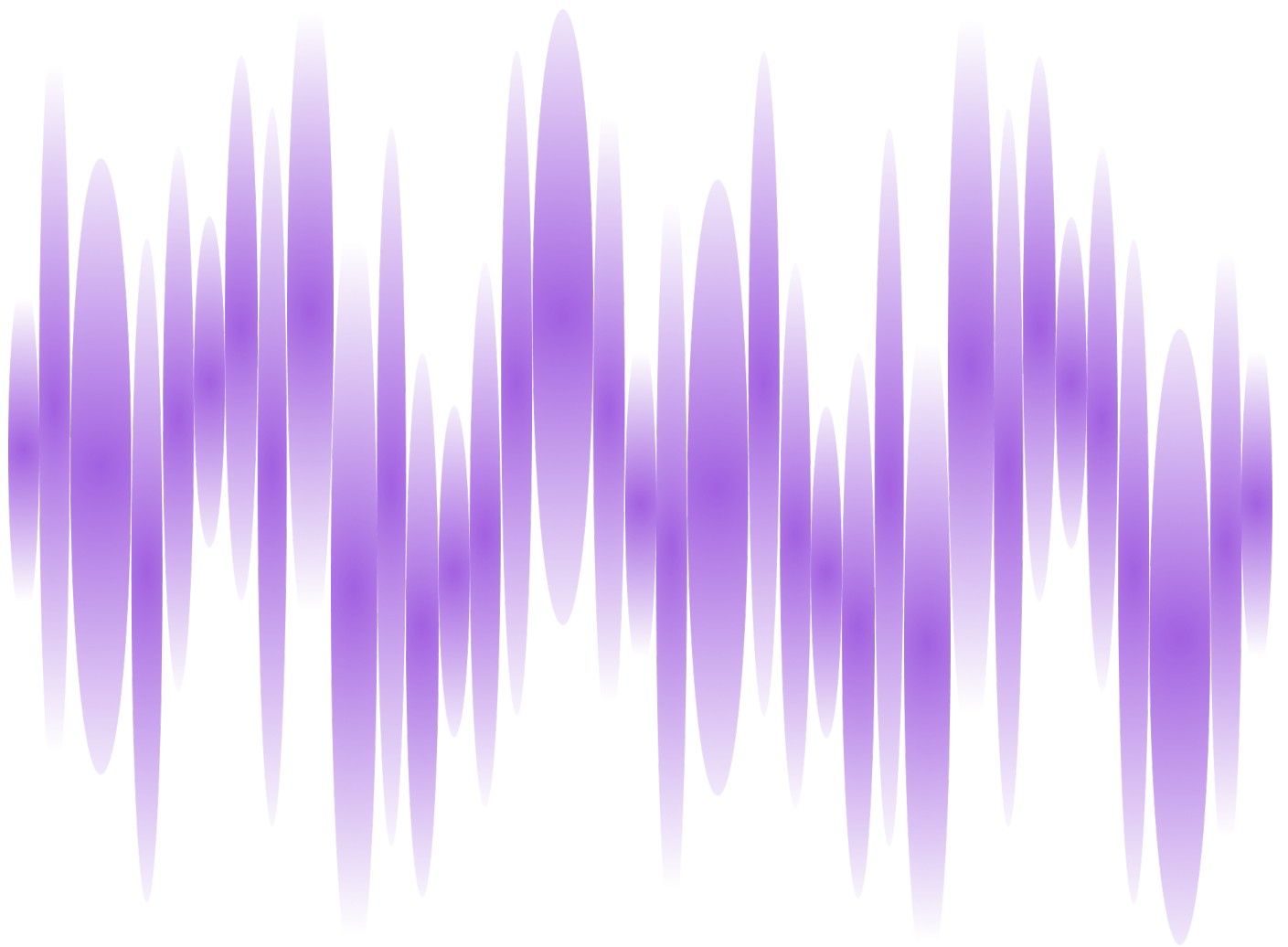# Elena

Advanced AI Card Support with
Intelligent Insights & Actions

zeta

Table of Contents

Card Support Needs an AI Revolution

Card issuers face a mounting customer support challenge: call volumes are projected to grow **20%** in three years[1], but traditional agent-based models can't scale to meet this demand.

The growing complexity of digital payments is placing higher demands on existing support infrastructure, This is compounded by inflexible capacities, talent shortages, and rising operational costs.

Issuers are addressing these challenges by investing in self-service capabilities and digital capabilities across chat, IVR and live agent channels. However, the response to these digital solutions has been mixed. Only **10%** of newly built digital platforms are fully scaled or adopted by customers[2] due to cognitive load, lack of personalization, and broken context across interactions (Image 1).

"

Today's customers want to be served by technology, but cared for by humans.

	Cognitive Load Due To			No Personalization	Static features/ limited by code	No context continuity
	Rigid menus	Complex steps	Not natural language			
IVR	✗	✗	✗	✗	✗	✗
Chat	✗		✗	✗	✗	✗
In-app Self-Service		✗	✗		✗	

Image 1: Shortcomings of digital interfaces for card support services

Some issuers are also experimenting with AI to enhance their support infrastructure. Their approach includes deploying predictive analytics for proactive customer engagement, AI-assisted agent tools for complex queries, and automated chatbots as first points of interaction in digital channels.

The more forward-looking issuers are integrating AI directly with their processing platforms to enable real-time data access, automated workflows, and intelligent service orchestration across channels. Through these implementations, issuers expect to automate **50-80%** of routine queries while reducing live agent dependence by up to **75%**[3].

