International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 945-951
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0136

NEXA: An Intelligent Conversational Framework for Automated Customer Support System

Vasavi Pasumarthi[1], *Sahasra Chamarti*[2], *Abhilash Chinthala*[3], *G.M. Padmaja*[4]

[1,2,3]*Students, Department of Computer Science and Engineering, SRK Institute of Technology, Vijayawada, India.*

[4]*Associate Professor, Department of Computer Science and Engineering, SRK Institute of Technology, Vijayawada, India*

Email ID: *vasavipasumarthi884@gmail.com*[1], *Padmaja.gmp@gmail.com*[2]

Abstract

Customer support is a critical component of modern digital businesses, where rapid response, personalization, and scalability are essential. Traditional support systems often struggle with high operational costs, limited scalability, and inconsistent user experience. This paper presents Nexa, a modern AI-powered customer support platform developed as a multi-tenant B2B SaaS solution integrating AI chatbots, voice assistants, real-time dashboards, and subscription-based access control. Nexa leverages a monorepo architecture using Turbo Repo, Next.js, Convex, Clerk, and Vapi, enabling seamless real-time communication and AI-driven automation. The platform supports multi-modal conversations through chat and voice, dynamic knowledge base creation via embeddings, and secure per-organization API key management using AWS Secrets Manager. Experimental deployment demonstrates that Nexa significantly improves response efficiency, operational scalability, and system reliability, making it suitable for real-world enterprise customer support applications.

Keywords: *AI Customer Support, Multi-Tenant SaaS, Voice Assistant, Convex, Knowledge Base, Real-Time Systems, AWS Secrets Manager, SaaS Architecture.*

1. Introduction

With the rapid digital transformation of businesses, customer expectations regarding service quality, availability, and responsiveness have increased significantly. Traditional customer support systems rely heavily on human agents, resulting in high operational costs and limited scalability. The introduction of AI-powered conversational systems has revolutionized this domain by automating routine queries and enabling 24/7 support. However, existing AI-based platforms face challenges related to multi-tenancy, secure API management, real-time data synchronization, and flexible deployment. To address these limitations, we propose Nexa, an AI-powered customer support platform designed as a scalable, secure, and extensible multi-tenant SaaS solution. Nexa integrates AI chat, voice agents, real-time operator dashboards, subscription enforcement, and embedded widget deployment into a single unified system. By leveraging modern full-stack technologies such as Convex for reactive databases, Clerk for authentication, and Vapi for voice assistants, Nexa enables organizations to deploy customized support systems with minimal overhead. This work focuses on designing, implementing, and evaluating a production-ready AI support platform that emphasizes modularity, security, scalability, and usability. [1]

2. Ease of Use

The Nexa platform is designed with a user-centric approach, ensuring minimal learning effort and smooth interaction across all components. For customers, Nexa provides an embedded chat widget and voice interface that requires no account creation. Session-based anonymous access allows users to initiate conversations instantly, reducing friction and improving engagement. The widget interface is intuitive, responsive, and lightweight, making it suitable for integration across various websites and devices. For operators, the real-time dashboard offers a unified view of all conversations, including user context, session history, and device metadata. Features such as infinite scrolling, status-based filtering, and manual takeover enable efficient handling of customer queries with minimal effort. The seamless transition between AI-handled and human-assisted conversations further improves operator productivity. From an organizational perspective, Nexa simplifies configuration through centralized authentication, billing, and plugin management. The ability to manage API keys, subscription plans, and widget customization through a single interface enhances usability for administrators. Overall, Nexa's modular design, clean user interface, and real-time responsiveness ensure a smooth and accessible user experience for customers, operators, and organizations alike. [2]

