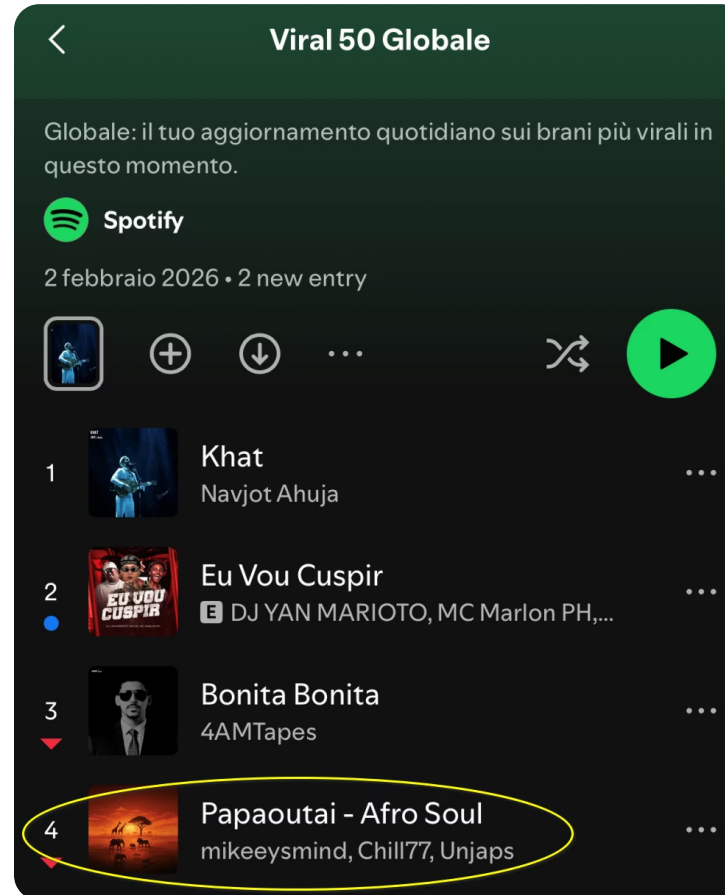


Audio e Intelligenza Artificiale

Come le macchine stanno imparando a comprendere
e creare il suono.



AI Music è in vetta nelle classifiche



Ma come funziona?

Sa inventare veramente?

Copia canzoni esistenti?

È uno strumento per l'artista?

È la fine della musica come arte?

Sappiamo riconoscerla?

Cosa Vedremo Oggi

01

Descrivere il suono

Comprendere le sfide nel tradurre la musica in parole

02

Come l'AI "vede" il suono

Trasformare onde sonore in immagini che le macchine possono analizzare

03

Creare una canzone da zero

Usare l'intelligenza artificiale per generare musica con semplici descrizioni testuali

04

Distinguere il vero dal fake

Riconoscere artefatti e limiti dell'audio generato dall'AI

05

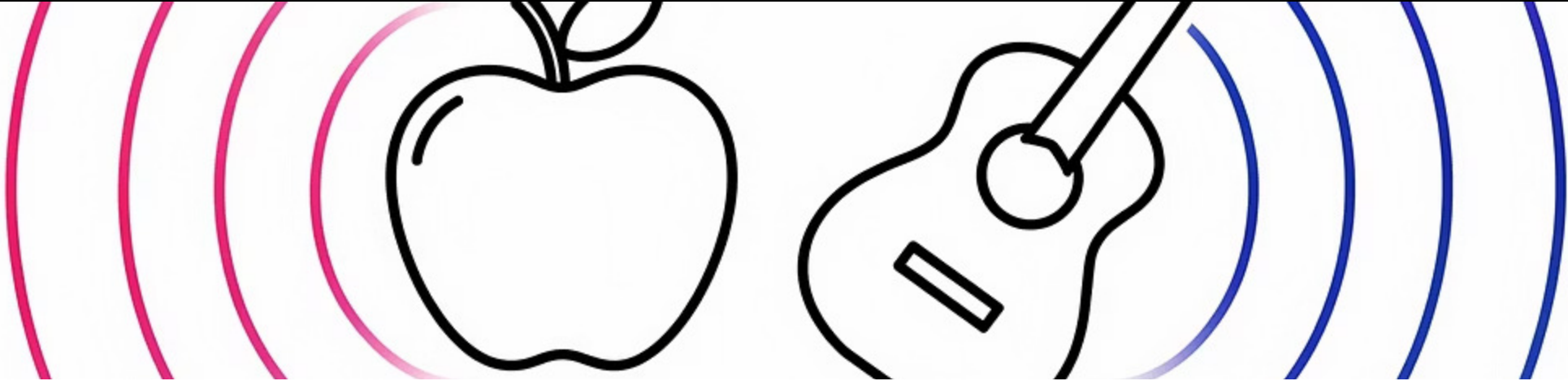
Come l'AI può evolvere per supporto all'arte