Despite the investments in AI however, issuers have yet to unlock its true potential in customer support. Chatbots - the most common AI implementation - demonstrate this gap clearly. They have increased the complexity of user interactions instead of simplifying them. Advanced Conversational AI can solve this, enabling natural interactions that transform how cardholders access support.

The Conversational AI Implementation Gap

While chatbots brought about a breakthrough in customer care through automation of responses they have largely failed to solve persistent challenges like personalization, adapting to human language, and gauging the context of a customer's query. Data on customer dissatisfaction with chatbots keeps surfacing, and a recent survey reported 68% consumers[4] having a bad experience when the chatbot was unable to understand their need or answer their question.
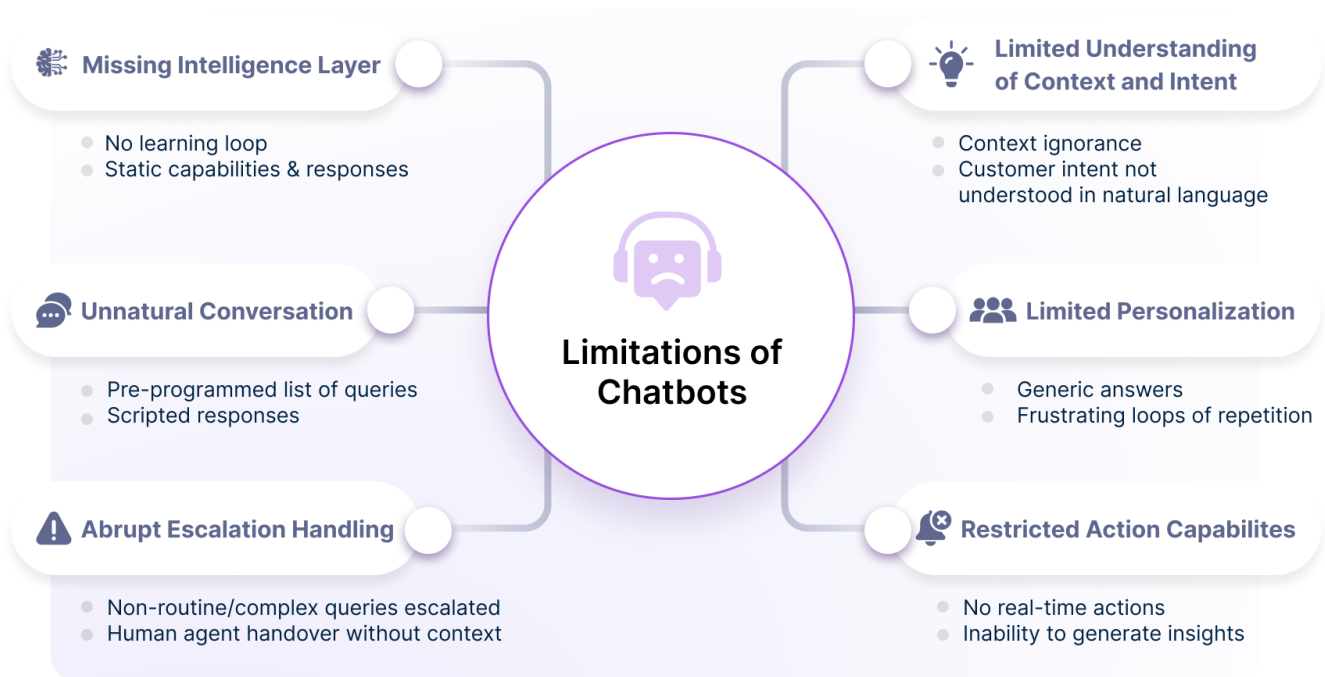
Image 2: Limitations of Chatbots in Customer Care

Combining the capabilities of Conversational and Generative AI underpinned by a next-generation processing platform can help issuers overcome these challenges. With these capabilities, an advanced Conversational AI card support solution can deliver:

- **A natural language interface** enabling intelligent, unscripted conversations

- **Intelligent automation** of routine query resolution at scale

- **Awareness of cardholder intents** and context continuity through the conversation

- **Personalization** of conversation through cardholders' historical and real-time data

- **Intelligent insights and action** to meet customer needs

- **Contextual escalation** to human agents

These capabilities form the foundation of Elena, Zeta's advanced Conversational AI that reimagines what card support can be.

Introducing Elena:
Advanced Conversational AI for Card

Zeta is pleased to introduce Elena, an advanced Conversational AI for cardholders that brings intelligent and personalized card support to life. Elena breaks through the limitations of conventional support solutions to deliver what cardholders truly need - instant and contextual engagement, relevant insights and automated actions - that scale seamlessly for issuers of all sizes and without the limitations of today's solutions.

This solution paper presents Elena's capabilities and impact as follows:

Part 1: Elena: AI-Powered Card Support at Scale

An introduction to Elena's core capabilities, its differentiators with respect to current chatbot models, and its enterprise-grade deployment capabilities

Part 2: System Design: Elena's Architecture and Orchestration

A view of the 4-layer architecture and system orchestration that enable dynamic cardholder experiences and rapid deployment of new workflows on Elena

Part 3: Building Trust: Elena's Banking-Grade Guardrails and Security

Measures to ensure accuracy and reliability of Elena's responses, prevent hallucinations and deliver quality and regulatory compliance

Part 4: Processing Foundation: The Role of Next-Gen Processing Systems in Conversational AI

Insights into why AI adoption on legacy processing systems delivers sub-optimal outcomes and how next-gen processing enables Conversational AI

Part 5 : Path to Value: Framework for Conversational AI Adoption

A 4-part adoption framework covering business, operational, compliance and security, and technology considerations to help issuers maximize value realization from Conversational AI

AI-Powered Card Support at Scale

Elena offers cardholders an intelligent, personalized and human-like experience at scale

Elena is an advanced Conversational AI platform for card support. With Elena, issuers can rapidly launch banking-grade intelligent card assistants with out-of-the-box capabilities for personalized support, insights and actions across the card lifecycle.

