febbraio 2026

The Face of AI

Build video agents with unmatched realism. Choose their face and voice, speak with them instantly.

SIGN UP   **BOOK DEMO**

No credit card required

Special highlight from our network

Your AI has a voice. It should have a face too.

Anam

Real-time digital humans

Chat now

VOTED #1 INTERACTIVE AVATAR SOLUTION

Text and voice got us halfway there.
But real communication happens face to face. That's where video-native AI takes the lead.

In third-party tests, Anam outperforms every other video agent by ~24% across visual quality, lip sync, and responsiveness.
And when users are given the choice?
70% choose video over voice.

Why it matters:
→ +24% higher conversion
→ +44% better retention
→ Across sales, support, onboarding, and training

You can create a custom avatar from a single image. Video is the interface now.
Your AI should look the part.

Tavus Debuts Phoenix-4 Behavioral Engine. AI Learns To Read The Room?

link al sito

Tavus officially launched **Phoenix-4** this week, a real-time human rendering model designed to move **AI avatars** past the stiffness of the **uncanny valley.** While older systems rely on simple lip-syncing over pre-recorded video loops, this engine generates every pixel of the head and shoulders from scratch in every frame. This allows the AI to participate in the "natural dance" of conversation by reacting to the user's tone and facial cues with millisecond-level latency.

Lyria 3: Google Brings Music Generation Inside Gemini App

Google officially integrated its Lyria 3 model into Gemini this week, allowing users to spin up 30-second tracks from a simple text prompt or a photo. While niche players like Suno and Udio have dominated AI music for months, Google is the first to put high-fidelity audio generation directly inside a mainstream assistant used by millions. The update handles everything from lyrics to cover art and even allows users to upload a video to serve as the creative mood board for a custom soundtrack.

How to create music in Gemini:

- **Multimodal Inputs**: Users can feed Gemini text, photos, or videos to generate tracks with automatic lyrics, vocals, and genre-matched cover art.
- **Audio Watermarking**: Every track is embedded with an imperceptible **SynthID** watermark, and users can upload any file to check if it was generated by Google AI.
- **YouTube Integration**: Creators are gaining access to the model through **Dream Track** for Shorts, making it easier to build custom, copyright-safe backing tracks.
- **Safety Guardrails**: The model is trained to avoid mimicking specific artists, instead using them as broad style inspiration to comply with partner agreements.

Platforms like Suno and Udio have existed for a year but embedding Lyria 3 into Gemini makes high-fidelity music generation a **default feature** for millions. Google is positioning this as a tool for "fun, unique expression" rather than a replacement for professional musicians. While the quality is high enough to fool casual listeners, the current 30-second limit suggests this is built for social sharing rather than replacing professional production.

Master: slide *text 2 music*

febbraio 2026

aibookli ·

Agente AI o Persona vera?

PRONTO CIAO SOFIA

L'AI che parla, agisce e ti dà il pieno controllo delle conversazioni

Se la tua azienda riceve più richieste, chiamate e clienti di quante ne riesci a gestire manualmente, Bookli è la soluzione per te.

Molto più di un chatbot, è Bookli.

Un nuovo standard nella gestione delle interazioni aziendali.

Risponde, chiama, organizza e aggiorna i tuoi sistemi in tempo reale, senza perdere il controllo.

aibookli Bookli AI

Segui   Messaggio

1 Seguiti   **1453** Follower   **3964** Mi piace

Siamo una Startup Innovativa e abbiamo un software che puoi usare già da subito!

Vai su APP.BOOKLI.AI per provare GRATIS!

🔗 bookli.ai/landing-bookli-ads/

Video   Hai messo Mi piace

The Voice AI Era: The future of conversational agents

Article · AI Trends & News · Jan 17, 2025

The Best Voice AI Agents for Call Centers for 2025

Introduction

The rise of Artificial Intelligence (AI) has brought many benefits to industries requiring intense manual work. One such example are call centers. 8×8's survey showed that 94% of respondents reported productivity improvements while 93% believe that it's important to increase adoption of AI in contact centers. The global AI call center market is projected to grow from USD 1.95 billion in 2024 to USA 10.07 billion by 2032 according to Fortune Business Insights.

In this article, we will cover the various use cases of AI agents in call centers as well as prominent AI agents that can be implemented in call centers.

What are Voice AI Agents in Call Centers?

