

## Project plan

### DATASET EXPLANATION:

**rag-mini-bioasq** from hugging face datasets --> enelpol/rag-mini-bioasq //cleaner version **specialized dataset for biomedical questions** as I'll build biomedical Ques ans sys its a reduced version of bioask dataset, already structured for training and testing RAG sys

## The dataset provides-->

---

- **Biomedical questions**
- **Relevant passages** from biomedical literature: info needed to answer questions
- **Expert-written answers** to these questions: standard responses

### MILESTONES

- **milestone 1**:data pre-processing and exploration
- **milestone 2**:Text embedding, and vector db — *Estimated: 1 day*
- **milestone 3**: implement retrieval part — *Estimated: 1 day*
- **milestone 4**: prompt construction& generation — *Estimated: 1 day*
- **milestone 5**: error handling, optimizations — *Estimated: 2 days*
- **milestone 6** : building basic gui-> streamlit (most probably) — *Estimated: 2 days*
- **Last step**: checking everything and doc --> wihin gui time

### Deadlines

- **estimated deadline**--> 7/5
- **Official deadline**--> 10/5 isa