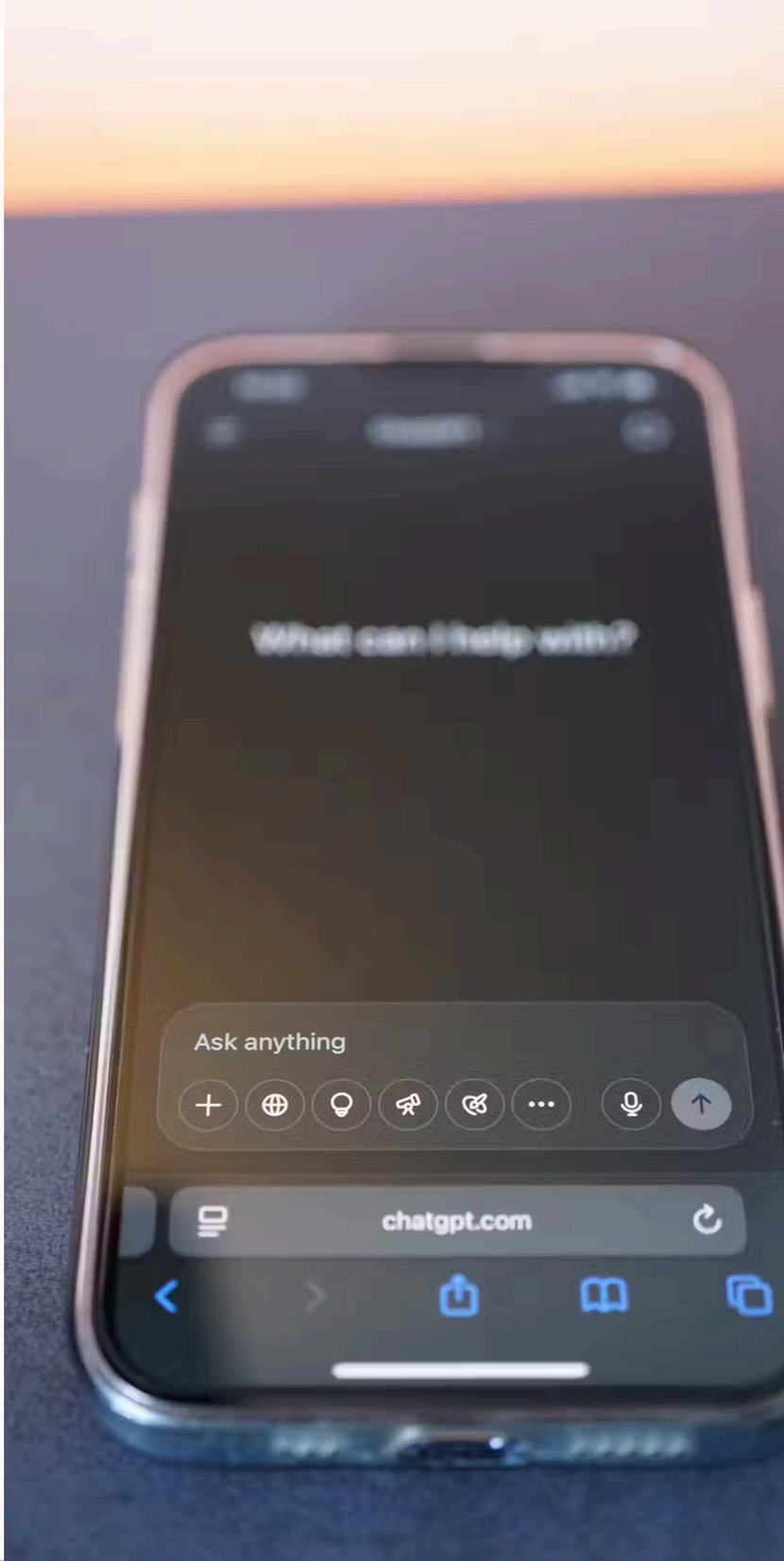
Enterprise Support: Strategic Advancement with Omnichannel AI

by Vijay Cinnakonda

May 30 2025



ServiceNow Virtual Agent Strategy

Acknowledging Progress, Identifying Critical Gaps

Migrating the VA's enterprise service desk (ESD) capabilities to the ServiceNow platform, including replacing the existing Abel chatbot with the ServiceNow Virtual Agent, represents a step forward in consolidating the system of record and streamlining IT service management (ITSM) processes. However, this strategy primarily addresses *where* the virtual agent resides, rather than evolving *what* it can achieve in terms of proactive support, intelligent automation, and cost efficiencies. True transformation necessitates a broader vision that extends beyond the capabilities of a single platform's embedded tools, requiring a strategic addition to complement and enhance the existing ServiceNow investment.

