

AI Security & Data Handling

EDPLAY INC.

A Delaware Corporation

2217 Pine Valley Drive, Houston, TX 77019, USA

Version 2.0 (last updated 13 July 2026)

Document: Appendix A - Edplay Platform Security and Data Handling

Version: 2.0

Last updated: 13 July 2026

An overview of the security architecture, AI processing controls, and data protection measures applied across the Edplay platform.

Highlights

Customer-selected data region: Core customer data is stored and processed in the customer's selected region - San Francisco (US), Frankfurt (EU) or London (UK) - as chosen in the Order Form, on DigitalOcean.

Workspace isolation: Strict tenant isolation per workspace across data, AI context and access controls.

PII redaction before AI: Microsoft Presidio, self-hosted by Edplay, redacts personal data before any AI provider call.

Prompt injection protection: Inbound prompts are scored and filtered with logged threat levels and matched patterns.

Encrypted AI chat histories: XChaCha20-Poly1305 with per-user DEKs wrapped by a Master Encryption Key.

Audit & monitoring: AI usage, encryption events, prompt safety and platform activity are continuously logged.

1. Platform Overview

Edplay is a cloud-hosted SaaS platform for creating and delivering AI-assisted educational content. The platform supports course generation, page and asset creation, AI tutoring, quiz generation, document-based knowledge retrieval, image generation and speech-to-text functionality.

The system is designed so that users interact with Edplay's application layer, rather than directly with third-party AI providers. AI requests are routed through Edplay's backend, where security, tenant isolation, personal data redaction, prompt safety and logging controls are applied before any request is sent to an AI provider.

2. Hosting and Infrastructure

Edplay is hosted on DigitalOcean. Core customer data is stored and processed in the region the Customer selects in the Order Form - San Francisco (United States), Frankfurt (Germany, EU) or London (United Kingdom). The platform uses a separately served frontend and backend architecture.

Component	Description
Frontend	Vite-based single-page application using Vue.js

Component	Description
Backend	Laravel 11 PHP application running behind Nginx
Containers	Docker-based deployment with PHP-FPM and Supervisor
Database	PostgreSQL 16
Object storage	DigitalOcean Spaces, S3-compatible storage
Queue / cache layer	Redis, with Laravel Horizon for queue monitoring
Vector database	Qdrant for embeddings and semantic search, hosted in the Customer's selected region consistent with primary data

Production and staging environments are separated.

3. Application Architecture

The backend is structured as a modular Laravel application. Core modules include authentication, workspaces, AI, course builder, billing, integrations, knowledge base, product, notifications and security. Traffic is routed through an Nginx reverse proxy to the Laravel application, which applies middleware controls for authentication, rate limiting, CORS, CSRF protection and workspace resolution.

Edplay uses a workspace-based multi-tenant model. Customer data is logically scoped to the relevant workspace using workspace-aware application logic, global query scopes and middleware. SAML SSO can be configured per workspace with isolated certificates and endpoints.

4. AI Processing Architecture

Edplay uses AI to support course creation and learning workflows, including course structure generation, educational page and content generation, interactive asset generation, image generation, retrieval-augmented generation, AI tutor functionality, quiz generation and speech-to-text transcription.

Edplay uses third-party AI providers including OpenAI, Anthropic, Google/Gemini, Stability AI and ElevenLabs. AI requests are brokered through the Edplay backend; users do not connect directly to these providers.

API data sent to AI providers is not used for model training by default under the relevant provider API terms. Edplay does not send individual user identifiers, email addresses, usernames, customer IDs or session identifiers to AI providers. AI providers receive only the sanitised content required for the AI task, system instructions, tool definitions and relevant redacted chat context.

5. Personal Data Protection Before AI Processing

Before user-submitted content reaches any AI provider, Edplay applies an automated sanitisation pipeline. This pipeline uses self-hosted Microsoft Presidio services within Edplay's own infrastructure to detect and redact personal data. It applies to chat messages, course-builder prompts, AI tutor interactions, admin chat interactions, uploaded PDFs, JSON content, audio and video transcripts, and document content used for knowledge-base and RAG workflows.

Detected personal data is replaced with generic entity markers such as <PERSON>, <EMAIL_ADDRESS>, <PHONE_NUMBER> or <LOCATION>. The original values are not sent to AI providers and are not recoverable from the redacted output.