Given the rapidly evolving nature of AI capabilities, Elena' addresses high frequency card support requests[5] out-of-the-box and also enables issuers to easily build custom intent flows and continuously evolve their AI-powered support capabilities over time. This allows them to stay ahead of the curve and differentiate their customer support.

Transformed Cardholder Experience

Elena offers cardholders an intelligent support experience through natural, human-like conversation across voice and chat. While conventional apps and automated voice responses force users through static menus, forms, and workflows, Elena leverages Generative AI models that enable intelligent, natural conversations, recognition of user intent, and context continuity in conversations without the constraints of pre-defined scripts.

Cardholders can request real-time insights on their transactions and ask for automated actions through these natural interactions. When situations require human expertise, Elena ensures seamless transitions by intelligently escalating to live agents with full context preserved.

Elena's ability to understand intent and resolve cardholder queries can be classified as follows:

1. Knowledge Driven Queries

Queries that can be resolved entirely from issuer knowledge base and product documentation.

Examples:

'How do I apply for a credit card'

'How do I increase my credit limit'

2. Data Driven Queries

Queries that require customer, account, card or transaction data along with the issuer knowledge base to resolve.

Examples:

'Why was my transaction declined'

'Why have I not received my card yet'

3. Insight Driven Queries

Queries that require customer, account, card or transaction data, analysis on this data as well as the issuer knowledge base to resolve.

Examples:

'Show my expenses by merchant category'

'Have I spent more money on groceries or gas last month'

4. Action Driven Queries

Queries that require a multi-step workflow leveraging customer, account, card or transaction data, its analysis and the issuer knowledge base to complete the desired cardholder action.
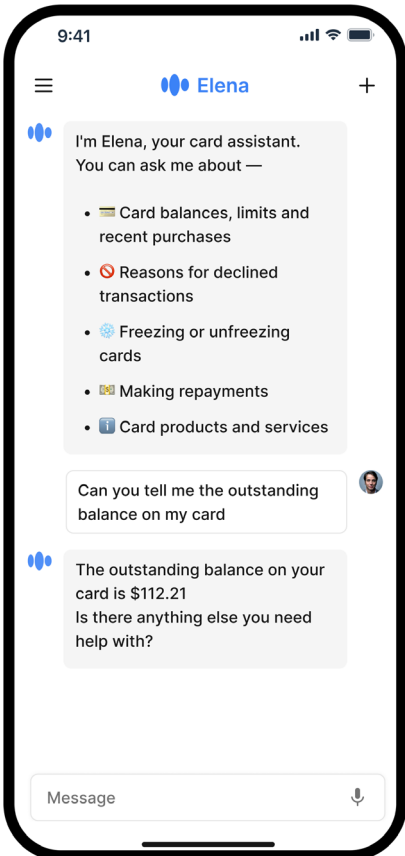
Examples:

'I'd like to file a dispute against the last transaction on my card'

'I need to make a payment'

Use Case: Cardholder Making a Payment Using Elena

The following example showcases Elena in action helping a cardholder make a payment against their outstanding balance.
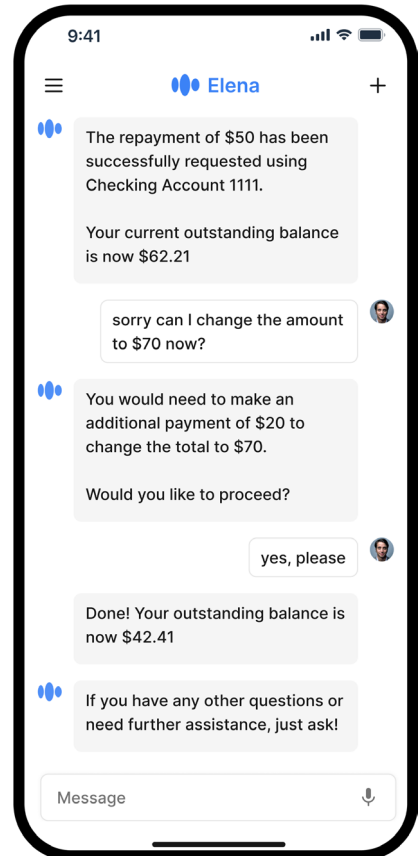
Elena accesses real-time account data to provide instant responses to cardholder queries

Elena executes a multi-step workflow to understand whether the cardholder wishes to pay their full or partial outstanding and the source of the payment
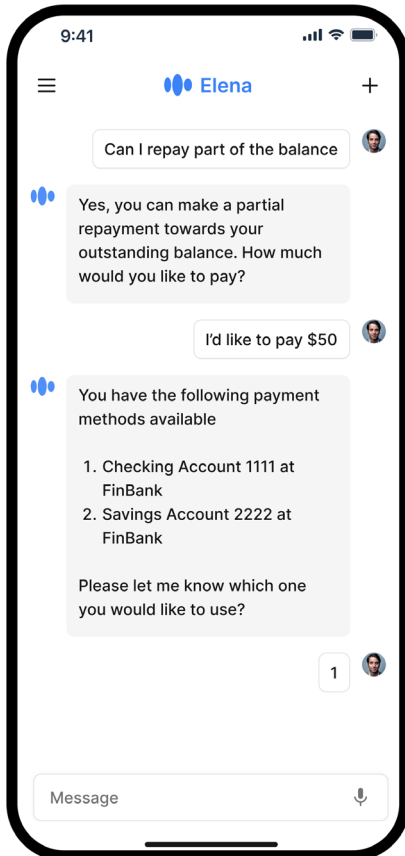
Elena maintains context even when the cardholder changes the amount to be paid and responds with accurate account balances in real time

