

AI CUSTOMER SUPPORT PLATFORM

Functional Blueprint

Version 1.0
January 2025

CONFIDENTIAL

Table of Contents

1. Executive Summary
2. Product Vision & Objectives
3. System Architecture Overview
4. Module Specifications
5. Data Architecture
6. Integration Points
7. User Roles & Permissions
8. Workflow Diagrams

- 9. Non-Functional Requirements
- 10. Implementation Roadmap
- 11. Success Metrics
- 12. Appendices

1. Executive Summary

1.1 Product Overview

The AI Customer Support Platform is an enterprise-grade solution designed to revolutionize customer service operations through intelligent automation. By combining advanced natural language processing with seamless human escalation capabilities, the platform delivers 24/7 customer support while significantly reducing operational costs.

1.2 Key Value Propositions

- 70%+ automation rate for routine customer inquiries
- Sub-30 second average response time
- 24/7 availability without staffing constraints
- Seamless AI-to-human escalation preserving conversation context
- Real-time analytics and actionable insights
- Enterprise-grade security and compliance

1.3 Target Metrics

Metric	Target	Measurement
AI Resolution Rate	>70%	Conversations resolved without human intervention
First Response Time	<30 seconds	Time from customer message to AI response
Customer Satisfaction	>85% CSAT	Post-conversation survey ratings
System Uptime	>99.5%	Platform availability
Escalation Efficiency	<2 min	Time to connect with human agent

2. Product Vision & Objectives

2.1 Vision Statement

To create the most intuitive and effective AI-powered customer support platform that empowers businesses to deliver exceptional customer experiences at scale, while freeing human agents to focus on complex, high-value interactions.

2.2 Strategic Objectives

2.2.1 Customer Experience

- Provide instant, accurate responses to common inquiries
- Maintain conversation context across channels and sessions
- Enable natural language interactions without rigid menu structures
- Ensure smooth transitions between AI and human support

2.2.2 Operational Excellence

- Reduce support ticket volume by automating routine queries
- Decrease average handle time for human agents
- Provide real-time visibility into support operations
- Enable data-driven optimization of support processes

2.2.3 Business Impact

- Lower cost per customer interaction by 40-60%
- Increase customer retention through improved satisfaction
- Scale support capacity without proportional headcount growth
- Generate insights for product and service improvements

3. System Architecture Overview

3.1 High-Level Architecture

The platform follows a modular microservices architecture, enabling independent scaling, deployment, and maintenance of individual components.

Layer	Components	Technology
Presentation	Web Widget, Agent Dashboard, Admin Portal	React, TypeScript, WebSocket
API Gateway	Authentication, Rate Limiting, Routing	Node.js, Express, JWT
Business Logic	AI Engine, Conversation Manager, Routing	Node.js, Python
AI/ML	Intent Recognition, Entity Extraction, NLG	OpenAI/Claude API, Custom Models

Data	Conversations, Knowledge Base, Analytics	PostgreSQL, Redis, Elasticsearch
Infrastructure	Container Orchestration, Monitoring	Docker, Kubernetes, Prometheus

3.2 Technology Stack

3.2.1 Frontend

- React 18+ with TypeScript for type safety
- Tailwind CSS for responsive styling
- WebSocket for real-time communication
- Progressive Web App (PWA) capabilities

3.2.2 Backend

- Node.js with Express for API services
- Python for AI/ML processing pipelines
- PostgreSQL for relational data storage
- Redis for caching and session management
- Elasticsearch for full-text search

3.2.3 AI/ML

- OpenAI GPT-4 or Anthropic Claude for language understanding
- Custom intent classification models
- Entity extraction with named entity recognition
- Sentiment analysis for escalation triggers

4. Module Specifications

4.1 Module Overview

Module	Description	Priority
Module 1: Web Chat Widget	Embeddable customer-facing chat interface	Must Have
Module 2: AI Engine	Intent recognition, response generation, NLU	Must Have
Module 3: Human Handoff	Escalation triggers, queue management, routing	Must Have
Module 4: Agent Interface	Agent workspace for handling conversations	Must Have
Module 5: Admin Dashboard	Analytics, reporting, system management	Must Have

Module 6: Knowledge Base	Article management, search, AI training	Must Have
Module 7: AI Configuration	Intent management, response templates	Should Have
Module 8: User Management	Roles, permissions, team structure	Should Have
Module 9: Notifications	Alerts, emails, push notifications	Should Have

4.2 Module 1: Web Chat Widget

The web chat widget is the primary customer-facing interface, providing an embeddable chat experience that integrates seamlessly with any website.

