

AI Intern Assignment

https://drive.google.com/drive/folders/1YfeyExTs8dJoeyv8NPlpwrk-MKjzGCwG?usp=drive_link

✓ Note:

- Feel free to use Q&A content from *any domain* you like — books, movies, tech, sports, philosophy, etc. Please don't spend more than **30 minutes** gathering or preparing your sample questions.

✖ Objective:

Develop an intelligent system that can extract questions from PDF documents and generate accurate answers, preserving the original format of the input (tables, checkbox forms, or plain text). Sample files are in the drive link above.

🔧 Your Task:

1. **Parse and analyze PDF files** containing questionnaires on various topics (e.g., security, books, authors, etc.).
2. **Identify and extract all question types**, regardless of format:
 - Structured tables
 - Checkbox-based questions with options
 - Descriptive/plain text questions
3. **Generate contextually accurate answers** to these questions using LLMs & Simple RAG (Retrieval-Augmented Generation)
4. **Output the answers in the same format as the input** — i.e., return tables with filled answers, checkboxes marked, or restructured paragraphs with answers.

✓ Deliverables:

A runnable solution (script, app, or notebook) that:

- Takes a PDF file as input
- Returns a filled PDF with answers in original format

📦 Submission Guidelines:

- You may use **any programming language** and libraries/frameworks of your choice (Python, Node.js, Java, etc.).
- Submit your solution as a GitHub/GitLab repository link. Include a **README.md** with setup instructions and a brief explanation of your approach.
- Sample input and output. One or Two files would do.