Phone 1 — Elena chat:

9:41

Elena

I'm Elena, your card assistant. You can ask me about —

- 💳 Card balances, limits and recent purchases
- 🚫 Reasons for declined transactions
- ❄️ Freezing or unfreezing cards
- 💵 Making repayments
- ℹ️ Card products and services

Can you tell me the outstanding balance on my card

The outstanding balance on your card is $112.21
Is there anything else you need help with?

Message

Phone 2 — Elena chat:

9:41

Elena

Can I repay part of the balance

Yes, you can make a partial repayment towards your outstanding balance. How much would you like to pay?

I'd like to pay $50

You have the following payment methods available

1. Checking Account 1111 at FinBank
2. Savings Account 2222 at FinBank

Please let me know which one you would like to use?

1

Message

Phone 3 — Elena chat:

9:41

Elena

The repayment of $50 has been successfully requested using Checking Account 1111.

Your current outstanding balance is now $62.21

sorry can I change the amount to $70 now?

You would need to make an additional payment of $20 to change the total to $70.

Would you like to proceed?

yes, please

Done! Your outstanding balance is now $42.41

If you have any other questions or need further assistance, just ask!

Message

Intelligence Beyond Scripted Chatbots

Most current implementations of chatbots fail to deliver intelligent, personalized customer experiences due to limited context awareness and poor integration with underlying core systems.

Elena's intelligence, allowing it to overcome these limitations, stems from two powerful integrations:

- Direct access to issuer knowledge bases ensuring responses align with institution-specific policies and operating procedures

- Native integration with Zeta's next-generation card processing platform, Tachyon to create, read, and update customer, account, card or transaction data in real-time.

With these integrations firmly in place, Elena can deliver human-like customer service through capabilities across three important dimensions:

1. Seamless context navigation

Elena maintains continuous conversation context, just like a human agent. Users can modify requests mid-conversation - for example, a user can initiate a request to pay $100 against their credit card bill and change the amount to $150 in the same conversation without restarting the process. When escalation is needed, the entire conversation context is transferred to human agents.

2. Banking-grade hallucination prevention

Elena manages the risk of incorrect or arbitrary responses through robust guardrails. Every response is grounded in issuer-specific knowledge bases and follows pre-approved workflows, ensuring accuracy and compliance.

3. Human-like conversations

Interacting with Elena reduces cognitive load. Cardholders can express what they want to know or do in everyday language, without worrying about the right terminology. It enables this through four capabilities:

- Advanced Speech Recognition (ASR) for accurate voice command interpretation
- Natural Text-to-Speech (TTS) for clear vocal responses
- Precise Intent Detection to understand user needs
- Contextual Dialogue Generation for natural conversation flow

Enterprise-Grade Deployment Capabilities

Elena's integration with a next-generation card processing core and the issuer knowledge base also enables enterprise-grade deployment capabilities like:

1. Omnichannel servicing capability

Elena provides consistent support across voice and chat platforms

2. Embeddability

Elena can be natively embedded into any modern mobile/web app or industry-leading contact center platform, for example, AWS Connect, Genesys, NICE or others

3. Monitoring and analytics

Issuers can access detailed performance data of all Elena interactions for in-depth analysis

4. Compliance and security

Elena has built-in guardrails to ensure it addresses only the intended issues while adhering to industry compliance and security standards

Part 2: System Design: Elena's Architecture and Orchestration

Architecture and System Orchestration

Elena's architecture enables dynamic cardholder experiences, continuous learning and agile deployment of new workflows

The digital interfaces enabling card support today have access to customer data but can't act on it intelligently. This is especially true in an omnichannel context across chat, IVR, and live agents. And every new feature demands custom development, hampering agility and increasing costs, often across each channel. Experienced IT teams understand this complexity all too well.

Elena is responsive, extensible and scalable by design. While its modern, 4-layer architecture enables rapid addition of new intent flows, its integration with Zeta's processing platform Tachyon allows swift deployment of new workflows. And its AI models support the ability to generate insights and take action without the need for specific code for each and can evolve independently to incorporate emerging capabilities across every relevant support channel.

Multi-Layer Intelligence Framework

Architecturally, Elena is composed of 4 distinct layers (Image 3):

1. A central conversation orchestration engine manages all conversations, whether voice or chat

2. Advanced AI models that understand and execute requests through
- Intent Detection
- Retrieval Augmented Generation
- Code Generation
- Other models including image recognition

3. Advanced toolkit to
- Run code
- Run Text-to-Speech, Automatic Speech Recognition
- Execute function calling and others

