

PROJECT TYPE

Governed GenAI product-description generation engine

TECHNOLOGIES

Python, LangChain, FastAPI, PostgreSQL, Qdrant, Azure OpenAI (GPT-4.1), RAGAS, React, TypeScript, Docker, Azure

DURATION

4 months

TEAM

1 AI Architect
 2 AI Engineers
 1 Backend Engineer
 1 Frontend Engineer
 1 Business Analyst

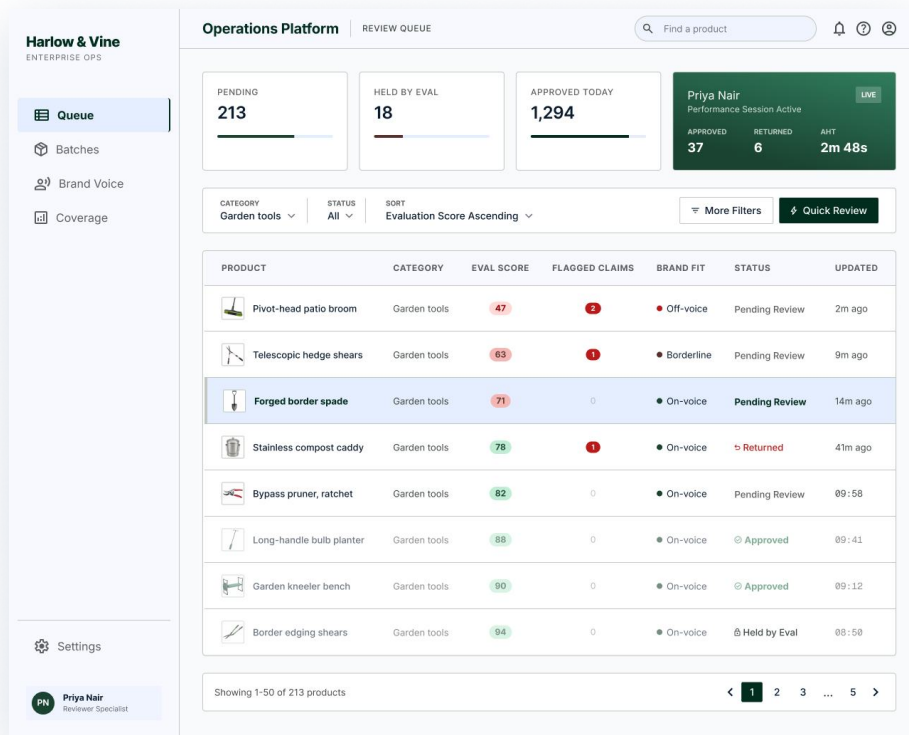
GenAI product-description engine for a multi-category online retailer

Description for the case page: A governed GenAI engine that generates SEO-ready product descriptions for a UK online retailer, grounded only in the retailer's structured catalog attributes, with brand-tone guardrails, automated claim checks, and human approval before anything publishes.

Project background

The Client is a UK mid-market online retailer with a multi-category catalog spanning home, garden, and lifestyle products, around 40,000 active SKUs sourced from several hundred suppliers. Product copy reached the catalog in inconsistent shapes: some products carried full descriptions, many carried a single supplier sentence, and a large share of newly onboarded SKUs went live with no description at all.

Manual copywriting could not keep pace with supplier onboarding, agency turnaround held up new-category launches by weeks, and inconsistent tone across thousands of listings weakened the brand on the storefront.



Project Distinctive Features

- ✓ Generation grounded only in the Client's structured catalog attributes, not open web data



