# **PolyglotAI – Phase 2 (Data & Setup) Timeline**

### **Week 1**

**Day 1–2: Data & Repo Setup**

* Gather project docs (PRD-lite, Decision Log).
* Create folder structure (docs/, code/, data/, tests/).
* Initialize GitHub repo.

**Day 3–4: Embedding Model Selection (EN/AR/FR)**

* Compare multilingual embedding models.
* Run sample tests in **all 3 languages**.
* Document choice + rationale.

**Day 5: Vector Store Decision**

* Compare FAISS, Chroma, Pinecone.
* Choose store and set up minimal instance.

### **Week 2**

**Day 6–7: Data Indexing**

* Store and clean project docs.
* Index them into the chosen vector store.
* Test retrieval with EN, AR, and FR queries.

**Day 8–9: Multilingual Testing (EN/AR/FR)**

* Build **trilingual test prompt set**.
* Validate embedding + vector store performance across all 3 languages.
* Test cross-language retrieval (e.g., FR query → EN doc).

**Day 10: Guardrails Prep**

* Draft guardrail logic for off-topic/irrelevant queries.
* Write test cases for multilingual scenarios.





