Requirements Document: Water LLM – Leak Detection in Water Networks

# 1. Document Control

Project Name: Water LLM – Water Networks & Distribution

Use Case Name: Leak Detection

Prepared By: Business Analyst, TCS

Date: July 21, 2025

Version: 1.0

Reviewed By: Solution Architect, GenAI Engineering Team

# 2. Purpose

This document defines the detailed requirements for a GenAI-enabled Leak Detection system within the Water LLM Engine. The goal is to detect possible leakages in water pipelines early using pressure and flow readings from sensors. This minimizes water losses and enables proactive maintenance, improving the operational efficiency of water utilities.

# 3. Scope

This solution targets both urban and rural water networks. The system continuously collects and analyzes sensor data, identifies potential leak events, and generates alerts and summaries. It integrates with existing systems such as SCADA (for sensor control), GIS (for pipe mapping), and work order tools (for field dispatch).

# 4. Actors & Stakeholders

|  |  |
| --- | --- |
| Actor | Role and Responsibilities |
| Field Engineer | Receives alerts, investigates on-site leaks, and performs repairs. |
| SCADA Operator | Monitors real-time sensor data and responds to alerts. |
| Water LLM Engine | Analyzes data and runs models to detect potential leaks. |
| Leak Detection Model | Uses AI/ML to estimate the likelihood of leakage events. |
| Supervisor/Manager | Oversees operations, validates alerts, and initiates field activities. |
| GIS System | Displays leak locations and pipe metadata for field planning. |
| Work Order System | Creates and manages maintenance tasks related to detected leaks. |

# 5. Use Cases

* UC1: Leak Detection via Flow-Pressure Anomalies

Triggered by incoming sensor data, the system compares actual values with historical baselines to detect discrepancies. If the pressure is unusually low and the flow is unusually high, it could indicate a potential leak. The system highlights such segments.

* UC2: GenAI Summary for Leak Insight

For any segment identified as suspicious, the GenAI component generates a narrative summary. This includes a possible cause of the anomaly, suggested urgency level, and historical context (e.g., recurring issues in that zone).

* UC3: Auto-Triage & Work Order Generation

When the leak severity crosses a configured threshold, the system automatically creates a work order and assigns it to a suitable field team.

* UC4: Leak Zone Risk Scoring

The system calculates a risk score for each zone based on pipe age, terrain, leak history, and material. This helps prioritize inspections.

* UC5: Technician Advisory

When a leak is detected, the system recommends a technician based on proximity, skillset, and current workload, helping speed up resolution.

# 6. User Stories

|  |  |  |  |
| --- | --- | --- | --- |
| ID | As a... | I want to... | So that... |
| US01 | SCADA Operator | receive real-time alerts on pressure drops | I can notify the field team promptly |
| US02 | Field Technician | view leak location and suggested action on a mobile device | I can prepare appropriate tools before arriving |
| US03 | Supervisor | get daily GenAI summaries | I can monitor risk zones and plan maintenance |
| US04 | Data Scientist | access structured leak history | I can train and validate better ML models |
| US05 | GenAI Engine | summarize causes and consequences of leaks | field staff can understand incidents quickly |
| US06 | GIS Analyst | overlay leaks on maps | field teams can navigate more effectively |
| US07 | Asset Manager | see which zones are most affected by leaks | I can plan replacements and investments |

# 7. Functional Requirements

* The system shall ingest real-time pressure and flow data from all connected sensors across the network.
* It shall compare actual flow and pressure values with expected thresholds calculated using historical baselines.
* If a discrepancy crosses a predefined threshold, the system shall flag the pipe segment for possible leakage.
* The GenAI engine shall generate a plain-language explanation for every flagged event, including cause and urgency.
* A work order shall be created automatically when a leak’s severity score exceeds a threshold.
* The GIS map interface shall highlight affected segments and provide contextual leak summaries on click.
* The system shall recommend the best-fit technician for leak repair based on skills, proximity, and workload.
* Historical leak incidents shall be exportable as structured CSV files for model improvement.
* The system shall send proactive alerts for high-risk zones before actual leaks are detected.

# 8. Non-Functional Requirements

* System availability shall be at least 99.9% to ensure continuous monitoring.
* Data processing and alert generation shall occur within 10 seconds of receiving sensor data.
* The alert delivery to SCADA or field teams shall not exceed 1 minute from detection.
* The system must scale to support 10,000+ concurrent sensors across distributed areas.
* All data handling shall comply with GDPR and ISO 27001 data security standards.
* The user interface must support both desktop and mobile form factors for on-site and control room use.
* All time-series data must be archived securely for a minimum of 2 years.

# 9. Data Requirements

|  |  |  |
| --- | --- | --- |
| Source | Type | Description |
| Pressure Sensor | Real-time | Captures the current pipeline pressure for every segment. |
| Flow Meter | Real-time | Records the volume of water flowing through pipeline junctions. |
| GIS | Static + Dynamic | Provides pipeline layout, elevation, and location context. |
| Leak History DB | Historical | Contains validated past leak cases and timestamps. |
| Work Orders DB | Operational | Tracks all scheduled and completed field repairs. |
| Weather API | Real-time | Provides rainfall and temperature data to correlate with leak incidents. |

# 10. Integration Requirements

- The system shall receive real-time inputs from SCADA using protocols like OPC-UA, Modbus, or MQTT.  
- GIS data layers must be accessible for mapping leaks on the user interface.  
- Work order integration shall support platforms like IBM Maximo, SAP PM, or IFS.  
- Notification services shall push alerts via SMS, email, or app notification APIs.  
- The platform shall be hosted on cloud infrastructure (Azure/GCP/AWS) with secure data access.

# 11. Open Questions

|  |  |  |
| --- | --- | --- |
| Question | Owner | Status |
| What SCADA protocol is used (Modbus, OPC-UA, MQTT)? | IT Ops | Pending |
| Will leak confirmations involve human validation? | Field Ops | Under Review |
| What frequency is acceptable for GenAI summaries? | Management | TBD |

# 12. Risks & Mitigations

|  |  |
| --- | --- |
| Risk | Mitigation Strategy |
| False positives from sensor noise or failure | Apply AI + rules + field validation for accuracy |
| Delayed response to critical leaks | Use severity scoring and auto-ticketing to prioritize |
| Missing or incomplete historical data | Use synthetic generation and expert tagging |
| Misinterpretation of GenAI narratives | Add field-level summaries and human review step |

# 13. AI Ethics & Governance

|  |  |
| --- | --- |
| Principle | Implementation |
| Bias Avoidance | Use diverse training data from various pipeline conditions and regions. |
| Transparency | Each leak alert is explained using sensor deviation logic and known context. |
| Accountability | All alerts can be overridden by a human supervisor with audit logging. |
| Explainability | Every GenAI summary is traceable back to data patterns and thresholds. |
| Roadmap | Begin with hybrid model, evolve toward fully autonomous GenAI insights. |

# 14. System Architecture Diagram

The architecture below shows how different components interact in the Water LLM Leak Detection system. It includes data sources, processing engines, AI/GenAI modules, and output systems such as GIS and Work Order tools.