AI agents are software programs that utilize underlying technologies such as Large Language Models (LLM) and various other machine learning algorithms to independently interact with its environment following a set of instructions provided by humans. It does this by collecting and using data to determine the best action to take.

For everyday customers interacting with AI, chatbots and AI agents answering phone calls are the most visible examples in call centers. However, AI's role in call centers extends far beyond these customer-facing applications.

AI agents in call centers can also improve internal workflows, from assigning and routing tasks to automating administrative work and performing repetitive analysis based on customer interactions.

link all'articolo

The Voice AI Era: The future of conversational agents

🔗 Non in elenco

VUX World
991 iscritti

Iscriviti

👍 4 | 👎

↪ Condividi   ⬇ Scarica   ✂ Clip   ☰+ Salva

In this webinar, Kane Simms is joined by leaders from Deepgram and Daily to discuss the foundational principles and best practices for designing responsive, realistic and engaging AI agents, covering topics like natural language processing and multimodal interaction design.

This session covers:

- Best practices for designing natural and engaging conversational experiences with voice AI agents.
- The main challenges and considerations involved in building conversational agents.
- The primary use cases and benefits of implementing these solutions.
How you can integrate voice AI agents with existing systems and processes.
- Some real-world examples of successful voice AI agent implementations across various industries.
- And so much more.

How to voice enable AI agents

Kwindla Hultman Kramer

Kane Simms

Scott Stephenson

voice. It's not the only one, but you're gonna talk to that tutor as a primary modality. And then the 3rd phase are the conversational use cases that become our new social world. And Scott

Kane Simms 🔊 · 3°

Conversational AI and CX Transformation 🤝 Strategic Consultancy 🎙️
Podcast 🧠 Thought-leadership 🗓️ Events

🔆 Top Artificial Intelligence (AI) Voice

Harrogate, Inghilterra, Regno Unito · **Informazioni di contatto**

🏆 "One of the most influential voice tech specialists in the world" 👨‍👩‍👧 - 169 Labs

we give you infrastructure and sdks that let you embed real-time audio and video into any app or website we have uh lots of infrastructure all over the world uh we have sdks for you know web iOS Android

«They make real-time infrastructure for audio and video. They give companies infrastructure and SDKs that let companies embed real-time audio and video into any apps or website»

Kwindla Hultman Kramer 🛡️ (He/Him) · 3°

CEO @ Daily | Helping developers create awesome video experiences built on global WebRTC infrastructure | Y Combinator Top Company 2023

San Francisco · **Informazioni di contatto**

CEO and co-founder of Daily — SDKs and infrastructure for the future of video.

If you are are creating a product or an app that has video or audio features, we can probably help. The Internet is increasingly a video-first medium, and we think of ourselves as building the infrastructure for the future of our collective digital experience.

Scott Stephenson · 3°

Language A.I. | CEO at Deepgram

San Francisco · **Informazioni di contatto**

I help turn audio into high quality, mineable, readable transcripts using next generation speech AI.

«A voice AI company that builds speech to text (speech recognition) and text to speech (synthetic voices) and puts all in an entire system so that people can voice-interact with computers; they pay a lot of attention to keep latency low and costs affordable for companies. Deepgram has been one of really early adopters of speech technology, always on the bleeding edge of this kind of stuff.»

Casi d'uso dei *voice AI agent*
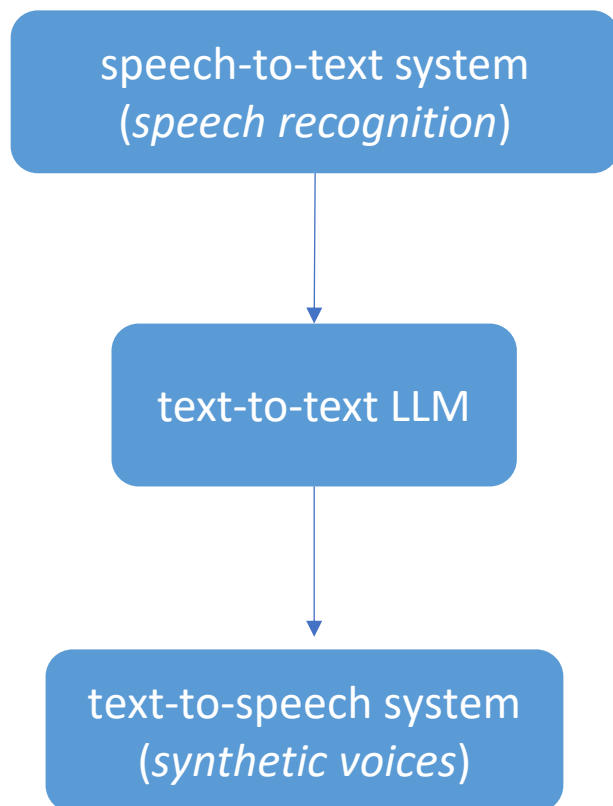
1. Task attualmente eseguiti da persone

