AIvengers – Humanizing Digital Conversations

Name	Role
Deepak Y	Conversational AI Developer
Harichandar D	AI Solution Architect
Harshini G	Innovation Specialist
Linu Joseph	Product Planner
Yogesh	Evaluation & Data Engineering Special
Stalin	Mentor

Problem Statement

- Deliver robotic, SOP-bound responses with no emotional intelligence.
- Collapse on complex / multi-step queries.
- Lose track of long contextual conversations.
- Slow down significantly when switching to reasoning models.

Industry Trends

- 64% of customers prefer companies not to use AI for customer service. (Gartner, 2024)
- 53% would switch to competitors if they discover AI is used. (Gartner, 2024)
- Empathy gap & lack of personalization remain the biggest blockers.

Solution Summary

- Unified Orchestrated AI Architecture: A central orchestrator analyzes intent, emotion, and context, selecting the best reasoning path to create a human-like, adaptive chatbot.
- Dual-Agent Processing Model: Fast SOP/DB Agent for quick responses + ReAct Agent for deep multi-step reasoning during complex troubleshooting.

Scope + Dataset

Scope

Telecom dataset – Broadband & Mobile (Postpaid & Prepaid).

Dataset Details

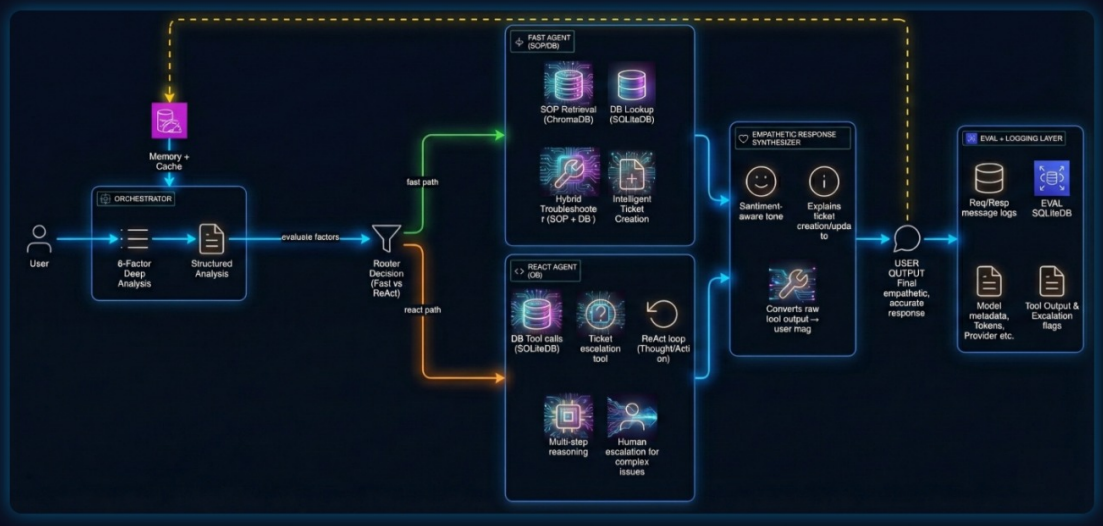
Synthetic dataset with 25 customers across 7 dimensions: product, customer, usage, issues, FAQ, plan, subscription.

Why We Consider

We chose telecom because customer frustration is high and current chatbots lack empathy. Since I have telecom experience, I know exactly where chatbots fail—latency, robotic tone, and hallucination when data is missing.

My role was to define the dataset and scope so we can measure empathy in a structured way. I created a synthetic telecom dataset with 25 customers and 7 key dimensions. This gives the chatbot realistic scenarios without noise, avoids hallucinations, and allows us to evaluate whether our model truly behaves more human-like.

The entire POC relies on this structured dataset to test empathy, tone, relevance, and accuracy in a fair and repeatable manner.