Proactive Deflection

Modern ESDs should resolve issues before they escalate, using AI to be proactive.

1

Voice Intelligence

The voice channel needs intelligent automation, not just call routing.

2

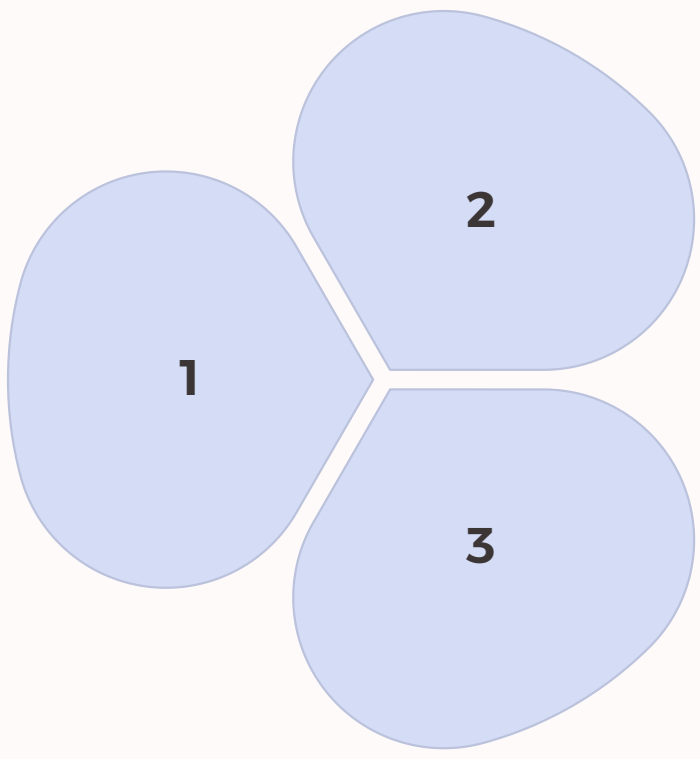
High Cost of Status Quo

The current voice support model has inefficiencies and above-market rates.

3

Crafting NextGen ESD

The vision is to create an ESD that offers users seamless, intelligent, and proactive support across all preferred channels, available 24/7. This "fit for purpose solution that can scale" 1 aims to resolve a high percentage of issues on first contact, often before the user even perceives a problem, while significantly reducing operational costs and freeing human agents to focus on complex, high-value tasks.



LLM Orchestration

Langchain, Azure Semantic Kernel

LLM Models

Azure OpenAI offers variety of LLMs (GCC/FedRAMP)

Channels

MS Teams, Copilot, Web Chat, Telephony, SMS

Core Architecture: The Pillars of Transformation

The proposed solution is founded on several key architectural components, designed to work in concert to deliver a comprehensive and intelligent support experience

1

Azure Communication Services (ACS) – FedRAMP Compliant:

ACS serves as the backbone for true omnichannel communication, natively supporting Voice (VoIP, PSTN), SMS, Web Chat, and integration with Microsoft Teams. It provides the essential "infrastructure for handling TTS/STT for LLM integration to surface virtual agents voice calls and SMS messages essential for delivering MFA/OTP codes for authentication". ACS also offers "call automation, and connectivity to NICE InContact ACD system used by ESD" 1, allowing for a phased integration or augmentation of existing telephony. Critically for the VA, ACS, as part of the Azure ecosystem, meets "GCC and FedRAMP compliance ensures deployment adheres to stringent security and regulatory standards".

2

Advanced LLM Orchestration (LangChain/Semantic Kernel):

This layer acts as the intelligent "brain," orchestrating the interactions between LLMs, VA-specific data sources (including ServiceNow as a system of record), enterprise systems, and user-facing channels. Frameworks like LangChain or Azure Semantic Kernel enable the development of "sophisticated customer-facing ITSM chatbots capable of understanding context, accessing knowledge, and automating tasks".1 This orchestration is vital for "intelligent information assimilation, safe moderation, and crucial RAG grounding of AI responses in our enterprise data, ensuring accuracy and relevance across all channels".1

3

Azure OpenAI (GCC) for Secure, Powerful NLU:

The solution will leverage Azure OpenAI Service, providing access to a "variety of LLMs (GCC/FedRAMP)".1 This ensures access to state-of-the-art NLU and generation capabilities within a secure, compliant government cloud environment. These models enable "advanced language capabilities: Generates coherent text, summarizes documents, translates languages, and demonstrates profound understanding of complex inquiries" 1, far surpassing the NLU limitations of traditional virtual agents.

