



# Offline Language-Based Support System for Tactical Decision-Making in Air Combat Operations



Joao P. A. Dantas

Institute for Advanced Studies, Brazilian Air Force

## Introduction

- Air combat missions depend on **fast, accurate access to tactical knowledge**, but this information is often **hidden inside long manuals**.
- Using these documents during **planning, debriefing, or flight** is slow, impractical, and not compatible with **time-critical decisions**.
- Because tactical manuals are often **classified**, tools that rely on **cloud services** cannot be used in operational or simulator environments.
- Advances in **Natural Language Processing (NLP)** and **Large Language Models (LLMs)** enable natural-language interaction with documents, but most advanced models (e.g., GPT-5) still require **online access**.
- New **open-weight LLMs**—such as **LLaMA 3, Mistral, and Qwen 2.5**—enable **secure, fully offline use**, creating opportunities for tactical support tools.

## Contributions

- **Offline Tactical Assistant:** A Retrieval-Augmented Generation (RAG) tool that answers tactical questions using doctrinal manuals with **no internet requirement**.
- **Open-Source Pipeline:** A fully **modular, offline, and reproducible** methodology combining document indexing, semantic retrieval, and local LLMs.
- **Expert Validation:** A structured evaluation with **Brazilian Air Force fighter pilots**, assessing **accuracy, clarity, and doctrinal adherence**.

## Related Work

- **DARPA TIPSTER:** introduced **large-scale text retrieval** and extraction for defense, supporting intelligence analysts in **processing high-volume documents and reports**.
- **TransApps / MiTAP:** delivered **mobile NLP tools** to deployed units, enabling **speech recognition, translation, and real-time intelligence support** during field operations.
- **Primer / COA-GPT:** employed **LLMs** to analyze large information streams and generate **military courses of action**, assisting staff officers in planning and decision-making.
- **DARPA ACE / F-35:** advanced **human-AI teaming** in air combat, exploring **adaptive autonomy, pilot assistance, and speech-enabled cockpit interfaces** (Figure 1).
- **Gap:** reliance on **cloud infrastructure** prevents use in classified environments.



Figure 1. DARPA ACE program: advancing human-AI teaming in air combat.