The redaction control operates on a fail-closed basis. If the PII redaction service is unavailable or returns an error while PII protection is enabled, the AI request is blocked rather than sent unredacted.

6. Document Processing and RAG Controls

Where customers upload documents for knowledge-base or retrieval-augmented generation use cases, Edplay extracts the relevant text and applies the same sanitisation pipeline before chunking, embedding or storage in the vector database. Document content is therefore redacted before it is embedded and before it can be retrieved into AI context.

Document & RAG processing flow: (1) document upload; (2) text extraction; (3) PII redaction + prompt-injection guard; (4) chunking & embedding; (5) Qdrant vector store; (6) RAG retrieval into AI context.

7. Prompt Injection Protection

Edplay applies prompt-injection controls to user messages and ingested document content before it reaches AI providers or is stored in the vector database. The guard checks for common attack patterns, including instruction override, role impersonation, jailbreak attempts, prompt extraction and delimiter injection, and applies a risk-scoring model:

Risk level	Action
Low	Message passes unchanged
Medium	Message is sanitised with a defensive instruction
High	Message is blocked before reaching the AI provider

This protection also applies to uploaded document content before it enters the knowledge base, reducing the risk of malicious instructions being introduced into RAG context. AI agents operate in a restricted environment: they do not have access to environment variables, API keys, database credentials or infrastructure secrets, and agent tools are purpose-built, schema-validated and limited to their intended functions.

8. Encryption and Data Security

Edplay protects data in transit using HTTPS across all endpoints. Secure cookies are enabled, and the platform applies HTTP security headers, including frame protection, content-type protection, HSTS in production, referrer-policy controls and restrictive permissions policies. Data-at-rest protections include database backup encryption, server-side encryption for object storage and application-level encryption for sensitive AI chat histories.

AI chat histories are encrypted using XChaCha20-Poly1305 with per-user Data Encryption Keys (DEKs). A Master Encryption Key wraps the per-user DEKs as wrapped Key Encryption Keys (KEKs), and key operations are logged. This design means that compromise of one user's encryption key does not expose other users' encrypted chat histories.

9. Identity and Access Controls

- OAuth2 using Laravel Passport.
- SAML SSO with per-workspace configuration.

- Google, Microsoft, LinkedIn and Slack OAuth login.
- MFA using authenticator apps, email OTP or phone OTP.
- Role-based access control with Owner, Admin, Member and Guest roles.
- Policy-based authorisation per workspace.
- CSRF protection.
- LTI token authentication for supported education integrations.

10. Monitoring, Audit and Logging

- Sentry for error and performance monitoring.
- Laravel Horizon for queue monitoring.
- Application logs and Slack alerts for critical events.
- AI usage logs recording provider, model, token usage, cost, operation type and user/workspace context.
- Encryption audit logs for key generation, rotation, bulk encryption activity and decryption failures.
- Prompt-injection logs recording threat level, score, matched patterns and request context.

Sentry is configured not to send default PII, and SQL query bindings are excluded from Sentry breadcrumbs.

11. Secrets and Provider Access

API keys and sensitive provider credentials are stored outside the application code. Edplay uses environment-based secret handling and a local-network vault. AI provider access uses organisation-level API keys rather than per-user tokens, meaning AI providers see Edplay as the API customer rather than individual platform users.

12. Third-Party Services

Edplay uses third-party services for hosting, AI processing, vector storage, payments, error monitoring, authentication, email, video integration and brand intelligence. Key providers include DigitalOcean, Qdrant, OpenAI, Anthropic, Google/Gemini, Stability AI, ElevenLabs, Stripe, Sentry, Pusher and supported email providers. Microsoft Presidio is used for PII detection and redaction but is self-hosted by Edplay, so the redaction process itself does not require sending content to an external Presidio service.

Summary

Edplay is designed as a backend-controlled AI SaaS platform with workspace-based tenant isolation, encrypted storage, controlled AI provider access, automated PII redaction, prompt-injection protection, role-based access controls and operational audit logging. The architecture is intended to ensure that customer content is processed through Edplay's security and privacy controls before it reaches third-party AI providers, while maintaining visibility over AI usage, encryption events, prompt-safety events and platform activity.

Version 2.0 · Last updated 13 July 2026. For questions, contact dpo@edplay.ai.