4. Integration layer that connects to
- Zeta's card processing core for API calls and operations
- The issuer document repository comprising its standard operating procedures, policies and knowledgebase.
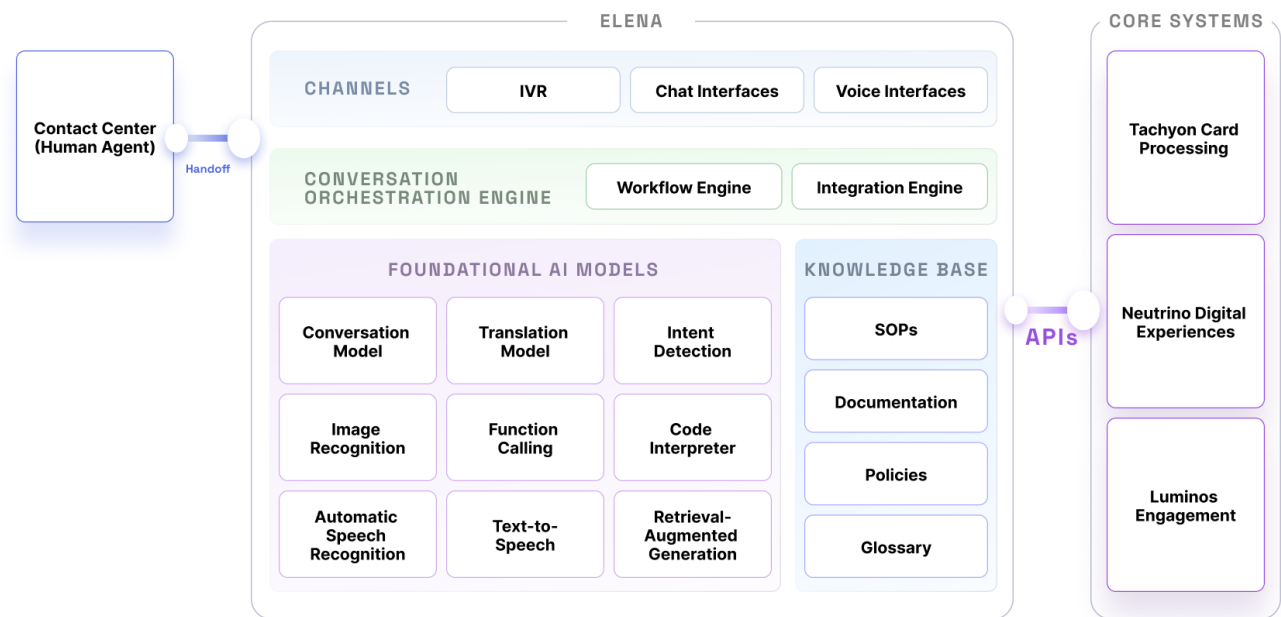

Image 3: Elena's 4-Layer Architecture

This architecture enables Elena to dynamically answer questions, generate insights, and execute actions without requiring specific code to be written for each.

Orchestrating Intelligent Insights and Actions

Let's see how it works with an example (Image 4). The conversation orchestration engine receives the cardholder request to view their recent purchases by category, visualized in a pie chart. This triggers the foundational models for Intent Detection, Retrieval Augmented Generation, Code Generation, etc., allowing Elena to understand and execute the request being made.
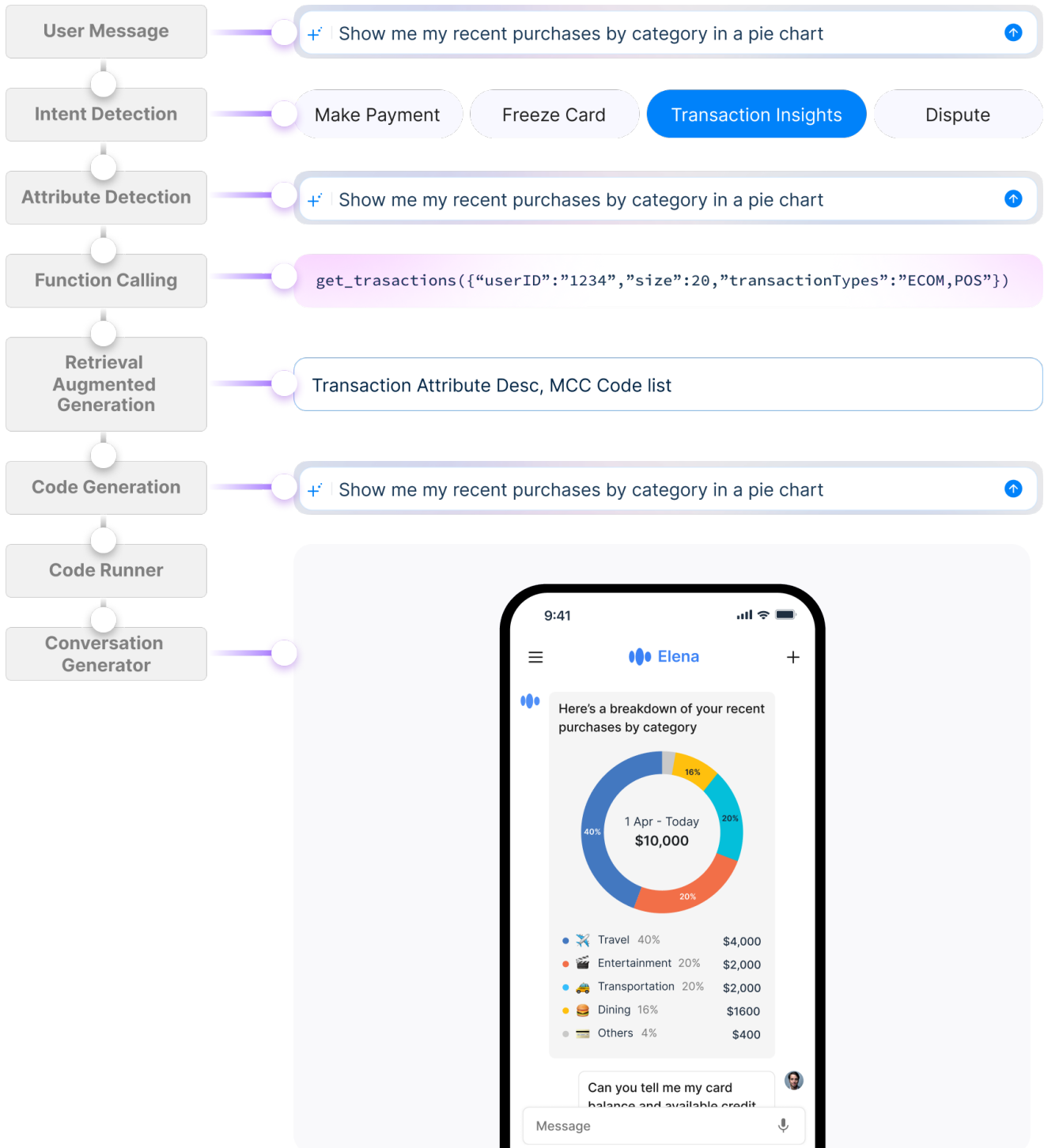

