How AI Voice Changed Learning and Development

About Me

Marketing Manager

Jeff is the Marketing Manager at WellSaid. He is passionate about the growing field of AI voice and its impact on content creators. He lives in Pittsburgh, PA.

Why Me

» I'm a content creator, and I instantly saw the value of AI Voice.

» I began working at WellSaid before the AI boom and have listened to hundreds of our customers describe how they use AI Voice.
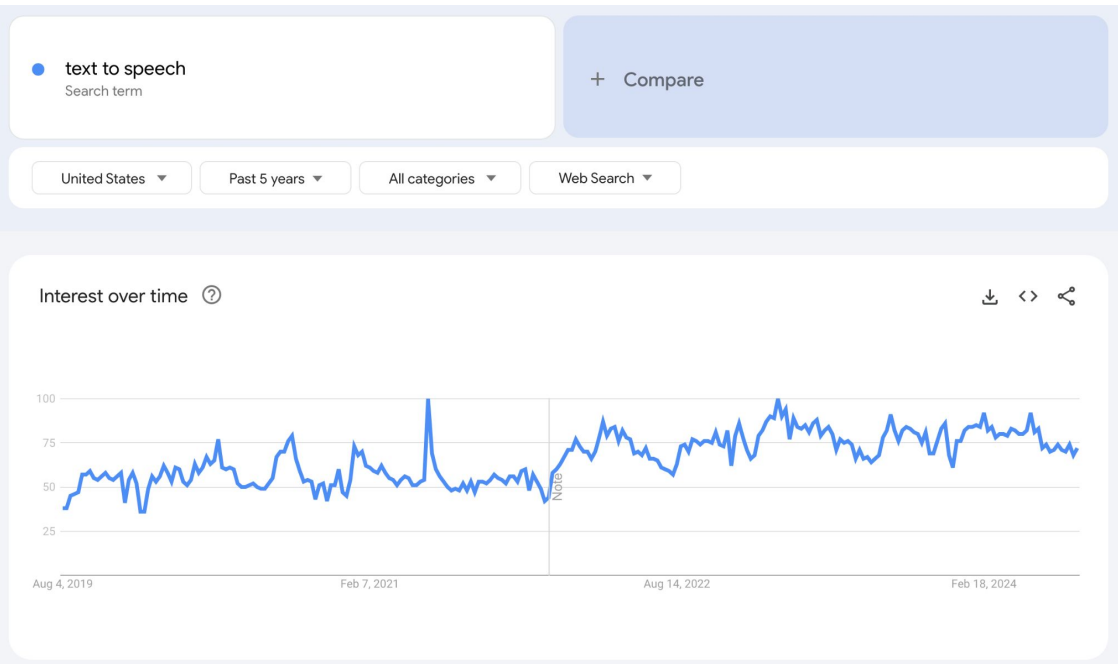
How People Search for a Solution Has Changed

"Whereas I used to say text to speech, now, it is AI text to speech or AI voice generator."

Text to Speech

The most common way people describe solutions for artificial voice has been text-to-speech.
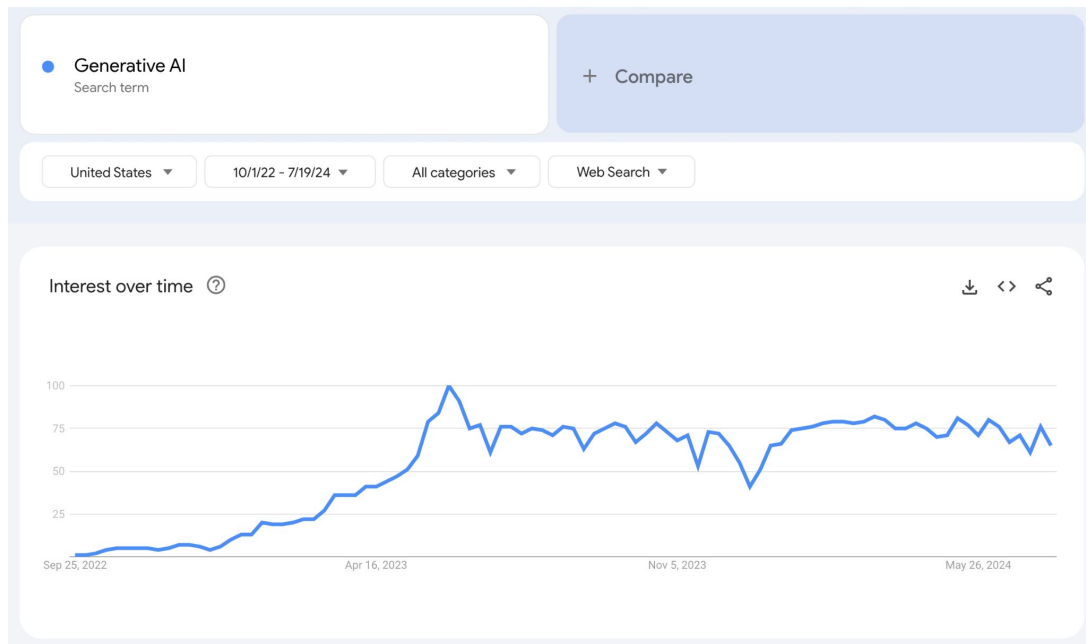
AI Boom

