

AI MINDS Theme Introduction

Modern users constantly capture fragments of information across many sources and formats: images, text, audio, documents, web pages, and visual content. Over time, this creates a large personal information space that is difficult to organize, search, or reuse. Important knowledge is lost not because it is unavailable, but because it cannot be understood or retrieved when needed. The system must therefore handle multimodal inputs and may respond using different output modalities when appropriate.

Objective

Build an intelligent system that converts raw personal data into a structured, searchable memory that can be interacted with through natural language. The system must interpret meaning, detect intent, and help the user recover useful knowledge from past data, regardless of the modality of the original information.

What the System Must Do:

1. Automatically ingest new personal data from multiple modalities without manual uploads.
2. Understand the content and purpose of each item independent of format.
3. Organize information into meaningful categories.
4. Produce concise summaries and extract actionable items.
5. Allow the user to ask questions and obtain grounded answers referencing past records.
6. Handle uncertainty and avoid confident incorrect answers.

Level of Complexity Expected

1. The system should operate continuously, not as a single request-response program.
2. It must maintain persistent memory across time.
3. It should reason about relevance rather than simple keyword matching.
4. It should verify its own answers before presenting them to the user.
5. It may output responses in different modalities when useful.

Limitations:

- Use of proprietary or hosted frontier models via external APIs is not allowed (e.g., OpenAI, Anthropic, Gemini, Grok, etc.).
- For LLM usage, teams must run a locally hosted open-source model with **fewer than 4 billion parameters**.

Goal of the Challenge

The aim is to design an AI that behaves like a reliable cognitive assistant — not only generating responses, but managing knowledge over time, deciding what matters, and helping humans think more clearly using their own data across modalities.