



CASE STUDY

replicant

Replicant is a Conversational AI technology that works out of the box to solve customer problems over the phone. The company was founded on the belief that machines are ready to have useful, complex conversations that will transform the way they interact with the world.

Founded
2017

Headquarters
San Francisco, CA

Replicant uses state-of-the-art voice technology and machine learning to solve customer problems, collecting knowledge from each interaction to become smarter over time. With Telnyx, Replicant delivers authentic conversations driven by sophisticated real-time analytics.

Why every millisecond counts

To create a realistic conversational AI experience, Replicant demands consistently low “talk time” latency (the time from when the user stops speaking to when the bot responds). Acceptable latency for their AI is 1 second or less.

But achieving this low latency was a challenge, as most providers could only reduce latency to three seconds.

Other AI platforms mask this kind of delayed response with “human” noises (e.g., typing sounds or audible “um” sounds), but Replicant wanted to be honest and upfront with their customers.

“People notice when you’re trying to cover up latency, and the call turns into a sluggish lecture instead of a conversation,” said Benjamin Gleitzman, Replicant CTO and co-founder. “It’s very important that we’re genuine and honest with our customers. It’s one of the reasons we focus on latency and engineer our systems to save every millisecond we can. **Because we’re able to get latency down to the level we want with Telnyx, we don’t have to bend the truth with our customers like some of our competitors.**”

Replicant has seen a huge reduction in latency levels with Telnyx, moving from 3 seconds plus to under 1 second.

Simplifying call flows and enhancing the customer experience with media forking

With their previous provider, Replicant had what they call an “ugly spider” of call flows. “We had all these different legs coming in, and there was so much that could go wrong,” said Gleitzman.

But with Telnyx as their carrier, Replicant has the ability to fork call media to multiple recipients in real time. “Media forking makes our developers’ lives easier because they’re not trying to manage about ten different call legs. Telnyx brings us far greater simplicity.”

Media forking also gives Replicant a substantial competitive advantage over other platforms by unlocking voice analytics, voice activity detection and mood detection — all in real time. As a result, the Replicant Thinking Machine gets smarter — and customer service gets better — by the second.

Money matters

With every business decision, price is always an important factor. By switching to Telnyx, Replicant reduced voice costs from 7 cents to 1 cent per minute — that’s an 86% reduction. The substantial cost difference means Replicant can offer more competitive pricing and scale efficiently without having to worry about spiraling costs.

Gaining real-time, interactive command and control

To create their conversational AI platform, Replicant required sophisticated features and granular control over their call flows. They chose to leverage the Telnyx Call Control API, which enables simultaneous bidirectional data transmission for an interactive, dynamic experience.

With the ability to play and gather tones, layer background audio and stop foreground and background music, Replicant can now deliver an authentic, customized conversational experience to meet each user’s unique needs.



Replicant lowers latency, costs and gains real-time media forking with Telnyx

The Results

3s+ to 1s

reduced latency between conversation turns

86%

cost reduction

Voice Analytics

Media forking to enable sophisticated voice analytics

Telnyx Products:

- Call Control
- Virtual Cross Connects
- Media Forking