Generative AI
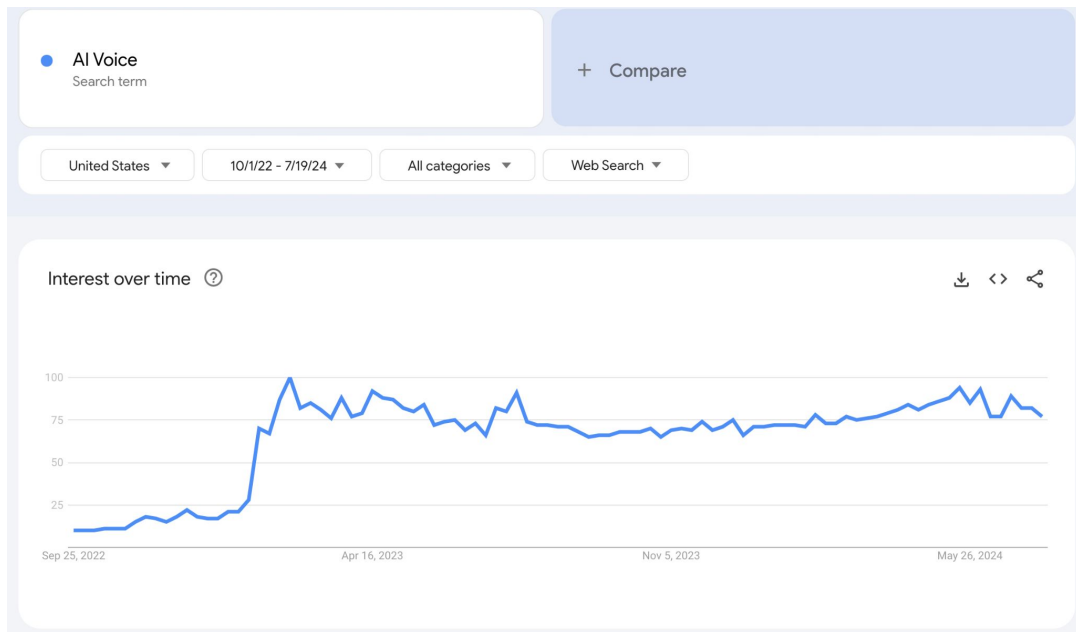
In the winter of 2023, Generative AI takes off.