4

Native Integration with Microsoft Teams & Copilot (Leveraging Existing M365 Investments):

A key strategic advantage is the ability to "strategically leveraging our existing O365 license - utilizing the free MS Teams / Copilot Chat (GCC) alongside ACS for broad reach".1 This delivers "a seamlessly integrated and highly capable AI support experience directly within the secure and familiar Microsoft ecosystem" 1, reducing adoption friction and maximizing the value of existing Microsoft investments.

These capabilities enable a more comprehensive and unified service desk experience compared to the limitations of a ServiceNow-centric approach, potentially deflecting from creating incidents and resolving issues that would typically require Tier 1 and above agent intervention.

Engagement via Voice/SMS & MS Teams Channels

Azure Communication Services provides infrastructure for handling TTS/STT for LLM integration to surface virtual agents voice calls and SMS messages essential for delivering MFA/OTP codes for authentication

Seamless Integration with Non-ServiceNow Systems

MCP allows AI to interact with diverse enterprise applications and data sources outside the ServiceNow platform. The current ServiceNow virtual agent

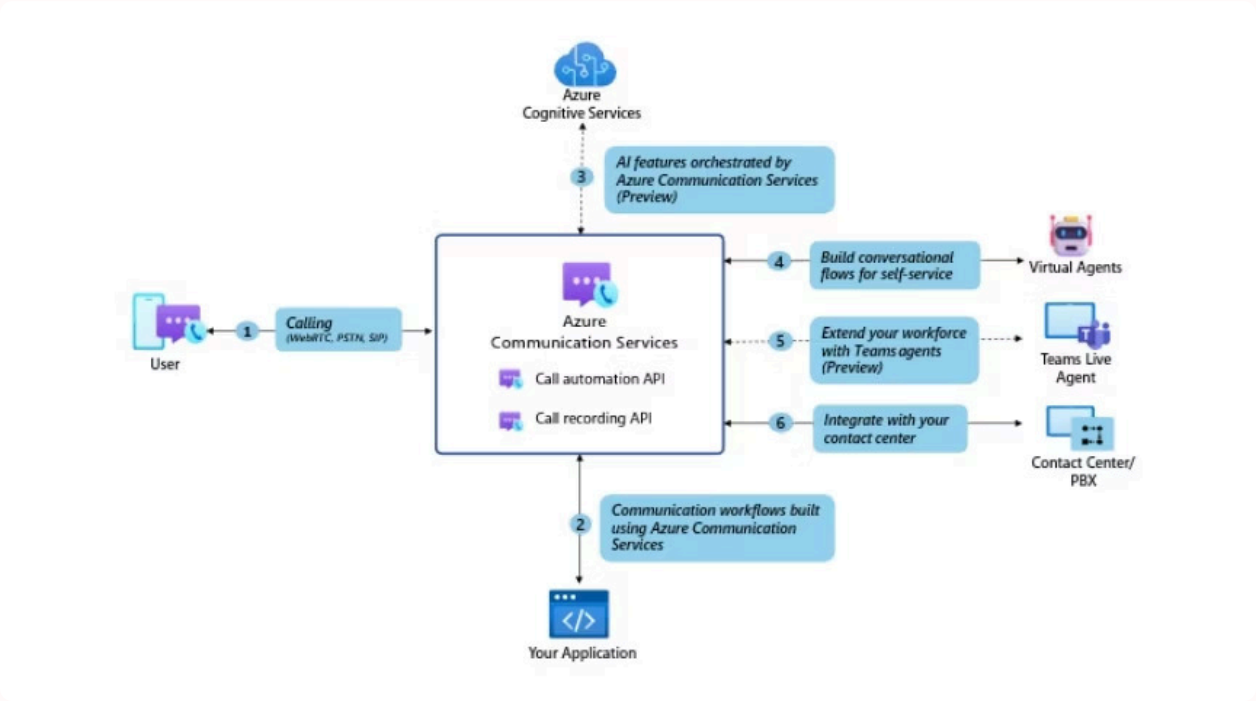
Multiturn Natural Language Conversations

LLMs can offer multiturn conversation handling without retraining and can enhance its intrinsic knowledge and intelligently adapt and push towards deflection rather than mere creation of tickets

Proactive Issue Detection and Resolution

By adding extrinsic capabilities like Tachyon or end point monitoring capabilities (InTune) it can integrate self heal capabilities (Group Policy changes etc).

