Updated Architecture Diagram – Full Feature Specification

This document provides a comprehensive, production-ready architecture specification for the Automotive and Diesel LMS platform. It integrates AI curriculum generation, media generation, simulation assets, and scalable infrastructure with explicit support for Stable Diffusion, Ollama, and Blender.

1. High-Level Architecture Overview

The system follows a layered architecture designed for scalability, GPU acceleration, and provider abstraction. Frontend clients interact with backend APIs, which orchestrate AI and rendering services and persist results to centralized storage.

2. Layered Architecture Breakdown

Frontend Layer

- Next.js LMS UI and Docs Site
- Course catalog and thumbnail rendering
- Instructor and admin dashboards
- Student content consumption

Edge & Routing Layer

- Nginx reverse proxy
- TLS termination
- Internal routing to backend APIs

Backend Application Layer

- Phoenix API (Elixir): authentication, job orchestration, metadata
- FastAPI (Python): AI/media gateway and provider routing
- Unified job status and retry logic

AI & Media Services Layer

- Ollama: LLM-based syllabus, lesson, and prompt generation
- Blender Worker: 3D models, lab environments, renders
- Stable Diffusion WebUI (AUTOMATIC1111): image and thumbnail generation via SDAPI

Storage Layer

- PostgreSQL: job state, assets, metadata
- S3 / MinIO: images, thumbnails, models, exported media

Infrastructure & Operations Layer

- Docker & Kubernetes orchestration
- GPU-enabled nodes for AI workloads
- Health checks, readiness probes, and monitoring

3. Stable Diffusion Integration (AUTOMATIC1111)

Stable Diffusion runs as a dedicated image-generation service exposing the SDAPI interface. It is responsible for generating course thumbnails, concept art, and documentation imagery.

Launch requirements:
- --api (mandatory)
- --medvram (recommended for 4GB GPUs)
- --no-half --precision full (prevents FP16 NaN errors on GTX 1650-class hardware)

4. Image Generation Provider Interface

The platform defines an abstract Image Generation Provider interface to decouple the LMS from specific rendering engines.

Provider contract:
GenerateImage(prompt, size, steps, cfg, sampler, seed) → PNG

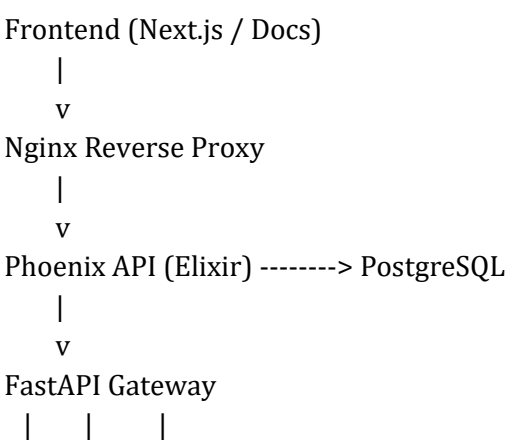
Supported providers:
- webui (AUTOMATIC1111 SDAPI)
- comfyui (future)
- cloud-based diffusion APIs (future)

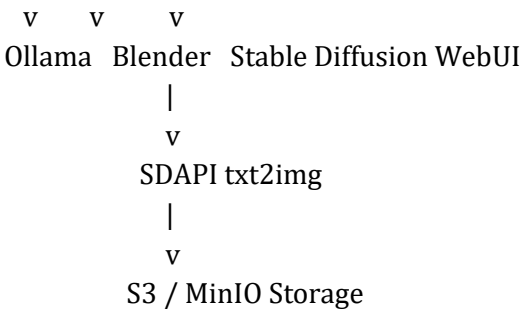
5. Kubernetes Deployment Model

In cluster deployments, Stable Diffusion runs on GPU-enabled nodes with node affinity. Services are exposed internally via ClusterIP.

Operational characteristics:
- Readiness probe: GET /sdapi/v1/options
- Liveness probe: TCP 7860
- GPU scheduling via nodeSelector / tolerations
- Horizontal scaling disabled (GPU-bound workload)

6. Visual Architecture Diagram (Logical)





7. Accreditation & Documentation Notes

This architecture supports accreditation and instructional documentation requirements by clearly separating concerns, ensuring reproducibility of generated assets, and maintaining auditable job metadata. The provider abstraction allows future upgrades without architectural changes.