ATD reported in July, "*More than half of instructional designers are using AI in their role, and of those who are using it, most are satisfied with it.*"
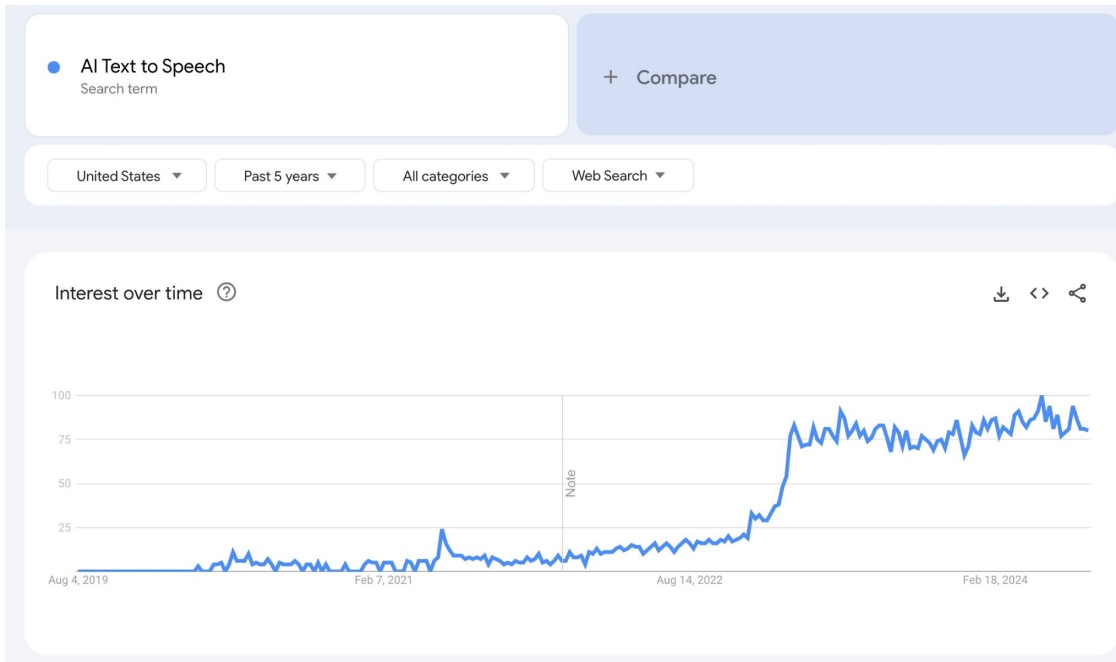
Google Trends

AI Voice

AI Voice is a subset of AI, and now half of the Fortune 500 are using WellSaid.
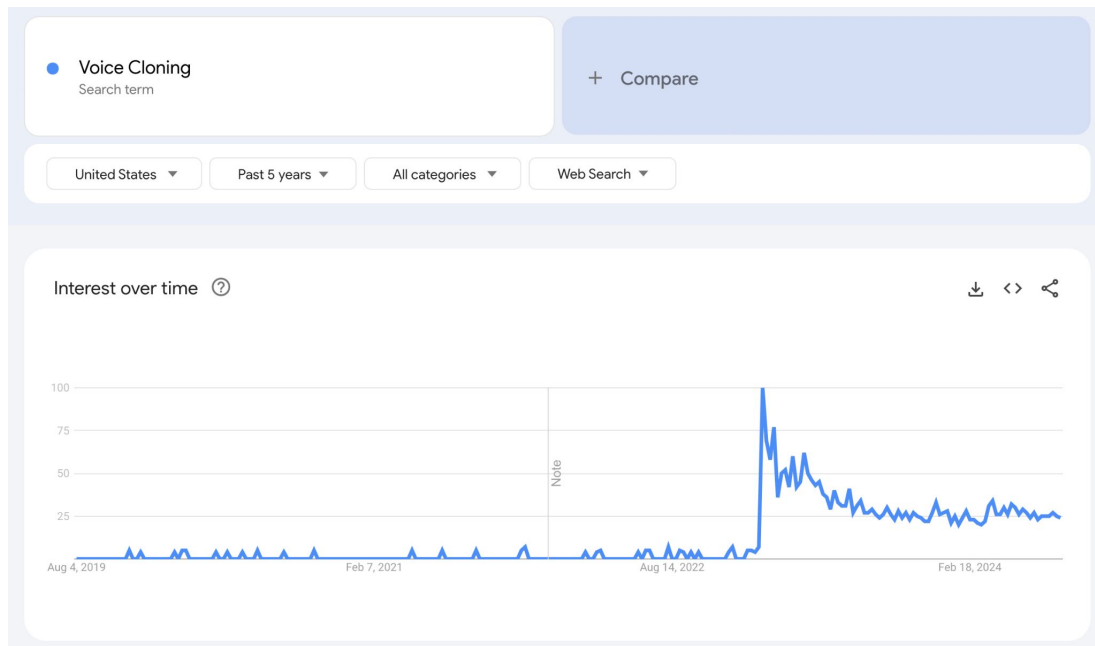
AI Text to Speech

AI Text to Speech
Search term

\+ Compare

United States ▾ | Past 5 years ▾ | All categories ▾ | Web Search ▾

Interest over time ?

100
75
50
25

Aug 4, 2019   Feb 7, 2021   Aug 14, 2022   Feb 18, 2024

Voice Cloning

Voice Cloning
Search term

+ Compare

United States ▼ | Past 5 years ▼ | All categories ▼ | Web Search ▼

Interest over time ⑦

100
75
50
25

Aug 4, 2019   Feb 7, 2021   Aug 14, 2022   Feb 18, 2024

Are You Using AI Voice?