3. Related Work

Recent advancements in artificial intelligence have led to widespread adoption of chatbots and voice assistants in customer service systems. Prior research highlights that AI-based chatbots can effectively automate routine interactions, improve response times, and influence user compliance through natural language interactions. However, studies also report limitations in trust-building and handling emotionally complex queries. Several works have explored the use of chatbots to enhance customer service efficiency by automating repetitive tasks. While these systems reduce workload and operational costs, they often struggle with contextual understanding and emotional intelligence, leading to reduced user satisfaction in complex scenarios. To address these challenges, hybrid intelligence models have been proposed, where AI systems collaborate with human agents. Such approaches enable mutual learning between humans and AI, although determining optimal control transfer remains challenging. More recent studies investigate the role of chatbots in business process management, emphasizing their potential to automate workflows such as customer support, order tracking, and service coordination. Despite these advantages, integration with complex enterprise systems and continuous maintenance requirements remain significant challenges. In contrast to existing approaches, Nexa combines AI chat, voice assistants, real-time human intervention, and secure multi-tenant SaaS architecture within a unified platform. By integrating retrieval-augmented generation, real-time dashboards, and subscription-based access control, Nexa addresses key limitations identified in previous studies, offering a scalable, secure, and user-friendly solution for modern customer support environments. [3]

4. Existing System

Although AI-powered customer service systems improve response speed and automate routine queries, they exhibit notable limitations. Many systems fail to build user trust or effectively handle complex and emotionally sensitive conversations due to limited contextual understanding and lack of empathy. Accuracy issues, including hallucinated or incorrect responses, are common in domain-specific scenarios, and integration with complex enterprise systems often requires continuous maintenance. Additionally, hybrid AI–human support models struggle with smooth control transitions, leading to inefficient handoffs and user frustration. These challenges highlight the need for more reliable, context-aware, and seamlessly integrated AI-driven customer support solutions. [4]

5. Proposed System

The proposed system, Nexa, is designed as a modern hybrid intelligence customer support platform that directly addresses the limitations of existing AI-based service systems. Nexa combines AI-driven automation with real-time human intervention to deliver accurate, scalable, and user-centric customer support. To improve conversation quality and ensure seamless handoffs, Nexa incorporates proactive escalation logic that continuously monitors user sentiment and automatically transfers conversations to human operators when frustration or explicit assistance requests are detected. During handoff, the operator dashboard provides complete conversational and contextual information, enabling informed responses without repetitive clarification. Nexa addresses accuracy and knowledge limitations through a retrieval-augmented generation (RAG) approach, grounding AI responses strictly in organization-specific documentation. Additionally, the platform supports a secure multi-tenant “Bring Your Own Key” (BYOK) model, ensuring data isolation, privacy, and ethical use of AI models across tenants. Figure 1 shows Layered Architecture

6. System Architecture and Core Components

Figure: 1- Layered Architecture of the Proposed

System The proposed system follows a layered architecture that ensures modularity, scalability, and secure integration of AI-driven services.

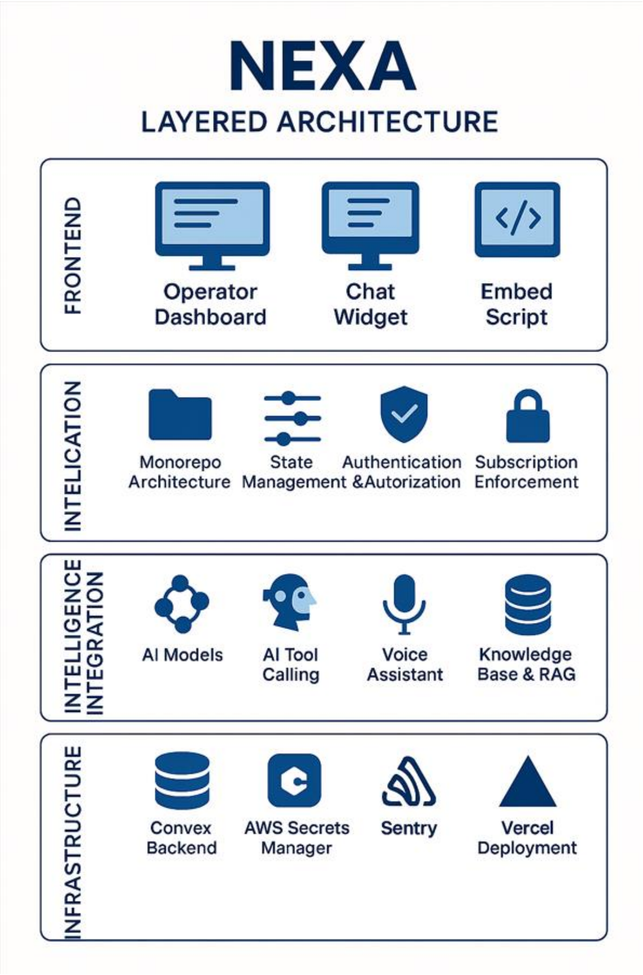
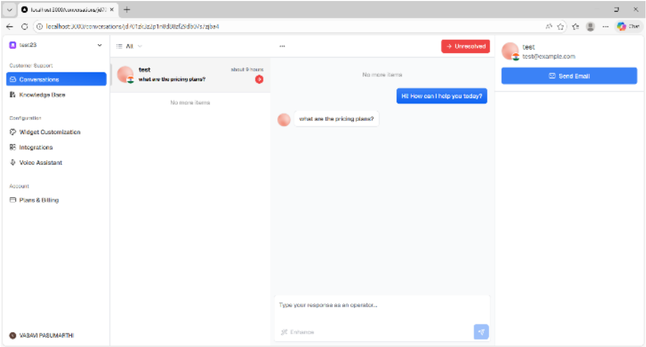