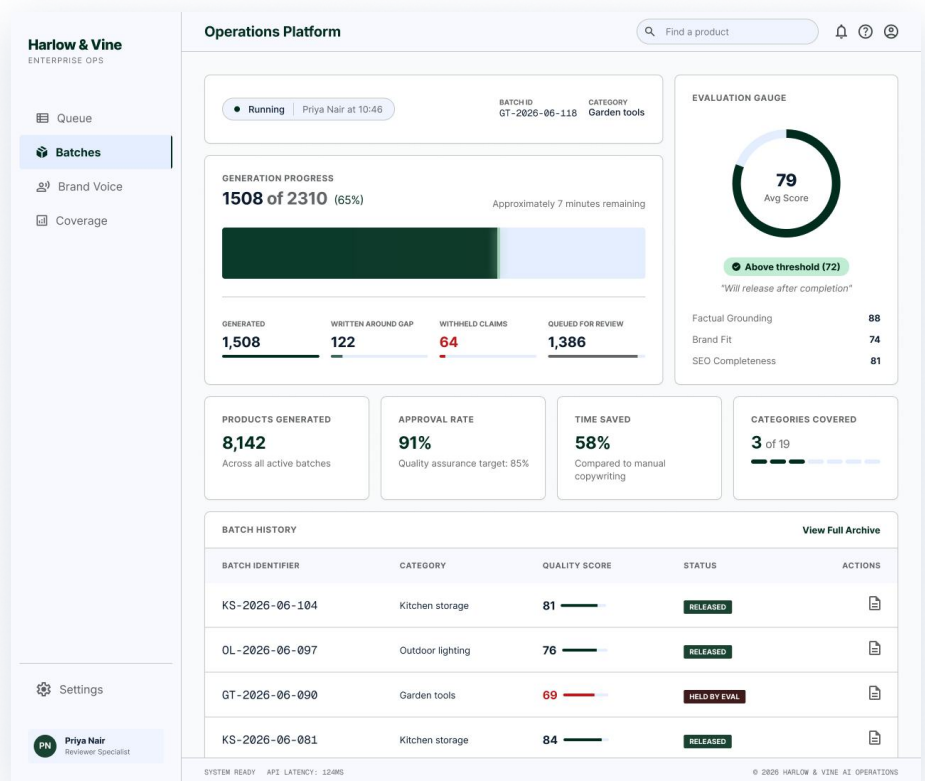
SumatoSoft gave us a way to scale catalog copy without losing control of our brand or our claims. The pilot proved the numbers before we committed to a rollout, and the approval workflow means our content team still signs off on everything. We are now extending the engine across the rest of our catalog with confidence.

CEO

- ✓ Brand-tone guardrails encoded from an approved Client style specification
- ✓ Automated claim verification against product attributes, with unsupported claims flagged or withheld
- ✓ Per-batch evaluation scoring before any draft reaches a human reviewer
- ✓ Human approval workflow with accept, edit, and reject on every generated draft
- ✓ SEO structure built into each description, covering title, key-feature list, and long copy
- ✓ VPC-isolated model deployment with zero data retention
- ✓ One attribute schema is reused across product categories

Business challenge

The Client wanted to generate consistent, on-brand, SEO-ready product descriptions at catalog scale without publishing any claim that the product data could not support, and without removing human sign-off before content went live.

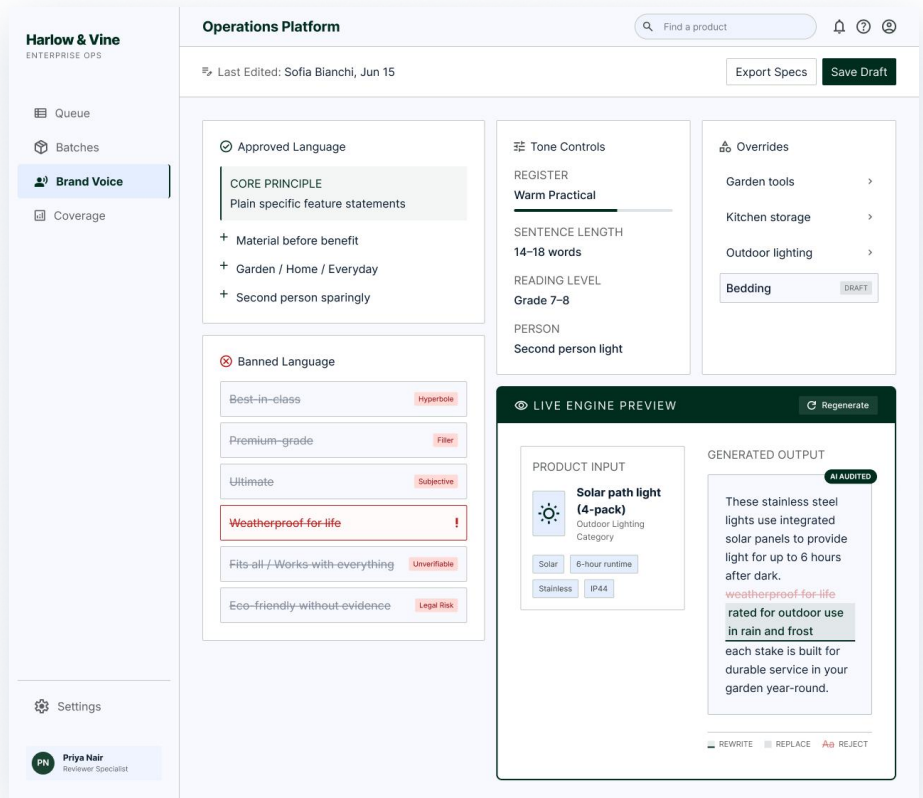


Additional challenges:

- ✓ Keep generated copy inside the Client's brand voice across every category
- ✓ Block product claims that the structured attributes do not substantiate
- ✓ Prove value on a measurable pilot before funding a full rollout

Our solution

SumatoSoft delivered a governed GenAI engine that drafts product descriptions from the Client's own catalog attributes, checks every draft against brand and factual rules, and routes it to a human reviewer before publishing. We built it under our Agentic Development Lifecycle (ADLC), our delivery model for AI systems that have to run inside enterprise guardrails. The pilot covered three categories so the Client could measure quality and cost before scaling.