Webinar Registration Poll

40% - Yes
60% - No

» How AI Voice is Made

» AI Voice vs. Voice Actor vs. Recording Your Voice

» The ROI Calculator of AI Voice

» AI Compliance, the Ethics of AI Voice

» Vendor Selection Tips

» How to Correct Pronunciation, add Emotion, Workflow Tips
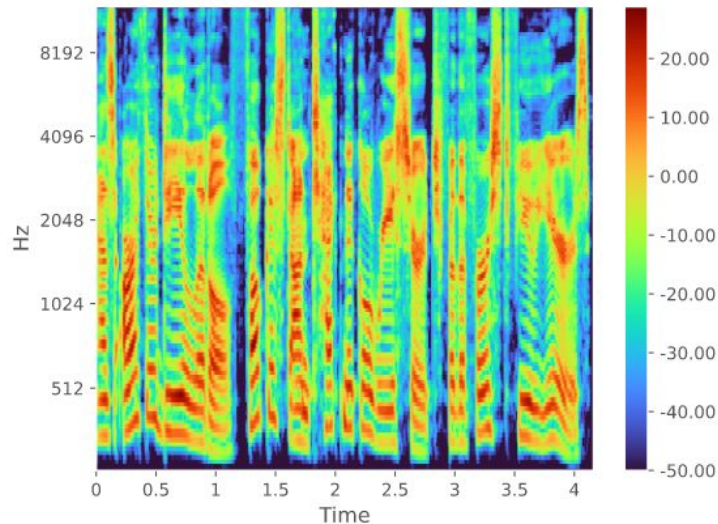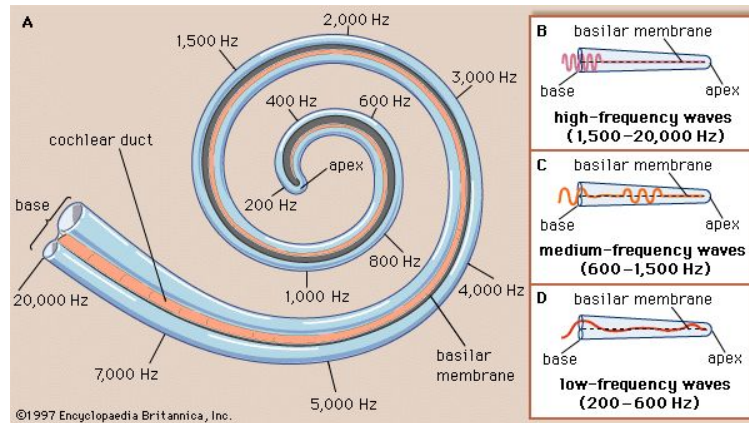
How AI Voice is Made

We start with spectrograms:

A multidimensional view of **sound**

Approximates the complexities of **human hearing**

When working with an appropriate **synthesis** model, matching the **spectrogram** often corresponds to matching the **sound** extremely well.
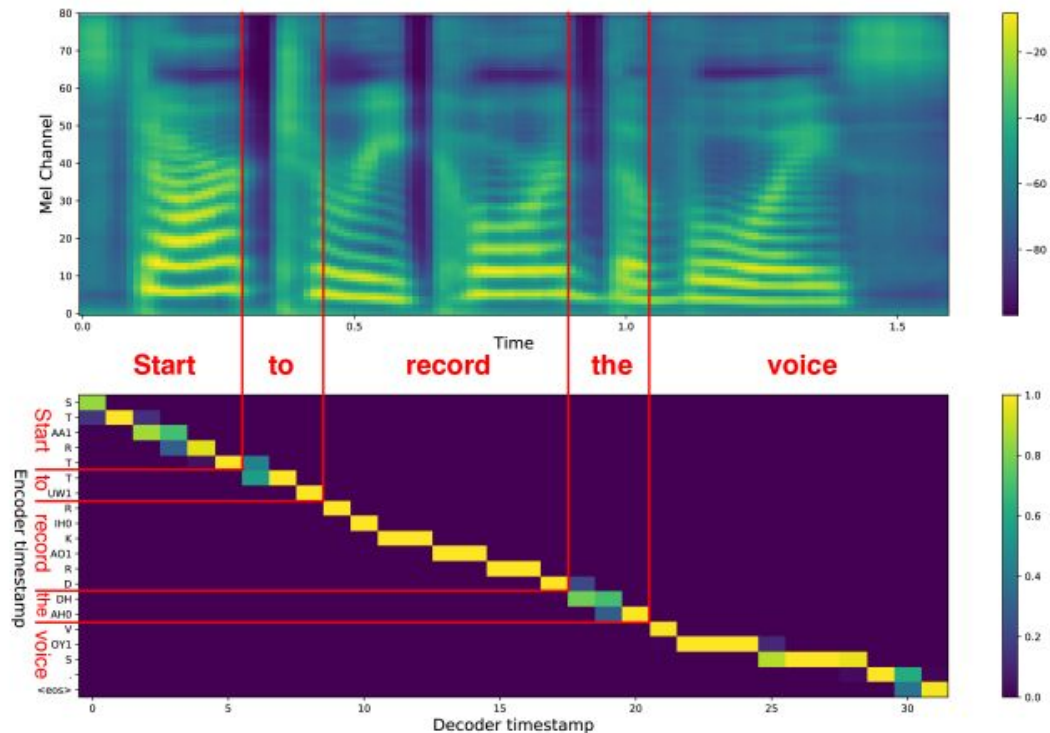
Then we create alignments:

We annotate the script data with a timestamp to mark the **beginning** and **end** of each word in the text.

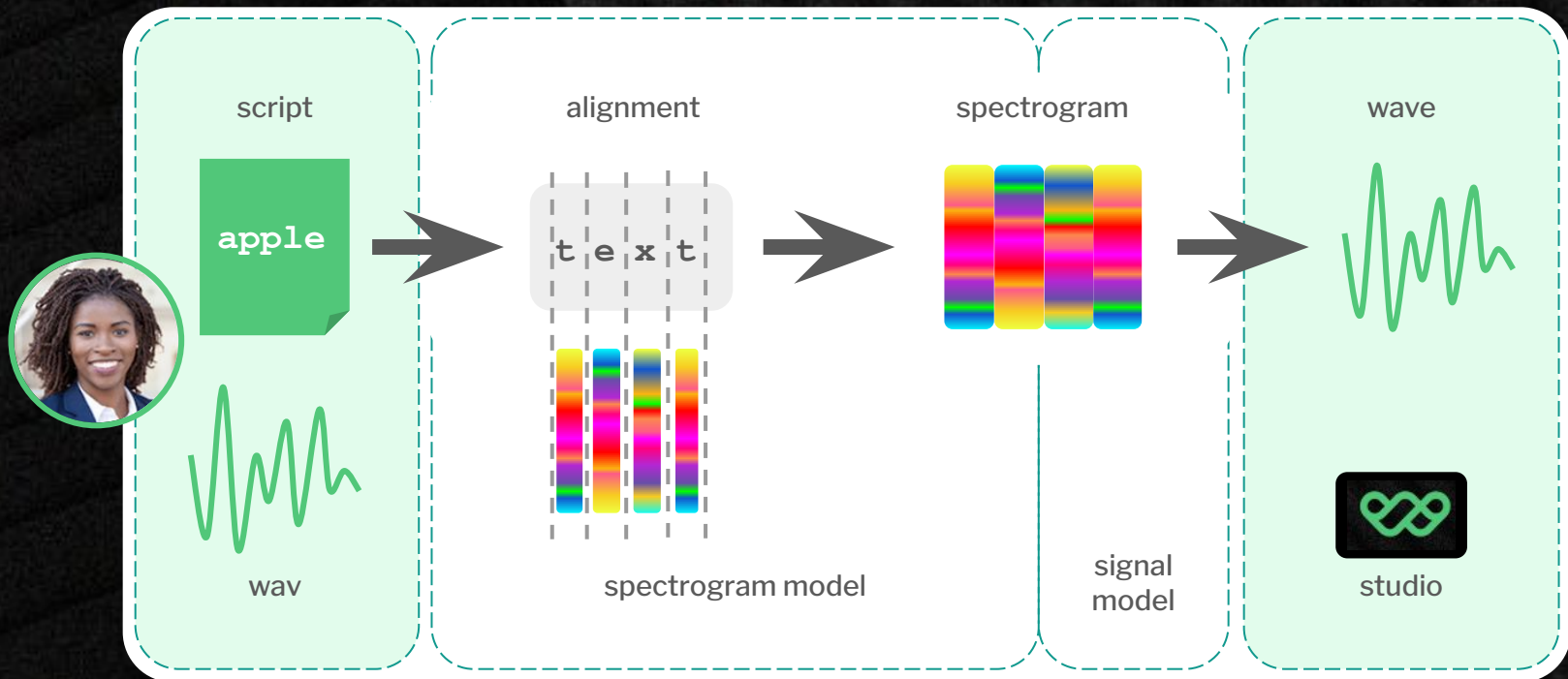This allows the model to **align** a timestamped spectrogram visual of the audio with the same timestamp of the words.