Omnichannel is the way



Moving beyond basic virtual agents, we can establish a superior, **fit-for-purpose** omnichannel ESM/ESD capability by strategically leveraging our existing O365 license – utilizing the *free* **MS Teams / Copilot Chat** (GCC) alongside **ACS** for broad reach, while underpinning the entire experience with robust **LLM Orchestration** essential for intelligent information assimilation, safe moderation, and crucial **RAG grounding** of AI responses in our enterprise data, ensuring accuracy and relevance across all channels.

The proposed solution is architected to directly address the gaps and supplant capabilities of the current Virtual Agent:

NLU Sophistication:

Instead of limited templates, LLM powered interaction provides access to profound language understanding and contextual awareness.

Voice Integration:

ACS offers native, comprehensive voice capabilities (VoIP, PSTN, call automation, TTS/STT) integrated with Azure OpenAI, contrasting sharply with ServiceNow's reliance on third-party integrations and limited voice topic creation.

Omnichannel Support:

ACS provides a unified platform for Voice, SMS, Teams, and Web Chat, unlike the separate channel development required with ServiceNow VA.

System Integration:

The LLM orchestration layer is designed for standardized integration with a wide range of enterprise systems, moving beyond ServiceNow's primary focus on its own ecosystem.

Proactive Capabilities:

The architecture supports integration with endpoint management and identity services for true proactive deflection and self-healing, capabilities not inherent in the standard ServiceNow VA.

Development and Cost Model:

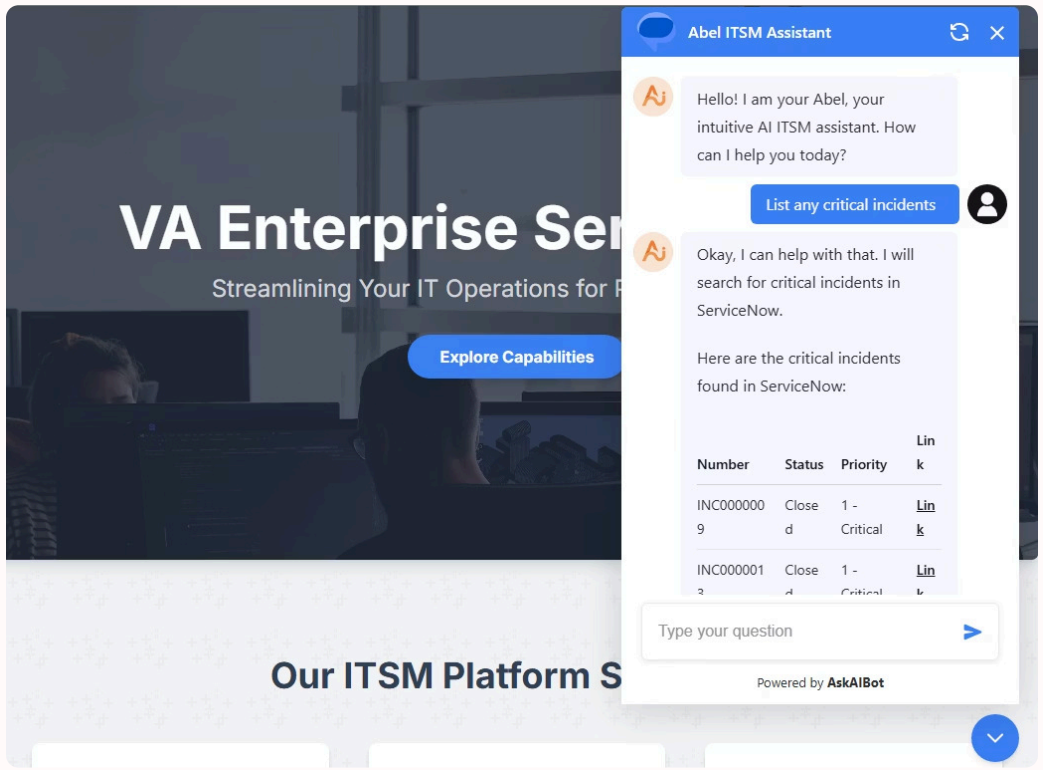
Leveraging open-source orchestration frameworks (LangChain) and existing Microsoft licenses (M365 for Teams/Copilot) offers a potentially more flexible and cost-effective model for advanced AI compared to proprietary add-ons like Now LLM.

Interactive Demo: Experience the Future Today

Place a voice call to **+1 (469) 530 9462** to simulate a AI powered voice agent

Connect with an interactive web chat at **vaesd.askaibot.com**

> Try the LLM Powered MCP in action at **vaesd.askaibot.com**. This working prototype demonstrates:
>
> a. LLM Orchestration (LangChain/Semantic Kernel) prowess
> b. Seamless integration with ServiceNow PDI
> c. Real-time execution of service desk actions
> d. Natural multi-turn conversations
> e. Contextual awareness using MCP



The barebone prototype handles user interactions more seamlessly than the currently proposed ServiceNow virtual agent.

