MATE: A Mission Assistant for Training and Evaluation Using an Offline Language Model in Air Combat Operations

Joao P. A. **Dantas** – Capt.
Institute for Advanced Studies (IEAv)
Brazilian Air Force

jpdantas@ita.br
www.joaopadantas.com

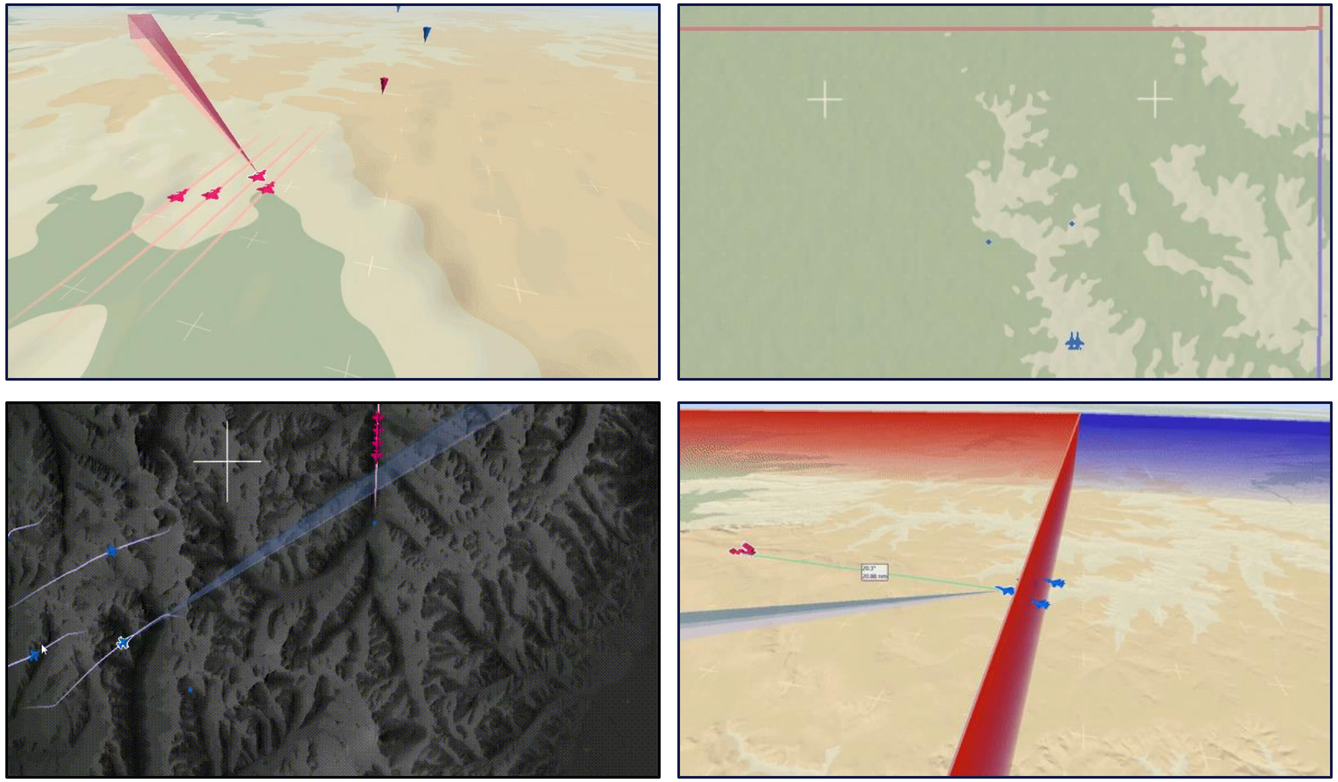
Motivation

Air combat missions demand **fast** and **reliable** access to tactical information.

Tactical knowledge is often stored in **lengthy** and **complex** manuals.

Manual consultation during **planning**, **debriefing**, or even **flying** is slow and impractical.

Offline access is **mandatory** in environments where internet connectivity is restricted and tactical manuals are classified



Simulated air combat and air defense missions.

MATE

Technological Context

Advances in **Natural Language Processing** (NLP) and **Large Language Models** (LLMs) enable natural-language interaction with documents.

Existing tools (e.g., GPT-5) depend on **cloud infrastructure**, unsuitable for **sensitive** or **disconnected** military environments.