"Start to record the voice."

Ma, Mingbo et al. "Incremental Text-to-Speech Synthesis with Prefix-to-Prefix Framework." Findings (2019).

Then we let the model make predictions...

script

apple

wav

alignment

text

spectrogram model

spectrogram

signal
model

wave

studio

Additional Resources

WellSaid Innovation: A Voice Model That Reads Like a Voice Actor,
Inside the Recording Studio: How WellSaid Avatars Are Built

#1 AI Voice Misconception

AI Voice adds an option to get more done.

AI Voice vs. Recording Your Voice

———

Webinar Registration Poll

58% - Recording Own Voice
34% - In House
9% - Voice Actor

AI Voice

» **2 Mins to Create in the Seattle Airport,** copy and paste the script, and select a voice actor.

Recording Your Own Voice

» Recorded in the house, everyone had to leave, complete silence.

» Lower clarity, compression, and loudness.

» 4 hrs to record 15 mins of scripts for five videos.

» Edits aren't possible without recording again.

» Uneasy about own voice.

AI Voice

» **On-demand voice talent 24/7.**

» **Consistent recording quality.**

» **Easy Edits, adjust your script.**

Recording Your Voice

» Inconsistent recording, due to voice mood, health, equipment, or room.

» Reformatting the script into a teleprompter style for reading.

» Editing clips into final production.

AI Voice vs. Voice Actor

Human vs AI

🔊 » **Audio Sample A**
SCRIPT: Here's what's in this for you here in the first chapter. If you ever find yourself playing a trivia game about liquid penetrant testing, you are definitely going to win because we're going to see how PT got from where it started, to what it is has become... I'm being told that I am in fact old, proved by the fact that no one plays trivia games any more. Ouch.

🔊 » **Audio Sample B**
SCRIPT: Today is July 18th, 2024. And now for Question #1. Are solvent suspendable developers applied by brushing, or spraying?   Question #2: True or False? The threshold of visual acuity for a person with 20 20 vision is about 0.3cm.

The ROI of AI Voice

AI Voice

» **Starting @ $49 Per Month,** an annual enterprise account with multiple seats is usually the equivalent of 1 hr of voice actor audio.

Voice Actor

» $35 - $55 per minute of finished audio.

$2,700 - 1 Hrs
$8,100 - 3 Hrs
$13,500- 5 Hrs
$27,000 - 10 Hrs

TTS vs. Speech to Speech

Text to Speech

» **Team oriented,** share projects, collaborate with scripts.

» **Enterprise,** works well in organizations that need to approve scripts.

» **Easy Edits,** adjust your text, render a new track in seconds, anywhere.

Speech to Speech

» Record your voice, and let AI adapt it to a new voice.

» It works well for solo content creators and those who record their voices already.

» Voice cloning widely varies in quality.

AI Compliance, the Ethics of AI Voice

AI Voice Compliance

» How are voices sourced?
Are you protected from copyright?

» Does the vendor use your information to train the AI?
Will your voice, voice examples, or scripts be used for training data?

» Does the vendor resell, or license technology, or use a proprietary approach?
Vendors that lease AI Voice could go out of business quickly based on the license.

Choosing a Vendor

» **Collecting Voice Samples, Ask Colleagues for Feedback**
Initiate a trial with vendors, render samples of your scripts with multiple voices, and set up a blind test with colleagues.

» **Render Long Form Content, Not Just Samples**
Make sure to render clips over one minute to account for mispronounced words; platforms that need the most cleanup work for pronunciation, emphasis, and emotion will be more frustrating in the long run.

» **Crosscheck the Workflow**
Write down all teammates who will share access to the AI voice studio. Many organizations will start small and add seats as they adopt AI voices in different departments and use cases.

Why Start with WellSaid

WellSaid

» **Closed loop AI training data.**

» **Voice actors are paid.**

» **Proprietary AI voice model.**

» **There are edits after the first render.**

» **UI built for teams and projects.**

Other Vendors

» Unknown data sources present copyright issues.

