QBurst | A High AI-Q™ Company

AI-Powered IRB Document Search and Compliance Platform

Transforming fragmented IRB operations with semantic search, intelligent summarization, and scalable regulatory data management.

Overview

- 60% reduction in IRB document retrieval and review time through AI-powered semantic search and intelligent summarization.
- 50% reduction in manual effort for IRB administrators and compliance teams through automated document processing and centralized access.
- Enabled near real-time access to study documentation with scalable indexing and search pipelines.
- Built a secure AI-powered IRB document platform combining semantic search, document summarization, auditability, and regulatory governance.

Client Profile

The client is a leading integrated academic health system focused on clinical research, regulatory compliance, and institutional governance. The organization manages multiple research programs and Institutional Review Boards responsible for overseeing human subject research, regulatory adherence, and study lifecycle governance.

Managing Fragmented IRB Documentation and Regulatory Workflows

The client required a centralized and searchable IRB document platform capable of improving operational efficiency, reducing manual review effort, and strengthening compliance management across research operations.

- IRB documents distributed across fragmented repositories, shared drives, and document management systems.

- Limited keyword-based search capabilities that could not effectively navigate complex regulatory documentation.
- Manual review processes required extensive effort from administrators and reviewers to validate compliance information.
- Difficulty maintaining audit readiness due to limited version tracking and incomplete audit trails.
- Delays and inefficiencies caused by inconsistent access to the latest document versions across teams.
- Increasing operational complexity due to growing volumes of research protocols, consent forms, amendments, and compliance documentation.

Building an AI-Enabled IRB Search and Compliance Platform

We designed and implemented a secure, scalable, and AI-powered IRB document search platform that centralized fragmented research documentation and transformed complex regulatory workflows into an intelligent, searchable, and compliant ecosystem.

The platform integrated document ingestion, preprocessing, semantic search, summarization, indexing, and governance into a unified workflow designed specifically for clinical research and IRB operations.

Centralized Document Ingestion and Processing

All IRB-related documentation including scanned PDFs, digital files, legacy records, protocols, consent forms, amendments, and approvals were consolidated into a centralized repository through an ingestion pipeline that included:

- Automated extraction of text from scanned and semi-structured documents.
- Metadata tagging based on study, protocol, and review stage.
- Data normalization pipelines standardizing multiple document formats.
- Structured transformation of unstructured content into searchable machine-readable formats.

AI-Powered Semantic Search

The platform leveraged natural language processing and vector-based search to significantly improve information discovery across large research document repositories.

- Conversational search using natural language queries instead of exact keywords.
- Context-aware retrieval for more accurate and relevant search results.
- Embedding-based similarity matching across regulatory and clinical documentation.
- Intelligent retrieval of study-specific information across protocols, amendments, approvals, and consent records.

Intelligent Summarization and Review Assistance

To reduce manual review effort and accelerate decision-making, the platform incorporated AI-driven summarization capabilities.

- Generated concise summaries of protocols, consent forms, and amendments.
- Highlighted key compliance and regulatory considerations.
- Enabled faster validation and review workflows for IRB administrators and reviewers.

Scalable Data Engineering and Search Architecture

A scalable data engineering backbone was implemented to support large-scale document processing, indexing, and retrieval operations.

- ETL pipelines for automated document transformation and processing.
- Incremental data ingestion workflows supporting efficient updates.
- Layered storage architecture separating raw, processed, and curated datasets.
- Real-time indexing pipelines continuously updating semantic search indexes.

Security, Governance, and Auditability

Given the sensitive nature of clinical research documentation, the platform incorporated enterprise-grade governance and compliance controls.

- Role-based access control (RBAC) for secure document access management.
- End-to-end encryption for data at rest and in transit.
- Comprehensive audit trails tracking document access, modifications, and approvals.
- Version control and immutable historical records supporting regulatory audits and inspections.

Lifecycle and Compliance Management

The platform supported the full IRB document lifecycle through:

- Approval and version tracking workflows.
- Historical record retrieval for compliance inspections.
- Audit-ready document access and governance capabilities.
- Centralized compliance management across research operations.

Technical Highlights

- AI-driven semantic search using NLP and vector embeddings.
- Agentic document parsing for scanned and unstructured IRB records.
- Centralized IRB document repository with layered lakehouse architecture.
- Automated ETL and incremental indexing pipelines.
- Intelligent summarization of regulatory and clinical research documents.
- Real-time indexing and contextual document retrieval.
- Role-based access control and audit-ready governance workflows.
- Immutable audit trails and version-controlled document lifecycle management.
- Scalable Databricks Lakehouse-based architecture for research data operations.

Impact

- 60% reduction in document retrieval and review time through AI-powered search and summarization.
- 50% reduction in manual effort for IRB administrators and compliance teams.
- Enabled near real-time access to study documentation and historical records.
- Improved audit readiness through centralized governance, version tracking, and immutable audit trails.
- Enhanced operational efficiency across IRB review, compliance, and regulatory workflows.
- Delivered a scalable and future-ready platform capable of supporting growing research data volumes and advanced AI-driven capabilities.