Terms of Reference (ToR)

Project Title: Development of SmartMaria Persona (YOUTHMED Version) – Full System Implementation

Client: NLCD / YOUTHMED Project Team

Duration: 18 to 20 weeks (from contract signing)

Development Model: Full System – Modular, Multi-Persona Ready, Hybrid AI Architecture (Cloud + Local GPU)

---

1. Objective

To fully develop the SmartMaria Persona system for the YOUTHMED project as an intelligent, multilingual, AI-powered network manager that enables structured dialogue, knowledge transfer, and coordination across youth organizations. The system will operate using a hybrid architecture (cloud-hosted UI + local GPU logic), and support multi-user moderated sessions, document-based idea extraction, and structured collective intelligence features.

---

2. Scope of Work

2.1 Functional Components (to be implemented)

- **Multilingual Chat Interface:** Supports English, Arabic, Spanish, and Greek input/output.
- **One-to-One and Group Dialogue Sessions:**
  SmartMaria guides or moderates according to four logic modes:
    1. Summarize after each message
    2. Summarize at the end
    3. Intervene only when requested
    4. Adaptive hybrid mode

- **Group Dialogue Management:**
  SmartMaria moderates multi-user cross-organization dialogue sessions. Participants interact in their native languages, while SmartMaria manages translation and logic flow. Sessions conclude with a generated English-language MoM for user validation.

- **Multilingual Translation Engine:**
  Supports both unified (all translated to a common language) and dual/paired translation (each participant receives replies in their own language), with shared English summary.

- **Smart MoM Generator:**
  AI drafts English-language meeting summaries identifying agreements, disagreements, and recommendations. Users must validate summaries.

- **Organizational Libraries and Knowledge Visibility:**
  Each organization has a dedicated library for uploading content (documents, media, ideas) classified as either private or public. Public items are used by SmartMaria to enrich relevant dialogues with other organizations.

- **Collective Memory and Knowledge Transfer Governance:**
  SmartMaria maintains a structured collective memory, storing validated content from sessions. Governance is handled via:
  1. **Automatic Filtering** based on predefined rules
  2. **Human Validation** by platform moderators
     Only approved, contextually relevant items are accessible in future interactions.

- **Admin Dashboard:**
  Allows administrators to:
  1. Manage users and sessions
  2. Moderate content
  3. Approve shared knowledge
  4. Adjust logic rules for memory and dialogue

- **Voice Interaction for Pro Users (Voice-to-Text and Text-to-Voice):**
  The system shall support full voice-based interaction for verified Pro Users. This includes:
  1. Voice-to-Text: Users can speak to SmartMaria via browser microphone; converted to multilingual text input in real time.

2. Text-to-Voice: SmartMaria replies in spoken voice using advanced multilingual TTS engine.
   This feature is restricted to Pro accounts and requires browser permission (WebRTC-based). It enhances accessibility, user experience, and inclusion.

2.2 Architectural Requirements

- **Hybrid Hosting Architecture:**
  The system will operate using a hybrid model:
  - **Frontend, UI, translation services:** Hosted on secure cloud (Google Cloud / VPS)
  - **AI logic engine:** Hosted locally on GPU at PDA/NLCD, exposed via API
  - **API Gateway:** Secure channel between frontend and local inference engine
  - **Fallback Logic:** If local GPU unavailable, revert to cloud inference
- **Modular Logic Loader:**
  Each session dynamically loads logic and memory based on persona_id.
- **Persona-specific Storage:**
  Data is separated by persona context (either via separate DBs or internal persona_id field).
- **Session Integrity:**
  Group sessions are bound to one persona context. All users in a session must share persona_id. Memory and moderation logic are consistent across session.

3. Technical Stack (Recommended)

- **Frontend:** React.js or Vue.js
- **Backend:** Node.js or Python (FastAPI or Express)
- **Database:** Firebase, Supabase, or MongoDB Atlas
- **AI & Translation APIs:** OpenAI (ChatGPT), DeepSeek, or local LLM instance

- **Hosting:** Google Cloud / Secure VPS + Local GPU server for inference

---

4. Deliverables

- Fully functional SmartMaria platform (hybrid architecture)
- Admin dashboard with rules and content control
- Local inference engine hosted on GPU
- Frontend with multilingual chat, upload, and memory views
- Documentation:
  - Code structure and setup guide
  - Persona loader and memory logic
  - Admin and deployment instructions
- Git-based source code repository
- Test credentials and admin access
- Optional walkthrough video

---

5. Timeline

Phase	Task	Duration
Phase 1	Analysis + Full System Design	3 weeks
Phase 2	Frontend + Backend + Library Setup	3 weeks
Phase 3	Dialogue Engine + Translation + MoM	3 weeks
Phase 4	Admin Dashboard + Rule Logic	3 weeks
Phase 5	Hybrid Architecture (API Gateway + Local GPU Logic)	3 weeks
Phase 6	QA, Testing, Documentation	3 weeks
Total		**18 weeks**

---

6. Budget Estimate

Estimated total cost: **18,000 EUR**
(Includes all platform components, admin logic, GPU setup, fallback architecture, voice module, and testing.)

---

7. Evaluation Criteria

- Demonstrated experience with GPT or LLM-based platforms
- Proven delivery of modular AI-based systems (hybrid/local setup)
- React/Node or Python backend with API integration
- Secure deployment and documentation
- Ability to meet full system logic and timeline

---

8. Contact and Coordination

All communications and deliverables shall be coordinated through NLCD / YOUTHMED digital team. Full access to architecture plans, files, and logic documents will be provided upon signature.

End of ToR