Open-weight LLMs such as **LLaMA 3 (Meta)**, **Mistral (Mistral AI, France)**, and **Qwen 2.5 (Alibaba Cloud)** support **local, offline deployment**, enabling **secure NLP tools** for sensitive or classified environments.

Proposed Solution

An **offline language-based support tool** using **Retrieval-Augmented Generation (RAG)**.

Combines **document retrieval** with a **local LLM** to answer user questions in plain language.

Built entirely with **open-source componentes**.

Operates **fully offline** — ideal for onboard systems, simulators, and training centers.

Main Contributions

Tactical Tool Development
Offline access to sensitive doctrinal data

Integration of air combat manuals using semantic retrieval and local language models

Validation with domain experts to evaluate response quality and operational usefulness

Support for training in air combat operations

Related Work

DARPA TIPSTER: early NLP for defense information retrieval.

TransApps / MiTAP: field NLP apps for speech, translation, and intelligence.

Primer / COA-GPT: NLP and LLMs for analysis and planning.

DARPA ACE / F-35: human–AI teaming and speech control in air combat.

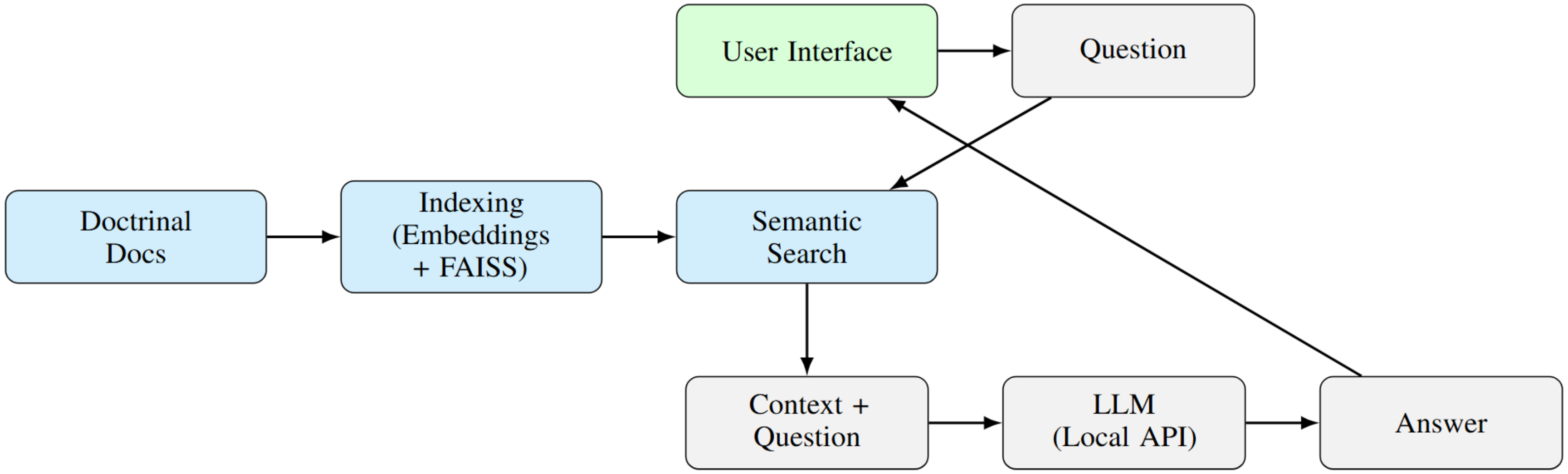
Gap: most systems rely on cloud.

MATE

Methodology

Goal: Enable offline, natural-language access to tactical manuals by combining document retrieval and local language generation.

Architecture: Modular design built with open-source tools and offline support.



