**🔹 Option C: *Two-Layer Style* (more advanced)**

You keep **FAQ/db answers as ground truth**, but Mistral acts as a **stylist**:

* Step 1 → Retrieve FAQ (rule/semantic/fuzzy).
* Step 2 → Build a prompt:

The official answer is:

"{faq\_answer}"

Rewrite this answer in a friendly tone, but DO NOT change the meaning.

* If the user asks a question that **isn’t in FAQ**, Mistral generates a **new answer** (like your Step 5).

👉 This ensures **accuracy (no hallucination)** + **friendliness**.

**⚖️ My Pick for You**

Since your chatbot is for **UnityToServe (nonprofit, user-facing)** →  
I recommend **Option C** (two-layer style):

* Keeps control (always anchored in FAQ).
* Always sounds friendly.
* Falls back to Mistral only when there’s no FAQ.

✨ Imagine:

* User asks: *“What is UnityToServe?”*
* FAQ raw: *“UnityToServe is a nonprofit organization.”*
* Mistral-enhanced: *“UnityToServe is a nonprofit organization dedicated to helping communities through service and support programs.”*

More upgrades:

User Query

↓

Contextual Input (from memory\_phase4)

↓

Intent Detection (spaCy)

↓

Semantic Search (SentenceTransformer + ChromaDB)

↓

Re-Ranking (CrossEncoder)

↓

Response:

• Exact Match → Rule Response

• Close Match → Semantic Response

• Low Confidence → RAG fallback (GPT with FAQ context)

↓

Logging + Continuous Learning