

השקנו את Gemini 3 Pro. אפשר לנסות אותו בחינם ב-Google AI Studio.
(<https://aistudio.google.com?model=gemini-3-pro-preview&hl=he>)

[Switch to English](#)

translated by Google
דף זה תורגם על ידי Cloud Translation API (cloud.google.com/translate/?hl=he/)

יצירת דיבור (המרת טקסט לדיבור)

בעזרת Gemini API אפשר להפוך קלט טקסט לאודיו עם דובר אחד או כמה דוברים באמצעות יכולות מובנות של המרת טקסט לדיבור (TTS). הפקת המרת טקסט לדיבור (TTS) היא ניתנת לשליטה (controllable), כלומר אפשר להשתמש בשפה טבעית כדי לבנות אינטראקציות ולהנחות את הסגנון, המבטא, הקצב והטון של האודיו.

יכולת ה-TTS שונה מיצירת דיבור שמתבצעת באמצעות Live API (<https://ai.google.dev/gemini-api/docs/live?hl=he>), שנועד לאודיו אינטראקטיבי ולא מובנה, ולתשומות ולתפוקות מולטימדיאליות. Live API מצטיין בהקשרים דינמיים של שיחות, אבל TTS דרך Gemini API מותאם לתרחישים שבהם נדרשת הקראה מדויקת של טקסט עם שליטה מדויקת בסגנון ובצליל, כמו יצירה של פודקאסט או ספר אודיו. במדריך הזה מוסבר איך ליצור אודיו עם קריינות של דובר אחד או כמה דוברים מטקסט.

תצוגה מקדימה: המרת טקסט לדיבור (TTS) מובנית נמצאת בתצוגה מקדימה (<https://ai.google.dev/gemini-api/docs/models?hl=he#preview>).

לפני שמתחילים

חשוב לוודא שאתם משתמשים בגרסה של מודל Gemini 2.5 עם יכולות מובנות של המרת טקסט לדיבור (TTS), כמו שמופיע ברשימה שבקטע מודלים נתמכים (<https://ai.google.dev/gemini-api/docs/speech-generation?hl=he#supported-models>). כדי לקבל תוצאות אופטימליות, כדאי לבחור את המודל שהכי מתאים לתרחיש השימוש הספציפי שלכם.

מומלץ לבדוק את מודלי ה-TTS של Gemini 2.5 ב-AI Studio (<https://aistudio.google.com/generate-speech?hl=he>) לפני שמתחילים לפתח.

הערה: מודלים של TTS מקבלים קלט של טקסט בלבד ומפיקים פלט של אודיו בלבד. רשימה מלאה של ההגבלות הספציפיות למודלים של TTS מופיעה בקטע Limitations (<https://ai.google.dev/gemini-api/docs/speech-generation?hl=he#limitations>).

המרת טקסט לדיבור של דובר יחיד

כדי להמיר טקסט לאודיו עם דובר אחד, מגדירים את אופן התגובה ל'אודיו' ומעבירים אובייקט `SpeechConfig` עם הגדרה של `VoiceConfig`. צריך לבחור שם קול מתוך [הקולות המובנים לפלט](#) (voices#).

בדוגמה הזו, האודיו שנוצר על ידי המודל נשמר בקובץ `wave`:

(rest#) [REST](#)(javascript#) [JavaScript](#)[Python](#)
(python#)

```
from google import genai
from google.genai import types
import wave

# Set up the wave file to save the output:
def wave_file(filename, pcm, channels=1, rate=24000, sample_width=2):
    with wave.open(filename, "wb") as wf:
        wf.setnchannels(channels)
        wf.setsampwidth(sample_width)
        wf.setframerate(rate)
        wf.writeframes(pcm)

client = genai.Client()

response = client.models.generate_content(
    model="gemini-2.5-flash-preview-tts",
    contents="Say cheerfully: Have a wonderful day!",
    config=types.GenerateContentConfig(
        response_modalities=["AUDIO"],
        speech_config=types.SpeechConfig(
            voice_config=types.VoiceConfig(
                prebuilt_voice_config=types.PrebuiltVoiceConfig(
                    voice_name='Kore',
                )
            )
        ),
    ),
)

data = response.candidates[0].content.parts[0].inline_data.data

file_name='out.wav'
wave_file(file_name, data) # Saves the file to current directory
```

דוגמאות נוספות לקוד זמינות בקובץ TTS - Get Started (התחלת העבודה עם TTS) במאגר המתכונים: 

[הצגה ב-GitHub](#)

https://colab.research.google.com/github/google-gemini/cookbook/blob/main/quickstarts/Get_started_TTS.ipynb?hl=he