Figure 1 Layered Architecture

As illustrated in Fig. 1.1, the architecture is divided into four logical layers: Frontend, Application, Intelligence Integration, and Infrastructure. The Frontend layer provides user-facing interfaces, including the operator dashboard, chat widget, and embeddable script. These components enable seamless interaction for customers, operators, and developers, supporting both web-based chat and voice-enabled communication. The operator dashboard allows real-time monitoring and manual intervention, while the embed script ensures easy integration of the support widget into external websites. The Application layer manages core platform functionality and enforces system-wide policies. It adopts a monorepo architecture to promote code reuse and maintain consistency across services. This layer handles state management, authentication, and authorization, ensuring secure access control for multi-tenant organizations. Subscription enforcement mechanisms are also implemented at this level to restrict premium features based on active plans. The Intelligence Integration layer is responsible for AI-driven capabilities. It integrates multiple AI models, supports AI tool calling for automated workflows, and incorporates voice assistants for multi-channel customer interaction. A knowledge base with retrieval-augmented generation (RAG) ensures that AI responses are grounded in organization-specific documentation, improving accuracy and reducing hallucinations. The Infrastructure layer provides backend services and deployment support. It includes the Convex backend for real-time data synchronization, AWS Secrets Manager for secure storage of per-organization API keys, Sentry for error monitoring, and Vercel for scalable deployment. Together, these services ensure reliability, security, and production readiness of the platform. Overall, this layered architecture enables clear separation of concerns, efficient development, and scalable operation, making the proposed system suitable for enterprise-grade AI-powered customer support applications. [5]

7. System Implementation

Nexa follows a monorepo-based modular architecture built using Turbo Repo and PNPM workspace management Summary of Nexa AI. The system consists of three main applications:



test
what are the pricing plans?
about 9 hours

No more items

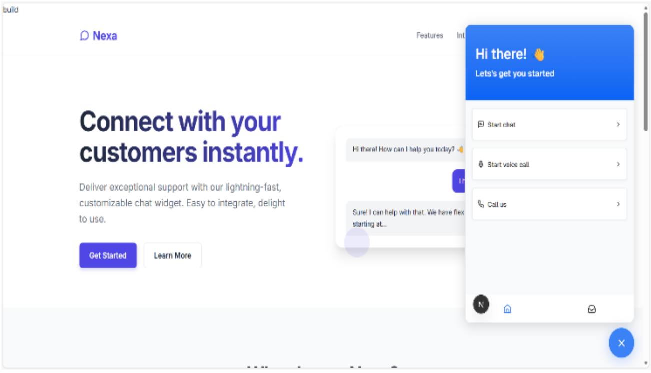
Hi! How can I help you today?

what are the pricing plans?

test
test@example.com

Type your response as an operator...

