EESTEC Hackathon – Patras, May 2025

# 🎯 Θέμα

Ανάπτυξη Έξυπνου AI Chatbot για Τεχνική Υποστήριξη με Βάση Έγγραφα PDF

# 🧩 Εισαγωγή

Στη σύγχρονη εποχή της πληροφορίας, η αναζήτηση τεχνικών γνώσεων σε μεγάλα και πολύπλοκα έγγραφα αποτελεί καθημερινή πρόκληση για φοιτητές, μηχανικούς, τεχνικούς και επαγγελματίες. Τα έγγραφα PDF, όπως εγχειρίδια χρήσης, τεχνικές προδιαγραφές ή πρότυπα, συχνά ξεπερνούν τις εκατοντάδες σελίδες και η ανάκτηση της σωστής πληροφορίας μπορεί να είναι χρονοβόρα και αναποτελεσματική.

Η **Τεχνητή Νοημοσύνη (AI)**, και ιδιαίτερα τα **έξυπνα chatbots με γλωσσικά μοντέλα (LLMs)**, μπορούν να μετασχηματίσουν αυτήν την εμπειρία: επιτρέπουν στον χρήστη να κάνει ερωτήσεις σε φυσική γλώσσα και να λαμβάνει άμεσα και στοχευμένα αποσπάσματα, ή και πλήρεις απαντήσεις, βασισμένες αποκλειστικά στο περιεχόμενο των εγγράφων.

# 🎯 Στόχος του Hackathon

Στο πλαίσιο αυτού του διαγωνισμού, ζητείται από τους συμμετέχοντες να αναπτύξετε ένα **AI-powered Chatbot** που λειτουργεί ως εικονικός τεχνικός βοηθός. Το σύστημα θα πρέπει να «κατανοεί» το περιεχόμενο που υπάρχει σε **έξι τεχνικά έγγραφα PDF** και να παρέχει:

1. **Ακριβή αποσπάσματα** σχετικών πληροφοριών.
2. **Ολοκληρωμένες απαντήσεις** σε ερωτήσεις, με αιτιολόγηση από τα έγγραφα.
3. **Διαδραστική διεπαφή χρήστη**, φιλική προς τον τελικό χρήστη.

# 📦 Τι σας παρέχουμε

* 5 έγγραφα PDF που αποτελούν τη "βάση γνώσης" του συστήματος.
* Ένα αρχείο Excel που θα χρησιμοποιηθεί για την τεκμηρίωση και αξιολόγηση των αποτελεσμάτων.

# 🛠️ Τεχνικές Οδηγίες

1. **Ανάκτηση Σχετικών Αποσπασμάτων**

* Εργαλείο: Python (.ipynb notebook)
* Περιγραφή:
  + Δεχόμαστε ως είσοδο μία ερώτηση σε φυσική γλώσσα.
  + Χρησιμοποιείται μηχανισμός αποθήκευσης (π.χ. **FAISS**, **ChromaDB**) για αναζήτηση στο περιεχόμενο των PDF.
  + Επιστρέφονται τα **5 πιο σχετικά αποσπάσματα (chunks)** από τα έγγραφα.

1. **Αρχική Διεπαφή Χρήστη (UI) για τα Αποσπάσματα**

* Εργαλείο: Streamlit (.py)
* Περιγραφή:
  + Φόρμα ερωτήματος από τον χρήστη.
  + Προβολή των 5 chunks που επέστρεψε ο μηχανισμός ανάκτησης.

1. **Ανάπτυξη AI Chatbot**

* Εργαλείο: Python (.ipynb)
* Περιγραφή:
  + Χρήση open source LLM models όπως Mistral, Llama, Qwen τα οποία θα μπορούν να τρέχουν τοπικά σε ένα τοπικό μηχάνημα.
  + Το chatbot δημιουργεί τεκμηριωμένη, φυσική απάντηση, βασισμένη στα αποσπάσματα που ανακτήθηκαν.
  + Δυνατότητα για “chain of thought” reasoning ή citations.

1. **Δεύτερη Διεπαφή Χρήστη για το Chatbot**

* Εργαλείο: Streamlit (.py)
* Περιγραφή:
  + Ο χρήστης υποβάλλει ερώτηση.
  + Το σύστημα απαντά απευθείας με τεκμηριωμένη πρόταση του chatbot.
  + Μπορεί να εμφανίζονται και οι πηγές (π.χ. με επισήμανση των PDF sections).

1. **Συμπλήρωση Αρχείου Αξιολόγησης**

* Αρχείο: Excel (επισυναπτόμενο αρχείο)
* Στήλες:
  + **Retrieved Chunks:** τα 5 αποσπάσματα που βρέθηκαν.
  + **System Answer:** η τελική απάντηση του chatbot.

# 📂 Σύνοψη Παραδοτέων

1. retriever.ipynb
2. retriever\_ui.py
3. chatbot.ipynb
4. chatbot\_ui.py
5. evaluation.xlsx

**ΣΗΜΕΙΩΣΗ:** Για τα **Α** και C, στα **2 αρχεία .ipynb**, θα πρέπει να φαίνεται ξεκάθαρα η λειτουργία και τα αποτελέσματα του μηχανισμού ανάκτησης και του chatbot **για όλες τις ερωτήσεις**. Σε κάθε περίπτωση πρέπει να συμπληρωθεί και το evaluation.xlsx.

# 🧰 Προτεινόμενα Εργαλεία & Τεχνολογίες

* Python libraries: langchain, PyPDF2, sentence-transformers, streamlit
* Vector Stores: FAISS, ChromaDB
* Language Models: HuggingFace (LLama, Mistral)
* PDF Parsing: PyMuPDF, pdfplumber