Product Requirements Document (PRD)

1. Introduction

1.1 Purpose

This Product Requirements Document (PRD) defines the functional and non-functional requirements for an **AI-Enhanced Interactive Electronic Technical Manual (IETM)** system augmented with **AR/VR and Agentic AI capabilities**. The document is intended for product managers, system architects, AI engineers, and academic evaluators.

1.2 Scope

The system aims to transform static technical manuals into an **intelligent, multimodal, context-aware assistant** capable of perception, reasoning, memory, and action across digital and physical environments.

1.3 Goals

- Enable multimodal understanding of text, images, diagrams, speech, and sensor inputs
- Support step-by-step maintenance and training via AR/VR
- Reduce cognitive load and repetitive actions for operators
- Improve task success rates compared to traditional GUI-based systems

---

2. Problem Statement

2.1 Core Problems

- **GUI grounding difficulties**: LLMs struggle to align visual UI elements with actions
- **Repetitive actions**: Lack of memory leads to repeated failures and loops
- **Noise sensitivity**: Unexpected windows, popups, or UI changes degrade performance
- **Poor adaptability**: Limited exploration and recovery strategies
- **Human performance gap**: Current agents lag behind human operators in long-horizon tasks

---

3. User Personas

- Maintenance Engineer

- AR/VR Trainee
- Field Technician
- System Administrator
- Incident Responder

---

4. Functional Requirements

4.1 Perception

- FR-P1: The system shall process text, images, diagrams, audio, video and sensor data
- FR-P2: The system shall support OCR for scanned manuals
- FR-P3: The system shall detect machine parts using vision models (SAM, Rex-Omni)

4.2 Reasoning & Planning

- FR-R1: The system shall support CoT, ToT, ReAct, and MCTS-based planning
- FR-R2: The system shall generate structured maintenance plans
- FR-R3: The system shall support reflection and self-correction

4.3 Memory

- FR-M1: The system shall support short-term and long-term memory
- FR-M2: The system shall retrieve relevant past cases using RAG
- FR-M3: The system shall store metadata (page, section, component)

4.4 Execution

- FR-E1: The system shall execute tool-based actions
- FR-E2: The system shall generate and run code securely for diagnostics
- FR-E3: The system shall support AR overlays and XR actions

---

5. Non-Functional Requirements

- Performance: Real-time or near real-time inference
- Reliability: Guardrails against hallucination and loops (Adding Rerankers & Validators)
- Scalability: Modular microservice architecture
- Security: Sandboxed code execution

---

6. Constraints & Assumptions

- Limited GPU availability
- Context window constraints

- Edge-device deployment limitations

---

7. Success Metrics

- Task completion rate
- Reduction in human intervention
- Accuracy of visual grounding
- Latency per interaction

---

8. Out of Scope

- Fully autonomous robotics (future work)
- Medical or safety-critical tasks

---

9. Risks

- Hallucination in multimodal reasoning
- High compute costs
- Dataset bias

---

10. Roadmap (High-Level)

- Phase 1: Text + Speech RAG
- Phase 2: Vision & Multimodal RAG
- Phase 3: AR/VR Integration
- Phase 4: Reflection & Autonomy