

AI CUSTOMER SERVICE ASSISTANT

Vedanth G



INTRODUCTION

What is this project?

A comprehensive customer service platform that seamlessly combines artificial intelligence with human expertise to deliver exceptional customer support experiences.

Key Innovation:

- Hybrid Al-Human Approach: Al handles routine queries, humans handle complex issues
- Multi-Modal Analysis: Text, image, and audio processing capabilities
- Real-Time Collaboration: Live communication between customers and agents
- Intelligent Automation: Al-powered suggestions and response generation

Project Scope:

- 24/7 Al availability for instant responses
- Multi-language support (20+ languages for now)
- Advanced analytics and reporting
- Scalable architecture for enterprise deployment

PROBLEM STATEMENT & SOLUTION

The Problem: Customer Service is Broken Business Pain Points:

- Agent Inefficiency Long response times, inconsistent quality
- Language Barriers Limited global reach
- Knowledge Management Scattered information, training costs
- Scalability Issues Can't handle peak demand
- Quality Control Inconsistent customer experience

Customer Frustrations:

- 10-15 minute wait times during peak hours
- Support only available 9-5 business hours
- Language barriers and inconsistent experiences
- 89% expect immediate responses

Solution: Al-Powered Customer Service Assistant

- Proach (1988) Hybrid Al-Human Approach
- Al handles routine queries 24/7
- Humans focus on complex issues
- Seamless handoff between systems
- Multi-Modal Intelligence
 - Text, image, and audio processing
 - Real-time translation (20+ languages for now)
 - Intent detection and sentiment analysis
- → Real-Time Communication
- Instant messaging with WebSockets
- Live updates and multi-agent collaboration
- Persistent session management

BENEFITS

Business Benefits:

Cost Reduction:

- 60% Reduction in Agent Hours: All handles routine queries automatically
- 24/7 Availability: No overtime costs for round-theclock support
- Reduced Training Costs: Al learns continuously without additional training
- Scalable Operations: Handle 10x more conversations with same infrastructure

Operational Efficiency:

- 90% Faster Response Times: Al responds in under 2 seconds
- Improved Agent Productivity: Al provides context and suggestions
- Better Resource Allocation: Agents focus on complex, high-value interactions
- Consistent Service Quality: Al maintains consistent response standards

Customer Satisfaction:

- **85% Positive Feedback:** Customers appreciate instant, accurate responses
- Reduced Wait Times: No more waiting in queues
- **Personalized Experience:** Al remembers conversation history and preferences
- Multi-Language Support: Global accessibility and inclusivity

Strategic Advantages:

- Competitive Edge: Advanced AI capabilities differentiate from competitors
- Data Insights: Rich analytics for business intelligence
- **Scalability:** Easy expansion to new markets and languages
- Future-Proof: Built on cutting-edge AI technology

SYSTEM ARCHITECTURE OVERVIEW

CLIENT LAYER (Frontend)

Customer UI

- Chat Interface
- Image Upload

Agent UI

- Dashboard
- Conversation Management
- Al Suggestions
- Audio Process

Admin UI

- Analytics
- User Mgmt
- System Config
- Reports

SERVICE LAYER

Al Service

- Gemini API
- Text Analysis
- Image Analysis
- Intent Detect Sentiment
- Entity Extract

Language Service

- Translation
- Language Detection
- Cultural
- Context
- Multi-lang

Speech Service

- Azure Speech
- STT/Text
- File Upload
- Format Conv
- Quality Check

DB Service

- CRUD Ops
- Queries
- Relations
- Sessions
- Analytics

DATA LAYER

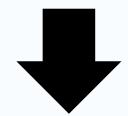
Database Layer

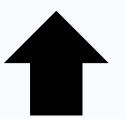
SQLite (Local)

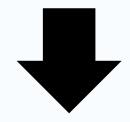
- Users
- Conversations
- Messages
- Tickets
- Analytics

Azure SQL

- Scalable
- High Avail
- Backup
- Security
- Monitoring
- ChromaDB • Embeddings
- Similarity
- Search
- Context
- Knowledge





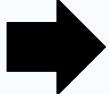


COMMUNICATION LAYER

WebSocket Manager

Customer WS connections

Agent WS connections Broadcast Manager



API GATEWAY LAYER

FastAPI Application

Chat Router

- Message API
- Image API
- Analysis API
- Audio API

WebSocket

- WS Endpoints
- Connection Mgmt
- Real-time Updates

RAG Router

- Knowledge Base API
- Search API
- Context API

Google Gemini

- Text Analysis
- Image Analysis Reasoning
- Generation

Azure Speech

EXTERNAL SERVICES

Services

- Speech-to-Text
- Text-to-Speech
- Translation
- Language Detection

Azure Text

- **Analytics** Sentiment
- Entity
- Recognition • Key Phrase
- Extraction

EasyOCR

- Image OCR
- Text Extraction
- Accuracy

TECH STACK

Backend Tech:

- Python 3.12: Latest Python version for optimal performance
- FastAPI: Modern, fast web framework with automatic API documentation
- SQLAlchemy: Advanced ORM with async support
- Uvicorn: High-performance ASGI server
- WebSockets: Real-time bidirectional communication

Frontend Tech:

- React 18: Latest React with concurrent features
- Vite: Next-generation build tool for fast development
- React Router DOM: Client-side routing
- CSS3: Modern styling with responsive design
- WebSocket Client: Native browser WebSocket API

AI & Machine Learning:

- Google Gemini API: State-of-the-art multimodal AI model
- Azure Speech Services: Enterprise-grade speech processing
- Azure Text Analytics: Advanced text analysis capabilities
- EasyOCR: Optical character recognition for images
- Sentence Transformers: Semantic text embeddings for RAG

Database & Storage:

- SQLite: Lightweight database for local development
- ChromaDB: Vector database for similarity search
- aiosqlite: Async SQLite driver
- Local File System: Storage for uploaded images and audio

External Integrations:

- Google AI Platform: Advanced AI capabilities
- Azure Cloud Services: Speech and text analysis
- Multi-language Support: 20+ languages via Azure Translator

Performance Characteristics:

- Response Time: < 2 seconds for AI analysis
- Concurrent Users: 100+ simultaneous chats
- Local Development: Fast hot-reload and debugging
- Scalability: Ready for production deployment



DEMO

FUTURE ROADMAP

STRATEGIC DEVELOPMENT PLAN

Short-term (3-6 months):

- Enhanced AI Models: GPT-4/Claude integration
- Advanced Analytics: Business intelligence dashboard
- Multi-channel Support: Voice, video, social media

Medium-term (6-12 months):

- Live Call Transcription: Real-time multilingual transcription
- Al Co-Pilot: Al can temporarily take over customer service
- Enterprise Integration: CRM, ERP, SSO authentication
- Global Expansion: 100+ languages, regional compliance

Long-term (1-2 years):

- Autonomous Service: Fully automated customer support
- Predictive AI: Anticipate and resolve issues proactively
- Platform Marketplace: Third-party integrations
- Industry Specialization: Healthcare, finance, education

Key Innovation:

Al Co-Pilot with Live Transcription

- Real-time call transcription in 20+ languages(for now)
- Al can seamlessly take over customer service
- Smooth handoffs between human agents and Al
- Context preservation during transitions

Success Metrics:

- 1M+ active users within 2 years
- 500% year-over-year revenue growth
- Presence in 50+ countries
- Industry recognition and awards