AI Customer Support Strategy

Tiago Ferreira da Silva

Part 1: Strategy Development

1. Opportunity Assesment

What is the problem?

MyServices' Customer Support is overwelmed: each agent handles 650 tickets per day and still only manages to address 60 % of total volume. Responses are delayed, quality is inconsistent, and Customer Satisfaction (CSAT) has dropped by 20 percentage points.

Where is the opportunity?

AI can help reduce agent overload, improve time & quality responses and enable 24/7 support- especially for repetitive or simple cases.

Method: How to assess it?

We approach the problem by identifying:

- Where AI brings the most value and has the highest potential impact.

- Which parts of the customer support process can be safely and effectively automated first.

Criteria for Assessment:

First, we break the problem down into two dimensions:

a) Ticket type – **What** the customer is asking for
b) Support channel – **Where** the interaction is happening

For each dimension, we evaluate AI suitability based on the following factors:

- **Volume**: How frequently does this type of ticket or channel occur?

- **Complexity**: How hard is the issue to resolve? How difficult is it to automate?

- **Sensitivity / Risk**: What is the impact of a poor or incorrect response?

- **AI Suitability** – Overall fit for automation, based on a trade-off between the above criteria.

Below is the evaluation we performed to assess AI suitability across ticket types and support channels.

Ticket type	Volume (%)	Complexity	Sensitivity / Risk	Automation Fit	Explanation Notes
General Info	48%	Low	Low	High	Repetitive: good for LLM Assistant + RAG
Sign Up Issues	24%	Medium	Medium	Medium-high	Chatbot with fallback
Cancel Subscription	11%	High	High	Medium-low	Very sensitive
Change Personal Info	8%	Medium	Medium	Medium	Automate with identity check
Complaints	4%	High	High	Low	Requires empathy, human interaction
Unsubscribe MKT	3%	Low	Low	High	Low effort
Other	2%	-	-	Medium	Case-by-case

Table 1 - AI Suitability by Ticket Type

Channel	Volume (%)	Real-time support	Complexity to implement AI	Sensitivity / Risk	Automation/ AI Fit	Explanation Notes
Chat	42%	Yes	Low	Low	High	best fit for AI Assistants / easy to integrate
SMS	38%	Yes	Medium	Low	High	also easy to integrate. Length constrains
Email	15%	No	Low	Medium	Medium-High	good for fallback automation for complex / long tickets
Phone	5%	Yes	High	High	Low	requires sophisticated voice interfaces (costly) with emotional tone detector

Table 2 - AI Suitability by Support Channel

From this assessment, we identify high-volume, low to medium risk ticket types such as:

- General Product Information (48 %)
- Marketing opt-outs (3%)
- Issues signing up (24 %)

These represent **75 % of all support tickets** making them strong starting points for AI automation. Their high volume means automating them would significantly reduce the workload on human agents. At the same time, their relatively low complexity and low senstivity reduce the risk of automation erros. These makes them ideal candidates for AI-based solutions such as LLM agents.

The best initial channels for implementation are **Chat** and **SMS**, which together handle **80 % of total traffic**. These channels are suited for text-based interactions allowing quicker response times. AI agents that can operate around the clock and handle simple replies is ideal here, as it reduces human workload without requiring advanced and costly voice recognition interfaces.

In contrast, **complaints**, **cancellations** and **phone-based support** should remain human-handled during early stages due to complexity and high business risk.

2. Solution Design

2.1 What are the fundamental aspects of your solution?

The core elements of the proposed solution are:

A. AI Agent

An AI-powered virtual assistant that uses a Large Language Model (LLM) to respond to customer questions in natural language. It is designed to handle common, low-risk tickets — like general information, sign-up help, and opt-out requests — through Chat and SMS, which represent 80% of support volume.

B. RAG with Internal Knowledge

To reduce hallucinations and ensure accurate, grounded answers, the AI agent uses Retrieval-Augmented Generation. This method fetches relevant information from MyServices' internal documentation (FAQs, product details, policies etc) before generating a response.