Figure 2 Operator Dashboard – for Monitoring Conversations and Managing Organizations



Connect with your customers instantly.

Deliver exceptional support with our lightning-fast, customizable chat widget. Easy to integrate, delight to use.

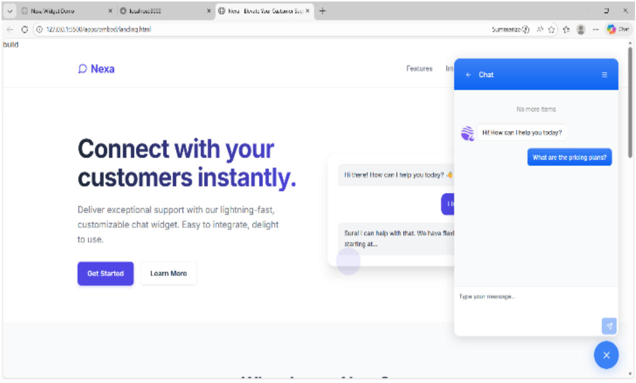
Hi there! How can I help you today?

Sure! I can help with that. We have flex starting at...

Figure 3 **Customer Widget** – Embeddable AI Chat and Voice Interface

7.1 AI Chat System

Nexa integrates AI chat using the Convex Agent component powered by OpenAI GPT-4-mini. Messages are stored and fetched using pagination with infinite scrolling support. Figure 4 shows AI Chat System [6]



Connect with your customers instantly.

Deliver exceptional support with our lightning-fast, customizable chat widget. Easy to integrate, delight to use.

Hi! How can I help you today?

What are the pricing plans?

Hi there! How can I help you today?

Sure! I can help with that. We have flex starting at...

Figure 4 **AI Chat System**

AI tool calling enables:

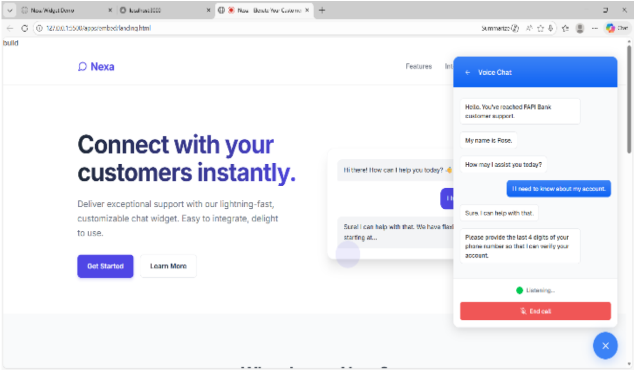
- Automatic escalation
- Conversation resolution
- Prompt enhancement [8]

7.2 Voice Assistant using Vapi

Vapi is integrated to build AI voice agents for customer calls. A sample banking assistant (“Rose”) supports:

- Balance lookup
- Transaction history
- Account verification

Each tenant can bring its own Vapi API keys, enabling white-labeled voice assistants with independent phone numbers and assistants



Connect with your customers instantly.

Deliver exceptional support with our lightning-fast, customizable chat widget. Easy to integrate, delight to use.

Hello. You've reached FAPI Bank customer support.

My name is Rose.

How may I assist you today?

I need to know about my account.

Sure, I can help with that.

Please provide the last 4 digits of your phone number so that I can verify your account.

Hi there! How can I help you today?

Sure! I can help with that. We have flex starting at...

Figure 5 **Voice Assistant**

7.3 Knowledge Base and RAG System

Nexa supports dynamic knowledge base creation through document embeddings using Convex’s Retrieval-Augmented Generation (RAG) component.