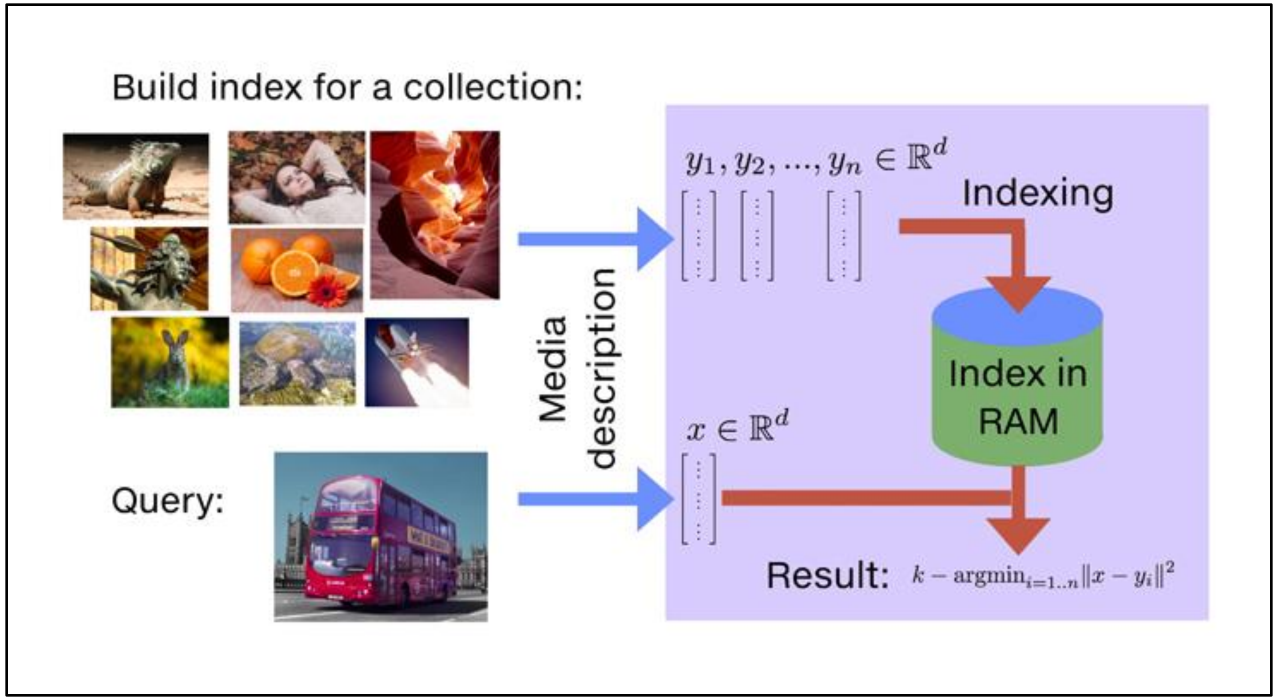
Overview of the offline tactical assistant system architecture.

Document Indexing

Manuals in **.txt**, **.pdf**, or **.docx** are converted into plain text for processing.

Each text segment is encoded into an **embedding** — a numerical vector representing its semantic meaning — using the **all-MiniLM-L6-v2** model (SentenceTransformers).

These embeddings are stored in a **FAISS (Facebook AI Similarity Search)** index, enabling fast similarity search across millions of text vectors.



Document indexing using FAISS.
Source: https://engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/

Semantic Retrieval

When a user submits a question, it is **converted into an embedding using the same model**.

The question embedding is **compared to the FAISS index** to find the top-k most relevant passages (k = 5 for MATE).

These retrieved passages are **concatenated into a context window**.

The context and user query are sent to the local language model, which generates an answer grounded in doctrinal content.

This process is known as **Retrieval-Augmented Generation (RAG)**.

Language Model Interaction

The system uses local large language models (LLMs) such as Mistral, LLaMA, or Qwen, deployed through the Ollama API.

Ollama is a lightweight runtime that **allows LLMs to run entirely on a local machine**, without requiring internet or cloud access.

It manages model loading, memory usage, and inference.