The workflow is as follows:

1. Convert user query into vector using embedding model
2. Search a vector database of embeded company documents
3. Retrieve the most relevant content
4. Feed that content into the LLM as context for response generation

C. Triage System (Classification Machine Learning Models)

Before a support ticket is routed to the AI agent, it is first processed by a set of three independent machine learning models:

1. **Intent Classifier** – Predicts what the user is asking (e.g., info request, sign-up help, complaint).
2. **Urgency Classifier** – Predicts how time-sensitive the request is (low, medium, high).
3. **Risk Classifier** – Predicts how sensitive or risky the ticket is (e.g., legal, financial, emotional), using a combination of machine learning and rule-based logic.

These three labels are then fed into a final **Routing Model**, which makes a binary decision:

- Send the ticket to the AI agent
- Escalate the ticket to a human agent

This two-step pipeline ensures that only safe, automatable tickets are handled by AI.

D. Escalation & Fallback Logic

The system includes a robust fallback mechanism that routes conversations to human agents when:

- The LLM's confidence is low
- The user requests a human agent
- The ticket is of high-risk (compaints, cancellations)

E. Feedback Mechanism

The system captures:

- User feedback (e.g., thumbs up/down, CSAT)
- Human agent corrections (after escalation)

This feedback is used to:

1. Improve the classification model
2. Fine-tune the LLM (if needed)
3. Identify gaps in the knowledge base and expand documentation

F. Multi-channel integration

The AI agent is initially integrated with **Chat** and **SMS** via their respective APIs. These channels are prioritized due to their high usage and suitability for text-based interactions.

Email will be integrated in a later phase (Phase 4 of Implementation Roadmap), once the system is stable and optimized for initial channels.

2.2. Solution Architecture

Figure 1 presents a high-level architecture of the proposed AI-powered customer support system. It illustrates the key components involved in ticket intake, triage, automated response generation, human fallback, and continuous improvement.

