Requirements Document: Water LLM – Leak Detection in Water Networks & Distribution

# 1. Document Overview

## 1.1 Purpose

This document outlines the functional and non-functional requirements for developing a Leak Detection System using Gen AI under the Water LLM initiative by TCS. The objective is to proactively identify and localize potential leaks in water distribution networks by analyzing pressure, flow, and sensor anomaly patterns using generative and predictive AI.

## 1.2 Scope

The system will ingest real-time SCADA and IoT sensor data from water distribution networks (flow rate, pressure, acoustic data), identify deviations, predict probable leak zones, and provide actionable insights. It will support leak minimization, water conservation, and operational efficiency in urban utilities.

# 2. Business Objectives

* Minimize Non-Revenue Water (NRW) due to leaks.
* Enable early detection of pipeline faults before major failures.
* Reduce manual inspection and repair time.
* Integrate with existing SCADA and GIS systems.
* Support regulatory compliance in leakage KPIs.

# 3. Actors / User Roles

|  |  |
| --- | --- |
| Role | Description |
| Water Network Operator | Monitors alerts, investigates leak predictions, validates system outputs. |
| Field Technician | Executes physical leak verification and repair at predicted locations. |
| Data Engineer | Integrates sensor data, manages ingestion pipelines and model retraining. |
| AI Analyst / Engineer | Oversees model performance, fine-tuning, prompt engineering. |
| Utility Manager | Reviews reports, KPIs, and strategic recommendations from the system. |
| System Admin | Manages user access, configurations, and system uptime. |

# 4. Use Cases

1. UC1: Real-Time Leak Detection  
   Description: Analyze pressure and flow anomalies to detect possible leak locations in real time.
2. UC2: Leak Prediction Using Historical Patterns  
   Description: Use historical leak records and sensor data to predict future high-risk zones.
3. UC3: GenAI Explanation of Leak Cause  
   Description: Provide a natural language explanation for the predicted leak based on patterns.
4. UC4: Field Dispatch Automation  
   Description: Generate work orders with location, leak description, and parts needed.
5. UC5: Leak Repair Confirmation and Feedback Loop  
   Description: Capture post-repair sensor data to confirm leak resolution and retrain model.

# 5. User Stories

* As a Water Network Operator, I want to receive real-time alerts of possible leaks based on pressure and flow pattern anomalies, so I can act quickly to prevent water loss.
* As a Utility Manager, I want to see a weekly risk map of potential leak zones so that we can prioritize preventive maintenance.
* As an Operator, I want the system to explain why it thinks there’s a leak at a location in plain English, so I can verify and escalate it.
* As a Field Technician, I want to receive automatic work orders with GPS-tagged leak locations, description, and required materials, so I can perform efficient repairs.
* As a Data Engineer, I want to use field validation data to retrain the AI model so it gets smarter over time.

# 6. Functional Requirements

1. The system shall ingest real-time pressure and flow data from SCADA/IoT sensors.
2. The system shall analyze incoming data using anomaly detection algorithms.
3. The system shall display leak alerts with location, severity, and confidence score.
4. The system shall provide a GenAI-based textual explanation of detected leaks.
5. The system shall allow field operators to mark a leak as verified or false positive.
6. The system shall generate and dispatch repair work orders to the field team.
7. The system shall learn from verified repairs to improve future detection accuracy.
8. The system shall support GIS overlay to visualize pipelines and leak locations.

# 7. Non-Functional Requirements

* System shall process new sensor data within 5 seconds of arrival.
* System shall have an uptime of 99.9%.
* GenAI outputs shall be explainable and auditable.
* Leak predictions shall be 85%+ accurate with <10% false positives.
* Integration with SCADA, GIS, and CMMS shall comply with existing APIs.
* System shall support user access control with role-based permissions.

# 8. Data Requirements

Input Data Sources:

* Pressure sensors (timestamp, zone ID, kPa)
* Flow meters (lps or m³/h)
* Historical leak logs (location, type, date)
* GIS pipeline map (pipe length, diameter, material)

Output Data:

* Leak alert with timestamp, location, severity
* Explanation text
* Risk map overlay
* Repair feedback

# 9. Integration Points

* SCADA system (for real-time telemetry)
* GIS systems (for mapping and visualization)
* CMMS (Computerized Maintenance Management System)
* Notification system (SMS, email, dashboard alerts)
* Database (PostgreSQL / TimescaleDB for time-series)

# 10. Sample GenAI Prompt Template

Input:  
“Pressure dropped by 25% in Zone C between 2AM–4AM, while inflow remained constant.”

Output:  
“⚠️ A pressure drop without a corresponding flow increase typically indicates a hidden leak in the zone. Based on historical behavior and similar incidents, there is an 87% chance of a medium-sized leak in Zone C.”

# 11. Reporting & KPIs

* Number of detected leaks
* Mean time to detect (MTTD)
* Leak volume saved (estimated)
* False positive / true positive rates
* Time-to-repair (TTR)
* Leak density by zone

# 12. Future Enhancements

* Acoustic sensor and satellite data integration
* Edge AI for decentralized leak prediction
* Customer complaint correlation
* Automated leak impact estimation on supply/demand