Title: A Domain-Restricted Local LLM for Meteorology Education: Faster, More Accurate Responses with Reduced Energy and Water Footprint

The rapid expansion of Large Language Models (LLMs) has created new opportunities for climate science education and decision support. However, general-purpose LLMs that rely on large-scale internet data may introduce inaccuracies, hallucinations, or content from unreliable sources. To address these challenges, we developed a fully local, domain-restricted Retrieval-Augmented Generation (RAG) system designed to operate exclusively on curated meteorological textbooks and peer-reviewed academic materials

The system processes selected PDF and EPUB books used in an undergraduate Introduction to Meteorology course. Text is extracted, filtered, segmented into semantically coherent chunks, and embedded using a multilingual transformer model. These embeddings are indexed via a FAISS vector database. During inference, user queries retrieve the most relevant passages, which are then provided as contextual input to a locally hosted open-weight LLM (via Ollama). The model is strictly instructed to generate answers grounded in the indexed literature, with explicit source citation.

Initial tests demonstrate that this domain-restricted system delivers faster and more accurate responses when compared to internet-dependent platforms such as ChatGPT or Gemini. Because it operates locally and accesses only a finite, curated knowledge base, it significantly reduces computational overhead associated with large-scale online searches. This translates into lower electricity demand and reduced water consumption in large data centers.

If replicated across universities, this approach could generate substantial global savings in energy and water use. Moreover, by restricting outputs to textbooks and peer-reviewed articles, the system minimizes conceptual errors, avoids unreliable platforms such as Wikipedia, and prevents AI-generated fabrications. The result is a scientifically rigorous, environmentally efficient, and scalable model for trustworthy AI integration in higher education.