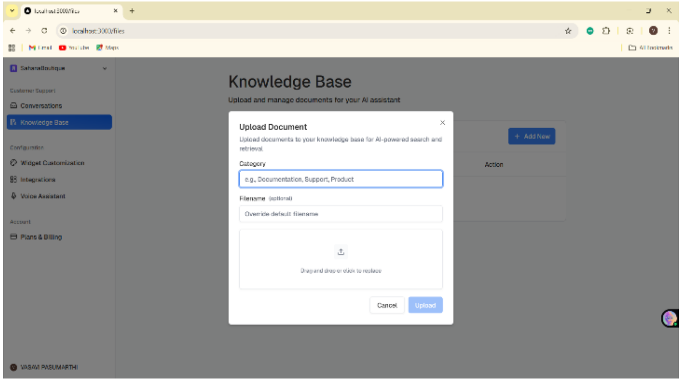
File Management [7]

Supported file types include:

- PDF, HTML, text
- JPEG, PNG, WEBP, GIF

Features include:

- Duplicate detection via content hash
- Metadata tracking
- Upload, list, delete APIs
- Secure tenant namespaces



Knowledge Base

Upload and manage documents for your AI assistant

Upload Document

Upload documents to your knowledge base for AI-powered search and retrieval

Category: e.g., Documentation, Support, Product

Filename (optional): Override default filename

Drag and drop or click to replace

Figure 6 **Knowledge Base RAG System**

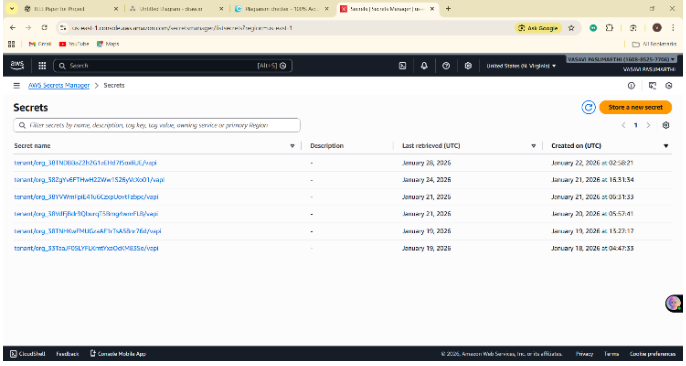
7.4 Secure API Key Management

To support white-labeled AI and voice services, Nexa integrates AWS Secrets Manager for per-organization API key storage.

Key Functionalities Include:

- Scoped IAM permissions
- Secret creation, update, retrieval
- Plugin-based UI (Vappy plugin)

- Secure connection and disconnection flows Figure 7 shows API Key Management



AWS Secrets Manager > Secrets
Store a new secret
Filter secrets by name, description, tag key, tag value, owning service or primary Region

Secrets

Secret name	Description	Last retrieved (UTC)	Created on (UTC)
tenant/org_38TND8Ba22h2G1aEHd7I5odiUE/vapi	-	January 28, 2026	January 22, 2026 at 02:58:21
tenant/org_38ZgYv6FTHwH22Ww1S26yVcXo01/vapi	-	January 24, 2026	January 21, 2026 at 16:31:34
tenant/org_38YVWmFpiE41u6CzxpUovtFzbpc/vapi	-	January 21, 2026	January 21, 2026 at 06:31:33
tenant/org_38VdFjBdr9QburqT58mg4wmFLB/vapi	-	January 21, 2026	January 20, 2026 at 05:57:41
tenant/org_38TNHKwFMUGzaAF1rTsA58rr76d/vapi	-	January 19, 2026	January 19, 2026 at 13:27:17
tenant/org_33TeaJF0SLYFLKmtYxaOdKM835a/vapi	-	January 19, 2026	January 18, 2026 at 04:47:33

Figure 7 Developer Toolkit – for Integration and Customization

This ensures isolation of sensitive credentials and prevents cross-tenant data leakage. [9]

7.5 Backend and Data Synchronization

Nexa uses Convex, a reactive database platform, which enables real-time synchronization between frontend and backend without manual WebSocket handling. Convex supports queries, mutations, and actions, simplifying state management while ensuring consistency across clients.

7.6 Authentication and Multi-Tenancy

Clerk is used for authentication and authorization with JWT templates configured for Convex integration. Each organization is treated as a tenant, with scoped data access enforced using organization IDs embedded in JWT claims. Organizations are enabled as premium features, allowing subscription-based access control.