» Risky business model through tech leases.

» Frequent editing and pronunciation issues cause workflow slowdown.
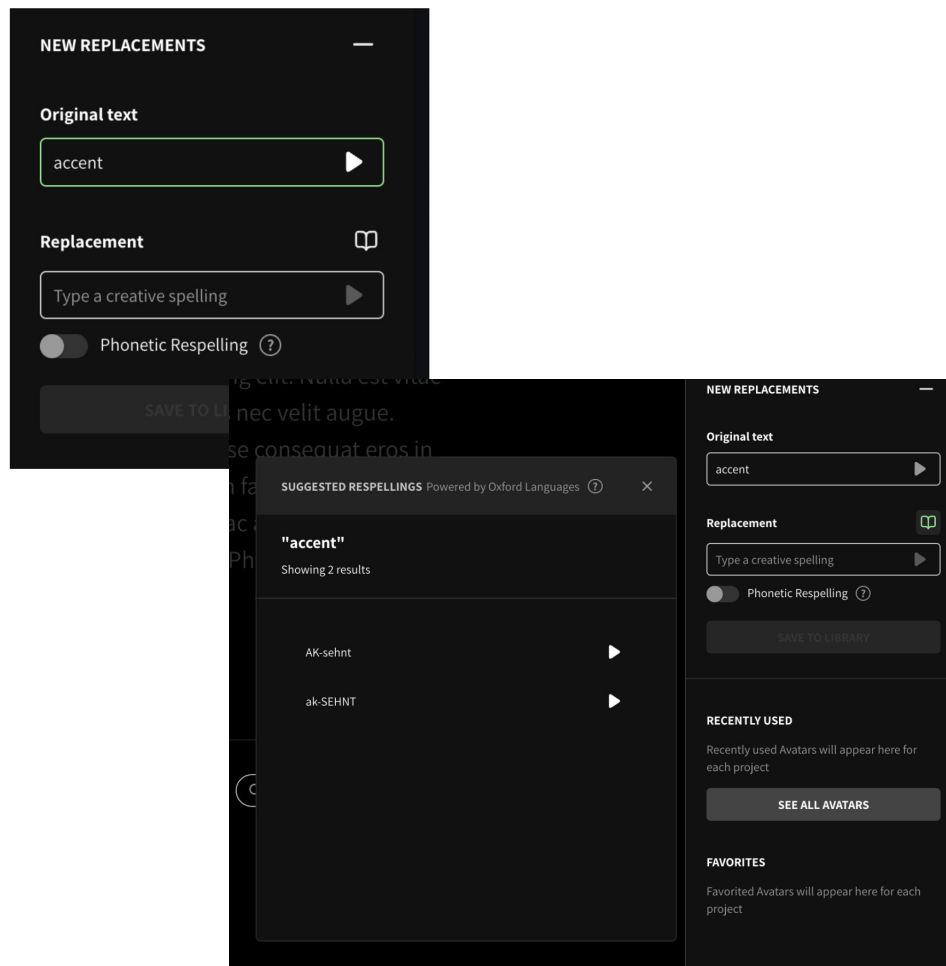
Pronunciation

Pronunciation

Phonetic spellings are the best way to help AI voices correctly pronounce medical jargon, industry slang, and other technical terms.

