Building AI Into a GEO Platform

An AI company focused on Generative Engine Optimization (GEO) helps brands make their products more discoverable, accurate, and persuasive across AI-driven answer engines such as ChatGPT, Perplexity, Gemini, and AI-powered search.

Its platform takes raw commerce catalogs, including product titles, descriptions, FAQs, and image alt text, and transforms them into AI-optimized content that remains grounded in source facts, follows each brand's voice and policies, and is structured for how AI systems retrieve and reason over information.

The platform uses multi-agent AI workflows, web-grounded research, brand-aware prompting, anti-hallucination controls, and continuous evaluation to optimize content across entire product catalogs.

THE CHALLENGE

The business needed to make AI-powered content generation work reliably at full catalog scale.
The challenge went beyond generating good content. The platform needed to manage LLM costs and latency, prevent unsupported claims, recover safely from failures, maintain consistent brand configurations, and continuously validate AI outputs as models and prompts evolved.
Key challenges included:

01 Managing LLM cost and latency across multi-agent workflows

02 Preventing hallucinated product claims and unsupported information

03 Recovering safely from partial failures and repeated requests

04 Keeping brand configurations consistent across multiple sources

05 Making AI scoring stable, explainable, and observable

06 Continuously validating AI behaviour as models and prompts changed

THE SOLUTION

Invimatic engineered the AI layer as a structured, production-ready system rather than relying on a single LLM prompt. The solution used specialized AI modules for research, media, content, FAQs, and catalog variants, with each module having defined responsibilities, model choices, and guardrails.

We also built controls around the AI workflows to:

- Ground customer-facing claims in source data or cited research
- Apply brand-specific rules and voice consistently
- Cache and reuse research to reduce unnecessary LLM calls
- Recover from partial failures without reprocessing completed work
- Track requests across AI calls, cache operations, and audit records
- Continuously evaluate model and prompt changes against a controlled dataset

This created a system where AI could operate at catalog scale while remaining controlled, traceable, and measurable.

HOW WE BUILT AI IN

We treated AI as a core product capability and built an engineering layer around the non-deterministic nature of LLMs. We also built controls around the AI workflows to:

Build

Specialized AI agents handled different parts of the optimization workflow, supported by model selection, structured outputs, brand-aware prompts, and defined guardrails.

Prove

Golden datasets, LLM-based evaluation, confidence metrics, contradiction checks, moderation, and regression thresholds continuously validated AI behaviour before changes reached customers.

Run

Production AI was continuously observed through end-to-end tracing, audit trails, cost tracking, cache performance, and confidence scores, making outputs traceable and issues easier to diagnose.

The result was an AI system designed not just to generate, but to generate, validate, observe, and continuously improve.

RESULTS

Faster catalog turnaround

Optimized product content could be generated in seconds per product instead of requiring days of manual copywriting for product batches.

Consistent brand voice

Brand configurations applied tone, messaging pillars, and preservation rules consistently across large catalogs.

Measurable content-quality improvement

GEO and semantic-similarity evaluations showed improvement across readability, semantic clarity, FAQ coverage, and query matching.

Safer AI outputs

Grounding, moderation, and anti-hallucination validation reduced unsupported claims and the need for manual review.

End-to-end observability

Every optimization could be traced across its AI calls, costs, cache activity, confidence scores, and audit history.

TECHNOLOGY STACK

python 3.11 · FastAPI · Pydantic AI · OpenAI · OpenAI ChatGPT 4.0 · GPT-4.1 mini · asyncpg · SQLModel · redis · amazon SQS · OpenAI Embeddings · PostgreSQL · SQLAlchemy · docker compose · docker · Flyway · amazon S3 · AWS Fargate · Terraform · langfuse · LocalStack · Boto 3 · pytest · AWS CloudWatch

KEY TAKEAWAY

Building AI into a product is not just about connecting an LLM to an application. It requires the engineering around it to make AI grounded, reliable, measurable, and scalable in production. Invimatic built that layer so the platform's AI workflows could operate reliably at catalog scale.