



## Techno-Commercial Proposal

For Water Solution Company (WSC)

**Project:** On-Premises AI Knowledge Management System

**Document Version:** iiOT-WSC-RAG-V.07

**Offer Submission Date:** July 29, 2025

**Offer Expiry Date:** August 29, 2025

IIoT Solutions, An Industry 4.0 Consultant.

Riyadh, Saudi Arabia

CR No 1010860724 - Mob No J549628363

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

شركة حوسبة وتقنيات المستقبل لتقنية المعلومات

س.ت 1010860724 - 0549628363

الرياض - المملكة العربية السعودية

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

# Table of Contents

Techno-Commercial Proposal	1
Table of Contents	2
Executive Summary	3
Strategic Overview & Proposed Solution	4
Introduction	4
Background: The Challenge of Untapped Knowledge	4
The Solution: A Sovereign RAG System	4
A Foundation for Future AI Capabilities	5
Project Scope & Solution Details	6
System Architecture Overview	6
A Note on Secure Remote Access	6
Proposed AI & Software Stack	6
Proposed Hardware Supply	7
Scope of Services	7
Commercial Proposal	9
Pricing Summary	9
Payment Terms & Conditions	9
Payment Milestones	9
Project Execution	11
Customer Responsibilities	11
General Exclusions	11
Service Level Agreement (SLA)	12
Term	12
System Availability	12
Support & Incident Response	12
Change Orders & Scope	12
SLA Exclusions	13

# Executive Summary

IIoT Solutions is proud to present this proposal to the Water Solution Company (WSC), a vital entity within the Public Investment Fund (PIF) and a cornerstone of the Kingdom of Saudi Arabia's national infrastructure.

This document outlines a project to deploy a state-of-the-art, **On-Premises AI Knowledge Management System**. The primary objective is to unlock the immense value contained within WSC's vast library of proprietary documents—including due diligence reports, technical studies, and strategic analyses—which are currently underutilized due to their sheer volume.

The proposed solution is a **Sovereign Retrieval-Augmented Generation (RAG) System**, engineered to operate entirely within WSC's physical premises. This approach guarantees that all sensitive data remains under WSC's exclusive control, addressing the critical requirement for absolute data security and sovereignty.

By leveraging powerful, open-source Large Language Models on dedicated, high-performance hardware, the system will provide WSC personnel with a secure and intuitive tool to query their document repository in natural language and receive accurate, instantly verifiable answers. This will transform data into a strategic asset, enabling faster, more informed decision-making across the organization.

The total investment for this turnkey solution is **SAR 400,859.80 (VAT Exclusive)**.

# Strategic Overview & Proposed Solution

## Introduction

IIoT Solutions is a pioneering technology provider specializing in custom Industrial Internet of Things (IIoT) and advanced AI system deployments. We understand the unique challenges faced by critical industries and are committed to delivering solutions that enhance efficiency, security, and operational intelligence. This proposal details a bespoke solution designed to address WSC's specific need for a secure and powerful knowledge management platform.

## Background: The Challenge of Untapped Knowledge

Following its acquisition by PIF, WSC has amassed a significant and strategically vital collection of documents. These reports, running into thousands of pages, contain a wealth of information that is essential for optimizing operations, assessing risks, and driving strategic initiatives. However, the manual effort required to search and cross-reference this data is prohibitive. Consequently, the ability to extract timely insights is severely limited, and the full value of these expensive studies cannot be realized. Furthermore, the sensitive nature of this data, which pertains to national security, strictly forbids the use of cloud-based AI systems that transmit information outside of the Kingdom.

## The Solution: A Sovereign RAG System

To solve this challenge, we propose the deployment of a **Sovereign Retrieval-Augmented Generation (RAG) System**. This is a self-contained AI platform that will be physically hosted within WSC's offices. The system works by:

- **Ingesting & Indexing:** Securely processing the entire library of WSC documents and converting them into a specialized, searchable knowledge base.
- **Retrieving:** When a user asks a question, the system intelligently finds the most relevant passages from the documents.
- **Generating:** A powerful, locally-hosted Large Language Model (LLM) then uses this retrieved information to generate a concise, accurate, and contextually aware answer, complete with citations pointing to the source documents. This ensures that all operations are performed locally, providing a "private GPT" for WSC's data, with no external data transmission.