User Interface

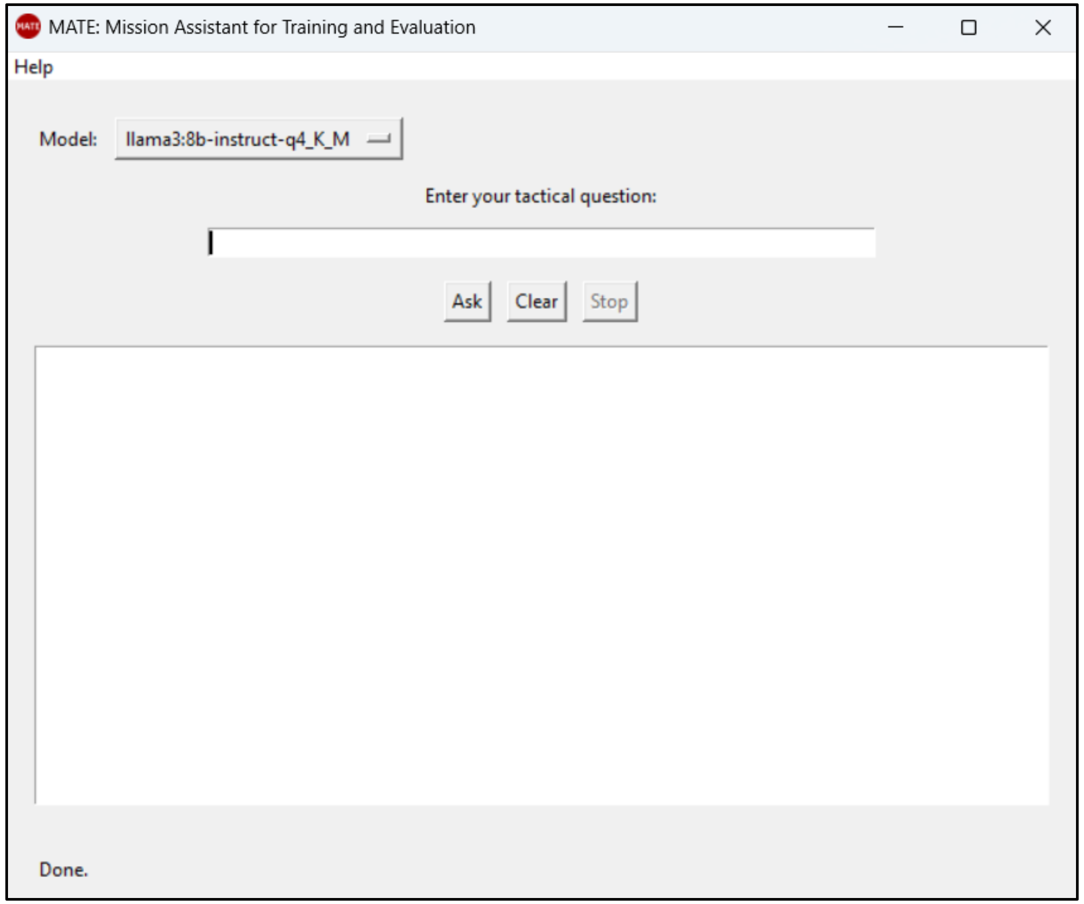
Built with **Tkinter** as a lightweight offline GUI.

Supports real-time responses from the local LLM.

Includes **Ask**, **Clear**, and **Stop** buttons for user control.

Connects to the FastAPI–Ollama backend via HTTP streaming.