Saving correctly pronounced words with team members helps workflow efficiency. We work with Oxford to make this as easy as possible.
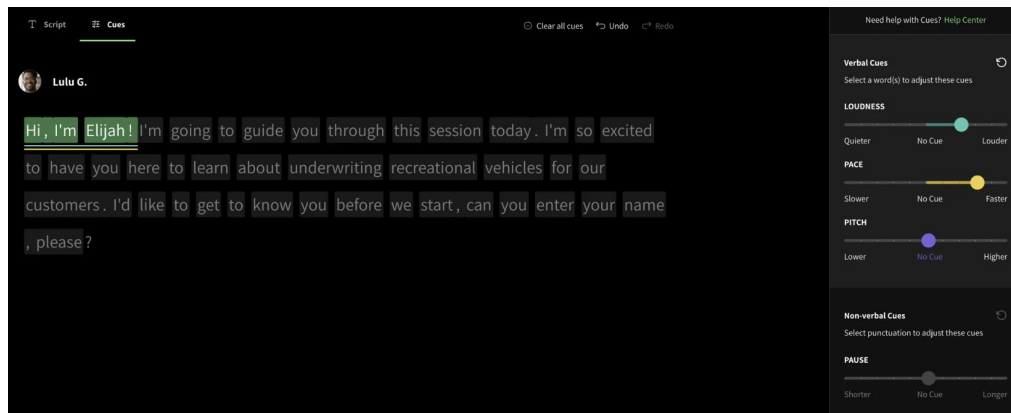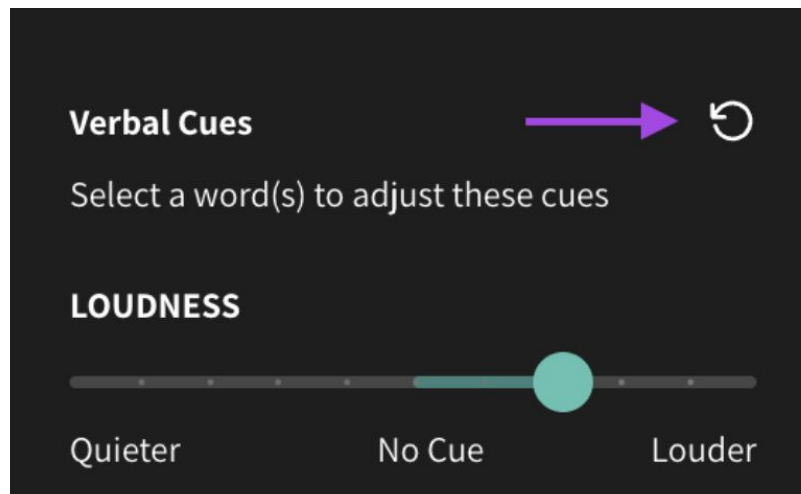
Emphasis

Emphasis

A voice's ability to project intent, intonation, and emotion without edits is the gold standard for AI voice.

With our AI voice technology, you have complete control. You can adjust verbal cues like loudness, pace, and pitch, and add text annotations to scripts, such as extra periods for pauses, pacing adjustments, and "" for adding emphasis to key phrases, giving you the power to create the perfect voice output.
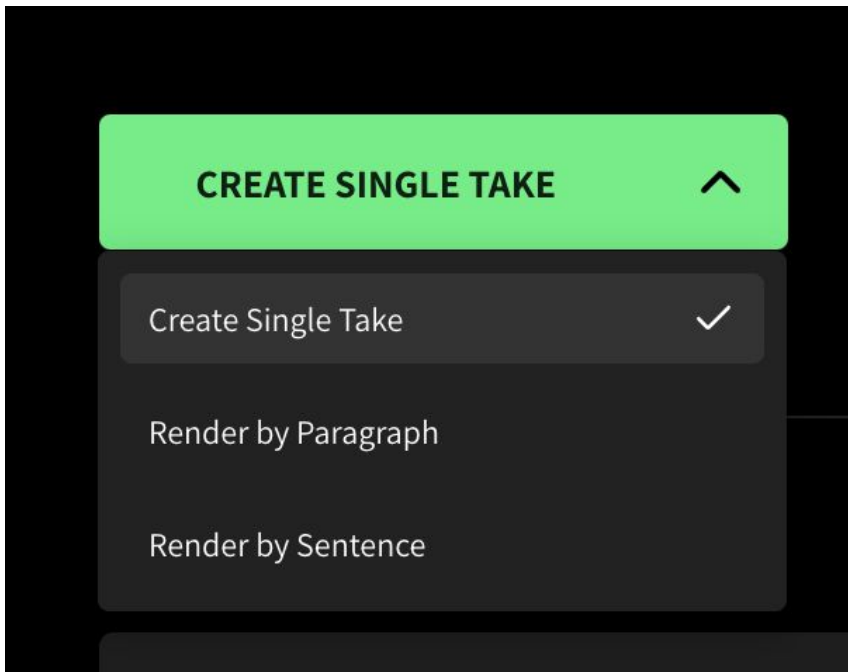
Workflow

Rendering

A page of text can be rendered into shorter clips, either by paragraph or sentence.

Our AI Voice Studio simplifies the editing process by making it easy to review voice-over tracks, make edits, and then combine them.

Combine, copy, delete, move clips to other projects, and search scripts.

CREATE SINGLE TAKE

Create Single Take ✓

Render by Paragraph

Render by Sentence

Combine

Search

Start Your Trial
www.wellsaidlabs.com

Thank You

Join Our Webinars