המרת טקסט לדיבור עם כמה דוברים

כדי להשתמש באודיו עם כמה רמקולים, צריך אובייקט `MultiSpeakerVoiceConfig` עם כל רמקול (עד 2) שמוגדר כ-`SpeakerVoiceConfig`. צריך להגדיר כל `speaker` עם אותם שמות שבהם השתמשתם בהנחיה `:(controllable#)`

(rest#) [REST](#)(javascript#) [JavaScript](#)[Python](#)
(python#)

```
from google import genai
from google.genai import types
import wave

# Set up the wave file to save the output:
def wave_file(filename, pcm, channels=1, rate=24000, sample_width=2):
    with wave.open(filename, "wb") as wf:
        wf.setnchannels(channels)
        wf.setsampwidth(sample_width)
        wf.setframerate(rate)
        wf.writeframes(pcm)

client = genai.Client()

prompt = """TTS the following conversation between Joe and Jane:
    Joe: How's it going today Jane?
    Jane: Not too bad, how about you?"""

response = client.models.generate_content(
    model="gemini-2.5-flash-preview-tts",
    contents=prompt,
    config=types.GenerateContentConfig(
        response_modalities=["AUDIO"],
        speech_config=types.SpeechConfig(
            multi_speaker_voice_config=types.MultiSpeakerVoiceConfig(
                speaker_voice_configs=[
                    types.SpeakerVoiceConfig(
                        speaker='Joe',
                        voice_config=types.VoiceConfig(
                            prebuilt_voice_config=types.PrebuiltVoiceConfig(
                                voice_name='Kore',
                            )
                        )
                    )
                ],
            )
        ),
    ),
```

```

types.SpeakerVoiceConfig(
    speaker='Jane',
    voice_config=types.VoiceConfig(
        prebuilt_voice_config=types.PrebuiltVoiceConfig(
            voice_name='Puck',
        )
    )
),
],
)
)
)
)
)

data = response.candidates[0].content.parts[0].inline_data.data

file_name='out.wav'
wave_file(file_name, data) # Saves the file to current directory

```

שליטה בסגנון הדיבור באמצעות הנחיות

אתם יכולים לשלוט בסגנון, בטון, במבטא ובקצב באמצעות הנחיות בשפה טבעית, גם בהמרת טקסט לדיבור עם דובר אחד וגם עם כמה דוברים. לדוגמה, בהנחיה עם דובר אחד, אפשר לומר:

Say in an spooky whisper:
 "By the pricking of my thumbs...
 Something wicked this way comes"

בהנחיה עם כמה דוברים, צריך לספק למודל את השם של כל דובר ואת התמליל המתאים. אפשר גם לספק הנחיות לכל דובר בנפרד:

Make Speaker1 sound tired and bored, and Speaker2 sound excited and happy:

Speaker1: So... what's on the agenda today?
 Speaker2: You're never going to guess!

כדי להדגיש את הסגנון או הרגש שרוצים להעביר, אפשר לנסות להשתמש באפשרות קוליות (voices#) שמתאימה להם. בהנחיה הקודמת, לדוגמה, יכול להיות שהצליל הנשימתי של אנקלדוס ידגיש את המילים 'עייף' ו'משועמם', בעוד שהטון העליז של יכול להשלים את המילים 'נרגש' ו'שמח'.

המערכת יוצרת הנחיה להמרה לאודיו

מודלים של TTS מוציאים רק אודיו, אבל אפשר להשתמש במודלים אחרים

המודל ה-TTS כדי שיקרא אותו בקול. (https://ai.google.dev/gemini-api/docs/models?hl=he) כדי ליצור תמליל קודם, ואז להעביר את התמליל הזה למודל ה-TTS כדי שיקרא אותו בקול.

(javascript#) JavaScriptPython
(python#)

```
from google import genai
from google.genai import types

client = genai.Client()

transcript = client.models.generate_content(
    model="gemini-2.0-flash",
    contents="""Generate a short transcript around 100 words that reads
    like it was clipped from a podcast by excited herpetologists.
    The hosts names are Dr. Anya and Liam.""").text

response = client.models.generate_content(
    model="gemini-2.5-flash-preview-tts",
    contents=transcript,
    config=types.GenerateContentConfig(
        response_modalities=["AUDIO"],
        speech_config=types.SpeechConfig(
            multi_speaker_voice_config=types.MultiSpeakerVoiceConfig(
                speaker_voice_configs=[
                    types.SpeakerVoiceConfig(
                        speaker='Dr. Anya',
                        voice_config=types.VoiceConfig(
                            prebuilt_voice_config=types.PrebuiltVoiceConfig(
                                voice_name='Kore',
                            )
                        )
                    ),
                    types.SpeakerVoiceConfig(
                        speaker='Liam',
                        voice_config=types.VoiceConfig(
                            prebuilt_voice_config=types.PrebuiltVoiceConfig(
                                voice_name='Puck',
                            )
                        )
                    )
                ]
            )
        )
    )

# ...Code to stream or save the output
```