Strumenti di controllo creativo per musicisti e artisti

06

Audio AI: oltre la musica

Applicazioni dell'intelligenza artificiale audio in altri campi

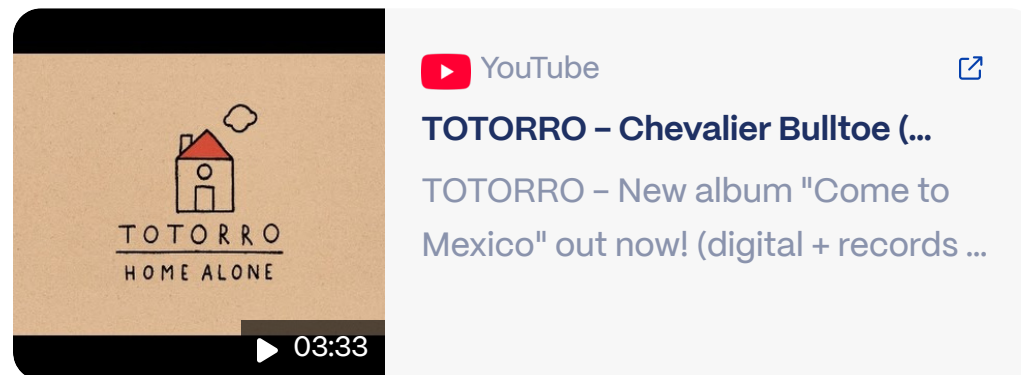


 INTRODUZIONE

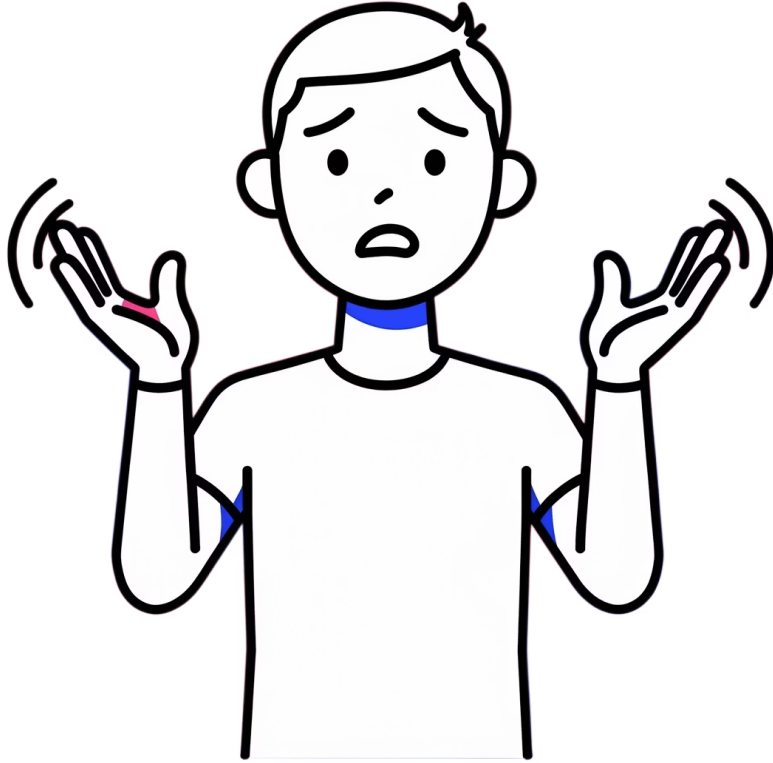
Il Suono: Vibrazione Fisica e Percezione

L'audio è fisicamente solo pressione dell'aria che si propaga, ma noi lo percepiamo ed elaboriamo come emozione, cultura e memoria.

La Sfida di Descrivere il Suono: Un Esempio



La Sfida di Descrivere il Suono



Il Problema Linguistico

Non abbiamo un lessico condiviso per descrivere il suono in modo oggettivo, quindi spesso ricorriamo a metafore:

- Un suono "caldo" (come una chitarra acustica)
- Un suono "freddo" (come un sintetizzatore)
- Un suono "brillante" (come i piatti della batteria)
- Un suono "scuro" (come un contrabbasso)
- Un suono "felice" (come una melodia in maggiore)
- Un suono "triste" (come un violino malinconico)
- Un suono "aggressivo" (come una distorsione heavy metal)

Decomposizione e Analisi Musicale

Ascolteremo tre clip di complessità crescente: Pop, Orchestrale ed Elettronica.

Pop

Batteria, basso, chitarra e voce.
Struttura relativamente semplice
con 4-5 elementi principali.

Orchestrale

Decine di strumenti sovrapposti:
archi, ottoni, legni, percussioni. Mix
complesso e stratificato.

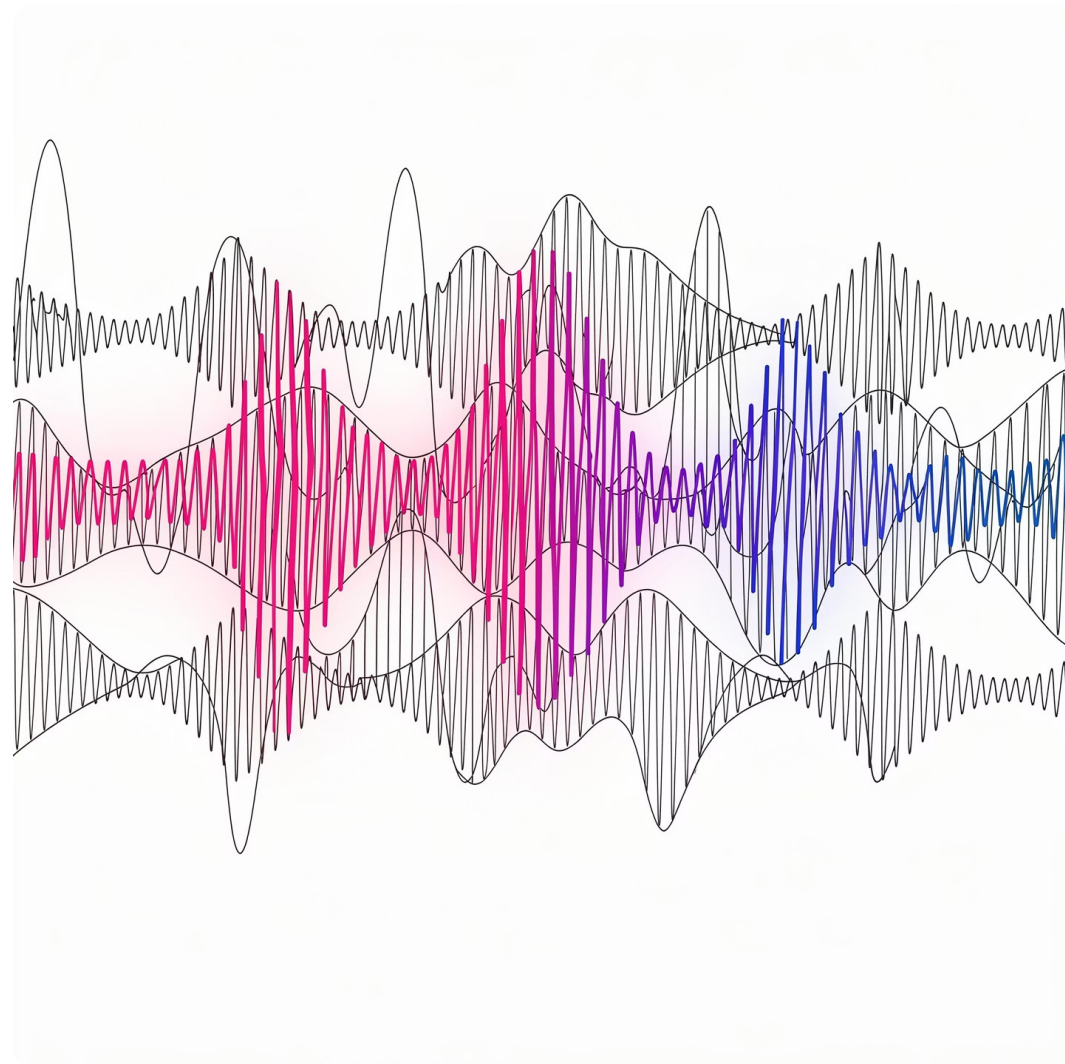
Elettronica

Sintetizzatori, campioni, effetti.
Layer multipli con texture sonore
artificiali e processate.