## Methodology

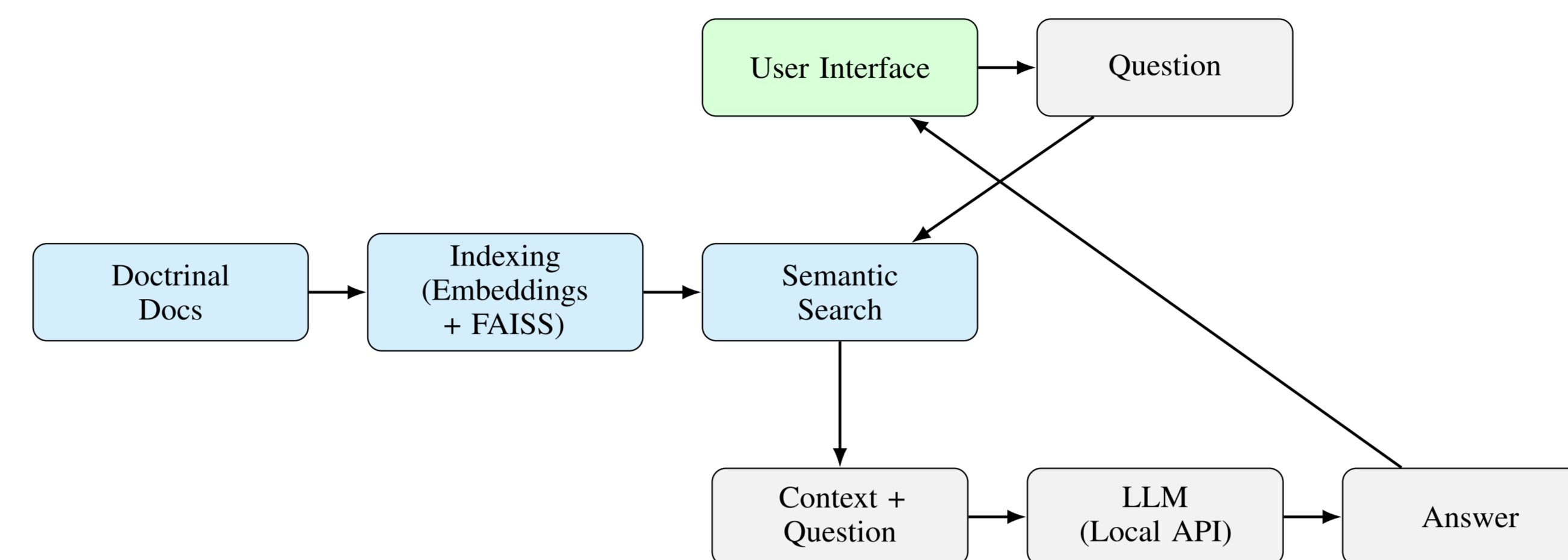


Figure 2. Overview of the offline tactical assistant system architecture.

- **Document Indexing**
  - Manuals (.pdf, .doc, .docx) are converted to plain text (.txt).
  - Each text segment is transformed into a semantic embedding using **all-MiniLM-L6-v2** from the SentenceTransformers library (Reimers and Gurevych, 2019).
  - Embeddings are stored in a **FAISS** index (Facebook AI Similarity Search) (Johnson et al., 2021), a fast vector search library used to retrieve the most relevant passages (Figure 3).