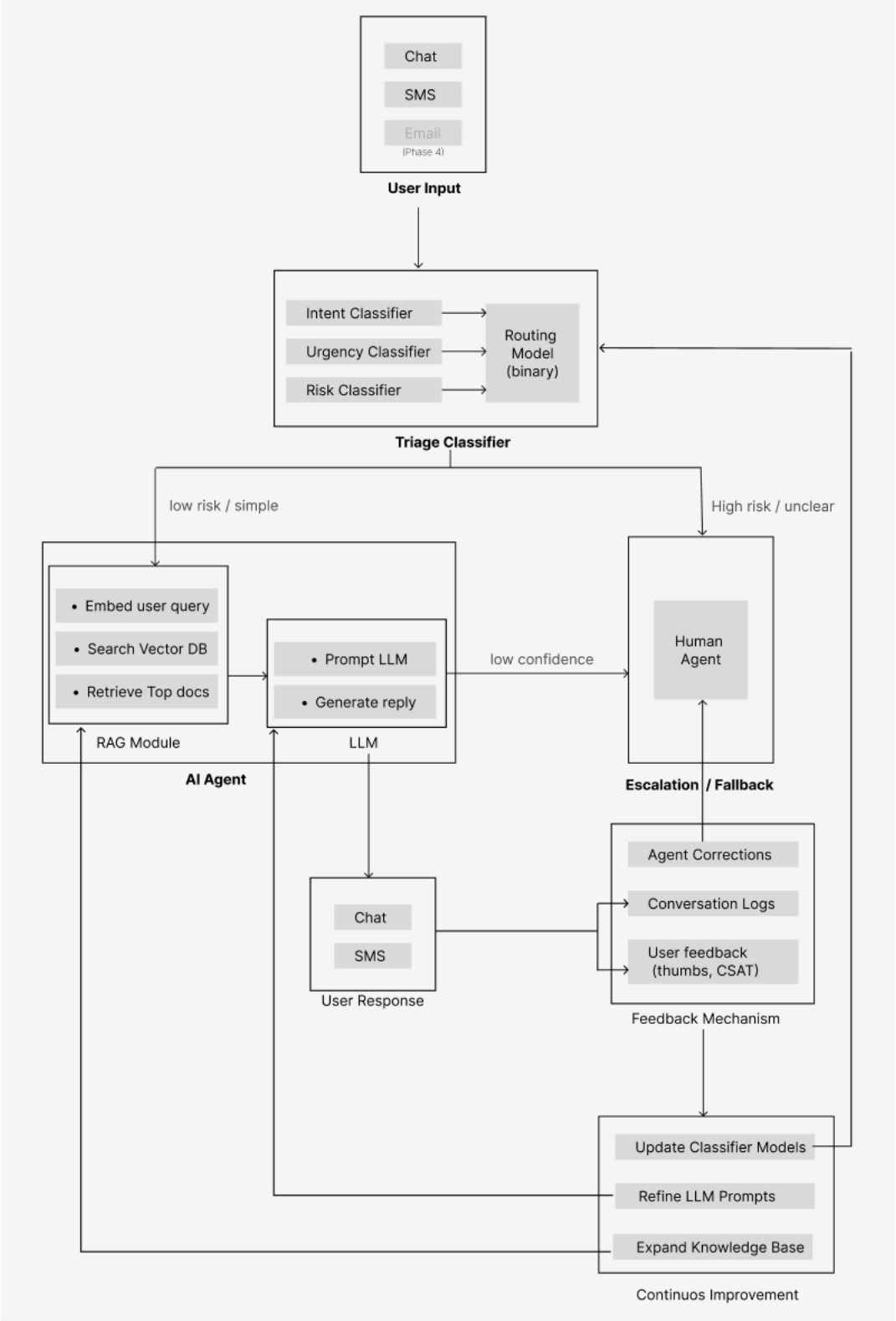

Figure 1 – Proposed Solution Architecture

2.3. Quality and Safety Guardrails

In order to ensure the AI system behaves reliably, safely and as expected, I have defined a set of safety mechanisms or guardrails listed bellow:

1. **Confidence Thresholds**

A confidence score is assigned by the model (LLM and Classifiers) that reflects how certain the AI is about its output / respose.

How it works:

- If confidence score > thershold -> AI response
- If confidence score < threshold -> escalate to human with fallback message.

Where does the confidence score come from?

- Intent, Urgency and Risk Classifiers -> vector of probabilities (e.g. softmax probabilities)
- LLM -> embedding similarity / RAG match score*

* embedding similarity is obtained by calculating the cosine similarity between user query vector and (internal) document vector

2. **Topic Restrictions**

In this guardrail, we define a list of “out-of-scope” / off-limits topics for the AI. If the classifier detects these, it will be immediatly routed to human agent.

List of topics example:

- Refund requests

- Legal complaints or mention of lawyers

- Billing disputes

- Request to change account information

- Data privacy

3. **Fallback prompts**

If the AI does not find relevant content (retrievel match) , is not confident about the response or the question is out-of scope, it should return a safe fallback message instead of attempting an uncertain answer.

These fallback prompts help prevent hallucinations, protect user trust and ensure smoth escalation to a human agent when needed.

Examples of fallback messages include:

- *"I'm not sure I can help with that — let me connect you to someone who can."*
- *"This might be better handled by a support specialist. I'll forward this now."*

4. **Human Escalation on Request**

This guardrail ensures users stay in control. If the user expresses a desire to speack with a human agent, the system should immediately escalate without delay.

How it works:

Intent classifier detects "human escalation"

AI responds "sure, I'll connect you with na agent right away"

AI agent stopps the conversation

Escalation of the case together with chat history

5. **Monitoring**

This guardrail helps understand how well the AI system is performing. A monitoring system tracks:

- Every conversation between users and the AI
- User feedback (like thumbs or CSAT)
- Agent corrections when human takes over and and rewrites AI's response

This guardrail is fundamental to:

- Help identify weak or wrong AI responses
- Highlight topics that often cause confusion
- Provide data to improve AI, prompts and knowledge base
- Update performance dashboards

3. Implementation Roadmap

3.1. Phased Rollout with Milestones

The solution will be rolled out step by step. In Phase 1, we focus on building the core system. In Phase 2, we test it in a controlled environment. Phase 3 is for improvement and expansion based on feedback. Finally, in Phase 4, we scale the solution to broader use.