Source Separation: Isolare le Sorgenti Sonore

Nelle tracce appena ascoltate, isolare mentalmente ogni strumento è una sfida non banale.

La **Source Separation** è la tecnica che permette all'intelligenza artificiale di analizzare un audio complesso e separare le singole sorgenti sonore. Questo è fondamentale per l'analisi musicale e diverse applicazioni audio.



Free AI Stem Splitter



EaseUS Vocal Remover and AI Instrumental Splitter

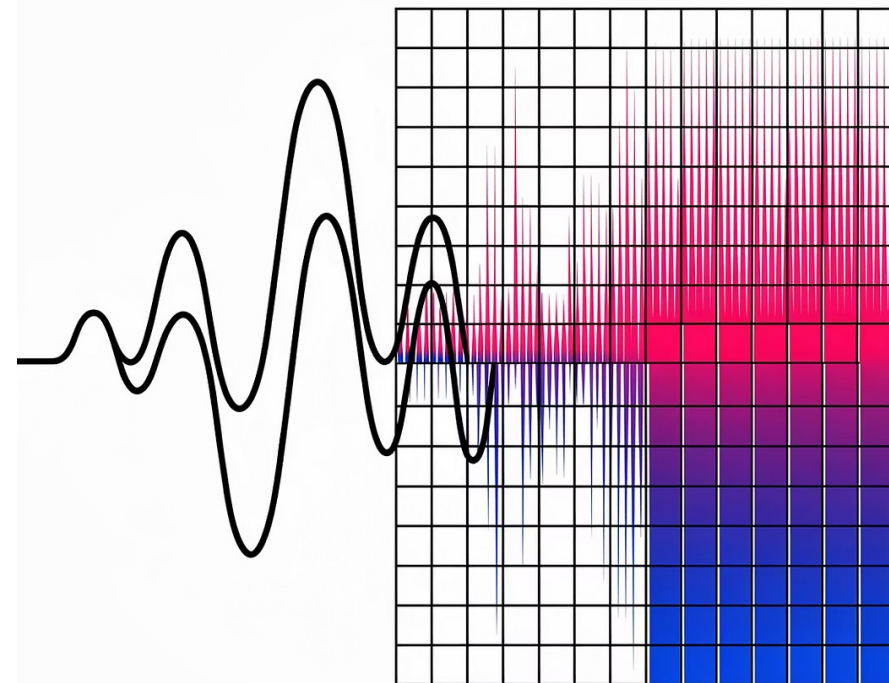


Free AI Stem Splitter

EaseUS AI Stem Splitter removes instrumentals from your favorite songs or videos using the advanced AI algorithm for remixing and karaoke creation.

∞ COMPUTER VISION

La Visione Macchina: Come l'AI "Vede" il Suono



Waveform vs Spettrogramma

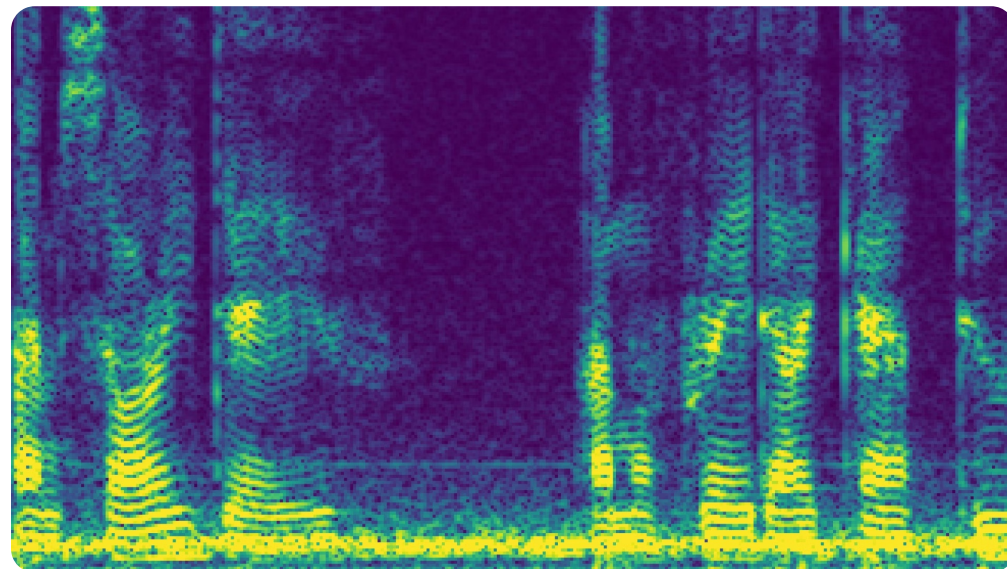
Waveform (Forma d'Onda)