4.2.1 Core Features

- Lightweight, asynchronous loading (<50KB initial payload)
- Customizable appearance (colors, position, branding)
- Responsive design for desktop and mobile
- Real-time messaging via WebSocket
- Typing indicators and read receipts
- Quick reply buttons for common actions
- File attachment support
- Conversation history persistence

4.2.2 Technical Specifications

Specification	Requirement
Initial Load Time	<1 second
Bundle Size	<50KB gzipped
Browser Support	Chrome 70+, Firefox 65+, Safari 12+, Edge 79+
Mobile Support	iOS Safari, Android Chrome
Accessibility	WCAG 2.1 AA compliant
Security	HTTPS only, XSS prevention, CSP compatible

4.2.3 Integration Options

- JavaScript embed code (single line)
- NPM package for framework integration
- REST API for custom implementations
- JavaScript SDK with 9 public methods

4.3 Module 2: AI Engine

The AI Engine is the intelligent core of the platform, responsible for understanding customer intent, extracting relevant information, and generating appropriate responses.

4.3.1 Core Capabilities

- Intent recognition with confidence scoring
- Entity extraction (dates, numbers, emails, custom entities)
- Context-aware response generation
- Knowledge base integration
- Multi-turn conversation handling
- Sentiment analysis for escalation

4.3.2 Default Intents

Intent	Description	Example Phrases
greeting	Welcome messages	Hi, Hello, Good morning
order_status	Order tracking	Where is my order?, Track package
refund_request	Return/refund inquiries	I want a refund, Return policy
shipping_info	Delivery questions	Shipping cost, Delivery time
account_help	Account management	Reset password, Update email
pricing	Cost inquiries	How much does it cost?, Pricing
contact_human	Escalation request	Talk to human, Real person
goodbye	Conversation end	Thanks, Bye, That's all
fallback	Unknown queries	Catch-all for unrecognized input

4.3.3 Confidence Thresholds

Level	Confidence	Action
High	≥80%	Respond directly with matched intent
Medium	70-79%	Respond with clarification option
Low	50-69%	Offer options or escalate after retry
Very Low	<50%	Immediate escalation to human

4.4 Module 3: Human Handoff

The Human Handoff module manages the seamless transition from AI to human agents when automation cannot fully resolve customer needs.

4.4.1 Escalation Triggers

Trigger Type	Condition	Priority
Low Confidence	AI confidence <50% on initial response	High
Repeated Low Confidence	Confidence 50-69% twice in a row	High
User Request	Customer asks for human agent	Critical
Consecutive Fallbacks	3 fallback responses in a row	High
Negative Sentiment	Frustration or anger detected	Critical
Complex Query	Multiple intents detected	Medium
Sensitive Topic	Billing disputes, complaints	High

4.4.2 Queue Management

- FIFO queue with priority override
- Real-time wait time estimation
- SLA monitoring and alerts
- Automatic load balancing across agents
- Business hours awareness
- Offline message handling

4.4.3 Context Preservation

- Full conversation history transfer
- AI-generated conversation summary
- Detected intent and entities
- Customer information and history
- Suggested responses for agent

4.5 Module 4: Agent Interface

The Agent Interface provides support agents with a powerful workspace to handle customer conversations efficiently.

