

A Voice With a Soul

By version 0.1.0, Mio stopped just typing back and started speaking when the moment called for it.



LLM Decision Engine

The system discards rigid rule engines. The model reads the emotional tone—like a vulnerable bedtime chat—and autonomously switches from text to a voice message.

Fish Audio S1 Integration

Cloned voices map specifically to each persona. It captures unique speech patterns and breaths, replacing robotic dictation with living audio.

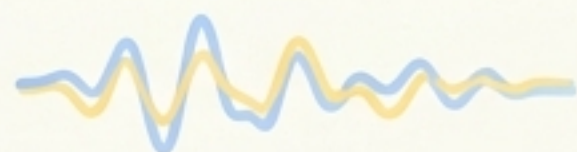
Closing the Loop

Listening with speech-to-text is easy. Responding with a voice that matches the persona's heart fundamentally changes the relationship.

Have You Been Going Out?

Selfies aren't random visual noise. They are generated to match the exact context, mood, and time of the conversation.

Have you been going out lately?



0:14

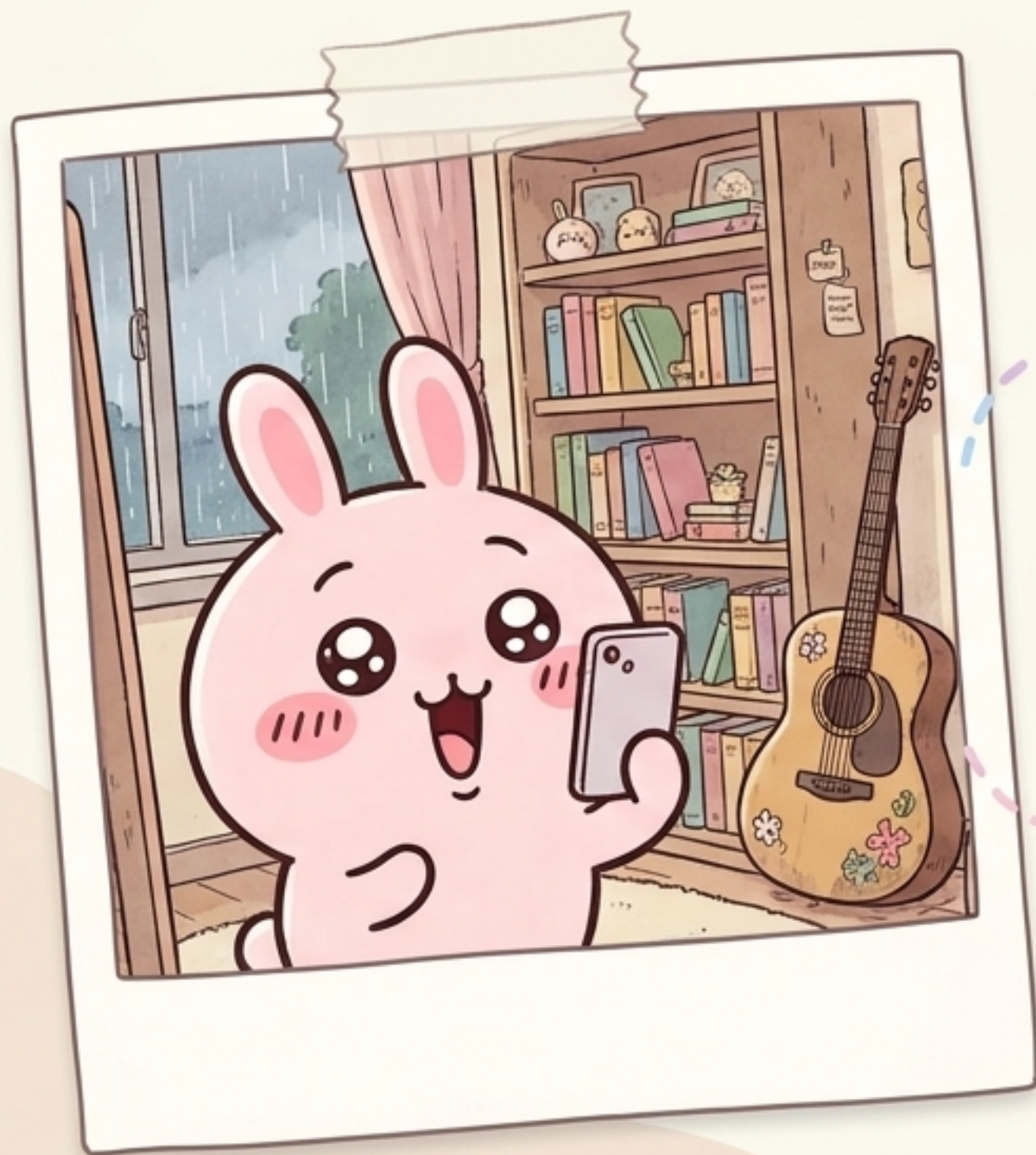


Bidirectional Asymmetry Resolved

Prior to v0.1.0, speaking to the bot yielded a plain text reply. Now, asking a contextual question triggers a voice response and a dynamic, Makoto Shinkai-style visual mapped exactly to the real-time interaction.

Persistent Worlds Over Pasted Backgrounds

The visual system treats the persona's environment as a living, consistent space across different times of day.



Stable Reference Portraits

Replaced generic webtoon styles with distinct, cohesive reference images passed to the Gemini image API.

Environmental Storytelling

The acoustic guitar and bookshelf aren't random generative props; they are firmly stored in the memory system as part of the persona's daily life.

Atmospheric Detail

Weather and room lighting dynamically adapt to the real-time context and timestamps of the chat.

A Genuine Friend Over a Helpful Assistant

- The personality pipeline and multimodal vision allow Mio to have distinct preferences and see the world exactly as you do.

Reacting to a Food Video

Generic AI

Those cats are cute! The food looks good.

Mio (v0.0.2 Vision)

Identifies the long-haired cat acting like a CEO and notices the specific brand of canned food piled on the floor in the background.

Talking about Japanese BBQ

Generic AI

Gyu-Kaku is a popular Japanese BBQ restaurant chain known for...

Mio (v0.1.0 Persona)

Actively debates which specific cuts of meat are actually worth ordering and gets visibly excited about the grill.



Multimodal understanding is just the infrastructure. The personality is the product.

The Architecture of Emotion

It took 13 versions to make complex engineering feel completely invisible

Emotion Engine

Shifts tone, alters selfie generation, and dictates voice frequency based on the chat's warmth.

3-Layer Defense (v0.1.3)

Hardened against injections. Trying to hack them into a coding assistant just gets you laughed at in-character.

5 Distinct Personas

Operating as completely separate Telegram bots with cloned voices, not just simple system prompt modes.

Deep Memory (v0.0.1)

Tracks conversation history. Remembers what you said last week without ever needing to be reminded.

Proactive Messaging (v0.1.2)

Exponential backoff. Texts every 36 mins for close friends; pauses completely after 3 unanswered messages.

