



How Rapyder's Generative AI Services Helped an EdTech Platform Cut Release Effort by 80%

Customer:

A Leading
EdTech Company

Industry:

EdTech – AI Learning

Offering:

AWS Bedrock +
Generative AI

AWS Services Used:

- » Amazon Bedrock
- » Lambda
- » OpenSearch
- » ECS/ECR
- » SQS
- » DynamoDB
- » API Gateway



Introduction:

Customer is a Chennai-based EdTech company building an agentic, generative-AI learning platform that personalizes education to each learner's style and pace. The platform generates custom study materials, tutorials, quizzes, and real-time doubt resolution, alongside comic-style visual narratives, podcast-style audio, animated explainers, and short instructional videos generated directly from learning content.

Business Challenges

» Manual server logins and ad-hoc deploy scripts made releases slow and inconsistent across every surface of the platform – RAG tutoring, comics, podcasts, notebooks, and video.

» Secrets were bundled with build artifacts or hardcoded in the application, creating security risk on every release and making key rotation difficult.

» No unified architecture existed for RAG tutoring, assessment generation, and video workflows. Scattered infrastructure slowed content generation and raised operating costs.

» Ad-hoc compute and storage for generative video workloads limited the scalability of animated and text-to-video learning content.

» Legacy and newer login modes coexisted without strict scoping, risking cross-account data bleed in saved learner progress.

Solution Implemented

» An automated CodePipeline → CodeBuild → CodeDeploy release flow with zero-downtime Application Load Balancer traffic shifting, enabling daily, predictable releases.

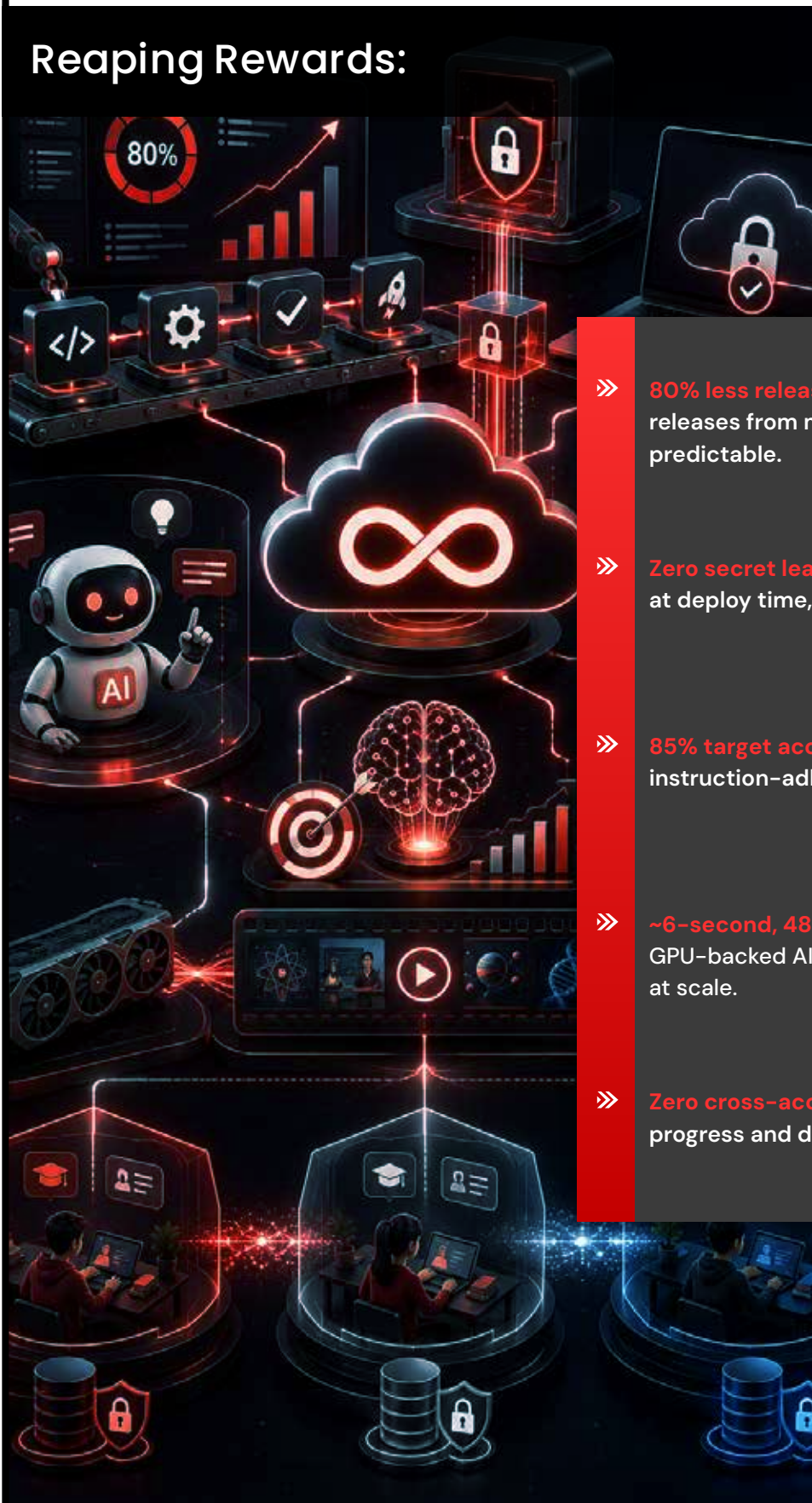
» Deploy-time secret injection from a dedicated, IAM-controlled S3 bucket – credentials never touch source control or build artifacts.

» A unified AWS Bedrock architecture: S3 and Lambda-based embeddings feed an Amazon OpenSearch knowledge base, powering grounded, context-aware RAG tutoring.

» A GPU-backed Amazon ECS pipeline running the WAN 2.1 model, with Meta Llama 3 70B for prompt optimization, Nova Canvas for imagery, and Amazon Polly for narration.

» JWT-backed sessions with strict per-account data scoping in Amazon DynamoDB, keeping learner progress and preferences correctly isolated.

Reaping Rewards:



- » **80% less release effort** – automated CI/CD took releases from manual and error-prone to fast and predictable.
- » **Zero secret leaks** – credentials are injected securely at deploy time, never stored in code.
- » **85% target accuracy** – AI tutoring delivers grounded, instruction-adherent answers for every learner.
- » **~6-second, 480p video clips generated reliably** – GPU-backed AI turns study content into video lessons at scale.
- » **Zero cross-account data bleed** – every learner's progress and data stay fully isolated.