אפשרויות קול

מודלים של TTS תומכים ב-30 אפשרויות הקול הבאות בשדה voice_name:

Charon – אינפורמטיבי	Puck – Upbeat	Zephyr -- Bright
Leda – צעיר	Fenrir – נרגש	Kore -- Firm
Callirrhoe -- Easy-going	Aoede – Breezy	Orus -- Firm
Iapetus – Clear	Enceladus -- Breathy	Autonoe -- Bright
Despina – Smooth	Algieba -- Smooth	Umbriel – נינוח
Rasalgethi -- Informative	Algenib – מחוספס	Erinome -- Clear
Alnilam -- Firm	Achernar -- Soft	Laomedeia -- שמח
Pulcherrima -- Forward	Gacrux -- Mature	Schedar – Even
Vindemiatrix -- Gentle	Zubenelgenubi – שגרתי	Achird -- Friendly
Sulafat -- חמה	Sadaltager -- Knowledgeable	Sadachbia -- Lively

אפשר לשמוע את כל האפשרויות של הקולות ב-[AI Studio](https://aistudio.google.com/generate-speech?hl=he) (https://aistudio.google.com/generate-speech?hl=he).

שפות נתמכות

מודלים של TTS מזהים את שפת הקלט באופן אוטומטי. הם תומכים ב-24 השפות הבאות:

שפה	קוד BCP-47	שפה	קוד BCP-47
ערבית (מצרית)	ar-EG	גרמנית (גרמניה)	de-DE
אנגלית (ארה"ב)	en-US	ספרדית (ארצות הברית)	es-US
צרפתית (צרפת)	fr-FR	הינדית (הודו)	hi-IN

שפה	קוד BCP-47	שפה	קוד BCP-47	
אינדונזית (אינדונזיה)	id-ID	איטלקית (איטליה)	it-IT	
יפנית (יפן)	ja-JP	קוריאנית (קוריאה)	ko-KR	
פורטוגזית (ברזיל)	pt-BR	רוסית (רוסיה)	ru-RU	
הולנדית (הולנד)	nl-NL	פולנית (פולין)	pl-PL	
תאילנדית (תאילנד)	th-TH	טורקית (טורקיה)	tr-TR	
ווייטנאמית (ווייטנאם)	vi-VN	רומנית (רומניה)	ro-RO	
אוקראינית (אוקראינה)	uk-UA	בנגלית (בנגלדש)	bn-BD	
אנגלית (הודו)	חבילה של en-IN ושל hi-IN		מראטהית (הודו)	mr-IN
טמילית (הודו)	ta-IN	טלוגו (הודו)	te-IN	

מודלים נתמכים

מערכת רמקולים	דובר יחיד	דגם
✓	✓	<u>תצוגה מקדימה של Gemini 2.5 Flash TTS</u> (https://ai.google.dev/gemini-api/docs/models?hl=he#gemini-2.5-flash-preview-tts)
✓	✓	<u>תצוגה מקדימה של Gemini 2.5 Pro TTS</u> (https://ai.google.dev/gemini-api/docs/models?hl=he#gemini-2.5-pro-preview-tts)

מגבלות

- מודלים של TTS יכולים לקבל רק קלט טקסט ולהפיק פלט אודיו.

- לסשן TTS יש מגבלה של 32,000 טוקנים בחלון ההקשר
(<https://ai.google.dev/gemini-api/docs/long-context?hl=he>)
- בקטע שפות (<https://ai.google.dev/gemini-api/docs/speech-generation?hl=he#languages>) מפורטות השפות הנתמכות.

מדריך לכתיבת הנחיות

מודל **Gemini Native Audio Generation Text-to-Speech (TTS)** שונה ממודלים מסורתיים של TTS בכך שהוא מבוסס על מודל שפה גדול (LLM) שיודע **לא רק מה לומר, אלא גם איך לומר את זה**.

כדי להשתמש ביכולת הזו, המשתמשים יכולים לדמיין שהם במאים שמגדירים סצנה לקריין וירטואלי. כדי ליצור הנחיה, מומלץ להשתמש ברכיבים הבאים: **פרופיל אודיו** שמגדיר את הזהות הבסיסית והארכיטיפ של הדמות, **תיאור סצנה** שמגדיר את הסביבה הפיזית ואת האווירה הרגשית, ו**הערות הבמאי** שמציעות הנחיות מדויקות יותר לגבי סגנון, מבטא ושליטה בקצב.