7.7 Real-Time Dashboard

Operators are provided with a real-time dashboard featuring: User session context

- Message escalation controls
- Infinite scrolling conversation inbox
- Status filters (unresolved, escalated, resolved)

This enables human agents to seamlessly intervene when AI confidence is low.

8. User Interface and Widget Embedding

8.1 UI Framework

Nexa uses Shadcn UI with Tailwind CSS v4 for consistent, accessible UI design. The dashboard includes:

- Organization switcher
- Sidebar navigation
- Gradient-based active states

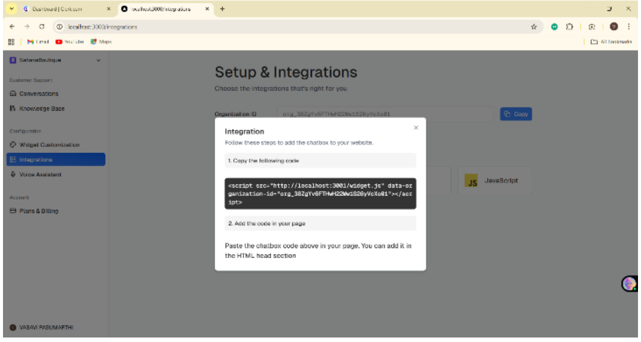
8.2 Widget Session Management

Anonymous users authenticate through session-based access valid for 24 hours. Metadata such as browser, location, and device info are captured and displayed in the operator dashboard.

8.3 Embed Script

A standalone Vite-based app generates an embeddable widget script supporting:

- HTML
- React
- Next.js
- JavaScript



Setup & Integrations

Choose the integrations that's right for you

Integration

Follow these steps to add the chatbox to your website.

1. Copy the following code

```
<script src="http://localhost:3001/widget.js" data-organization-id="org_38ZgYv6FTHwH22Ww1S26yVcXo01"></script>
```

2. Add the code in your page

Paste the chatbox code above in your page. You can add it in the HTML head section

Figure 8 Setup and Integrations

The script loads the widget inside an iframe with configurable position and exposes a global API for control. [10]

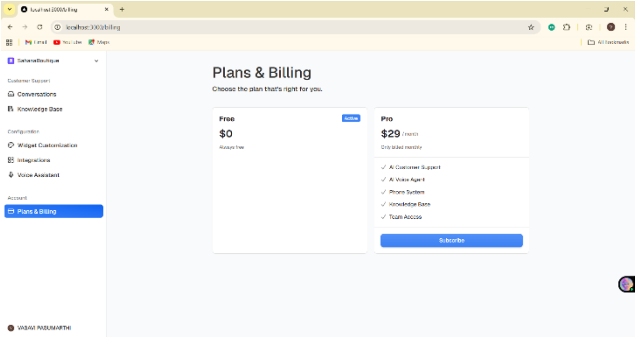
8.4 Subscription Enforcement and Billing

Nexa integrates Clerk Billing to support Free and Pro subscription plans.

- Real-time validation of subscription status at both the UI and backend levels prevents unauthorized access to premium features, improving system security.
- The flexible subscription model enables organizations to scale their usage seamlessly as their customer support demands grow, supporting long-term SaaS sustainability.

Table 1 Plan and Features

Plan	Features
Free	Single user, limited AI and voice features
Pro	Multiple users, full AI & voice support



Plans & Billing

Choose the plan that's right for you.

Figure 9 Plans and Billing

Backend APIs enforce subscription status, protecting:

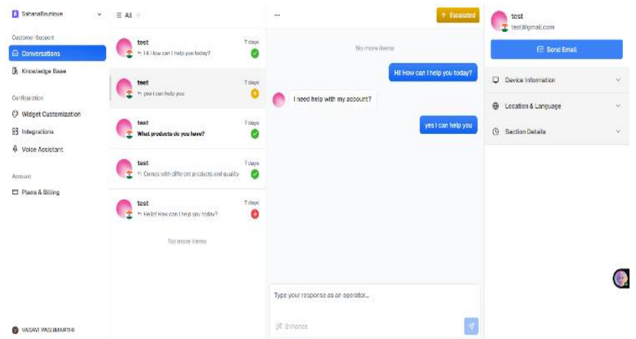
- AI generation
- File uploads
- Response enhancement

Webhook events update subscription status and organization member limits dynamically.