## A Foundation for Future AI Capabilities

It is important to view this RAG system not as an isolated project, but as the foundational first step in WSC's journey to becoming a fully AI-enabled organization. The hardware, software, and expertise established through this project create a powerful and reusable **AI Center of Excellence** within WSC. This sovereign AI infrastructure can be leveraged in the future to develop and deploy a wide range of high-impact applications, ensuring a significant and compounding return on this initial investment.

# Project Scope & Solution Details

## System Architecture Overview

The solution is built on a high-performance, single-server architecture designed for intensive AI workloads. The server will be equipped with a next-generation GPU providing 96GB of VRAM, enabling it to run state-of-the-art open-source LLMs efficiently. All software will be containerized using Docker to ensure stability and reproducibility.

## A Note on Secure Remote Access

For authorized WSC employees who may require remote access to the system, it is a strict security requirement that this access be facilitated exclusively through WSC's existing corporate **Virtual Private Network (VPN)**. The provision, configuration, maintenance, and administration of the corporate VPN and its user accounts are the sole responsibility of the WSC IT department and are outside the scope of this project.

## Proposed AI & Software Stack

Component	Recommendation	Justification
Generator LLM	Llama 3.1 70B / Qwen3:32B	Apex open-source models for reasoning and instruction-following.
Embedding Model	Qwen3-embedding-8b	A powerful, large model to ensure accurate semantic search.
Re-ranking Model	Qwen3 4B/8B Re-ranker	Critically refines search results to improve the precision of retrieved information.
Vector Database	ChromaDB / Qdrant	High-performance, specialized database for storing and querying document embeddings.
Model Serving	Ollama	Simplifies the local management and serving of the LLMs.

<b>Backend API</b>	FastAPI	A modern, high-performance framework for building the application's API.
<b>Orchestration</b>	Custom IngestionAgent	A bespoke Python-based agent to manage the entire data ingestion and processing workflow.
<b>OS &amp; Frameworks</b>	Linux (Ubuntu), PyTorch, Docker	Industry-standard, stable, and high-performance tools for AI development and deployment.

### Proposed Hardware Supply

#	Component	Model Description	Quantity
1	<b>Case</b>	SilverStone Seta H2	1
2	<b>Motherboard</b>	ASUS Pro WS WRX90E-SAGE SE	1
3	<b>CPU</b>	AMD Threadripper PRO 7965WX	1
4	<b>CPU Cooler</b>	Noctua NH-D9 TR5-SP6 4U	1
5	<b>RAM</b>	G.SKILL G5 Neo 128 GB DDR5-6400 CL32 RDIMM	1
6	<b>GPU</b>	NVIDIA RTX PRO 6000 Blackwell Edition (96 GB)	1
7	<b>PSU</b>	CORSAIR AX1600i 1600W Titanium	1
8	<b>SSD (System)</b>	Crucial T705 Gen5 - 2TB NVMe	1
9	<b>SSD (Storage)</b>	Crucial T705 Gen5 - 4TB NVMe	1

### Scope of Services

#	Description
---	-------------

IIoT Solutions, An Industry 4.0 Consultant.

Riyadh, Saudi Arabia

CR No 1010860724 - Mob No J549628363

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

شركة حوسبة وتقنيات المستقبل لتقنية المعلومات

س.ت 1010860724 - 0549628363

الرياض - المملكة العربية السعودية

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

1	RAG Pipeline & Model Integration: Development and integration of the core AI pipeline.
2	Custom Backend API & IngestionAgent Development: Building the application and data ingestion agent.
3	Vector DB Setup & Configuration: Deployment and optimization of the vector database.
4	Server Assembly, OS & Driver Installation: Physical assembly and software setup of the server.
5	Model Serving & Containerization: Configuration of local model serving and application containerization.
6	Project Management & Client Coordination: End-to-end project management and communication.
7	System Testing, Documentation & User Training: Full system testing, documentation, and user training.