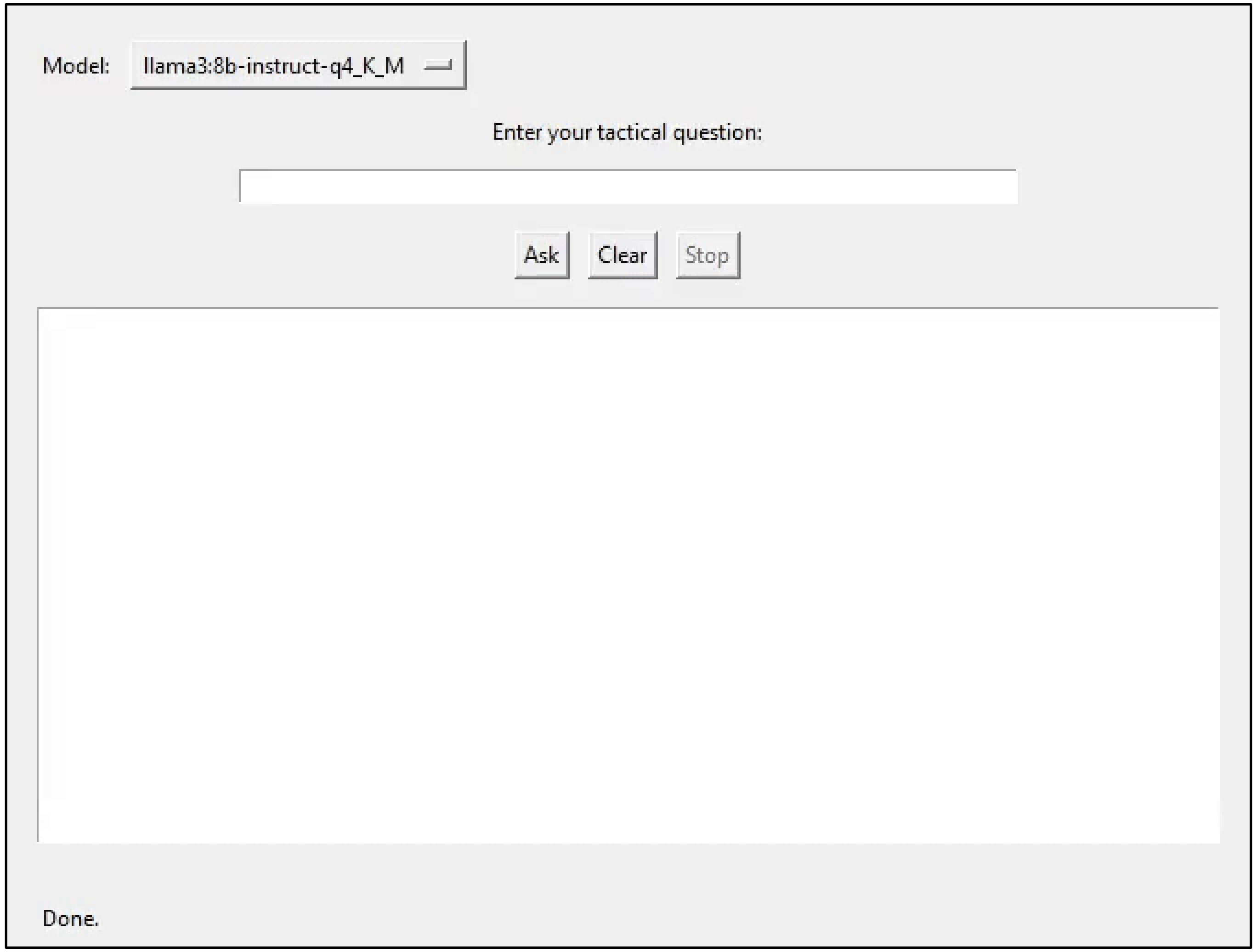
Voice interaction (**PyAudio + Whisper**) was temporarily disabled in this version to speed up testing and evaluation.



Offline Deployment

Runs entirely offline (embedding, FAISS, and LLM stored locally).

Packaged with **PyInstaller** for standalone execution in **secure, self-contained environments**.



