

LOCKED TASK — SOHAM KOTKAR

Role: LM + Multilingual Stability Task

Focus: Make Gurukul's intelligence layer reliable, indigenous, multilingual, and demo-safe

Objective (Non-Negotiable)

Ensure the Language Model → KSML → Multilingual output layer works correctly, consistently, and fast across the Gurukul platform, with Arabic as priority, and without relying on Grok or any external fallback.

This is stabilisation + grounding, not experimentation.

Scope (What Soham touches)

- ✓ Language Model
 - ✓ Multilingual routing
 - ✓ KSML tagging / alignment
 - ✓ MCP-fed knowledge
 - ✓ Response correctness
 - ✗ Frontend UI
 - ✗ Infra / Docker
 - ✗ Auth
 - ✗ TTV internals
-

TIMEBOX

Total: 2.5 days

- Day 1: Core stabilisation
 - Day 2: Multilingual hardening
 - Day 3 (half): Verification + handover
-

DAY-BY-DAY EXECUTION CHECKLIST

DAY 1 — LM & KSML STABILITY (8–9 hours)

Goal: Answers must be correct, structured, and predictable.

1. Disable all fallback logic
 - Ensure NO Grok / external LM is used anywhere.
 - If fallback exists → hard fail (log error, don't answer).
2. Verify LM response pipeline
 - Input → LM → KSML → final text
 - Ensure no missing fields:
 - intent
 - language
 - karma_state (basic)
 - clean output text
3. Fix response quality issues
 - No hallucinated facts
 - No broken sentences
 - No language mixing unless requested
4. Standardise answer formats
 - Explanation questions → structured paragraphs
 - MCQs → numbered with clear options

- Short answers → ≤ 5 lines
- Deliverable (Day 1):
- lm_stability_report.md
 - 10 questions
 - Expected vs actual output
 - Known limitations (if any)
-

DAY 2 — MULTILINGUAL HARDENING (8–9 hours)

Goal: Arabic-first multilingual reliability.

1. Arabic language pass (highest priority)
 - Ask 10 curriculum-style questions in Arabic
 - Verify:
 - Response is fully Arabic
 - No English leakage
 - Correct grammar (basic)
2. Secondary languages sanity
 - Hindi
 - English
 - One low-resource language (e.g. Bhojpuri or Dogri)
3. Language locking
 - Input language = output language
 - No auto-translation unless explicitly asked
4. KSML language tagging
 - Ensure source_lang and target_lang are accurate
 - Log mismatches if detected

- Deliverable (Day 2):
- multilingual_smoke.md

- Language
 - Prompt
 - Pass / Fail
 - Notes
-

DAY 3 (HALF DAY) — VERIFICATION & HANDOVER (4–5 hours)

Goal: Others can trust this layer without asking Soham again.

1. Edge-case testing
 - Very short question
 - Long explanation
 - Follow-up question (context)
2. Latency check
 - Average response time
 - Flag anything > 3 seconds
3. Prepare handover note
 - What works perfectly
 - What is acceptable but not ideal
 - What must NOT be touched before demo

Final Deliverables:

- handover_soham.md
 - Updated repo (no broken branches)
 - Zero open TODOs
-

ACCEPTANCE CRITERIA (STRICT)

Soham's task is **DONE** only if:

- Arabic responses are clean and reliable

- No external LM fallback exists
- KSML fields are always present
- Other devs can use the LM without clarification
- Demo questions do not break the system