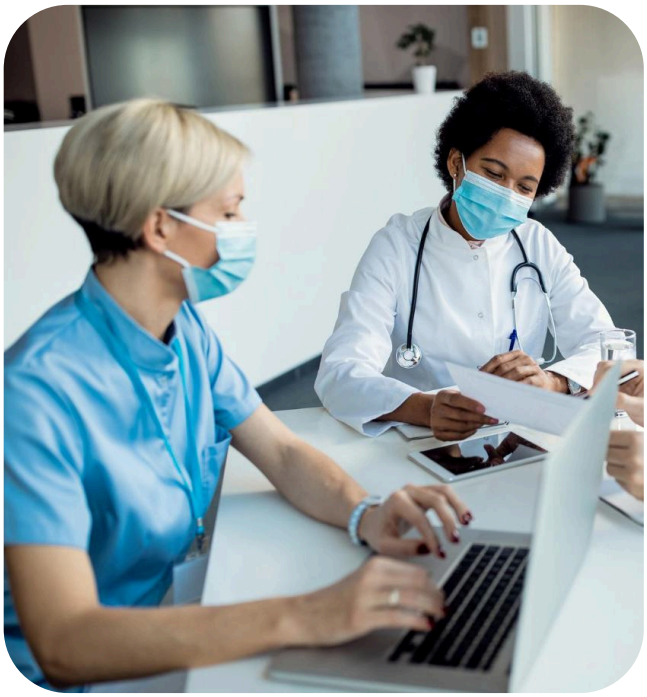
QBurst | A High AI-Q™ Company

Scalable AI Validation Framework for Trusted Contract Intelligence

Transforming AI reliability in contract lifecycle management with a scalable validation framework that ensures accuracy, compliance, and trust in automated outputs.

Overview

- Implemented a scalable 3-layer AI validation framework to continuously evaluate metadata extraction, clause verification, and semantic search accuracy.
- Reduced LLM evaluation cost by 40–60% through a deterministic-first validation approach.
- Achieved 100% automated validation coverage and reduced validation cycles from hours to minutes.



Client Profile

US-based integrated academic health system with a workforce of over 40,000 healthcare professionals, educators, and researchers. The organization operates multiple inpatient and outpatient facilities while advancing patient care, clinical research, and medical education through technology-driven innovation.

Challenges: Validating AI Outputs at Scale

- Long-running validation cycles caused by excessive reliance on LLM-based evaluation.
- Validating diverse metadata types including dates, entities, booleans, and unstructured text.
- Ensuring semantic search accuracy for complex natural-language queries without static ground truth.
- Handling partial matches, formatting inconsistencies, special characters, and lack of measurable confidence scoring.

QBurst Solution: 3-Layer AI Validation Framework

We designed and implemented a scalable validation framework that combines deterministic logic, contextual rule evaluation, and LLM-based semantic assessment to maximize AI accuracy while minimizing operational cost. The platform integrated automated validation workflows with live APIs and operational dashboards to enable continuous testing, monitoring, and AI quality assurance.

Layer 1: Deterministic Validation

Uses regex checks, format validation, and direct text matching to resolve straightforward validation scenarios without invoking LLMs.

Layer 2: Baseline Rules

Applies context-aware business rules to dynamically derive expected metadata and clause values directly from contract content.

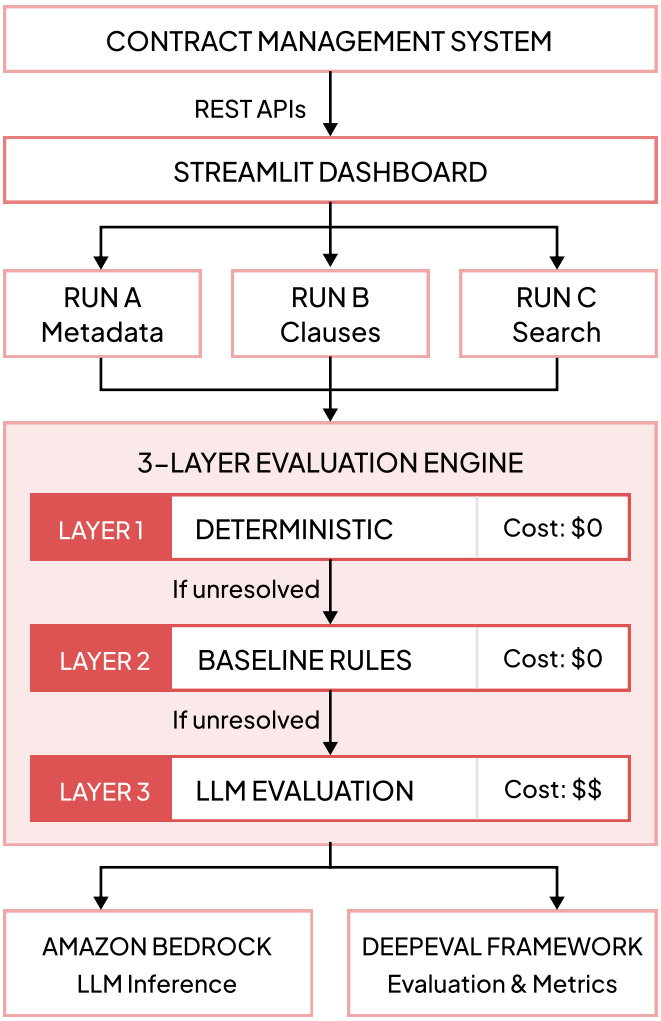
Layer 3: LLM Evaluation

Uses DeepEval-based semantic assessment to validate ambiguous outputs, contextual accuracy, and semantic search relevance.

The framework supported separate validation workflows for metadata validation across 15+ fields, clause verification, and semantic search relevance evaluation.

Technical Highlights

- Search evaluation using a 150-query dataset with combined scoring of LLM evaluation, precision, and recall.
- Fast-path optimization bypassed LLM evaluation for high-confidence scenarios.
- Confidence scoring assigned measurable reliability levels across validation workflows.
- Built using DeepEval, Amazon Bedrock, Streamlit, and REST APIs for scalable validation orchestration and real-time operational visibility.



Impact

- Achieved 100% automated validation coverage, eliminating dependency on manual spot-checking.
- Reduced LLM evaluation cost by 40–60% through a deterministic-first validation approach.
- Reduced validation time from hours to minutes through automated execution workflows.
- Improved AI observability with measurable metrics including precision, recall, and confidence scoring.
- Enabled repeatable validation workflows for early detection of AI regressions.