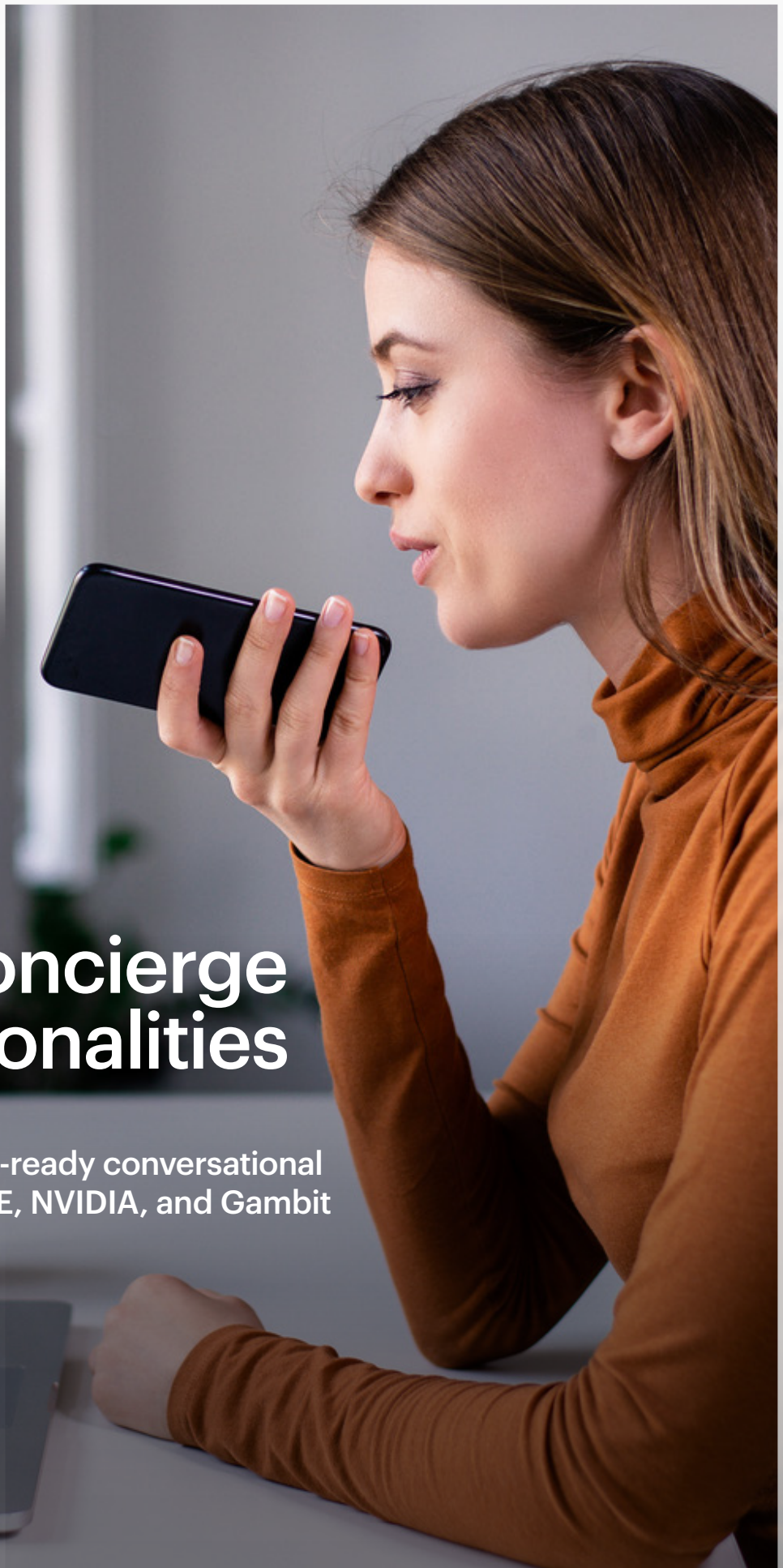
Solution brief



AI concierge personalities

Production-ready conversational AI from HPE, NVIDIA, and Gambit

Market dynamics: From chatbots to trusted AI personalities

Enterprises and public sector organizations are rapidly adopting conversational AI—yet most deployments remain transactional, scripted, or disconnected from real operational systems.

Traditional chatbots and virtual assistants often fail because they:

— Lack real-time integration with enterprise data
— Cannot sustain multi-turn, contextual dialogue
— Generate inconsistent responses across channels
— Struggle to scale securely in regulated environments
— Require unpredictable public cloud spending for high concurrency

At the same time, organizations face growing demand for always-on, 24x7 engagement without expanding staff, along with seamless multilingual support and real-time decision assistance. They are expected to deliver personalized experiences that feel natural and intuitive while ensuring AI systems remain tightly governed, secure, and fully controlled within enterprise and public sector environments.

The shift is clear: Customers no longer want generic bots. They want trusted, domain-specific AI personalities that operate reliably in production environments.