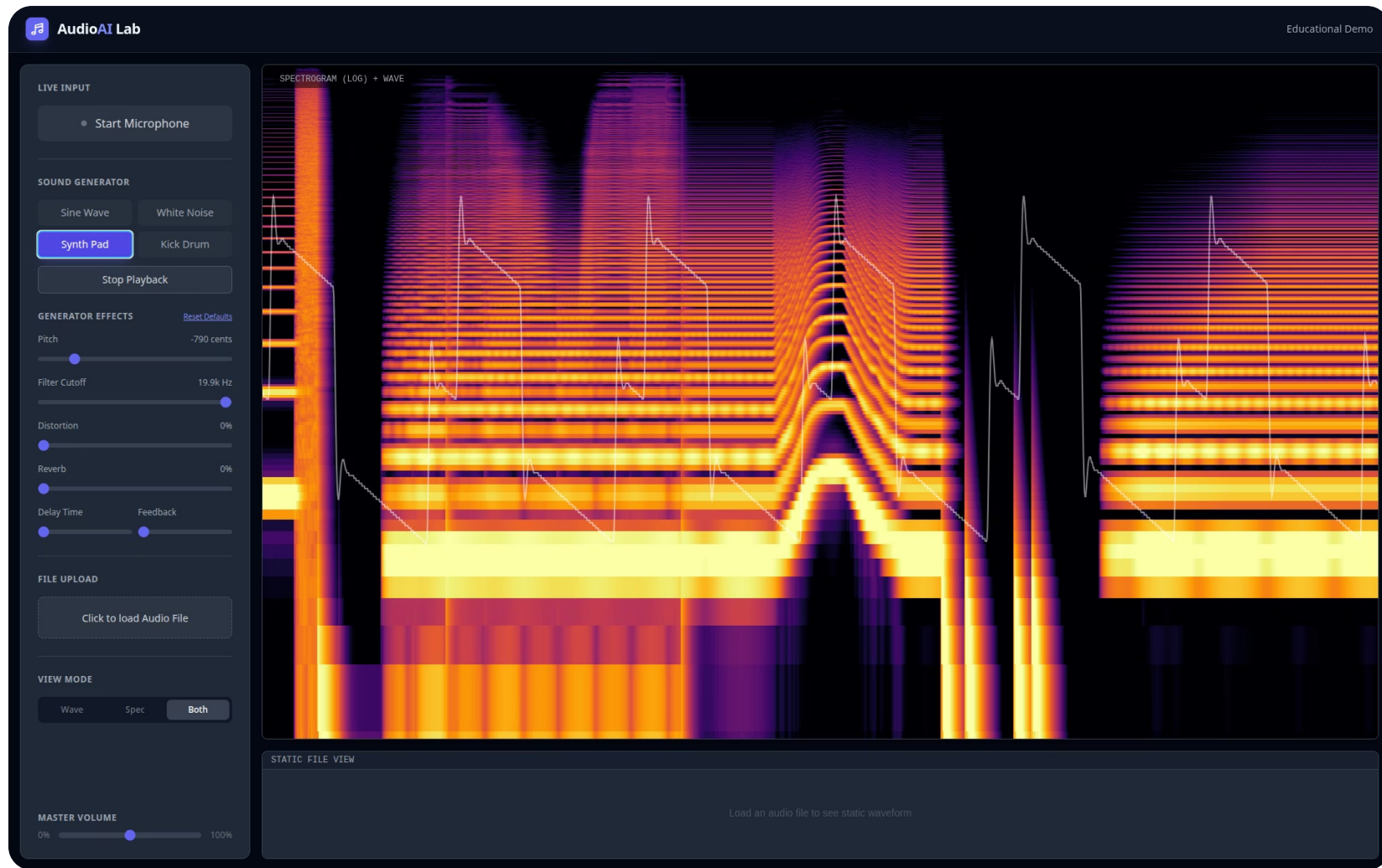
Mostrando l'ampiezza (volume) nel tempo.

È utile per identificare picchi e valli di volume, ma non rivela le specifiche frequenze (note) o gli strumenti presenti nell'audio.

Spettrogramma

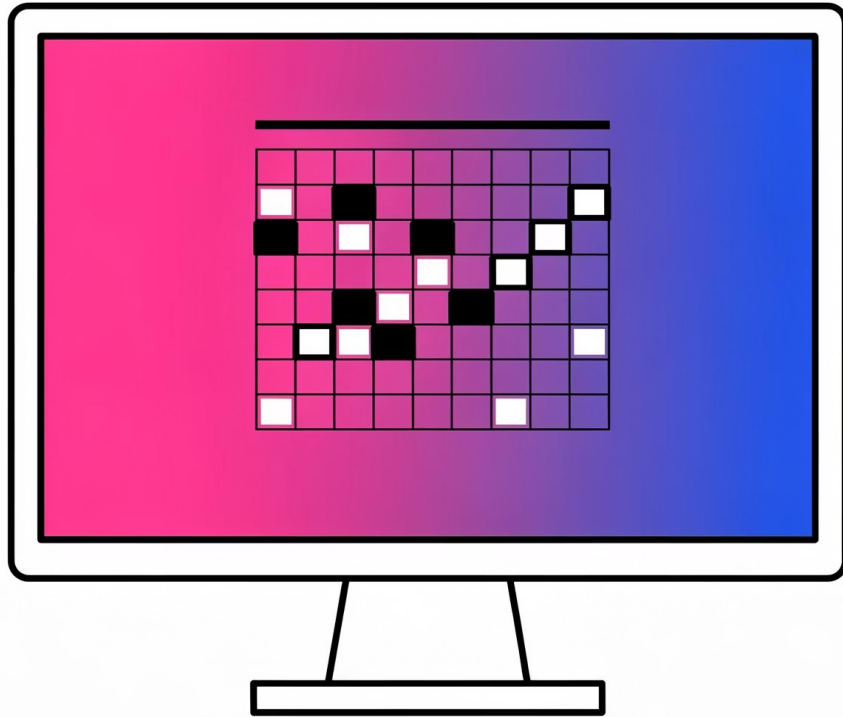


L'asse orizzontale indica il tempo, quello verticale le frequenze (altezza delle note), e l'intensità del colore mostra il volume di ogni frequenza. Questo permette all'intelligenza artificiale di "vedere" e analizzare tutte le componenti sonore.



[Audio AI Visualizer](#)

