**📄 RAG PDF Q&A με Quantized DeepSeek-7B**

Αναφορά: Δυνατότητες, Περιορισμοί και Προτάσεις Βελτίωσης

Εφαρμογή από: [Το Όνομά σας]

Ημερομηνία: Μάιος 2025

# RAG PDF Q&A με Quantized DeepSeek-7B – Τεχνική Αναφορά

## Δυνατά Σημεία (Strengths)

- Εύχρηστο Streamlit interface για χρήστες χωρίς τεχνικές γνώσεις.  
- Χρήση ChromaDB για ταχύτατη αναζήτηση συμφραζομένων.  
- DeepSeek-7B με quantization 4-bit για απόδοση ακόμα και σε A10g GPUs.  
- Υποστηρίζει ελληνικά και αγγλικά, με αυτόματη αναγνώριση γλώσσας (langid).  
- Αυτόματος καθαρισμός retriever και vectorstore για κάθε νέο PDF.  
- Υψηλής ποιότητας embeddings με MiniLM.

## Περιορισμοί (Limitations)

- Δεν υποστηρίζει OCR για εικόνες εντός PDF.  
- Δεν γίνεται paraphrasing για ερωτήσεις στα ελληνικά.  
- Αν το PDF είναι πολύ μικρό ή με λίγα δεδομένα, το LLM μπορεί να απαντήσει "Δεν γνωρίζω."  
- Δεν υποστηρίζει ακόμη πολλαπλά PDFs ή συνδυασμό πηγών.

## Περιπτώσεις για Έλεγχο (Cases to Test)

- PDF με πολύ μεγάλο όγκο κειμένου (>50 σελίδες).  
- Ερωτήσεις που απαιτούν κατανόηση πολλών chunks.  
- Ερωτήσεις σε διαφορετικές γλώσσες (GR/EN).  
- PDF με περίπλοκη μορφοποίηση (πίνακες, λίστες).

## Προτάσεις για Μελλοντικές Βελτιώσεις

- Προσθήκη paraphrasing και για ελληνικά (αν υπάρξει διαθέσιμο μοντέλο).  
- Υποστήριξη για πολλαπλά PDFs (multi-document retrieval).  
- Προσθήκη caching embeddings για επαναλαμβανόμενα έγγραφα.  
- Ανάπτυξη fallback OCR pipeline (προαιρετικά, αν ζητηθεί στο μέλλον).  
- Βελτιστοποίηση chunking με adaptive size (ανάλογα το περιεχόμενο).  
- Υποστήριξη χρηστών με πολλαπλές ερωτήσεις/συνεδρίες.