Our Solution – High-Level Architecture

- **1. Intelligent Orchestrator (Brain)**
  Performs 8-factor deep analysis across intent, emotion, and context, routing queries to the optimal reasoning path.
- **2. Fast Agent – SOP / DB Automation**
  Uses SOP-RAG routing, ChromaDB retrieval, SQL rules, and hybrid troubleshooting for instant and consistent telecom answers.
- **3. ReAct Agent – Deep L3 Reasoning Engine**
  Executes multi-step reasoning, Thought→Action loops, DB/tool calls, error handling, and escalation workflows.
- **4. Empathetic Response Synthesizer**
  Applies tone calibration, emotional rewriting, and trust-building phrasing to deliver human-like, supportive responses.
- **5. State, Memory & Context Layer**
  Maintains long-term conversation memory, continuity, and safety signals for consistent and personalized user experience.
- **6. Logging, Telemetry & Evaluation**
  Tracks complete interaction logs, model metadata, escalation flags, and tool outputs for auditability and improvement.
- **7. Continuous Learning Loop**
  Evaluation signals refine orchestration logic, reduce hallucinations, and improve deep reasoning accuracy.
- **8. Unified Outcome**
  A chatbot that feels human, solves complex issues intelligently, maintains context, and delivers real-time reliable support.

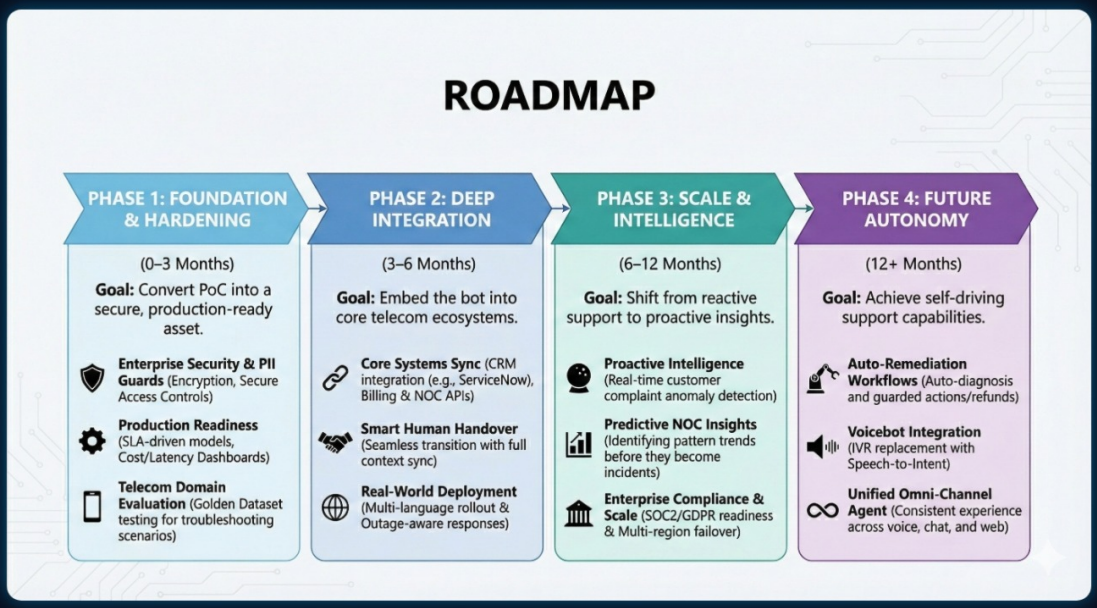
“A unified architecture combining speed, intelligence, empathy, and reliability — built for telecom.”

Key Features & Business Benefits

- **1. Human-Like, Empathy-Driven Support**
  Emotion-aware tone for supportive responses.
- **2. Fast & Accurate Troubleshooting (<300ms)**
  SOP flows + long-context memory.
- **3. Deep Reasoning for Complex Issues**
  ReAct multi-step reasoning.
- **4. Operational Efficiency**
  40–60% fewer escalations and reduced SLA penalties.
- **5. Enterprise-Grade & Scalable**
  Secure, compliant, and modular architecture.

