# **PolyglotAI — Test Plan Skeleton (EN/AR/FR)**

## **1) Scope**

Validate Phase-2 setup for **multilingual Q&A (EN/AR/FR)**, **cross-language retrieval (RAG)**, **per-chat memory**, **citations**, and **basic guardrails**.

## **2) Test Matrix (high level)**

| **Area** | **Goal** | **Pass Criteria** |
| --- | --- | --- |
| Multilingual Q&A | Same answer quality across EN/AR/FR | ≥90% semantic parity on a shared prompt set |
| Cross-language RAG | Retrieve sources regardless of query/doc language | Top-K includes a correct source ≥80% of trials |
| Memory (per chat) | Context persists within a chat only | Follow-ups resolved correctly; no leakage across chats |
| Citations | Provide supporting sources when RAG is used | Each factual answer includes at least one correct citation |
| Guardrails | Avoid irrelevant/off-scope responses | Off-topic prompts are declined with a short, neutral reply |

## **3) Test Cases**

### **A. Multilingual Q&A Parity**

| **ID** | **Description** | **Preconditions** | **Steps** | **Expected** |
| --- | --- | --- | --- | --- |
| MQ-EN-001 | Answer a planning question in English | RAG index built; EN doc exists | Ask: “What are the Phase 2 deliverables?” | Accurate list; fluent EN; no hallucinations |
| MQ-AR-002 | Same question in Arabic (MSA) | Same as above | اسأل: «ما هي مخرجات المرحلة الثانية؟» | Same content as EN; fluent MSA; consistent structure |
| MQ-FR-003 | Same question in French | Same as above | Demander : «Quels sont les livrables de la phase 2 ?» | Same content; fluent FR; consistent structure |

### **B. Cross-Language Retrieval**

| **ID** | **Description** | **Preconditions** | **Steps** | **Expected** |
| --- | --- | --- | --- | --- |
| XR-AR→EN-004 | AR query, EN source | EN doc indexed; AR query phrased | Ask in AR about an item that only exists in EN docs | Correct EN passage retrieved; answer in AR; citation to EN doc |
| XR-FR→AR-005 | FR query, AR source | AR doc indexed; FR query phrased | Ask in FR about item only in AR docs | Correct AR passage retrieved; answer in FR; citation to AR doc |
| XR-EN→FR-006 | EN query, FR source | FR doc indexed; EN query phrased | Ask in EN about item only in FR docs | Correct FR passage retrieved; answer in EN; citation to FR doc |

### 

### **C. Memory (Per-Chat, Short-Term)**

| **ID** | **Description** | **Preconditions** | **Steps** | **Expected** |
| --- | --- | --- | --- | --- |
| MM-001 | Follow-up within same chat | Memory window enabled | Q1: “List Phase 2 tasks.” → Q2: “Group them by week.” | Q2 uses Q1 context correctly |
| MM-002 | No cross-chat leakage | Two chats open | Ask context in Chat A, then refer to it in Chat B | Chat B asks for context; no leakage from A |

### 

### **D. Citations & Traceability**

| **ID** | **Description** | **Preconditions** | **Steps** | **Expected** |
| --- | --- | --- | --- | --- |
| CT-001 | RAG answers must cite | RAG enabled | Ask fact requiring doc lookup | At least one correct citation attached |
| CT-002 | Non-RAG (general) answers | RAG optional | Ask general question (no doc ground truth) | Clear response; either no citation or a generic note (“no sources used”) |

### 

### **E. Guardrails (Relevance/Scope)**

| **ID** | **Description** | **Preconditions** | **Steps** | **Expected** |
| --- | --- | --- | --- | --- |
| GR-001 | Irrelevant prompt | RAG enabled | Ask: “Tell me a joke about cats” (if out of project scope) | Short neutral refusal or steer back to project scope |
| GR-002 | Ambiguous prompt | — | Ask: “Do it” | Ask for 1-line clarification or offer options |

## 

## **4) Sample Trilingual Prompt Set (starter)**

**EN**

* “Summarize the goals of Phase 2.”
* “Which embedding model are we considering and why?”
* “List risks and mitigations for multilingual retrieval.”

**AR (MSA)**

* «لخّص أهداف المرحلة الثانية.»
* «ما نموذج التضمين المقترح ولماذا؟»
* «اذكر المخاطر وخطط التخفيف لاسترجاع متعدد اللغات.»

**FR**

* «Résume les objectifs de la phase 2.»
* «Quel modèle d’embedding envisage-t-on et pourquoi ?»
* «Énumère les risques et mesures d’atténuation pour la recherche multilingue.»

**5) Acceptance Criteria (Phase-2 readiness)**

* ≥10 docs indexed; at least one doc per language (EN/AR/FR).
* Cross-language retrieval passes **XR-004/005/006** in ≥80% trials.
* Parity (**MQ-EN/AR/FR**) judged “equivalent” by checklist in ≥90% cases.
* Memory tests **MM-001/002** pass consistently.
* Citations present for all RAG-based answers (**CT-001**).
* Guardrails respond appropriately (**GR-001/002**) in ≥95% of attempts.