Solution overview

Gambit delivers AI-powered voice concierge personalities—domain-trained, guardrailed AI agents that engage users through voice and text in a natural, emotionally intelligent manner.

Deployed on HPE powered by NVIDIA accelerated computing, these AI personalities combine:

— Real-time data integration
— Structured reasoning and decision logic
— Secure large language model (LLM) inference
— Multichannel delivery (voice + SMS)
— Analytics and governance

Unlike generic assistants, Gambit's personalities are purpose-built for specific domains—healthcare, tourism, logistics, aging services, and more—with carefully designed tone, scope, and behavioral guardrails.

HPE provides the secure, scalable on-premises compute backbone required to support concurrent voice sessions, LLM workloads, and enterprise-grade availability. NVIDIA accelerated computing enables high-performance inference at scale.

The Unleash AI advantage

Developed through the HPE Unleash AI program, this solution represents a validated, solution-led deployment model—enabling customers to move from AI experimentation to operational, production-grade conversational systems.

Unleash AI delivers:

— Jointly integrated and validated solutions from top ISV and AI innovators that reduce integration risk and accelerate AI deployments
— Scalable AI solutions on proven HPE infrastructure that are production-ready across on-premises, hybrid, and edge environments
— A consistent platform with enterprise-grade performance and reliability for repeatable, scalable AI architectures across more than 75 validated use cases

Key business outcomes

Organizations deploying AI concierge personalities from HPE, NVIDIA, and Gambit can expect measurable operational and experience gains.

Operational efficiency: Reduce time and labor costs

— Reduced inbound call volume to service centers
— Automated triage of routine inquiries
— 24x7 engagement with no staffing increase
— Structured conversational routing based on user intent

Optimized experience and engagement: Increase revenue and retention

— Multilingual support across 50+ languages
— Consistent, branded personality-driven interactions
— Increased engagement rates compared to static digital channels
— Higher trust and repeat interaction through personality continuity

Scalable AI deployment: Control infrastructure costs and optimize ROI

— Concurrent voice session support
— High availability for public-facing services
— Secure deployment environments aligned to compliance requirements
— Predictable infrastructure costs compared to public cloud–only approaches

Architecture and technical capabilities

The combination of Gambit AI concierge personalities and enterprise-grade HPE AI infrastructure, accelerated by NVIDIA, enables always-on AI voice systems that operate reliably in regulated and mission-critical environments.

AI concierge personalities can be deployed on-premises, as part of a hybrid model, or at the edge. Supported platforms include:

- HPE Private Cloud AI, co-developed with NVIDIA as part of the NVIDIA AI Computing by HPE portfolio. A turnkey enterprise AI factory solution, HPE Private Cloud AI unites NVIDIA accelerated computing, NVIDIA networking, NVIDIA AI Enterprise software, and services to simplify AI deployment and management.
- HPE ProLiant Compute servers with NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPUs for a secure, optimized, and automated compute foundation purpose-built for enterprise AI workloads.
- GambitOS adds an enterprise AI orchestration layer that connects LLM reasoning, real-time data feeds, structured decision logic, multichannel delivery, and analytics and monitoring.

The platform is designed for enterprise-scale performance, supporting high levels of concurrent voice and conversational sessions while maintaining consistent response quality. It integrates with real-time data sources to deliver up-to-date information dynamically, enabling intelligent, context-aware interactions.

Built on a high-availability architecture, the system operates continuously to support mission-critical, customer-facing use cases. GPU-accelerated LLM inference provides low-latency responses and scalable performance across production environments.

Full stack architecture

Experience layer: GambitOS	AI and orchestration layer: GambitOS	Infrastructure layer: HPE and NVIDIA
— Voice interface (phone-based AI concierge) — SMS delivery of links, directions, and resources — Language auto-detection and switching	— Domain-trained AI personality models — Guardrails for scope, tone, and geography — Decision logic routing — Real-time API integrations	— HPE Private Cloud AI with NVIDIA accelerated computing and NVIDIA software: • NVIDIA® L40S GPUs for efficient scaling. • NVIDIA Inferencing Microservices (NIMs) for secure, reliable deployment of high-performance AI inferencing models • NVIDIA Triton™ Inference Server to deploy any AI model from multiple deep learning and machine learning frameworks • NVIDIA AI Enterprise to deliver optimized performance, robust security, and stability for production AI deployments or HPE ProLiant Compute infrastructure with NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs — Secure, compliant deployment environments — Enterprise observability and resource monitoring — High-availability architecture

Key use cases

Municipalities/smart cities:
Chloe for the Town of Vail

Challenge: High parking call volumes, congestion, multilingual visitor needs, limited staff

Outcome: 24x7 AI voice concierge delivers real-time parking guidance, event promotion, and congestion reduction through sensor-integrated routing.

Healthcare:
AskEllyn for breast cancer support

Challenge: Patients navigating complex emotional and informational journeys with limited access to consistent, trusted support

Outcome: AI companion delivers empathetic, domain-specific engagement for breast cancer patients. Licensed by a major U.S. insurer for digital patient engagement.