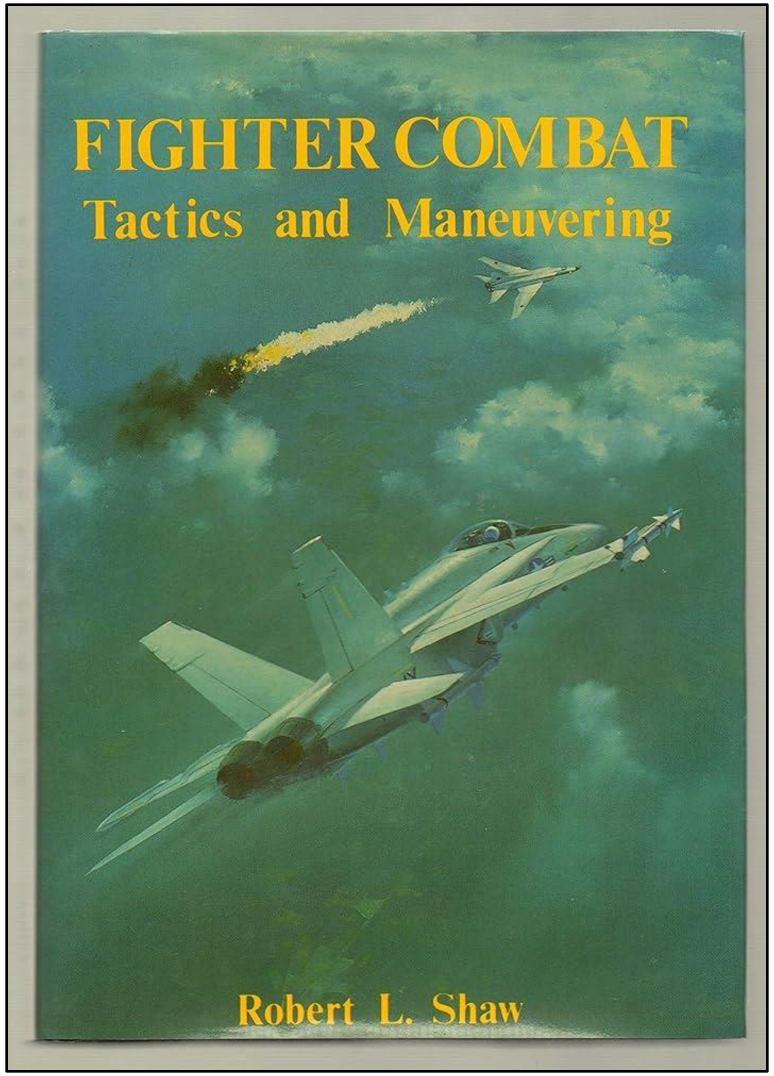
Validation Proposal

Goal: Assess the system’s ability to support tactical decision-making in air combat.

Reference Source: Fighter Combat: Tactics and Maneuvering (Robert L. Shaw, 1985) – **Chapter 5 “Section Tactics: Two-versus-One**”.

Reason: Avoid classified material while using a recognized doctrinal manual.

Approach: Evaluate how well the tool retrieves and generates accurate, context-aware responses to tactical questions.



MATE

Evaluation Procedure

Participant Info: Collect operational background and flight experience.

Tactical Evaluation: 5 air combat questions per model.

Models Tested: Llama 3.1 8B, Qwen 2.5 14B, Mistral NeMo 12B.

Expert Review: 3 Brazilian Air Force fighter pilots assess responses.

General Feedback: 5 tool-level questions + suggestion field.

Rating Scale: 1 – Very Inadequate to 5 – Very Adequate.

Example

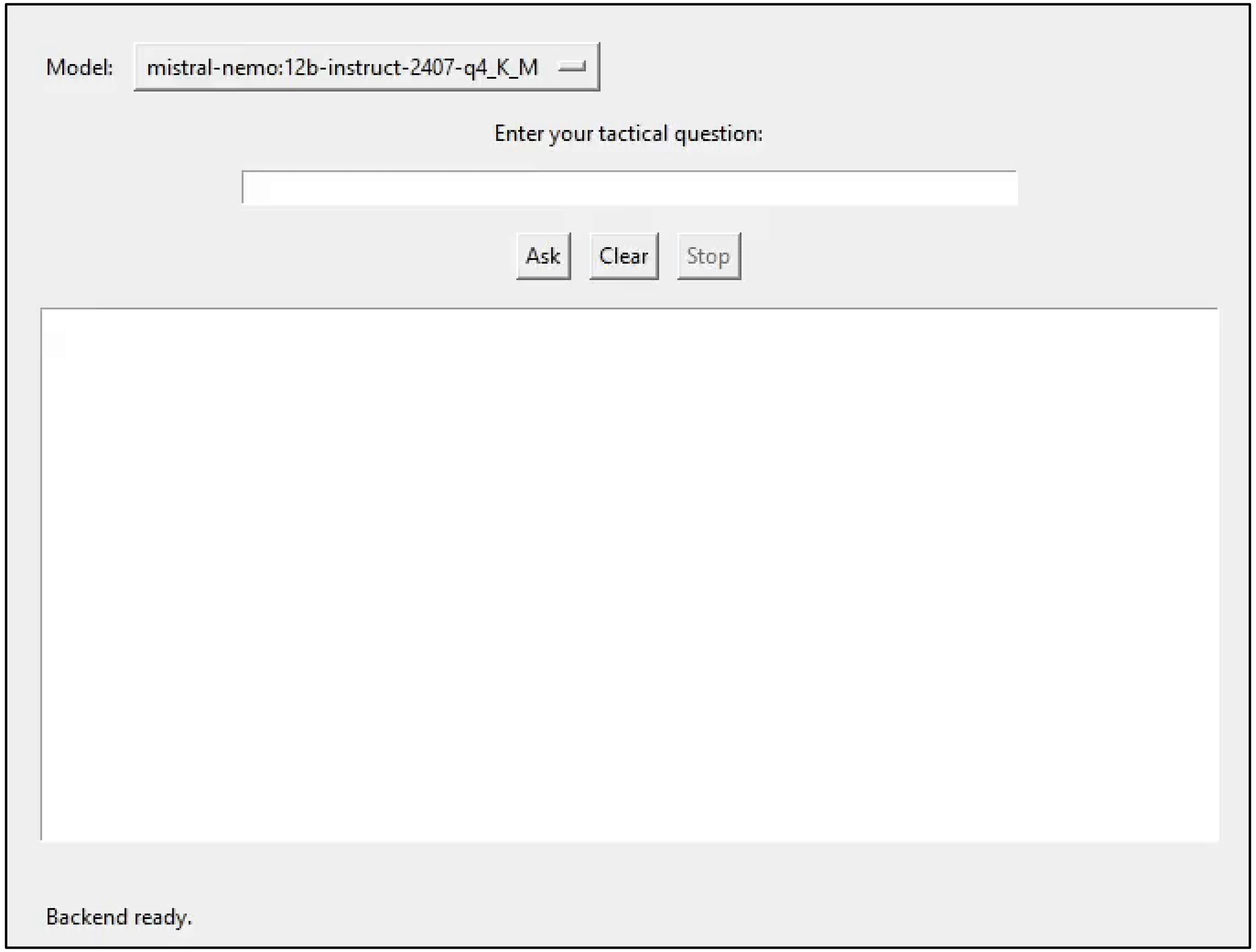
What are the main advantages of the Fighting Wing formation for new pilots?