- **1. Dataset gap:** No telecom noise, multilingual data.
- **2. Troubleshooting gap:** No live CRM/NOC billing APIs.
- **3. Scale gap:** ReAct untested under outage traffic.
- **4. Compliance gap:** Real PII validation missing.
- **5. Evaluation gap:** No golden or hallucination tests.
- **6. Workflow gap:** Ticketing & handover not integrated.
- **7. Observability gap:** No latency/cost/error dashboards.

Roadmap & Deployment Plan



Conclusion Summary

- ✓ **1. Unified AI balancing speed, empathy, and deep reasoning.**
  Human-like, accurate, context-aware telecom support.
- ✓ **2. Enterprise-ready architecture with measurable value.**
  Secure, modular, scalable — with strong CSAT and cost improvements.
- ✓ **3. Proven feasibility with a clear path to production.**
  Humanized interactions + responsible AI + strong evaluation metrics.

Live Golden KPI Monitor

CareConnectAI Chatbot – KPI Metrics

Stronger connection and kinder conversation – measuring how empathetic, accurate, and stable your LLM customer experience truly is.

GOLDEN BENCHMARK	LOGS MATCHED
95 test cases	150 entries

SENTIMENT ACCURACY
44.0%
How often the bot gets the customer's feeling right.
Target ≥ 80% — Needs empathy tuning

ROUTING ACCURACY
9.3%
Correctly selecting between fast/React/agent flows.
Target ≥ 85% — High misrouting risk

TOOL SELECTION ACCURACY
36.7%
Picking the right resolver vs FAQ vs other tools.
Target ≥ 90% — Tool strategy review

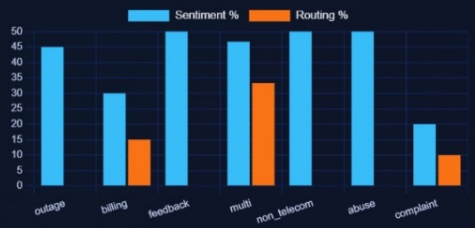
ESCALATION ACCURACY
86.7%
Handing off to humans at the right moment.
Target ≥ 90% — Close to ideal

OVERALL KPI SCORE
44.2%
Weighted blend of sentiment, routing, tools & escalation.
Target ≥ 85% — Scaling not yet safe

AVG TOKENS / RESPONSE
342.9
Response length & cost footprint per interaction.
Ideal band 120–260 — Could be more concise

Model Usage Distribution
gpt-4o vs gemini-2.5-pro

Per-Intent Performance
Sentiment vs Routing accuracy



Model Usage (Counts)
From 150 matched logs

Model Name	Responses
gpt-4o-2024-11-20	42
gemini-2.5-pro	33

Per-Intent KPI Snapshot
Golden benchmark coverage

Intent	Test Cases	Sentiment %	Routing %
outage	20	45.0%	0.0%
billing_issue	20	30.0%	15.0%
feedback	40	50.0%	0.0%
multi_problem	30	46.7%	33.3%
non_telecom	20	50.0%	0.0%
abuse	10	50.0%	0.0%
complaint	10	20.0%	10.0%

Sample Golden vs Actual (First 20 Logs)
Use this slice to visually inspect where the model deviates from the golden expectations.

Diagnostics · Outage & Billing focused

User Query	Expected Intent	Expected Sentiment	Actual Sentiment	Expected Routing	Actual Model	Expected Tool	Approx. Actual Tool	Expected Escalation	Actual Escalation
My internet has been down all morning	outage	frustation	Frustration	react	gemini-2.5-pro	resolve_issue	resolve_issue	0	0
My internet has been down all morning	outage	frustation	None	react	None	resolve_issue	faq_engine	0	0
The broadband connection stopped working today	outage	frustation	Frustration	react	gemini-2.5-pro	resolve_issue	resolve_issue	0	0
The broadband connection stopped working today	outage	frustation	None	react	None	resolve_issue	faq_engine	0	0
Why is my WiFi not working since earlier?	outage	frustation	Frustration	react	gemini-2.5-pro	resolve_issue	resolve_issue	0	0