9. Experimental Results and Observations

Nexa was deployed on Vercel with independent deployments for:

- Dashboard
- Widget
- Embed script



test
Hi! How can I help you today?

test
yes I can help you

test
What products do you have?

test
Comes with different products and quality

test
Hello! How can I help you today?

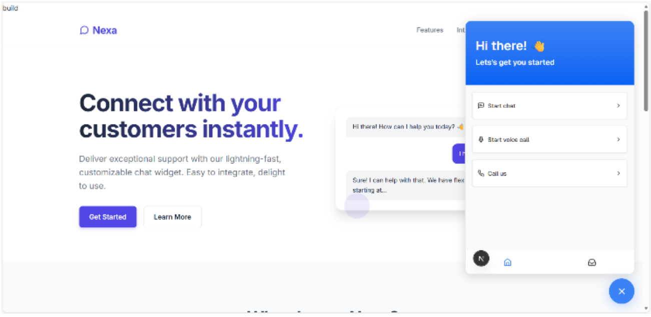
Hi! How can I help you today?

I need help with my account?

yes I can help you

Type your response as an operator...

Figure 10 Experimental Results



Connect with your customers instantly.

Deliver exceptional support with our lightning-fast, customizable chat widget. Easy to integrate, delight to use.

Hi there! How can I help you today? 👋

Sure! I can help with that. We have flex starting at...

Hi there! 👋

Lets's get you started

Figure 11 Connect with Customer

Key Observations:

- Real-time sync eliminated latency issues common in REST-based systems
- AI-assisted workflows reduced human agent workload significantly
- Subscription enforcement effectively restricted premium features
- Widget embedding worked seamlessly across different web frameworks

The use of Convex and AI tools drastically simplified real-time and AI integration complexity compared to traditional architectures.

Conclusion and Future Work

This paper presented *Nexa*, a production-ready AI-powered multi-tenant customer support platform integrating AI chat, voice agents, real-time dashboards, knowledge base search, and subscription enforcement in a unified SaaS solution.

Nexa demonstrates how modern full-stack technologies can be combined to build scalable, secure, and extensible AI systems for real-world enterprise use. The platform ensures:

- Strong tenant isolation
- Secure credential management
- Real-time responsiveness
- High usability

Future Enhancements

- Integration of multimodal AI (image + voice + text)
- Explainable AI for transparency
- Edge deployment for low-latency regions
- Fine-tuned domain-specific AI models
- Advanced analytics and sentiment detection

References

[1] Adam, M., Wessel, M., & Benlian, “AI-based chatbots in customer service and intelligence their effects on user compliance”, 2020.

[2] Følstad & Brandtzæg, “Users’ experiences with chatbots: Findings from a questionnaire study”, 2020.

[3] Chaves & Gerosa, “How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design”, 2021.

[4] Gartner Research, “AI and automation in customer service: Trends and insights” 2021.

[5] Martins De Andrade & Tumeler, “Increasing customer service efficiency through artificial chatbot” ,2022.

[6] Wiethof & Bittner, “Toward a Hybrid Intelligence System in Customer Service: Collaborative Learning of Human and AI”, 2022.

[7] Md. Arman & U.R. Lamiya, “Exploring the Implication of ChatGPT AI for Business: Efficiency and Challenges”, 2023.

[8] McKinsey & Company, “The Future of AI in Customer Support: A Hybrid Model”, 2023.

[9] Melnyk et al, “Prospects of business process management based on chatbots”, 2024.

[10] Kedi, Kamukama, & Mugisha, “AI Chatbot integration in SME marketing platforms: Improving customer interaction and service efficiency”, 2024.