Image 4: Elena Executing a Cardholder Query for Transaction Insights

Banking-Grade Guardrails & Security

With advanced guardrails to prevent arbitrary responses or hallucinations, Elena uses issuer-specific knowledge and customized workflows to ensure accurate handling of queries.

The accuracy and integrity of information provided to cardholders is non-negotiable in card services: a single error can damage customer trust and issuer reputation. The risks are amplified with modern AI technologies; while Large Language Models (LLMs) offer powerful conversational capabilities, they can also generate convincing but incorrect information (aka hallucinations), make unauthorized decisions, or breach compliance guidelines.

Elena's built-in safeguards ensure that every interaction meets strict standards for safety, accuracy, and compliance.

Intelligence Trained on Issuer Knowledge

Elena builds its knowledge foundation exclusively from issuer-approved sources:

- Detailed knowledge repositories
- Standardized operating procedures
- Approved workflow documents
- Compliance guidelines

This focused training ensures Elena operates strictly within defined boundaries, eliminating the risk of unauthorized actions or incorrect information.

Multi-Layer Hallucination Prevention

Elena employs four critical components to prevent AI hallucinations:

1. Intent-focused action generator

Elena's Action Generator validates user intent before directing conversations through pre-approved flows, keeping the risk of hallucinations near zero.

2. Enterprise-grade language model

Elena's response generation system uses an enterprise-grade large language model independently verified for its accuracy and minimal hallucination risk.

3. Robust prompt engineering and guardrails

Elena's comprehensive guardrails combine precise prompt engineering with strict boundary enforcement. Every response is validated in real-time against approved topics and functions, ensuring conversations remain focused and accurate.

4. Rigorous input management

Each interaction undergoes systematic validation through sophisticated anti-jailbreak protections to maintain robust security against potential system manipulations.

Quality Assurance

Elena maintains peak performance through continuous response monitoring and systematic scenario testing. Edge cases undergo careful evaluation, with immediate corrective action capabilities ensuring consistent accuracy. This proactive approach to quality control ensures Elena consistently meets the exacting standards of card issuers while delivering the fluid, natural interactions cardholders expect.

Regulatory Compliance

Elena is engineered to support card issuers' compliance requirements. Every implementation adheres to critical banking and payments industry standards like:

1. Payment Card Security

- Built to PCI-DSS specifications for card data handling
- End-to-end encryption of sensitive information
- Secure data transmission protocols
- Controlled access to cardholder data
- Compliant with operational security protocols

2. SOC 2 Type II certified infrastructure and processes

- Regular security audits and assessments
- Robust data privacy protocols
- Secure cloud infrastructure

Its integration with the issuer processing platform ensures that it inherits the platform's banking compliance framework and meets the data retention and privacy requirements.

Role of Next-Gen Systems in Conversational AI

Delivering personalized and intelligent cardholder support demands a next-generation processing platform built for real-time data access and action across systems.

Since its first introduction in banking around a decade ago, AI-powered customer support has struggled with managing high expectations in terms of query resolution rates, response times, and productivity. The emergence of Generative AI has further fueled expectations, with success stories of early adopters now routinely making headlines[6]. However, for most issuers, the technical complexities of deploying and scaling advanced AI solutions with technical and operating debt due to their legacy systems remains a key challenge.

The ability of an AI-powered card solution like Elena to evolve and scale effectively depends on how the underlying card processing platform enables real-time data access and rapid feature updates.

The 4 Barriers to Conversational AI Adoption on Legacy Systems

Advanced AI requires instant access to card, account and transaction data through direct card processor integration. Legacy card processors, with their batch processing and rigid workflows, can't support the real-time operations that a Conversational AI demands.

While legacy card systems have multiple constraints, the following 4 limitations critically impact AI-powered support performance.

1. Fragmented Data and Systems Integration

Legacy systems store data in silos across multiple subsystems. This fragmentation makes it difficult for Conversational AI to aggregate information and generate coherent, context-aware responses. While middleware can consolidate data between systems, it adds significant costs and delays, creating bottlenecks that slow down AI implementation.

2. Outdated Messaging Protocols Leading to Reduced Interoperability

Legacy messaging protocols don't align with modern API architectures, creating barriers between AI agents and critical subsystems like payment processing, fraud detection, and authentication. This prevents AI from coordinating across systems to deliver comprehensive responses.

3. Lack of Real-Time Event Handling

Legacy systems have limited real-time processing capabilities. This limits the AI's ability to respond dynamically to state changes and user interactions, often resulting in delayed responses and outdated information.