Similar capabilities can be stood up with Azure Semantic Kernel and surfaced on web or *Copilot Chat (GCC)* to *Microsoft Teams (GCC)*

Business Case: Driving Efficiency, Reducing Costs, Enhancing Service

The adoption of the proposed AI-powered omnichannel solution presents a compelling business case for the VA, underpinned by significant potential for cost reduction, operational efficiency gains, and enhancements to user experience and compliance.

A. Significant Cost Reduction: A Compelling Financial Narrative

The financial benefits alone offer strong justification for this strategic shift.

Voice Channel Transformation

The most immediate and substantial cost savings are anticipated from revolutionizing the voice channel.

Current High Cost

Onshoring voice channel costs approximately $8-9 per call (User Query).

AI-Driven Cost

Industry benchmarks for AI-powered automated voice interactions range from $1-$2 per conversation. Some analyses suggest that AI voice agents can achieve over 90% savings compared to live answering services.

Dramatic Savings

If 50% of 200,000 annual voice calls are automated at $1.50/call, the savings could exceed $650,000 annually.

Reduced Call Times

AI can reduce call times by up to 40%, further lowering costs by 15-35% in the first year.

Deflection Benefits

Every incident deflected before reaching a human agent translates to direct cost avoidance.

Reduced Reliance on Tier 1 Support via Deflection

Every incident or request successfully deflected by the AI solution before it reaches a human agent translates into direct cost avoidance.

Impact of Agentic AI

Advanced "agentic AI" systems are demonstrating significantly higher deflection rates, with B2C companies seeing 55% improvements and public companies achieving 86% higher deflection.

Resolution Cost Reduction

A majority (63%) of companies using agentic AI report that their resolution costs are trending downwards. While specific VA figures would need to be modeled, one industry example (Aisera) with 50 agents at an $80,000 average salary handling 200,000 tickets a year at a $20 cost per ticket, projected potential savings of $3.125 million through AI automation. While the exact figures are illustrative, the principle of substantial savings through AI-driven deflection is well-established.

B. Operational Efficiency Gains: Doing More with Less

Beyond direct cost savings, the proposed solution drives significant operational efficiencies:

Increased First Call Resolution (FCR)

AI's ability to understand intent accurately and access a broad knowledge base can dramatically improve FCR rates. World-class FCR is considered 80% or higher. For every 1% improvement in FCR, a typical midsize call center can realize annual operational savings of approximately $286,000.

Reduced Average Handle Time (AHT)

AI can process information and execute simple tasks much faster than human agents, leading to a reduction in AHT for automated interactions. As noted, AI systems can cut call times by up to 40%.

Higher Agent Productivity and Satisfaction

By automating repetitive and mundane inquiries, which can constitute up to 80% of total support tickets, human agents are freed to focus on more complex, engaging, and high-value work. This not only boosts productivity but also improves agent morale and job satisfaction.

Conclusions and Recommendations

Rather than hitting the easy button and be walled in by single platform capability - embracing a fit for purpose solution that can scale can be achieved in a cost effective means by leveraging LLM Orchestrations, Azure Open AI, ACS, and the MCP. It offers a compelling alternative to the current plan focused solely on the ServiceNow Virtual Agent.

The proposed solution provides the necessary foundation to handle a wide range of service desk inquiries beyond the reach of the current strategy. The financial benefits alone, with potential for significant cost savings and strong ROI, provide compelling justification for this strategic shift.

Feature	Current Virtual Agent Strategy	Fit for purpose AI
Toolset	ServiceNow closed source. Separate License for Now LLM (does not have any customer interaction capability currently)	LLM Orchestration - Langchain (open source) / Semantic Kernel (Azure) MS Teams / Copilot Chat - M365 (existing licenses)
NLU Depth	Limited pre-built templates, requires constant feedback loops and is being phased out as it struggles with complex intent	Access to cutting-edge models, excels in understanding complex language, intent, and context
Voice Integration	Relies on third-party integrations	Native, comprehensive voice capabilities integrated with Azure Open AI
System Integration	Primarily ServiceNow ecosystem	Standardized integration with wide range of enterprise systems
Vendor Affiliation	Closed source. Workflow and source tied to ServiceNow	Solution can extend beyond one platform and can scale better

Thanks! Questions / Details/ Demo?

Contact the author - Vijay Cinnakonda - vijay.cinnakonda@va.gov /vijaycinn@gmail.com