L'Audio viene trattato come Immagini

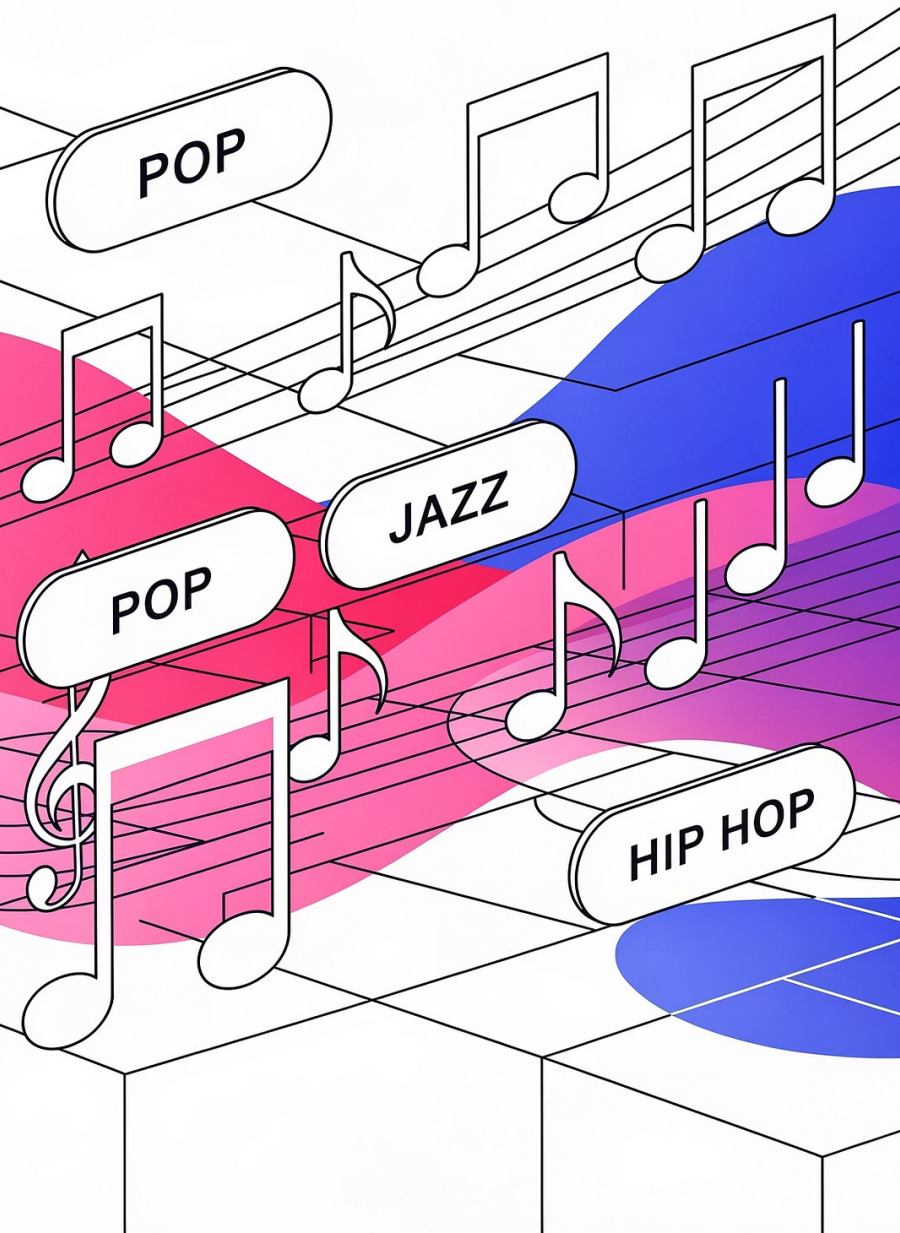


La Rivoluzione Concettuale

Le moderne AI generative non elaborano regole armoniche o teoria musicale. Trattano l'audio come immagini.

Un Do Maggiore diventa un pattern riconoscibile di pixel colorati nello spettrogramma, proprio come un gatto è un pattern riconoscibile in una fotografia.

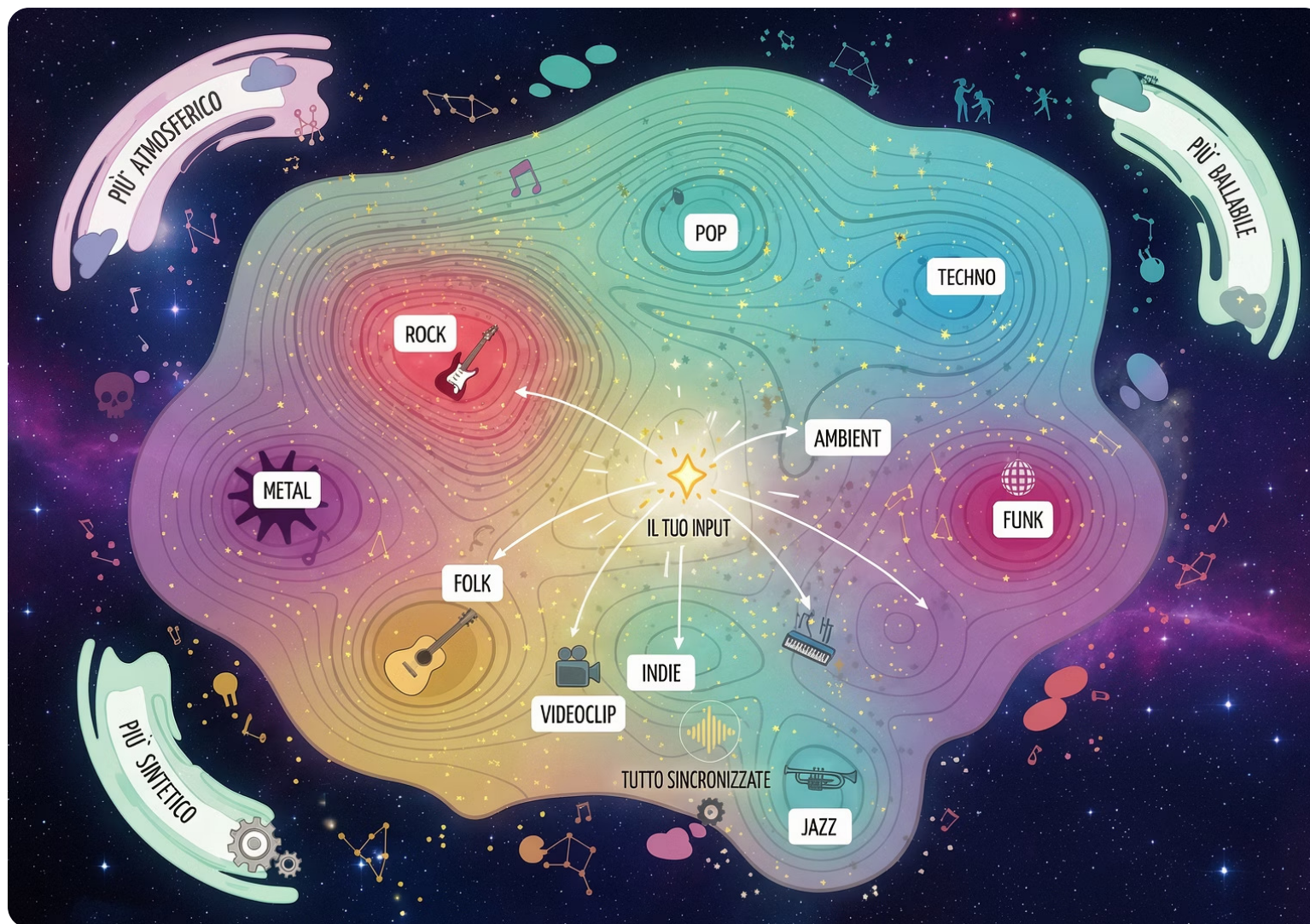
Questo approccio ha rivoluzionato completamente il campo dell'AI audio negli ultimi anni.



✦ GENERAZIONE

Generazione Generativa: Lo Spazio Latente

Come l'AI pensa alla musica.



Dimostrazione: Text-to-Audio

01

Scelta dei Generi

La classe propone due generi musicali molto contrastanti (es. Jazz e Techno)

03

Generazione

Il modello elabora il prompt e genera l'audio in tempo reale

02

Scrittura del Prompt

Creiamo insieme una descrizione testuale che combini entrambi gli stili

04

Analisi Critica

Identifichiamo eventuali artefatti o "glitch" come errori di ricostruzione spettrale

La Tua Formula per Creare Musica

 **Formula del Prompt Perfetto:**
[Genere] + [Strumenti] + [Vibe/Mood] + [Descrizione Tecnica]

Ecco alcuni esempi pratici per iniziare:

Trap anni 2016, bassi distorti,
atmosfera malinconica, riverbero in
una grande stanza

Jazz acustico, pianoforte e
contrabbasso, mood rilassato e
notturno, registrazione intima

Math Rock, chitarre twinkle e
batteria complessa, energicaa,
tempi dispari

Ora tocca a voi! Usate questa formula per creare la vostra traccia su Suno AI

Quiz di Ascolto Critico

Ascolteremo tre brevi clip audio. Riuscite a identificare i difetti dell'AI?