Discovery and evaluation phase

Before any generation code, our business analyst and AI architect ran a three-week discovery with the Client's merchandising and brand teams. We mapped the attribute data available per category, captured the brand voice as a written style specification with approved and banned language, and agreed an evaluation framework with the content team that scored drafts on factual grounding, brand fit, and SEO completeness. We selected three categories covering roughly 8,000 SKUs for the pilot, chosen for attribute quality and launch priority.

Grounded generation engine

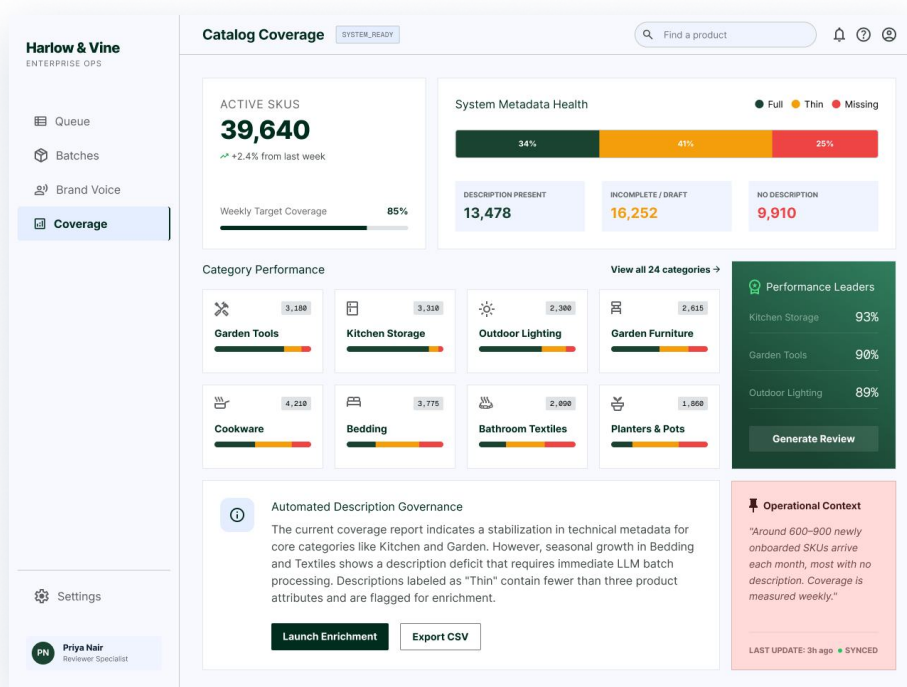
The engine generates each description from the product's structured attributes, retrieving matching brand-voice exemplars and category rules from a Qdrant index so the model writes in the right register. Generation runs on Azure OpenAI GPT-4.1 inside the Client's tenant. Each description follows a fixed SEO structure: a concise title, a key-feature list drawn from attributes, and long-form copy. When an attribute is missing, the engine writes around the gap rather than inventing a value, grounding output in what the data supports.

Brand-tone and claim guardrails

Two guardrails run on every draft before a human sees it. The brand-tone guardrail checks the copy against the approved style specification and rewrites or rejects off-voice phrasing. The claim guardrail compares every factual statement in the draft against the source attributes and withholds any claim the data does not support, flagging it for the reviewer with the attribute it failed against. A per-batch evaluation step using RAGAS-style scoring rates factual grounding and brand fit, and batches below the agreed threshold are held back from review.

Human approval workflow

Reviewers work in a web console where each draft arrives with its evaluation score and any flagged claims. The content team can accept, edit, or reject a draft, and nothing publishes without that sign-off. Every accept, edit, and reject is recorded, and the editing patterns feed a fortnightly refinement of the style specification and the generation prompts, so the engine moves closer to the team's own standard over the pilot.



Governance and deployment

The engine runs inside the Client's Azure tenant with a VPC-isolated model endpoint and zero data retention, keeping catalog and supplier data within the Client's perimeter. We integrated with the Client's product information management system for attribute intake and approved-copy export, and used role-based access aligned to the merchandising and content teams. Generation throughput, approval rate, and evaluation scores were reviewed weekly with the Client's e-commerce lead through the pilot.

Customer's benefits

During the pilot the engine generated descriptions for around 8,000 products across three categories and cut average copy production time per product by 58%, with 91% of drafts approved by the content team with light or no edits. No flagged unsupported claim reached the live storefront.

What's happening with the project right now?

The Client has approved scaling the engine to the remaining catalog categories and is extending it to a second language for its EU storefront, with the same guardrail and approval workflow applied to the new content.