<u>Phase 1:</u> Setup of core System | Month 1 - 2

In this preliminar phase, we build the main functionalities of the system (pilot version) and prepare it for internal testing. This stage is planned to take place during the first two months. See Table 3 for tasks and corresponding milestones.

Task	Milestone
Set up API's for Chat and SMS integration	Chat and SMS integrated and ready for tests.
Build the Classifiers (Intent, Risk and Urgency)	Classifiers achieve > 80 % accuracy on historical tickets
Create 1st version of internal knowledge base	Knowledge base (v1) ready and covers > 80 % of general info + MKT out-out queries
Set up RAG to let the LLM use internal documentation	AI respondes to test queries based on retrieved documents
Add fallback logic and escalation triggers	System can safely escalate unclear or risky queries
Run internal QA with historical tickets	All core system components (classifier, RAG, fallback and escalation) test and approved by team

Table 3 – Phase 1: Tasks and Milestones

<u>Phase 2</u>: Launch Pilot with Real Users | Month 3

At this stage, we are ready to launch the first version of the AI system. During Month 3, the pilot will be tested in a safe and controlled environment in a limited scope with human oversight. See Table 4 for the key tasks and corresponding milestones.

Task	Milestone
Launch pilot on Chat & SMS (80 % of tickets)	AI Assistant is live in Chat & SMS channels
Handle low-risk ticket types only: - General Product info (48 %) - MKT opt-out (3%)	AI attempts to answer 50 % of support queries (safe tickets)
All AI responses reviewed (and corrected) by human agents	Human in-the-loop system in place
Begin collectiing feedback data: - User feedback (thumbs and CSAT) - Agent corrections and escalations - System fallback triggers	First feedback and escalation data collected that will support continuos system improvement
Weekly reviews of pilot performance with team	First improvements made based on pilot feedback and performance.

Table 4 – Phase 2: Tasks and Milestones

<u>Phase 3:</u> Expansion of AI system (controlled) | Month 4 - 5

At this stage, the AI system is ready to take on more types of tickets with less help from humans. In Phase 3, the AI will start replying directly when it's confident, and support more tickets like sign-up issues. Table 5 shows the main tasks and goals for this phase, which should take about two months (months 4 and 5).

Task	Milestone
Add more medium-risk ticket types: - Sign-up issues (24 %)	AI handles > 75 % of total ticket volume
Improve triage Classifiers using data from Phase 2	Escalation Rate < 25 %
Expand knowledge base (for new ticket types and failed pilot cases)	Knowledge base covers 90% of the top 50 frequent queries
Set LLM + Classifiers confidence thresholds	First tickets routed automatically without human review
Keep complaints, cancellations, personal info changes manual	High-risk tickets still handled by agents

Table 5 - Phase 3: Task and Milestones

<u>Phase 4:</u> Full Launch of AI Support | Month 6

In Phase 4, the AI starts working on all safe ticket types and also begins handling emails. It gives answers on its own when it’s confident. The team now watches how well it works and makes sure everything runs smoothly. Table 6 shows the tasks and goals for this phase, which should take one month (month 6).

Task	Milestone
Launch Email Support (15 % Volume) for simple queries	AI handles Chat, SMS and Email for safe use cases
Fine-tune fallback logic based on Phase 3 data	Fallback rate < 15 % Escalation rate < 20 %
Track live AI performance metrics: - CSAT - AI resolution rate - Escalation Rate - Fallback rate	System is considered stable when following KPIs are met for 2 consecutive weeks: - CSAT > 85 % (on AI-resolved tickets) - AI resolution rate > 70 % - Fallback rate < 15 % - Escalation rate < 20 %

Table 6 - Phase 4 : Tasks and Milestones

3.2. Test and Validate the Solution before Full Deployment

To ensure a safe full launch, I would follow a 3-step testing and validation approach:

1. Internal Validation (Pre-Pilot)

This takes place during Phase 1 of the Implementation Roadmap. The goal is to confirm that all technical components are working as expected.