4.5.1 Core Features

- Unified inbox for all assigned conversations

- Real-time message updates via WebSocket
- Customer information panel with history
- AI context and suggested responses
- Canned response library
- Internal notes (agent-only)
- Conversation transfer capabilities
- Multi-conversation handling (up to 5 concurrent)

4.5.2 Agent Actions

Action	Description
Reply	Send message to customer
Add Note	Add internal note (not visible to customer)
Transfer	Transfer to another agent or team
Resolve	Mark conversation as resolved with summary
Tag	Add tags for categorization
Snooze	Temporarily hide conversation

4.5.3 Agent Status

Status	Description	Receives Work
Online	Available for new conversations	Yes
Busy	At maximum capacity	No
Away	Temporarily unavailable	No
Offline	Not working	No

4.6 Module 5: Admin Dashboard

The Admin Dashboard provides comprehensive visibility into support operations with real-time analytics and management capabilities.

4.6.1 Dashboard Metrics

Metric	Description	Update Frequency
Active Conversations	Currently open conversations	Real-time
Resolved Today	Conversations closed today	Real-time

AI Automation Rate	% handled by AI without human	5 minutes
Avg Response Time	Time to first response	5 minutes
Queue Length	Customers waiting for agent	Real-time
CSAT Score	Customer satisfaction rating	Daily
Agent Utilization	Active time / Available time	Hourly

4.6.2 Reporting Features

- Customizable date range filtering (Today, 7 days, 30 days, Custom)
- Conversation volume trends
- Status distribution analysis
- Channel performance comparison
- Agent performance leaderboard
- AI intent accuracy reports
- Export capabilities (CSV, PDF, Excel)

4.6.3 Management Functions

- View and manage all conversations
- Assign escalated conversations to agents
- Monitor agent status in real-time
- Configure system settings
- Access audit logs

4.7 Module 6: Knowledge Base

The Knowledge Base serves as the central repository for information that powers AI responses and agent assistance.

4.7.1 Article Management

- Rich text editor with formatting options
- Category and tag organization
- Version history and rollback
- Draft and published states
- Search engine optimization
- Article performance analytics

4.7.2 AI Integration

- Automatic article indexing for AI retrieval

- Semantic search for relevant content
- Article suggestion during conversations
- Training data extraction

4.7.3 Article Structure

Field	Type	Required
Title	Text (max 200 chars)	Yes
Content	Rich text	Yes
Category	Select from list	Yes
Tags	Multi-select	No
Status	Draft/Published	Yes
Visibility	Public/Internal	Yes

4.8 Module 7: AI Configuration

AI Configuration enables administrators to customize the AI behavior without requiring technical expertise.

4.8.1 Intent Management

- Create, edit, and delete custom intents
- Add training phrases for each intent
- Configure response templates with variables
- Set confidence thresholds per intent
- Test intents with the built-in console

4.8.2 Entity Management

- Configure custom entity types
- Define extraction patterns
- Set validation rules

4.9 Module 8: User Management

- Create and manage user accounts
- Assign roles (Admin, Supervisor, Agent)
- Organize users into teams
- Configure permissions per role
- Track user activity and audit logs

4.10 Module 9: Notifications

- In-app notifications for agents and admins
- Browser push notifications
- Email notifications for important events
- Configurable notification preferences
- Queue alerts and SLA warnings

5. Data Architecture

5.1 Core Entities

Entity	Description	Key Fields
User	System users (agents, admins)	id, email, role, status, team_id
Customer	End customers	id, email, name, external_id, metadata
Conversation	Support conversations	id, customer_id, status, channel, assigned_to
Message	Individual messages	id, conversation_id, sender_type, content, timestamp
Article	Knowledge base articles	id, title, content, category_id, status
Intent	AI intents	id, name, description, is_active
TrainingPhrase	Intent training data	id, intent_id, phrase
IntentResponse	Response templates	id, intent_id, response_text, variables

5.2 Entity Relationships

- User → Team (many-to-one)
- Conversation → Customer (many-to-one)
- Conversation → User/assigned_to (many-to-one)
- Message → Conversation (many-to-one)
- Article → Category (many-to-one)
- TrainingPhrase → Intent (many-to-one)
- IntentResponse → Intent (many-to-one)