Expected Answer:

Page 198:

“Fighting wing allows inexperienced pilots to engage in combat under the tutelage of a veteran leader at reduced risk. Actual combat is the best teacher, but historically the highest attrition rate for fighter pilots has occurred during their first few combat missions.”

How do you rate the model’s answer?



Average Pilot Profile

Aviation	Average Operational Experience (years)	Average Flight Hours	Average Familiarity with LLMs	Commonly Used Models
Fighter Aviation	6.7	933	Moderate (occasional use, basic knowledge)	ChatGPT (OpenAI), Gemini (Google), Llama (Meta)

MATE

Tactical Questions

Tactical Question	Llama 3.1 8B (μ ± σ)	Qwen 2.5 14B (μ ± σ)	Mistral NeMo 12B (μ ± σ)
Q1 – *What is the primary advantage of flying as a section instead of a single fighter?*	4.67 ± 0.58	4.00 ± 1.00	**4.67 ± 0.58**
Q2 – *What are the main advantages of the Fighting Wing formation for new pilots?*	4.67 ± 0.58	4.33 ± 0.58	**4.67 ± 0.58**
Q3 – *What are the main weaknesses of the Fighting Wing doctrine?*	3.00 ± 0.00	3.00 ± 0.00	**3.00 ± 0.00**
Q4 – *What tactical innovation did the Double Attack introduce compared to Fighting Wing?*	3.33 ± 0.58	4.00 ± 1.00	**4.67 ± 0.58**
Q5 – *How does the Loose Deuce differ from the Double Attack in offensive engagement?*	5.00 ± 0.00	4.67 ± 0.58	**5.00 ± 0.00**

MATE

General Questions

Evaluation Question	Mean ± Std. Dev (μ ± σ)
Relevance of Retrieval — *Did the information used in the answer correspond to the correct content from the reference material?*	3.67 ± 0.58
Conceptual and Terminological Accuracy — *Did the model use correct tactical terms consistent with operational doctrine?*	4.33 ± 1.15
Integration between Retrieved Context and Reasoning — *Did the model properly incorporate the retrieved text into its explanation, showing understanding of the material?*	3.67 ± 0.58
Clarity and Coherence of the Response — *Was the answer clear, logical, and well-structured, facilitating understanding of the tactical concept?*	**4.33 ± 0.58**
Response Time — *Was the generation time appropriate compared to the time it would take to manually locate the information in the reference text?*	**5.00 ± 0.00**

MATE

Conclusion and Future Work

Offline tactical assistant combining document retrieval and a **local LLM** for air combat operations.

Works in **disconnected** or **classified** environments with doctrinal manuals.

Open-source, modular design for use in simulators and mission planning.

Expert validation phase with fighter pilots showed **promising** results.

Next: expand **pilot evaluations**, improve **response time** with better hardware (better GPU), enhance **interactivity**, and optimize the **voice interface** (currently too heavy for daily use).

MATE: A Mission Assistant for Training and Evaluation Using an Offline Language Model in Air Combat Operations

Joao P. A. **Dantas** – Capt.
Institute for Advanced Studies (IEAv)
Brazilian Air Force

jpdantas@ita.br
www.joaopadantas.com