Activities:

- Test the performance of the triage classifier using historical ticket data
- Run end-to-end RAG + LLM flows with example queries to check if relevant content is retrieved from the knowledge base
- Confirm that fallback, escalation and logging mechanisms work properly

Validation criteria:

- Classifier achieves > 80 % accuracy on test data
- RAG fetches relevant content in > 80 % of general info + MKT out-out queries
- All high-risk rickets are escalated (not answered by AI)

2. Pilot with Real Users

This step happens during Phase 2**.** The AI is introduced in a live environment but limited to low-risk tickets (e.g. general info and opt-outs) and Chat and SMS channels**.** Human agents supervise all replies.

Activities:

- AI suggests responses, but agents approve or reject them before sending
- Collect feedback: thumbs, CSAT, agent corrections and escalation data.
- Monitor fallback rate and identify knowledge base gaps.

Success criteria:

- AI handles 50% of ticket volume in scope
- CSAT on AI-assisted replies ≥ 80%
- Escalation rate <30%

3. AI Gradual Takeover

This takes place during Phase 3. The AI begins replying directly to users, but only when confidence is high. Fallback and escalation are still active for unclear cases.

Activities:

- Set confidence thresholds for both the LLM and classifiers.
- Use dashboards to monitor CSAT, fallback rate, escalation rate, and AI resolution rate
- Weekly reviews to fine-tune thresholds, KB content and fallback logic

Success criteria:

- CSAT ≥ 85%
- AI resolution rate ≥ 70%
- Fallback rate ≤ 15%
- Escalation rate ≤ 20%
- Metrics are stable for at least 2 consecutive weeks.

Once these criteria are met, the system is considered stable and ready for full deployment across all safe tickets and supported channels in Phase 4.

3.3. Success Measurement

Success is measured across three dimensions: User Satisfaction, AI Performance and Operational Efficiency. Table 7 shows the main metrics for each dimension and the recommended target values.

Category	Metric	Target
User Satisfaction	CSAT on AI-resolved tickets	> 85 %
AI Performance	AI Resolution Rate	> 70 % of tickets handled by AI
	Fallback Rate	< 15 %
	Escalation Rate	< 20 %
	Classifier accuracy	> 80 % on historical data
Operational Efficiency	Reduction in agent workload	> 50 % of tickets passed to AI
	Response time improvement	> 30 % faster

Table 7 - Key Success Metrics

Part 2: Prompt Engineering

1. System prompt for na AI agent handling “general information about XYZ Company”

“ You are an **AI customer assistant for XYZ Company**.
You handle **only general questions** about the company. Answer the questions in a polite, friendly, concise and professional way.

Use only information from the internal knowledfe base, such as FAQs, product description and company policies.
Do not guess or use external information.

If the user asks something that is out of scope (e.g.,personal account issueas, cancellations, complaints or billing issues):
- Inform them that this question requires a human agent
- Trigger the fallback response and escalate

Use simple language, short sentences and helpful attitude.

If you are not sure of the answer, **do not guess**. Trigger a fallback message.”

Note: *This was written using a format similar to a Python script.*

2. System prompt for handling “issues signing up”

“You are an AI assistant for XYZ Company.
Your job is to help users who are having trouble signing up for the service.

How to behave:

1. Help user troubleshoot **common, low-risk sign-up issues**
Examples: “I didn´t get the confirmation email” or “Where do I enter my code?”

2. Do not handle:

- Account-specific problems (e.g. password reset, email mismatch)
- Technical errors you cannot confirm (e.g., “I clicked sign up and nothing happened”)
- Legal, payment or identity check questions

If the issue seems unclear, start by asking a **brief clarifying question** (one sentence).
If the user is still stucked or frustrated, triugger fallback and escalate to a human.

Information source:

- Use only the knowledge base
- Do not guess or suggest unsupported actions.

Tone and style:

- Helpful, friendly and concise

Fallback message template:

“It lloks like I can’t fix this directly. I’m forwarding your request to a support agent right away””

Note: *This was written using a format similar to a Python script.*