Clip 1: Audio Perfetto

Questo è il nostro punto di partenza: suono naturale e pulito

Clip 2: Glitch nei Transienti

Cosa notate di strano nella batteria? Ascoltate attentamente l'attacco dei colpi

Clip 3: Voce Robotica

Concentrate l'attenzione sulla voce: suona naturale o artificiale?

Cosa Cercare

- Distorsioni o "glitch" nei transienti percussivi
- Innaturalità nel riverbero o nella spazializzazione
- Incoerenza timbrica tra sezioni diverse
- Artefatti nella voce (respirazione innaturale, transizioni brusche)

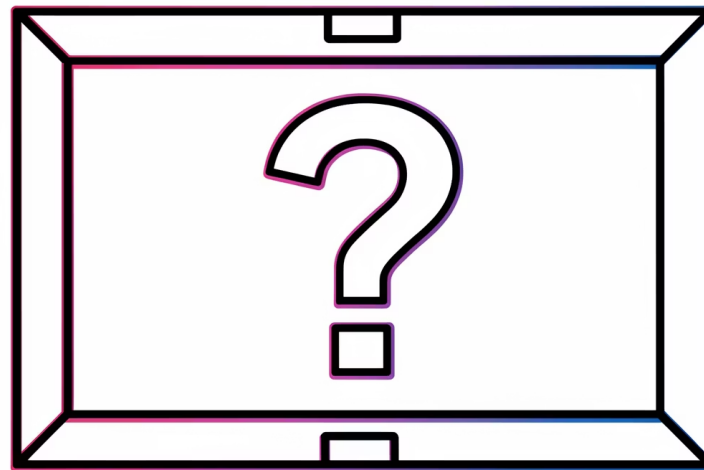
Il Problema della 'Scatola Chiusa'

Suno è Magico... Ma

- Genera musica completa in secondi
- Risultati impressionanti
- Facile da usare

Però:

- Non puoi modificare quanto è presente il basso nel mix
- Non puoi cambiare il tono di un singolo strumento
- Non puoi aggiustare il finale
- È tutto o niente



Ecco perché serve il **Controllo Attivo**: strumenti che ti permettono di modificare ogni singolo elemento del suono

→ CONTROLLO

Sintesi Neurale e Controllo Attivo

Passiamo dalla generazione passiva (prompt testuale) al controllo attivo in tempo reale.



DDSP e Timbre Transfer



Input Vocale

Fischio o humming umano con
espressività naturale



DDSP Processing

Differentiable Digital Signal
Processing analizza e trasforma



Output Strumentale

Saxofono o violino che mantiene
l'espressività originale

L'espressività umana viene mantenuta mentre il timbro cambia completamente. L'AI diventa un **processore di segnale controllabile**, non un sostituto del musicista.



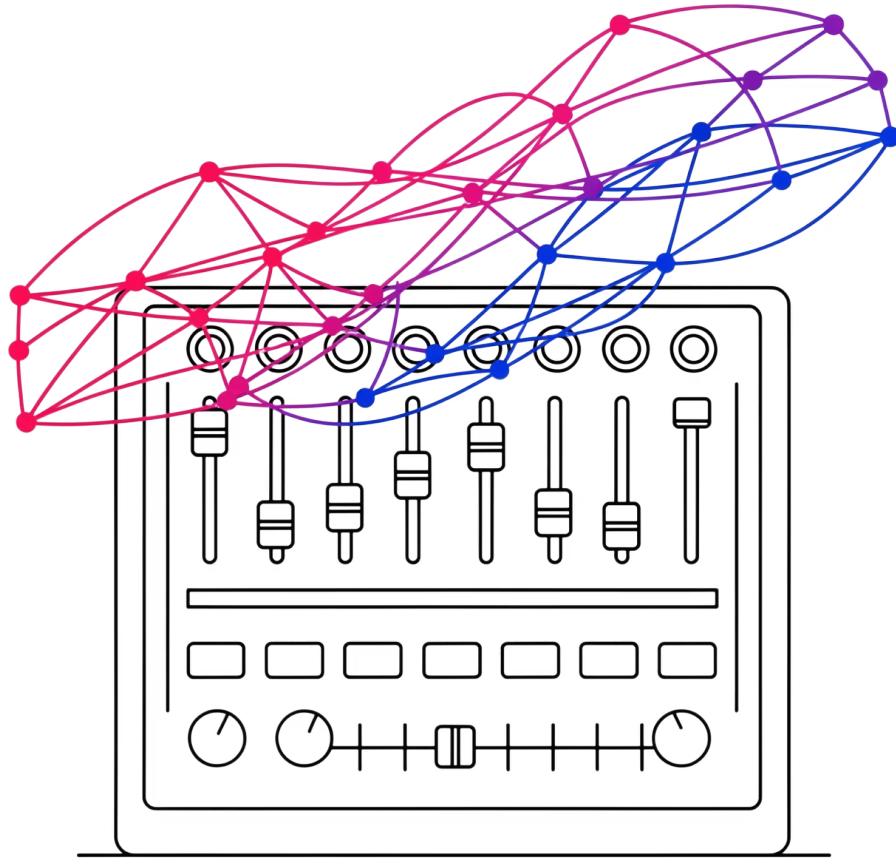
 sites.research.google

Tone Transfer — Magenta DDSP

Play around with different inputs and outputs to see how machine learning transforms sounds.



Controllo per i Musicisti



Parametri Controllabili

- **Riverbero:** spazialità e profondità del suono
- **Inviluppo:** attacco, decay, sustain, release
- **Timbro:** caratteristiche frequenziali dello strumento
- **Dinamica:** variazioni di intensità ed espressività

Questi controlli trasformano l'AI da "black box" misteriosa a strumento musicale professionale.