5.3 Data Retention

Data Type	Retention Period	Notes
-----------	------------------	-------

Conversation Messages	2 years	Configurable per compliance requirements
Audit Logs	5 years	Immutable, no deletion
Analytics Data	1 year (detailed)	Aggregated data kept longer
Session Data	30 days	Automatic cleanup
Knowledge Base	Indefinite	Version history maintained

6. Integration Points

6.1 External Integrations

Integration	Purpose	Protocol
OpenAI/Anthropic	LLM for NLU and response generation	REST API
SMTP Server	Email notifications	SMTP
OAuth Providers	SSO authentication	OAuth 2.0
Webhook Endpoints	Event notifications to external systems	HTTP POST
CRM Systems	Customer data sync (future)	REST API

6.2 API Endpoints

The platform exposes a comprehensive REST API for custom integrations.

6.2.1 Public API Categories

- Authentication: `/api/auth/*` (login, logout, refresh)
- Conversations: `/api/conversations/*` (CRUD, messages)
- Customers: `/api/customers/*` (CRUD, lookup)
- Knowledge Base: `/api/articles/*` (CRUD, search)
- Analytics: `/api/stats/*` (metrics, reports)
- Webhooks: `/api/webhooks/*` (subscribe, manage)

6.2.2 WebSocket Events

Event	Direction	Description
message:new	Bidirectional	New message in conversation
conversation:assigned	Server → Client	Conversation assigned to agent

conversation:resolved	Server → Client	Conversation resolved
typing:start	Bidirectional	User started typing
typing:stop	Bidirectional	User stopped typing
agent:status	Server → Client	Agent status change

7. User Roles & Permissions

7.1 Role Definitions

Role	Description	Typical Users
Administrator	Full system access, configuration	IT Admin, Support Manager
Supervisor	Team oversight, reporting	Team Lead, QA Manager
Agent	Handle customer conversations	Support Agent, CSR

7.2 Permission Matrix

Feature	Admin	Supervisor	Agent
Dashboard - Full	✓	✓	Personal only
Conversations - All	✓	✓	Assigned only
Knowledge Base - Edit	✓	✓	View only
AI Configuration	✓	View only	No access
User Management	✓	View team	No access
System Settings	✓	Limited	Personal only
Audit Logs	✓	Team only	No access

8. Workflow Diagrams

8.1 Customer Message Flow

1. Customer sends message via widget
2. Message received by API Gateway

3. If new conversation → Create conversation record
4. Route message to AI Engine
5. AI Engine processes: Intent → Entities → Confidence
6. If confidence $\geq 70\%$ → Generate response from template
7. If confidence $< 70\%$ or escalation trigger → Route to Human Handoff
8. Response sent back to customer via WebSocket

8.2 Human Escalation Flow

1. Escalation triggered (low confidence, user request, etc.)
2. AI sends pre-escalation message to customer
3. Conversation added to escalation queue with priority
4. Available agents notified via all channels
5. Agent claims or is assigned the conversation
6. AI context summary provided to agent
7. Agent handles conversation with AI assistance
8. Agent resolves and closes conversation
9. CSAT survey sent to customer

8.3 Conversation Resolution Flow

1. Agent clicks 'Resolve' button
2. Agent selects resolution type (Resolved/Closed/Spam)
3. Agent adds optional resolution note
4. System sends closing message to customer
5. CSAT survey triggered (if enabled)
6. Conversation status updated to 'Resolved'
7. Metrics updated (handle time, resolution count)
8. Conversation archived for reporting

9. Non-Functional Requirements

9.1 Performance Requirements

Metric	Requirement	Measurement Method
API Response Time	<200ms (95th percentile)	APM monitoring
Widget Load Time	<1 second	Lighthouse score
AI Processing Time	<2 seconds	Request logging
WebSocket Latency	<100ms	Ping monitoring
Dashboard Load	<3 seconds	Browser metrics
Search Results	<1 second	Query logging
Concurrent Users	10,000+ simultaneous	Load testing

