

# Generative AI Research for Multimodal Document Understanding

## Project Description:

This project is organized as a portfolio of independent, high-impact subprojects under one shared goal: enabling strong multimodal LLM capabilities for Arabic/English Visual Document Understanding and interaction, from PDFs and scans to charts, diagrams, maps, complex layouts, tables/forms, and historical documents. Interns will choose a subproject that matches their interests and strengths and own it end-to-end, while still integrating with the broader platform. Subprojects span both research and engineering, including improving Arabic OCR, building document embeddings/retrieval, LLM/SLM fine-tuning for grounded document QA, developing robust evaluation for citation-accurate answers, and creating UI/UX for interactive document experiences (evidence highlighting, smart navigation, visual translation). Students may also specialize in focused domains such as maps, legal, historical archives, or Qur'anic collections, contributing deep expertise while producing a concrete, standalone deliverable that strengthens the overall system.

## Sub-projects Include

- Improving QA components (NLP research)
- Improving Arabic OCR for handwriting and historical documents (Language-Vision Research)
- Domains of interest can be picked by the interns:
  - Quranic/Hadith QA
  - Charts and diagrams
  - Map interactions
  - Legal QA
- Improving visual translation models (NLP-Vision Research)
- UI/UX dev for interaction with collections of documents.

## Required Skills:

Any one of the following is sufficient (teams are multidisciplinary):

- **ML/AI:** (L)LMs, RAGs, Agents, Multimodal models, Training/evaluation
  - **Engineering:** Python, APIs, data pipelines, backend/frontend development
  - **UI/UX:** interaction design and prototyping for document experiences (highlighting, navigation, overlays)
  - **Core:** strong programming fundamentals and ability to collaborate in a team setting
-

## Preferred Intern Academic Level:

All levels are welcome:

**BSc, MSc, or PhD** in Computer Science or related fields and linguistics

(strong coding skills expected; research experience is a plus but not required).

---

## Learning Opportunities:

- Contributing to Fanar GenAI platform (national project)
  - Publishing the work in AI conferences
  - Ownership of an end-to-end subproject: **define** → **build** → **evaluate** → **demo** → **integrate** into a larger platform.
  - Practical multimodal benchmarking and enhancement
  - Real-world Arabic document intelligence experience: OCR/layout, tables/forms, retrieval + embeddings, and evidence-based interaction design.
- 

## Expected Team Size:

This is a meta project that will be organized into small teams of 2-4, each owning one subproject, with opportunities for cross-team integration.

## Mentor

**Name:** Ehsaneddin Asgari

**Email:** [easgari@hbku.edu.qa](mailto:easgari@hbku.edu.qa)