4. Limited Availability of Purpose-Built APIs

Without comprehensive APIs, Conversational AI cannot deliver personalized support or automated workflows. It needs real-time access to cardholder data to provide tailored responses and alerts. Lacking these APIs, it can only offer generic answers, forcing users to rely on traditional support channels for anything beyond basic queries.

The 4 Next-Generation Processing Enablers of Conversational AI

Delivering AI-powered personalized, and conversational cardholder support demands a next-generation processing platform built for real-time data access and action across systems – such as Zeta's Tachyon processing platform.

Tachyon delivers four key enablers for advanced Conversational AI solutions like Elena to deliver on their promise:

1. A comprehensive API catalog that supports every create, read, update operation in scope for the AI solution.

For example, Elena's access to Tachyon's extensive API library enables it to perform actions typically handled by agents, including payment processing, account freezing/unfreezing, transaction searches, and adding payment methods.

2. Real-time events stream that allow the AI to learn and adapt its actions and decisions.

Elena adapts continuously through real-time events and conversation context, enhancing decision-making at scale. For example, during payment conversations, Elena can display balances due after accounting for partial payments in real-time.

3. Open data stack for continuous, supervised learning of the AI models on comprehensive user data available in structured data lakes.

Elena analyzes varied phrases used in real user conversations to request payments, such as 'clear my balance' or 'settle my dues' to boost its recognition of payment intents in natural language.

4. Hyper-personalization and configurability of the AI's conversational flows, enabling unique and context-aware responses to cardholder needs.

For example, a cardholder can ask Elena to freeze a lost card in natural conversational style, eliminating the need to locate specific app functions.

Framework for Conversational AI Adoption

Leveraging Conversational AI for card support represents a mature, adoption ready use case for issuers.

For issuers exploring use cases where AI can be introduced for tangible benefit, cardholder support is a clear winner. In its recent report on AI-enabled customer service, McKinsey highlights two key areas of operational impact[7]:

- Enhanced self-service automation with contextual in-app nudges and controls (~50-80% contacts automated), and

- Reduction in call volumes escalated to human agents by installing Conversational AI at each entry point (>80% of all actions automated)

Zeta adopts a similar framework to drive improved operational outcomes for issuers.

1. Improved effectiveness of in-app self-service

Next-gen capabilities like granular card and transaction controls and advanced security features can reduce dependence on live agent card support by up to 75%[8]

2. Automation of routine queries

Elena's natural language interfaces across chat and voice can automate 20% of the remaining 25% calls, ensuring that only the most complex and critical calls are escalated to live human agents

3. Enhanced operational efficiencies

Elena leverages cloud-native infrastructure to enable flexible processing capacity management, 24/7 availability and 99.9% uptime with faster response times and higher accuracy

However, realizing value with a Conversational AI assistant for card support requires a structured framework (Image 5) that evaluates key considerations along 4 critical dimensions:

- Business Considerations
- Operational Readiness
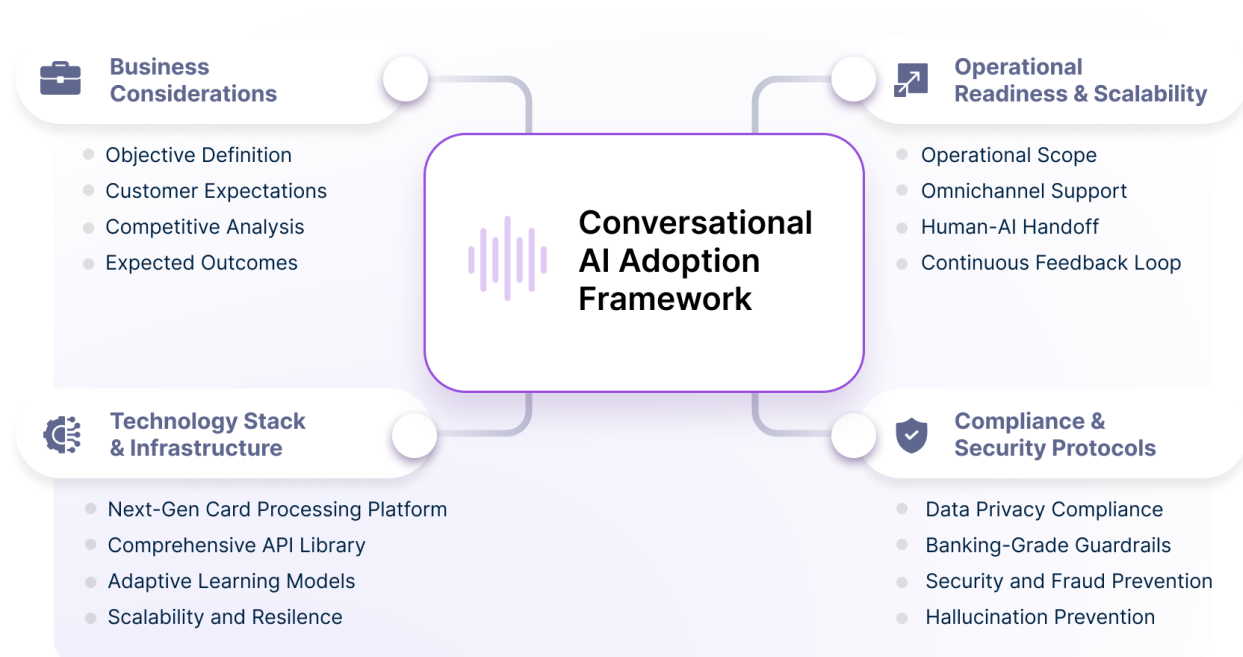- Compliance
- Technological

Image 5: Conversational AI Adoption Framework

The following section elaborates the essential considerations under each dimension.

