context

DLP Scans Content

Inspects file content and text payloads for PII, source code, credentials, financial data

Policy Action

Allow, block, alert, or quarantine based on DLP match and severity

DLP Detection: PII (SSN, credit cards, emails) | Source Code (regex patterns) | Financial Data | Credentials (API keys, tokens) | Custom Dictionaries

Layer 4: AI Governance

Shadow AI Dashboard

App inventory with sanctioned/shadow classification. Model detection per app. Department usage analytics. Newly discovered apps timeline.

Foundation Model Detection

3-stage inline pipeline: Provider ID, Model Extraction, Normalization. First-to-market SASE capability. No competitor does this.

AI Safety Scoring

Versa AI Safety Rating (0-100, A-F grades). Per-model risk dimensions. Policy enforcement by safety grade (e.g., block models rated below C).

Prompt Injection Detection

Dual-engine: PromptGuard 2 + DeBERTa v3 consensus. Detects jailbreak attempts, prompt injection attacks, and alignment failures.

Code Shield

Regex + Semgrep scanning of AI-generated code responses. Detects 50+ CWE vulnerabilities before code reaches the developer.

Per-Model Policy Engine

Block or allow specific foundation models (e.g., allow GPT-4o, block GPT-3.5). Risk-grade based policies. Unified governance rules.