

CASE STUDY

Supporting Realtime AI Avatars & Voice Agents in Janus based WebRTC application



A customer had already built and deployed their real-time communication platform around Janus as the WebRTC media server and wanted to introduce conversational AI capabilities, including a voice agent that could join sessions, listen, respond naturally, and interact through a photorealistic or animated avatar. However, available AI services did not support direct integration with Janus-based WebRTC platforms, meaning the customer would need to replace or significantly rework their existing media infrastructure to adopt modern AI agent capabilities. This would create unnecessary migration effort, architectural complexity, and potential platform lock-in for a system already running successfully in production. The customer approached SpringCT to bring conversational AI and digital avatars to their existing Janus deployment without re-platforming the entire WebRTC infrastructure. SpringCT evaluated the architecture and applied its reusable Avatar Framework to integrate conversational AI and avatar capabilities directly into the customer's Janus-based platform. The framework provides session management, conversational turn-taking, and a provider-plugin architecture for integrating AI, speech, and avatar services, allowing SpringCT to adapt the framework to the customer's specific requirements while preserving the existing media infrastructure already deployed in production.

Product Features of Avatar Framework

- **Pluggable Provider Architecture**
Allow swapping provider for speech-to-text, LLM, text-to-speech, VAD, or avatar with a one-line config change. Ships with Deepgram, OpenAI, ElevenLabs, Cartesia, Silero, Anam, Tavus, and a lightweight GIF-based audio visualizer.
- **Auto-Join Dispatch**
Fully integrated with Janus event-handler system, joins any newly created room with no manual invitation step.
- **Full Turn-Taking Engine**
Supports dual VAD + STT endpoint, preemptive LLM generation, and configurable barge-in (VAD, STT, or adaptive mode) with noise-filtering thresholds so a stray sound doesn't cut the agent off.
- **Realtime Media Without a Managed Platform**
Encapsulates the full WebRTC media path — Janus SFU signalling, SDP/ICE negotiation, RTP encode and decode.
- **Multi-Participant Rooms**
Supports multiple human to avatar interactions.
- **Hosted & Local Avatars**
Anam and Tavus photoreal avatars, or a lightweight GIF audio-visualizer for deployments without a hosted-avatar budget — auto-sized to the source asset's own aspect ratio so nothing gets letterboxed.

Technologies Used

- **Python / asyncio:** The framework's core runtime and concurrency model.
- **aiortc:** WebRTC peer connections, SDP/ICE negotiation, and RTP media encode/decode.
- **Janus Gateway (videoroom plugin):** The SFU and signalling backbone the framework connects to directly over janus-protocol, with no intermediary platform.
- **Deepgram, OpenAI, ElevenLabs, Cartesia:** Pluggable speech-to-text, LLM, and text-to-speech providers.
- **Silero VAD (ONNX Runtime):** Inline voice-activity detection, fast enough to run without a thread hop.
- **Anam, Tavus:** Hosted photoreal avatar providers integrated as first-class avatar plugins.
- **aiohttp:** The auto-join dispatcher's HTTP server and every provider's WebSocket/HTTP client.
- **React + Vite:** The reference browser client used to develop and verify every feature live.

Results

SpringCT's implementation of the Avatar Framework enabled the customer to successfully introduce conversational AI and digital avatar capabilities into their existing Janus platform while retaining their production WebRTC infrastructure.

- **Rapid, Zero-Touch Room Participation:** The AI agent can automatically join newly created Janus rooms in approximately 0.3 seconds, without requiring browser-side invitations or additional participant setup.
- **3–4x Faster Spoken Responses:** By eliminating hidden TTS reconnection overhead, SpringCT reduced median voice response latency from approximately 1050ms to under 300ms, enabling a significantly more natural conversational experience.
- **Reliable Multi-Participant Conversations:** Improved turn-taking and hand-off handling allows the agent to immediately respond to a participant's live speech, rather than processing stale audio that may have accumulated while the participant was inactive.
- **Flexible AI and Avatar Integration:** The reusable provider-plugin architecture allowed the customer to integrate the AI, speech, and avatar services required for their application without changing the underlying Janus media infrastructure.
- **Production-Ready Foundation for Future Applications:** The Avatar Framework provides the customer with a reusable foundation for building additional AI and avatar-driven applications, without having to solve real-time media handling, session management, and conversational turn-taking again for each new use case.

Conclusion

By applying its reusable Avatar Framework to the customer's existing Janus deployment, SpringCT enabled the customer to add modern conversational AI and digital avatar capabilities without replacing or re-platforming their WebRTC media infrastructure.

The solution allowed the customer to preserve their existing investment in Janus while gaining a flexible and reusable AI agent foundation. With real-time media handling, session management, conversational turn-taking, and AI provider integration handled by the SpringCT framework, the customer's application can focus on its specific conversational and business requirements rather than rebuilding the underlying AI and media infrastructure.