Hospitality and sports:
AskAmber for NASCAR

Challenge: Engaging large fan bases in real time during live events

Outcome: AI fan engagement platform launched at the Daytona 500, drove a 36x engagement spike on race day.

Supply chain:
Harlo for a logistics alliance

Challenge: Manual coordination of inbound freight across fragmented systems

Outcome: Enterprise logistics AI operating system automates inbound freight workflows, delivering time and cost efficiencies.

Government and aging:
AskAshleigh for elderly caregiver assistance

Challenge: Caregiver support across distributed area agencies on aging

Outcome: AI support tool scalable to 600+ agencies nationally delivers structured guidance and 24x7 accessibility for caregivers.

Learn more at

HPE Unleash AI program

HPE Private Cloud AI

NVIDIA AI Computing by HPE

HPE ProLiant Compute

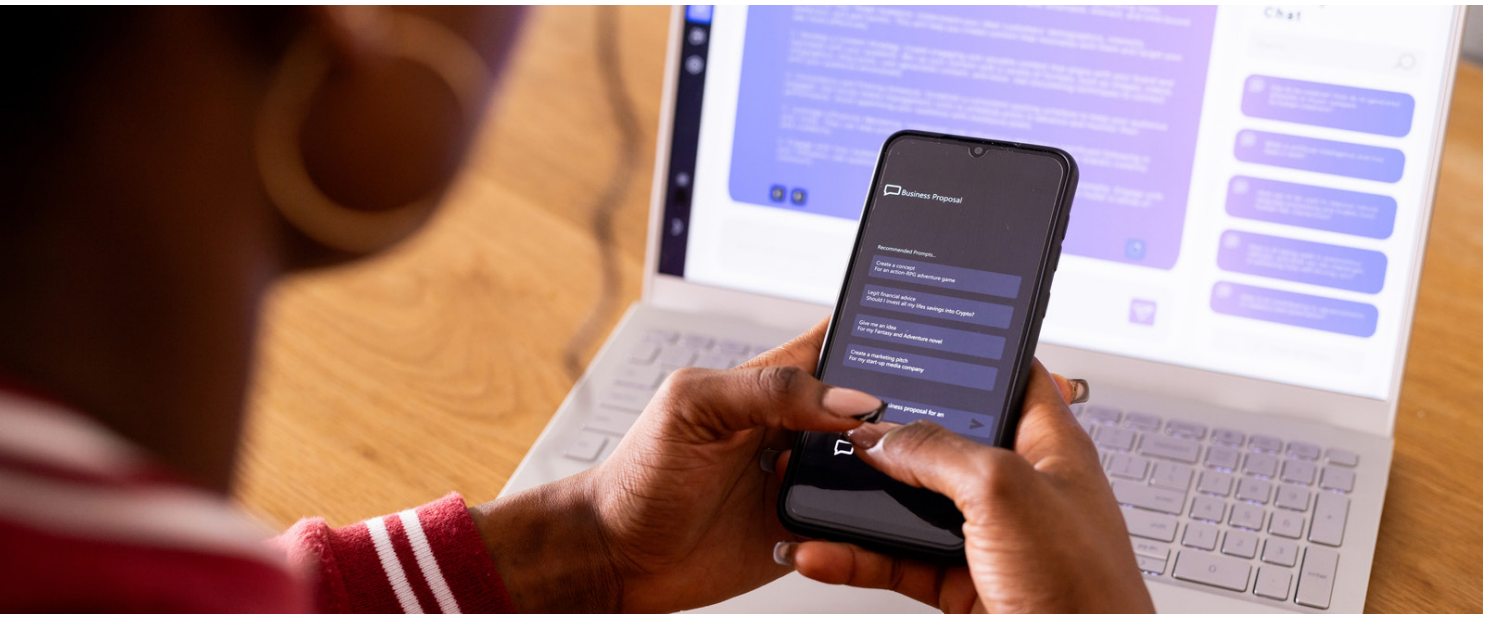
Visit HPE.com

Why choose the AI concierge personality-validated Unleash AI solution?

More than a conversational interface, this solution is a production-ready, enterprise-validated AI system designed for real-world outcomes. Through the Unleash AI program, you gain a jointly integrated and validated solution from HPE, NVIDIA, and Gambit, reducing integration risk and accelerating deployment timelines. Rather than stitching together infrastructure, models, and orchestration layers independently, you benefit from a tested, repeatable architecture built for performance, governance, and scale.

HPE provides secure on-premises and hybrid infrastructure with predictable economics and enterprise observability. NVIDIA delivers accelerated computing and optimized inference for high-performance AI workloads. Gambit contributes domain-trained AI personalities with structured guardrails and production deployment expertise across healthcare, government, logistics, tourism, and enterprise environments.

The result is AI that is not experimental, but operational—not transactional, but trusted. And not siloed, but architected for long-term scale and real-world impact.



Chat now

a00158268ENW

HEWLETT PACKARD ENTERPRISE

hpe.com