9.2 Availability & Reliability

- System Uptime: 99.5% SLA
- Disaster Recovery: RPO 1 hour, RTO 4 hours
- Automated failover for critical services
- Geographic redundancy for database
- Graceful degradation during AI service outages

9.3 Security Requirements

- All data encrypted in transit (TLS 1.3)
- All data encrypted at rest (AES-256)
- JWT-based authentication with refresh tokens
- Role-based access control (RBAC)
- SQL injection and XSS prevention
- Rate limiting on all endpoints
- Audit logging for all administrative actions
- GDPR and CCPA compliance

9.4 Scalability

- Horizontal scaling via container orchestration
- Auto-scaling based on load metrics
- Database read replicas for query distribution
- CDN for static asset delivery
- Message queue for async processing

10. Implementation Roadmap

10.1 MVP Phase (Weeks 1-12)

Phase	Weeks	Deliverables
Foundation	1-3	Architecture setup, database schema, authentication
Core Modules	4-7	Chat widget, AI engine basics, human handoff
Agent Tools	8-10	Agent interface, basic dashboard
Integration	11-12	Testing, bug fixes, deployment

10.2 MVP Scope

Included

- Web chat channel only
- Basic AI with 10-15 intents
- Human escalation and routing
- Agent conversation handling
- Simple analytics dashboard
- Basic knowledge base

Excluded from MVP

- Multi-channel support (email, WhatsApp, SMS)
- Advanced analytics and reporting
- Workflow automation
- Custom integrations
- Mobile applications

10.3 Post-MVP Phases

Phase	Timeline	Features
Phase 2	Months 4-6	Email channel, advanced analytics, API v2
Phase 3	Months 7-9	WhatsApp/SMS, workflow automation, integrations
Phase 4	Months 10-12	Mobile apps, advanced AI, enterprise features

11. Success Metrics

11.1 Key Performance Indicators

KPI	Target	Measurement
AI Resolution Rate	>50% (MVP), >70% (Year 1)	Conversations resolved by AI / Total conversations
First Response Time	<30 seconds	Avg time from customer message to response
Customer Satisfaction	>4.0/5 (MVP), >4.5/5 (Year 1)	Post-conversation CSAT surveys
System Uptime	>99.5%	Monitoring and incident tracking
Agent Utilization	>75%	Active time / Available time
Escalation Rate	<30%	Escalated / Total conversations
Average Handle Time	<5 minutes	Time from escalation to resolution

11.2 Business Impact Metrics

- Cost per conversation reduction: Target 40-60%
- Support capacity increase: Target 3x without headcount
- Customer retention improvement: Target +5%
- Agent satisfaction score: Target >4.0/5

11.3 Monitoring & Reporting

- Real-time dashboard for operational metrics
- Daily automated reports via email
- Weekly trend analysis
- Monthly business review package
- Quarterly strategic assessment

12. Appendices

12.1 Glossary

Term	Definition
CSAT	Customer Satisfaction Score (typically 1-5 scale)
Intent	The purpose or goal behind a customer's message
Entity	Specific data extracted from a message (dates, numbers, etc.)

Escalation	Transfer of conversation from AI to human agent
NLU	Natural Language Understanding
NLG	Natural Language Generation
SLA	Service Level Agreement
FIFO	First In, First Out (queue ordering)

12.2 Reference Documents

- MVP Requirements Document (MVP_Requirements.md)
- Developer Technical Specifications (MVP_Developer_Requirements.md)
- Functional Requirements Document (MVP_Functional_Requirements.md)
- Module 1: Web Chat Widget Requirements
- Module 2: AI Engine Requirements
- Module 3: Human Handoff Requirements
- Module 4: Agent Interface Requirements
- Module 5: Admin Dashboard Requirements
- Interactive Screen Mockups (HTML files)

12.3 Document Control

Version	Date	Author	Changes
1.0	January 2025	Product Team	Initial release

— End of Document —