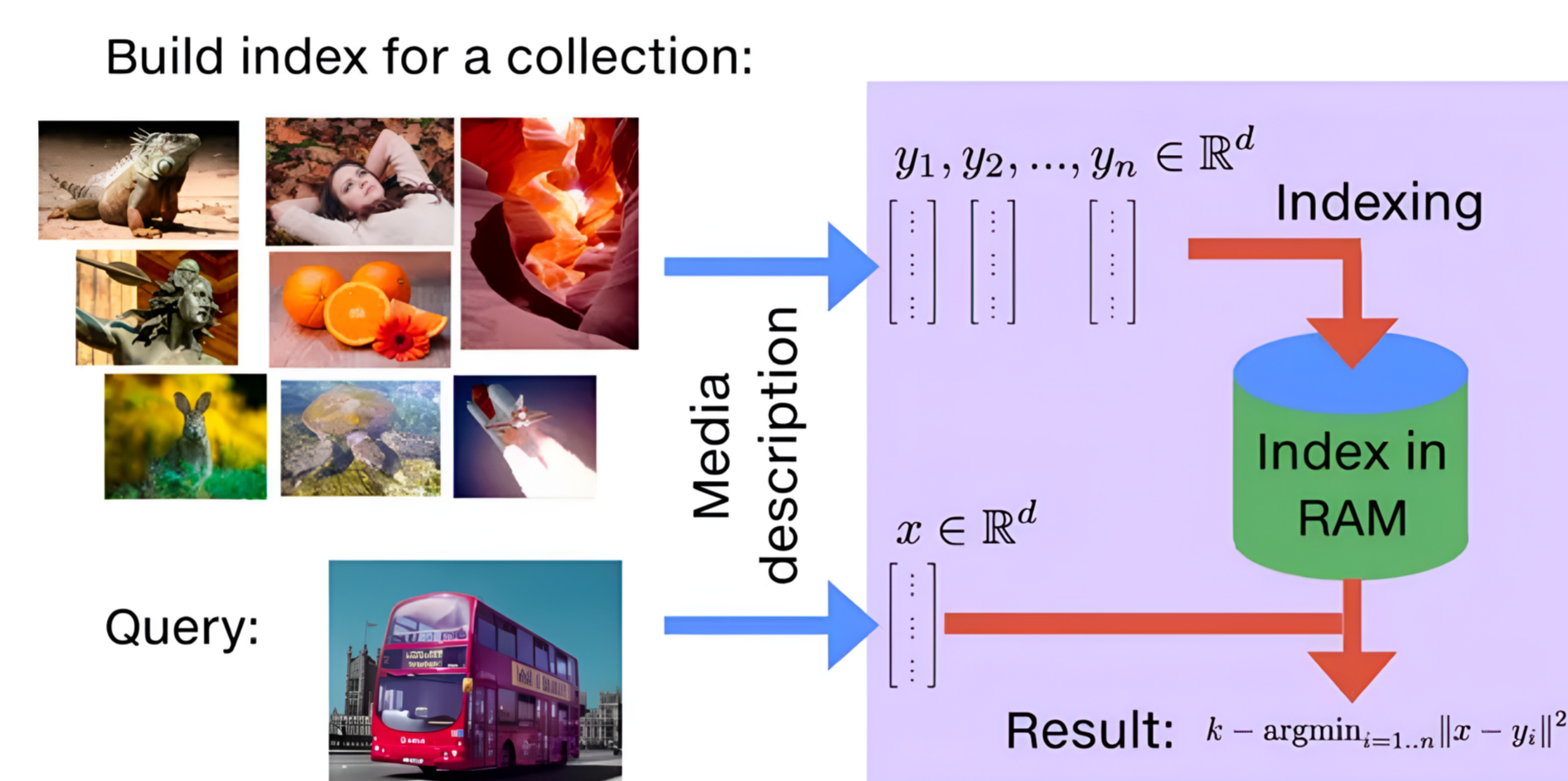


Figure 3. Document indexing using FAISS.

Source: <https://engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/>

- **Semantic Retrieval**
  - User queries are embedded using the same model.
  - The FAISS index retrieves the **top-k** most relevant passages.
  - Retrieved passages are concatenated into a context window and combined with the query.
  - The local LLM generates an answer based on the retrieved doctrinal passages (RAG).
- **Language Model Interaction**
  - Local LLMs (Mistral, LLaMA, Qwen) run through the **Ollama** API.
  - Ollama handles model loading, inference, and memory management entirely offline.
- **User Interface**
  - Lightweight **Tkinter** GUI with real-time responses and basic controls (Ask, Clear, Stop).
  - Communicates with the FastAPI-Ollama backend via HTTP streaming.
  - Optional voice input (Whisper + PyAudio), available depending on user preference.