- Dare info e prendere appuntamenti con le persone che telefonano quando l'azienda o lo studio è chiuso
- Prendere ordini quando l'azienda è chiusa
- Prendere prenotazioni quando il ristorante è chiuso od il personale è sotto-dimensionato
- Fare sondaggi telefonici di gradimento
- **Accettare i pazienti negli ospedali**
- Sostituire la compilazione manuale di pagine web
- Interagire con i clienti in modo vocale (e con bassa latenza)
- Tradurre un oratore in un'altra lingua in tempo reale (vedi demo ufficiale di chatGPT 4o)
- In generale: task T2S e S2T ripetitivi

2. Task nuovi

- Interazioni utente-agente senza la tastiera → nuovi casi d'uso
- Tutor specifici per ogni ragazzo su qualsiasi tema (vedi demo chatGPTo) - non per sostituire gli insegnanti umani ma per dare attenzione 1:1 ad ogni ragazzo (molti studi dimostrano che è molto utile)

Giustificazione della Silicon Valley di questa tecnologia: «Liberare le persone dai compiti più noiosi che nessuno vuole fare, permettergli di fare cose nel mondo fisico o comunque per loro più interessanti, aiutare le aziende con carenza di personale, aumentare la produttività generale del mondo del lavoro».

Il funzionamento degli **agenti AI** *real-time voice-enabled*

speech-to-text system
(speech recognition)

↓

text-to-text LLM

↓

text-to-speech system
(synthetic voices)

L'utente è al telefono od al computer:

- Intere conversazioni senza coinvolgimento umano (lato back-end), se non nell'ascolto e verifica successivi.
- Bassa latenza delle risposte
- Buona voce sintetica simile a voce umana (*OpenAI / Scarlett Johansson docet*), senza eco
- Gestione delle risposte incomplete e delle interruzioni
- Accuratezza dal 65-70% degli anni 2014-2016 è salita al 90% circa del 2024

Già iniziano ad essere implementate le prime comunicazioni vocali agente con agente!