1. Business Considerations

- **Objective Definition:** Prioritize business objectives, e.g., reducing customer service costs, enhancing user satisfaction, or improving resolution times.

- **Customer Expectations:** Identify customer needs and preferences for personalization, quick resolutions, and accessible support across multiple channels.

- **Competitive Analysis:** Benchmark capabilities of Conversational AI solutions offered by leading issuers to ensure competitiveness of your solution.

- **Expected Outcomes:** Quantify expected outcomes. e.g. reduced agent-handled queries, faster response times, and higher customer satisfaction ratings.

2. Operational Readiness

- **Scope:** Ring fence the scope of service that your Conversational AI can provide. E.g., 1) Information like balances, statements, or billing cycle 2) Financial transactions like repayments 3) Advanced insights like spend categorization.

- **Omnichannel Service Delivery Support:** Evaluate what is needed for the assistant to be seamlessly accessible via the mobile app, website, and IVR, catering to different customer preferences.

- **AI Agent - Human Agent Handoff:** Implement a smooth handoff process to human agents when queries go beyond the scope of the AI agent, with the context preserved in transitions.

- **Continuous Feedback Loop:** Integrate performance analytics and customer feedback to iteratively improve the assistant's responses and expand its capabilities over time.

3. Compliance and Security

- **Data Privacy Compliance:** Adhere to CARD Act, CCPA, and other regulations for customer data, ensuring consent and robust data encryption.

- **Banking-Grade Guardrails:** Prevent unauthorized actions or responses by limiting the AI's access strictly to predefined, bank-approved workflows and knowledge sources.

- **Security and Fraud Prevention:** Employ strict anti-fraud measures, including transaction verification, and ensure the Conversational AI detects any signs of suspicious activity.

- **Hallucination Prevention:** Integrate multiple layers of control to ensure accurate responses, such as intent-focused action generators, input validation, and anti-jailbreak protections to maintain reliable, secure interactions.

4. Technology & Infrastructure

- **Next-Gen Card Processing Platform:** Leverage a cloud native, real-time card processing platform to handle API requests, track transactions, execute payments, and fetch details to ensure quick and accurate responses.

- **Comprehensive API Catalog:** Ensure the availability of an extensive API library to support all types of customer interactions, from simple queries to complex requests like payment processing and account management.

- **Scalability and Resilience:** Design the architecture to scale on demand, with cloud-native infrastructure capable of handling high volumes and concurrency without degradation in response time.

The successful implementation of an advanced Conversational AI like Elena for card support represents a significant leap forward in modern banking services. By following this comprehensive framework - encompassing business considerations, operational readiness, compliance protocols, and technological infrastructure - financial institutions can deliver a solution that not only meets current customer needs but also scales for future demands.

Next Steps

Launch Intelligent Card Assistants with Elena

Accelerate AI adoption by launching banking-grade intelligent assistants for your card programs.

Scan here to Contact Sales

Bibliography

AI Models & Components

1. Conversation Model: Component that produces natural language responses based on processed data and context. This is an enterprise-grade large language model verified for accuracy and performance.

2. Intent Detection Model: Component that processes user inputs to determine the requested action or query type

3. Attribute Detection Model: Component that identifies specific entities within user inputs to enable accurate request processing

4. Function Calling Model: Component that selects appropriate API calls and system functions from a predefined API catalog based on user request

5. Code Generation Model: Component that generates code for data processing and analysis tasks

6. Conversation Orchestration Engine: System that coordinates the sequence of interactions across AI models and components

7. Retrieval Augmented Generation (RAG): Framework that enhances AI responses by incorporating relevant information from issuer knowledge bases

Technical Capabilities

1. Advanced Speech Recognition (ASR): System for accurate interpretation of voice commands and spoken queries

2. Text-to-Speech (TTS): Technology that converts text responses into natural-sounding voice output

3. Anti-jailbreak Protections: Security measures preventing manipulation of AI responses or unauthorized access

4. Content Verification Systems: Controls ensuring responses align with approved topics and issuer guidelines

References

[1] **McKinsey**, Where is customer care in 2024 | March 2024

[2] **McKinsey**, The state of customer care in 2022 | July 2022

[3] **McKinsey**, The next frontier of customer engagement: AI-enabled customer service | March 2023

[4] **Verint**, The State of Digital Customer Experience 2024 | June 2024

[5] Some advanced capabilities are in late-stage development

[6] **Klarna**, Klarna AI assistant handles two-thirds of customer service chats in its first month | February 2024

[7] **McKinsey**, The next frontier of customer engagement: AI-enabled customer service | March 2023

[8] Zeta internal estimates based on live customer data

About Us

Headquartered in San Ramon, California, Zeta was named by Celent in 2023 as being among the likeliest partners for US banks and credit unions looking to modernize to a next-gen processing platform. Zeta's modern platform is proven in production with major US financial institutions, delivering high-performance credit card programs at scale

Globally, Zeta is recognized as a next-gen banking technology company. Our platform enables financial institutions to launch extensible and compliant banking asset and liability products, across cards, loans and deposits, rapidly. Our cloud-native and fully API-enabled stack supports processing, issuing, lending, core banking, fraud, loyalty, digital banking apps, and many other capabilities.

Zeta has 1700+ employees with over 70% in technology roles across locations in the US, Middle East, and Asia - representing one of the largest and most capable teams ever assembled in banking tech.

Globally, customers have issued 25M+ cards on our platform.
Visit us at www.zeta.tech or follow us on LinkedIn, YouTube, and X.