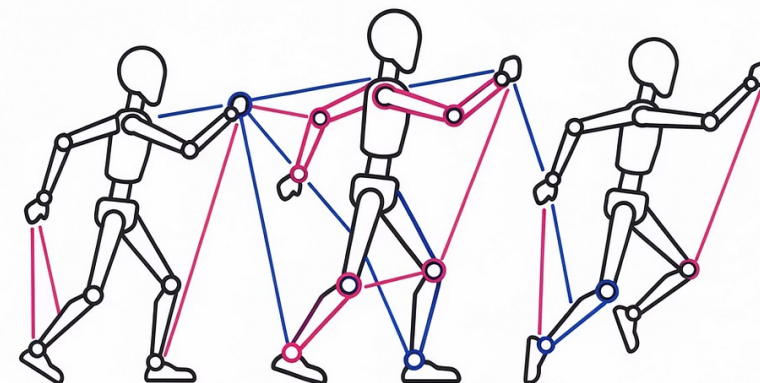
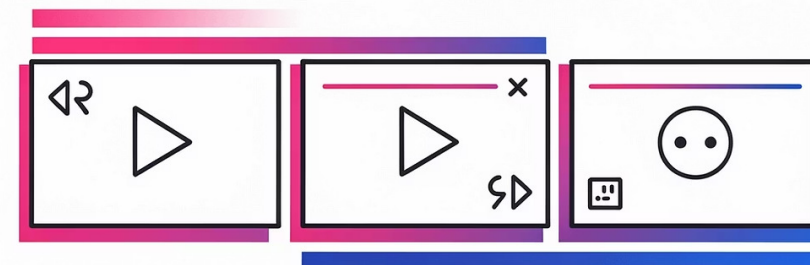
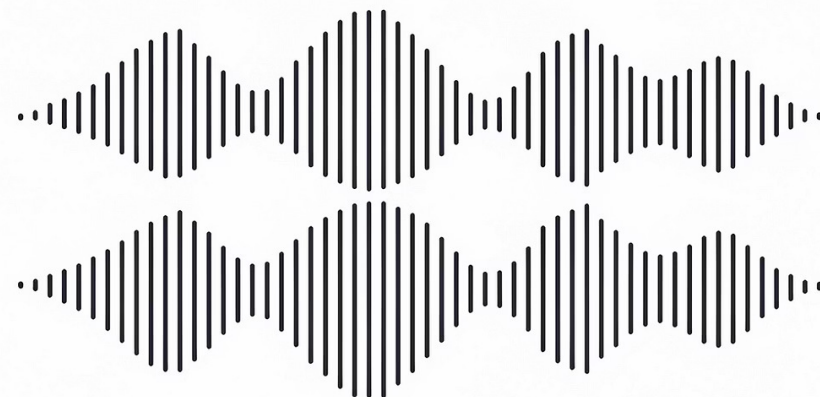


[PromptDJ](https://promptdj.com)

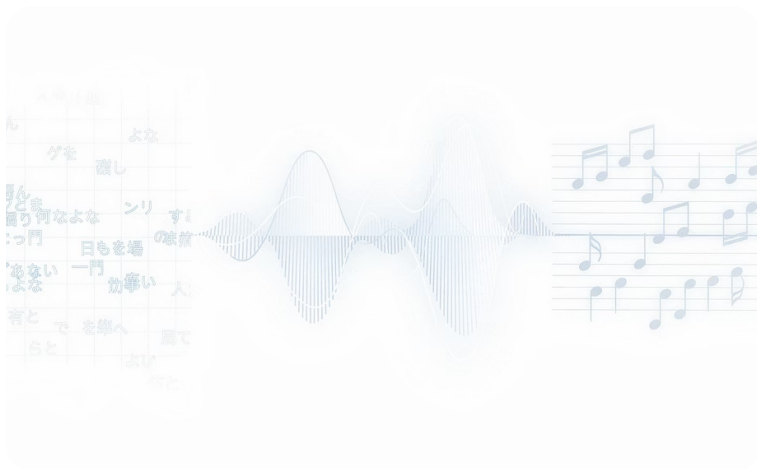
MULTIMODALITÀ

L'incontro tra Audio, Video e Motion

Con l'IA non creiamo solo una canzone: da una singola idea possiamo generare contemporaneamente la musica, il video musicale e la copertina dell'album, tutto perfettamente coordinato.

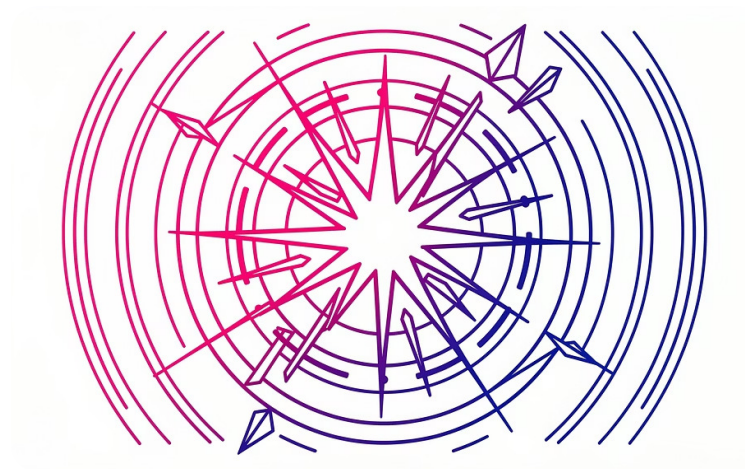


Applicazioni Audio AI



Text-to-Speech

L'AI parla seguendo le indicazioni testuali o crea un dialogo tra più voci.



Video-to-Audio

L'AI genera automaticamente effetti sonori coerenti analizzando il contesto visivo del video.

Style Instructions

Read aloud in a warm and friendly tone:

Text

Ciao! Sono un'AI parlante!

0:02 / 0:02

Run Ctrl ←

Gemini 2.5 Pro Preview TTS

gemini-2.5-pro-preview-tts

Our 2.5 Pro text-to-speech audio model optimized for powerful, low-latency speech generation for more natural outputs and easier to steer prompts.

No API Key

Switch to a paid API key to unlock higher quota and more features.

Mode

Single-speaker audio

Multi-speaker audio

Model settings

▼

Voice

🌐 Zephyr

▼

[Generate Speech - Google AI Studio](#)

Video-to-Audio

Text-to-Audio

hkchengrex MMAudio 925

MMAudio — Video-to-Audio Synthesis

Project page: <https://hkchengrex.com/MMAudio/>
Code: <https://github.com/hkchengrex/MMAudio>

Ho Kei Cheng, Masato Ishii, Akio Hayakawa, Takashi Shibuya, Alexander Schwing, Yuki Mitsufuji

University of Illinois Urbana-Champaign, Sony AI, and Sony Group Corporation

CVPR 2025

NOTE: It takes longer to process high-resolution videos (>384 px on the shorter side). Doing so does not improve results.