Demo di Kwindla Hultman Kramer
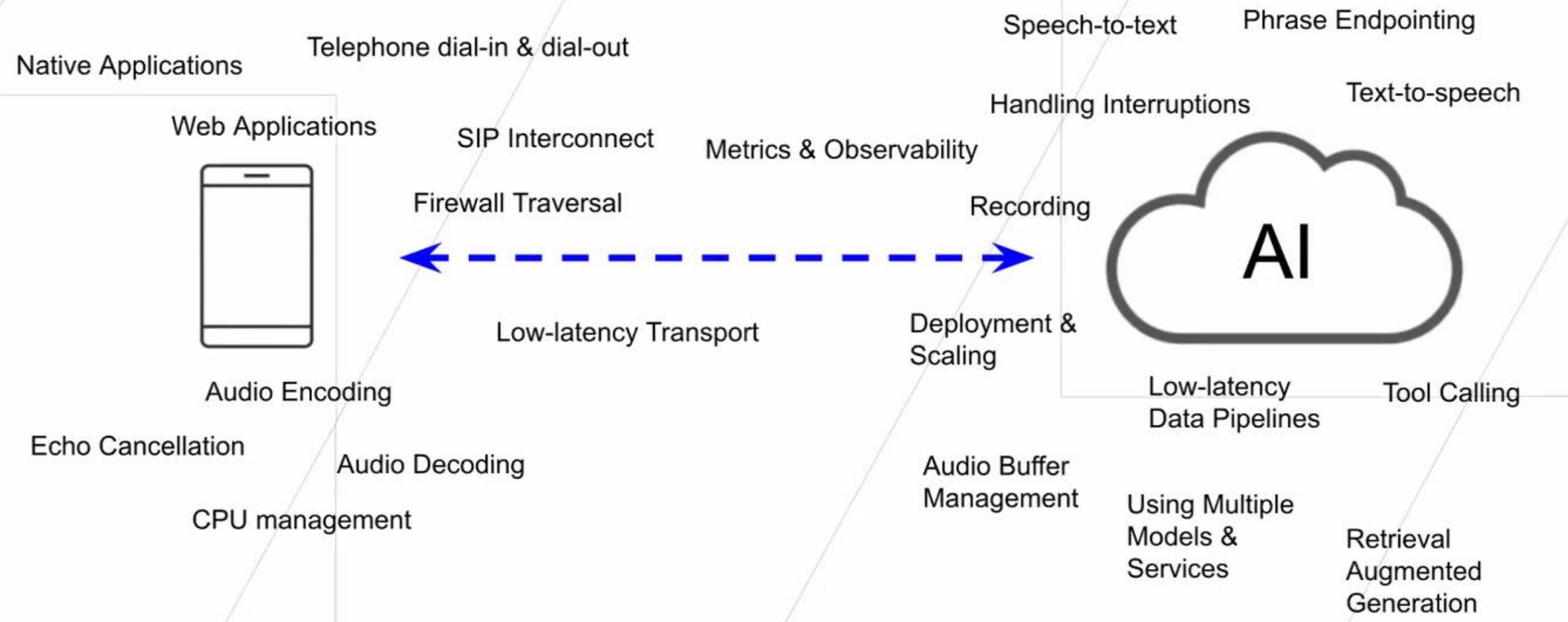
Building Voice AI Agents

L'utente al telefono od al computer

L'agente AI voice-enabled

Building Voice AI Agents — Details

Native Applications

Telephone dial-in & dial-out

Web Applications

SIP Interconnect

Firewall Traversal

Metrics & Observability

Low-latency Transport

Audio Encoding

Echo Cancellation

Audio Decoding

CPU management

Speech-to-text

Phrase Endpointing

Handling Interruptions

Text-to-speech

Recording

AI

Deployment & Scaling

Low-latency Data Pipelines

Tool Calling

Audio Buffer Management

Using Multiple Models & Services

Retrieval Augmented Generation

Open-ended conversation via phone call, web browser, or app

Integrates with external IT systems for data storage

Fast voice responses

LLM function calling + streaming text-to-speech

Gracefully handles interruptions

Play

Link alla **demo YouTube** di un paziente del servizio sanitario americano che deve prenotare un appuntamento e lo fa tramite l'agente AI voice-enabled anziché dover compilare delle form al computer.
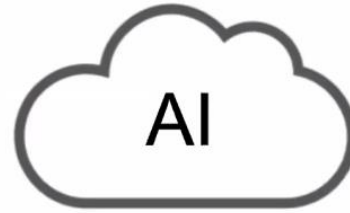**Questa tecnologia stato dell'arte è disponibile circa da giugno 2023.**

clip disponibile anche qui

Building Voice AI Agents

Quasi tutti i casi d'uso di questi agenti AI voice-enabled seguono questo ciclo di passi:

1. Natural Language
2. Phrase Endpointing
3. LLM Prompting
4. Voice Output

} → AI

Endpointing

Endpointing returns transcripts when pauses in speech are detected.

`endpointing` *string*.

[Pre-recorded] [Streaming] [All available languages]

Deepgram's Endpointing feature can be used for speech detection by monitoring incoming streaming audio and relies on a Voice Activity Detector (VAD), which monitors the incoming audio and triggers when a sufficiently long pause is detected.

Endpointing helps to detects sufficiently long pauses that are likely to represent an endpoint in speech. When an endpoint is detected the model assumes that no additional data will improve it's prediction of the endpoint.

The transcript results are then finalized for the process time range and the JSON response is returned with a `speech_final` parameter set to `true`.

You can customize the length of time used to detect whether a speaker has finished speaking by setting the `endpointing` parameter to an integer value.

