AI Log Insights Copilot - Design Document

# 1. Overview

The AI Log Insights Copilot is a generative AI-powered assistant that integrates with Splunk to accelerate log analysis, incident response, and troubleshooting. It offers automated log summarization, anomaly explanation, predictive insights, and conversational querying to help engineers understand and act on log data faster.

# 2. Goals and Objectives

- Accelerate Root Cause Analysis

- Enhance Observability

- Simplify Splunk Usage

- Proactive Incident Management

# 3. Scope

In Scope

- Splunk log ingestion and embedding

- LLM-based summarization and explanation generation

- Conversational interface for natural language queries

- Predictive anomaly detection and incident forecasting

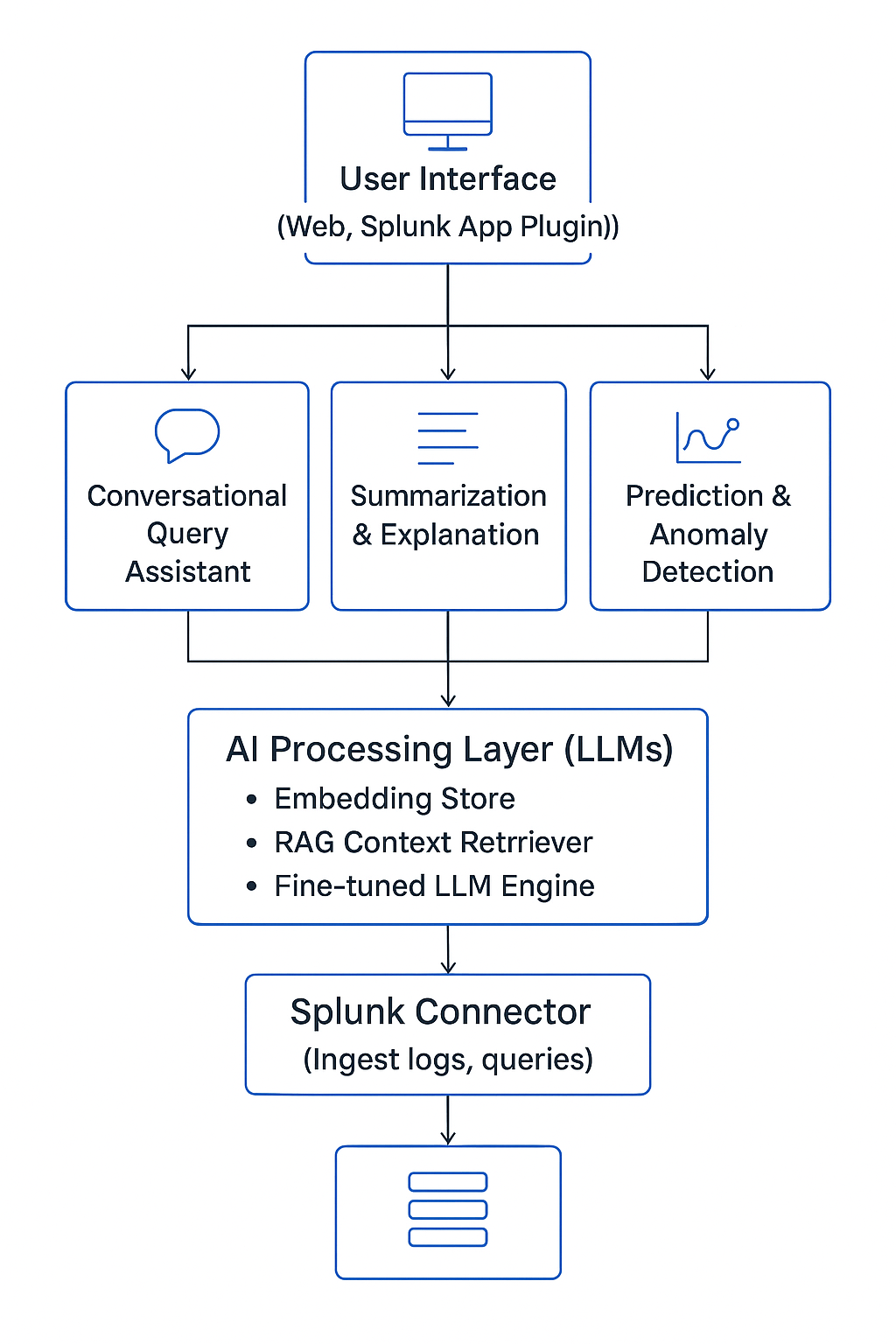
Out of Scope (Phase 1)

- Direct incident remediation automation

- Multi-cloud/multi-platform log ingestion outside Splunk

# 4. Architecture

Visual Architecture Diagram:



# 5. Components

## 5.1 Splunk Connector

- Interfaces with Splunk REST API.

- Executes SPL queries.

- Streams real-time or historical log data.

## 5.2 AI Processing Layer

- Embedding Store: Vector DB (e.g., Pinecone, FAISS)

- Context Retriever: Fetches relevant logs/runbooks for RAG

- LLM Engine: Fine-tuned LLM (e.g., Llama3, GPT-4 Turbo)

## 5.3 Conversational Query Assistant

- Natural language → SPL translation

- Clarification dialog when queries are ambiguous

## 5.4 Summarization & Explanation Module

- Auto-generate summaries of log clusters

- Generate human-friendly explanations for detected anomalies

## 5.5 Prediction & Anomaly Detection Module

- Time-series anomaly models (LSTM, Prophet)

- Predict future incidents based on historical patterns

## 5.6 User Interface

- Splunk App Plugin (primary)

- Optional standalone web dashboard

# 6. Data Flow

- User Input (via UI or Splunk App) → Conversational query OR trigger on log event

- Splunk Connector → Pulls relevant logs or accepts real-time stream

- AI Processing Layer → Embeds, retrieves context, sends to LLM

- Output Generation → Summary, explanation, query suggestion, or prediction

- Response to User Interface → Display results (narrative, charts, SPL, alerts)