# Commercial Proposal

## Pricing Summary

Category	Price (SAR)
Total Material (Hardware & Software)	117,409.80
Total Engineering Services (Includes 1-Year SLA)	283,450.00
<b>TOTAL PROJECT PRICE (Excluding VAT)</b>	<b>400,859.80</b>

## Payment Terms & Conditions

- **Currency:** All payments shall be made in Saudi Riyal (SAR).
- **Payment Method:** Payments are to be made via bank transfer to the account specified on our official invoice.
- **Payment Due Date:** Full payment for each invoice is due within thirty (30) calendar days from the date of invoice receipt.
- **Taxes:** All prices quoted are exclusive of Value Added Tax (VAT). VAT at the prevailing rate (15%) will be added to each invoice.

## Payment Milestones

Milestone	% of Total	Description
1. Advance Payment	30%	Due upon signing of the contract and issuance of the Purchase Order (PO).
2. Hardware Procurement	30%	Due upon confirmation of all hardware orders being placed.
3. System Development & Backend Readiness	20%	Due upon demonstration of the functional backend API and data ingestion pipeline.
4. On-Site Deployment & Training	10%	Due upon successful installation of the server at WSC premises and completion of user training.

5. Project Go-Live & Final Acceptance	10%	Due upon successful go-live of the full RAG system and formal final acceptance by WSC.
---------------------------------------	-----	--

IIoT Solutions, An Industry 4.0 Consultant.

Riyadh, Saudi Arabia

CR No 1010860724 - Mob No J549628363

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

شركة حوسبة وتقنيات المستقبل لتقنية المعلومات

س.ت 1010860724 - 0549628363

الرياض - المملكة العربية السعودية

[www.IIoTSolutions.sa](http://www.IIoTSolutions.sa)

# Project Execution

## Customer Responsibilities

To ensure the successful and timely completion of this project, WSC will be responsible for:

- **Access to Data:** Providing secure access to the corpus of documents to be ingested into the system.
- **Technical Point of Contact:** Designating a primary technical contact to facilitate communication.
- **Site Preparation:** Ensuring a suitable and secure location for the server with adequate power, cooling, and network connectivity.
- **Personnel Availability:** Ensuring relevant personnel are available for consultation and training sessions.
- **User Management & Access Control:** Managing user accounts and access levels will be the sole responsibility of the WSC IT department.

## General Exclusions

The following are expressly excluded from the scope of this proposal:

- Ongoing data curation or management of the document corpus post-initial ingestion.
- Model re-training or fine-tuning beyond the initial deployment scope.
- Maintenance or security of the Client's internal IT network, including the corporate VPN.
- Provision of any software, hardware, or services not explicitly listed in this proposal.
- Any civil works required for site preparation.

# Service Level Agreement (SLA)

This Service Level Agreement (SLA) outlines the level of service that IIoT Solutions will provide to WSC for the On-Premises AI Knowledge Management System.

## Term

The support and service levels defined in this SLA are valid for a period of **one (1) year** from the Project Go-Live & Final Acceptance date.

## System Availability

- **Target Uptime:** IIoT Solutions guarantees a **98% uptime** for the core application services during standard KSA business hours (Sunday - Thursday, 8:00 AM - 5:00 PM).
- **Scheduled Maintenance:** Maintenance will be scheduled outside of business hours with at least 48 hours prior notification.

## Support & Incident Response

Priority Level	Definition	Response Time (Business Hours)	Resolution Target
P1 - Critical	The system is completely unavailable.	2 Hours	8 Hours
P2 - High	A major feature is non-functional.	4 Hours	24 Hours
P3 - Medium	A non-critical feature is impaired.	1 Business Day	5 Business Days
P4 - Low	General inquiries or cosmetic issues.	2 Business Days	As scheduled

- **Support Channels:** Support requests can be initiated via email or a dedicated support line.

## Change Orders & Scope

- **New Features & Enhancements:** Any requests for new features or functionality not included in the original scope will require a separate Change Order and will be quoted accordingly.

- **Model Retraining:** The retraining of AI models with new data or for performance tuning is not covered under this SLA and will require a Change Order.

### SLA Exclusions

This SLA does not cover unavailability resulting from:

- Failures in WSC's internal network, power, or cooling infrastructure.
- Hardware failures, which are covered under the manufacturer's warranty.
- Issues arising from changes made to the system by non-IIoT Solutions personnel.
- Problems with the corporate VPN or remote access infrastructure managed by WSC.

Thank you for the opportunity to present this proposal.

**Aadil Feroze**  
Business Lead

**Ayman Al Johani**  
Chief Executive Officer