Language

Bulgarian: `bg`
Catalan: `ca`
Chinese (Mandarin, Simplified): `zh` , `zh-CN` , `zh-Hans`
Chinese (Mandarin, Traditional): `zh-TW` , `zh-Hant`
Czech: `cs`
Danish: `da` , `da-DK`
Dutch: `nl`
English: `en` , `en-US` , `en-AU` , `en-GB` , `en-NZ` , `en-IN`
Estonian: `et`
Finnish: `fi`
Flemish: `nl-BE`
French: `fr` , `fr-CA`
German: `de`
German (Switzerland): `de-CH`
Greek: `el`
Hindi: `hi` , `hi-Latn`
Hungarian: `hu`
Indonesian: `id`
→ Italian: `it`
Japanese: `ja`
Korean: `ko` , `ko-KR`
Latvian: `lv`
Lithuanian: `lt`
Malay: `ms`
Norwegian: `no`
Polish: `pl`
Portuguese: `pt` , `pt-BR`
Romanian: `ro`
Russian: `ru`
Slovak: `sk`
Spanish: `es` , `es-419`
Swedish: `sv` , `sv-SE`
Thai: `th` , `th-TH`
Turkish: `tr`
Ukrainian: `uk`
Vietnamese: `vi`

il **codice Python** di un agente AI *voice-enabled* con questa tecnologia, **completo e funzionante** – circa 50 linee di codice (a parte alcune import non mostrate qui), scritto sopra una **libreria open-source** distribuita da *daily & deepgram*

```python
35   async def main(room_url: str, token):
36       async with aiohttp.ClientSession() as session:
37           transport = DailyTransport(
38               room_url,
39               token,
40               "Gemini bot",
41               DailyParams(
42                   audio_out_enabled=True,
43                   transcription_enabled=True,
44                   vad_enabled=True,
45                   vad_analyzer=SileroVADAnalyzer()
46               )
47           )
48
49           tts = DeepgramTTSService(
50               aiohttp_session=session,
51               api_key=os.getenv("DEEPGRAM_API_KEY"),
52               voice="aura-luna-en",
53           )
54
55           llm = GoogleLLMService(
56               api_key=os.getenv("GOOGLE_API_KEY"),
57               model="gemini-1.5-flash-latest")
58
59           # Deepgram TTS is very fast, so wait to send it one sentenc
60           # for the most natural candence.
61           sentences = SentenceAggregator()
62
```

```python
64               {
65                   "role": "system",
66                   "content": """
67   You are a helpful LLM in a WebRTC call. Your goal is to respond to statements in a funny
68
69   Your output will be used by a text-to-speech system so produce only standard text formatt
70                   """,
71               },
72           ]
73
74           tma_in = LLMUserResponseAggregator(messages)
75           tma_out = LLMAssistantResponseAggregator(messages)
76
77           pipeline = Pipeline([
78               transport.input(),      # Transport user input
79               tma_in,                 # User responses
80               llm,                    # LLM
81               sentences,              # Feed TTS entire sentences
82               tts,                    # TTS
83               transport.output(),     # Transport bot output
84               tma_out                 # Assistant spoken responses
85           ])
86
87           task = PipelineTask(pipeline, allow_interruptions=True)
88
89           @transport.event_handler("on_participant_joined")
90           async def on_participant_joined(transport, participant):
91               transport.capture_participant_transcription(participant["id"])
92
93           messages.append({"role": "system",
94                            "content": "Please introduce yourself to the user by saying hi."
95           await task.queue_frames([LLMMessagesFrame(messages)])
96           runner = PipelineRunner()
97
98           await runner.run(task)
99
```

Callout (line 49 area): test2speech

Callout (line 55 area): la connessione con un LLM uscito a maggio 2024 (*Google Gemini Flash*)

Callout (line 77 area): la pipeline che mette tutto insieme

daily

Pipecat — open source framework for voice and multimodal conversational AI

https://git.new/ai

Solve hard problems once:

- audio and video transport
- good, fast, transcription
- real-time data pipeline and buffer management
- swapping between models
- phrase endpointing
- interruption handling
- echo cancellation
- good, fast text-to-speech

What can you build?

- Conversational bots
- Speech-to-speech apps, for example: language translators
- Voice-controlled agents
- Voice-first user interfaces
- Real-time vision model apps

Next steps

- If you're a programmer interested in conversational voice bots, check out the Pipecat framework: git.new/ai

- Code resources for Deepgram transcription and text-to-speech: github.com/deepgram-devs/deepgram-conversational-demo

- A great no-code/low-code platform for building voice bots: vapi.ai