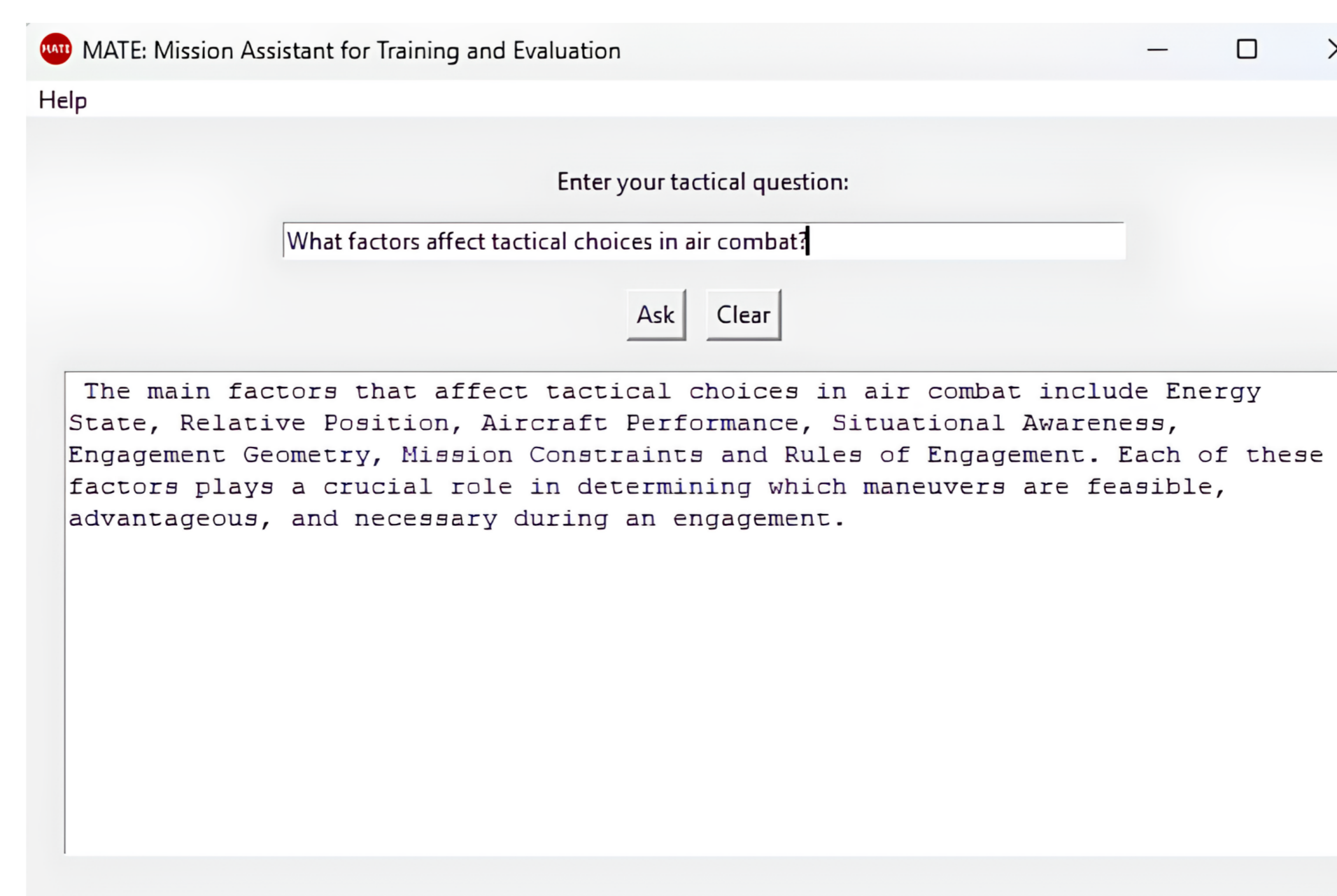


Figure 4. Screenshot of the offline assistant's graphical interface.

## Validation Process

- **Goal:** Assess whether the tool can support **tactical decision-making** in air combat.
- **Reference Source:** Shaw (1985), **Chapter 5: "Two-versus-One"** (Figure 5).
- **Procedure:** The tool used **three LLMs** (Llama 3.1 8B, Qwen 2.5 14B, Mistral NeMo 12B) to answer **5 tactical questions** derived from the book.
- **Participants:** **Brazilian Air Force fighter pilots** with operational experience.
- **Expert Review:** **3 pilots** rated the LLM answers on a **1–5 scale** and provided feedback on the responses and the tool itself.



Figure 5. Evaluation reference: Shaw (1985).

Aviation	Avg. Oper. Exp. (yrs)	Avg. Flight Hours	LLM Familiarity	Common Models
Fighter Aviation	6.7	933	Moderate (occasional use, basic knowledge)	ChatGPT, Gemini, Llama

Table 1. Average pilot profile.

	Tactical Question	Llama 3.1 8B	Qwen 2.5 14B	Mistral NeMo 12B
Q1	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	Main advantages of the Fighting Wing formation for new pilots?	4.67 ± 0.58	4.33 ± 0.58	4.67 ± 0.58
Q3	Main weaknesses of the Fighting Wing doctrine?	3.00 ± 0.00	3.00 ± 0.00	3.00 ± 0.00
Q4	Tactical innovation of the Double Attack 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

Table 2. Ratings for tactical air combat questions (1–5 scale).

Evaluation Question	Mean ± Std. Dev.
Relevance of Retrieval – correctness of sourced information	3.67 ± 0.58
Terminological Accuracy – correct tactical concepts used	<b>4.33 ± 1.15</b>
Integration of Retrieved Context – correct grounding	3.67 ± 0.58
Clarity and Coherence – structure and readability	<b>4.33 ± 0.58</b>
Response Time – compared to manual search	<b>5.00 ± 0.00</b>

Table 3. General evaluation of the tool.

## Conclusion and Future Work

- Built an **offline tactical assistant** using local LLM + document retrieval for use in **disconnected or classified settings**.
- Fully **open-source and modular**, suitable for simulators and mission planning.
- **Pilot validation** indicated strong accuracy and operational relevance.
- **Next steps:**
  - Expand pilot evaluations to broader tactical scenarios.
  - Improve response time with enhanced hardware (e.g., more capable GPU).
  - Increase interactivity and usability in the interface.
  - Optimize the voice interaction module, currently too heavy for routine use.

## References

Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2021. Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data* 7, 3 (2021), 535–547. doi:10.1109/TBDATA.2019.2921572

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Hong Kong, China, 3982–3992. <https://aclanthology.org/D19-1410>

Robert L. Shaw. 1985. *Fighter Combat: Tactics and Maneuvering*. Naval Institute Press, Annapolis, MD, USA.