The model has been trained on 8-second videos. Using much longer or shorter videos will degrade performance. Around 5s~12s should be fine.

Video

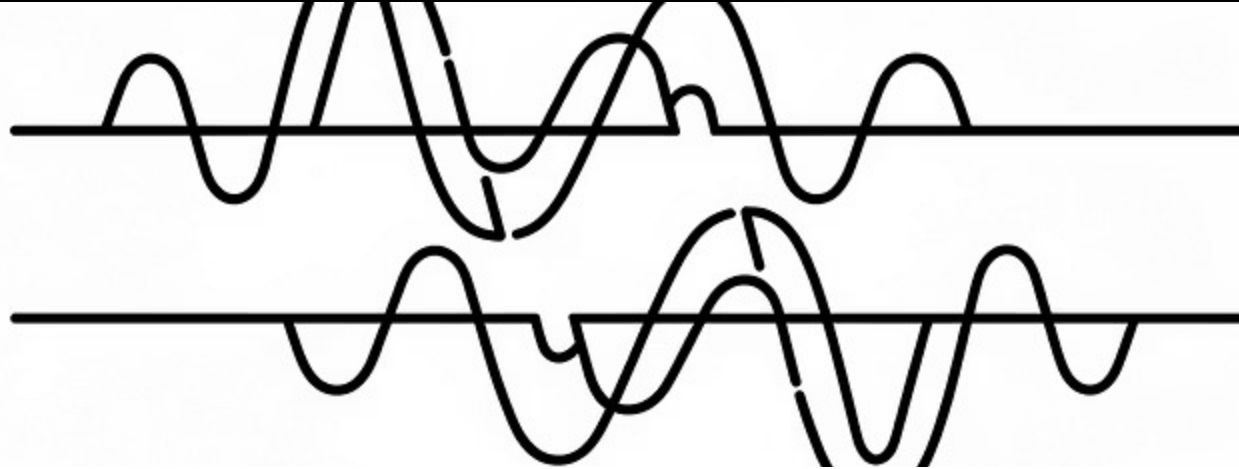


output

Share via Link

Prompt

[MMAudio — Video to Audio Synthesis](#)



⚠ ETICA

Voice Cloning ed Etica

Faremo ora un "Turing Test" audio: ascolteremo due clip di una voce nota. Una è reale, l'altra è sintetica. Riesci a distinguerle?

▶ SOCIAL MEDIA

L'AI nel Tuo Feed

Avete mai visto questi trend su TikTok?



Cover AI Virali

Peter Griffin che
canta canzoni pop,
personaggi dei
cartoni che rappano



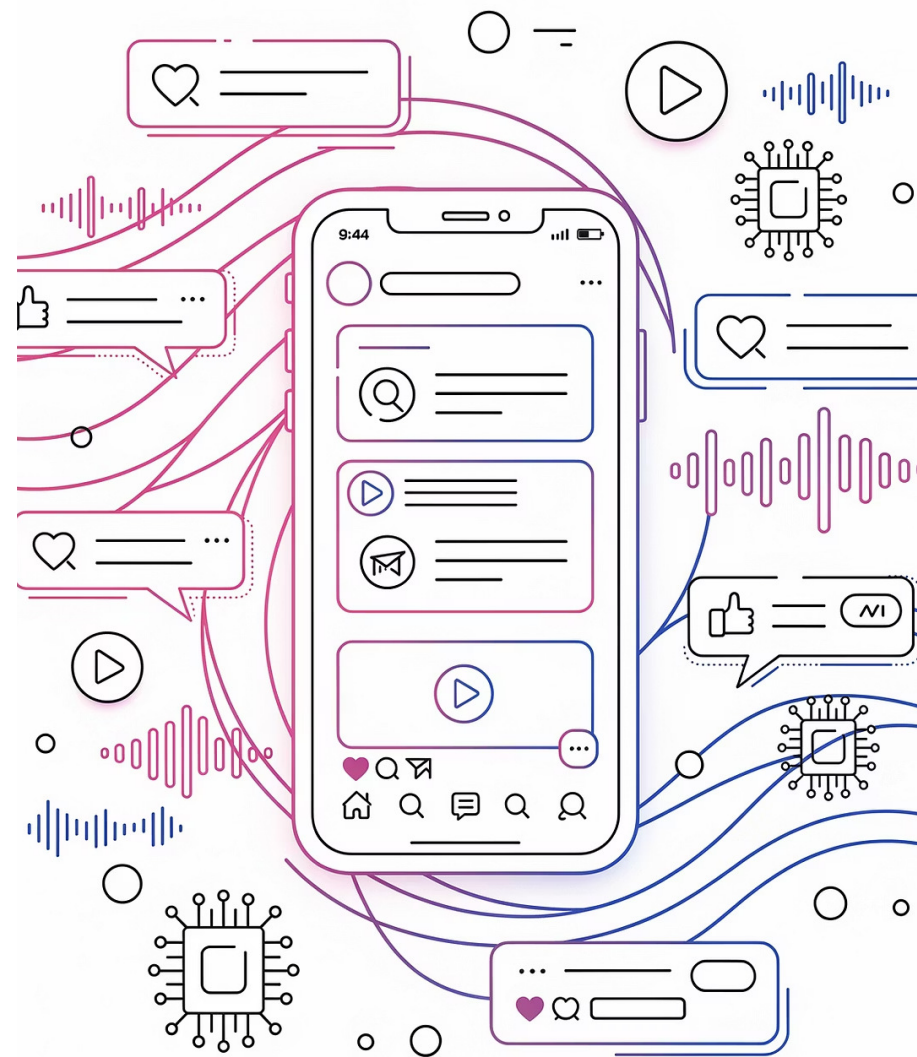
Voci Clonate di Celebrità

Politici che giocano ai
videogiochi, attori
che leggono meme



Remix Impossibili

Collaborazioni tra
artisti che non si sono
mai incontrati



Dibattito: Chi Deve Essere Pagato?

Avvocati dell'AI

- L'AI ha creato qualcosa di nuovo
- È come imparare da un maestro
- La tecnologia è libera

Avvocati dell'Artista

- Lo stile è proprietà intellettuale
- Senza artisti non esisterebbe
- Il lavoro originale va protetto



 RIEPILOGO

Il Nostro Viaggio nell'AI

Abbiamo esplorato le fondamenta, le capacità generative, il controllo artistico, le intersezioni multimodali e le implicazioni etiche dell'Intelligenza Artificiale nel mondo dell'audio.

Le domande che rimangono aperte sono:

- **Dove ci porterà questa rivoluzione?**
- **Quali sono i prossimi traguardi raggiungibili?**