המשתמשים יכולים להשתמש במודל כדי ליצור ביצועים אודיו דינמיים, טבעיים ומלאי הבעה, באמצעות מתן הוראות מפורטות כמו מבטא אזורי מדויק, מאפיינים פרא-לשוניים ספציפיים (למשל, נשימה) או קצב. כדי להשיג ביצועים אופטימליים, מומלץ שההנחיות לבימוי יתאימו **לתמליל**, כך שההנחיה 'מי אומר את זה' תתאים להנחיות 'מה נאמר' ו'איך זה נאמר'.

מטרת המדריך הזה היא לספק הנחיות בסיסיות ולעורר רעיונות לפיתוח חוויות אודיו באמצעות יצירת אודיו ב-Gemini TTS. אנחנו סקרנים לראות מה תיצרו!

מבנה ההנחיות

הנחיה חזקה כוללת באופן אידיאלי את הרכיבים הבאים, שמשולבים יחד כדי ליצור ביצועים מצוינים:

- **פרופיל אודיו** – הגדרת פרסונה לקול, כולל זהות הדמות, ארכיטיפ ומאפיינים אחרים כמו גיל, רקע וכו'.
- **סצנה** – מגדירה את הבמה. מתאר את הסביבה הפיזית ואת האווירה.
- **הערות הבמאי** – הנחיות לגבי הביצועים שבהן אפשר לפרט אילו הוראות חשובות לכישרון הווירטואלי. דוגמאות: סגנון, נשימה, קצב, הבעה ומבטא.
- **הקשר לדוגמה** – מספק למודל נקודת התחלה הקשרית, כך שהשחקן הווירטואלי ייכנס לסצנה שהגדרתם באופן טבעי.
- **Transcript (תמליל)** – הטקסט שהמודל יקריא. כדי להשיג את הביצועים הכי טובים, חשוב לזכור שהנושא של התמליל וסגנון הכתיבה צריכים להיות קשורים להוראות שאתם נותנים.

הערה: אתם יכולים לבקש מ-Gemini לעזור לכם ליצור את ההנחיה. פשוט תנו לו תבנית ריקה כמו שמופיעה בהמשך ותבקשו ממנו ליצור סקיצה של דמות בשבילכם.

דוגמה להנחיה מלאה:

AUDIO PROFILE: Jaz R.
"The Morning Hype"

THE SCENE: The London Studio

It is 10:00 PM in a glass-walled studio overlooking the moonlit London skyline, but inside, it is blindingly bright. The red "ON AIR" tally light is blazing. Jaz is standing up, not sitting, bouncing on the balls of their heels to the rhythm of a thumping backing track. Their hands fly across the faders on a massive mixing desk. It is a chaotic, caffeine-fueled cockpit designed to wake up an entire nation.

DIRECTOR'S NOTES

Style:

* The "Vocal Smile": You must hear the grin in the audio. The soft palate is always raised to keep the tone bright, sunny, and explicitly inviting.
* Dynamics: High projection without shouting. Punchy consonants and elongated vowels on excitement words (e.g., "Beauuutiful morning").

Pace: Speaks at an energetic pace, keeping up with the fast music. Speaks with a "bouncing" cadence. High-speed delivery with fluid transitions – no dead air, no gaps.

Accent: Jaz is from Brixton, London

SAMPLE CONTEXT

Jaz is the industry standard for Top 40 radio, high-octane event promos, or any script that requires a charismatic Estuary accent and 11/10 infectious energy.

TRANSCRIPT

Yes, massive vibes in the studio! You are locked in and it is absolutely popping off in London right now. If you're stuck on the tube, or just sat there pretending to work... stop it. Seriously, I see you. Turn this up! We've got the project roadmap landing in three, two... let's go!

שיטות מפורטות ליצירת הנחיות

בואו נפרט כל רכיב בהנחיה.

פרופיל אודיו

תאר בקצרה את הפרסונה של הדמות.

- **שם.** כדאי לתת לדמות שם כדי להקנות למודל בסיס טוב יותר ולשפר את הביצועים. כשגדירים את הסצנה וההקשר, כדאי להתייחס לדמות בשם שלה.
- **תפקיד.** הזחות העיקרית והארכיטיפ של הדמות שמופיעה בסצנה. לדוגמה: תקליטן ברדיו, פודקאסטר, כתב חדשות וכו'.

דוגמאות:

```
# AUDIO PROFILE: Jaz R.  
## "The Morning Hype"
```

```
# AUDIO PROFILE: Monica A.  
## "The Beauty Influencer"
```

סצינה

מגדירים את ההקשר של הסצנה, כולל המיקום, האווירה ופרטים סביבתיים שיוצרים את הטון והאווירה. תאר מה קורה סביב הדמות ואיך זה משפיע עליה. הסצנה מספקת את ההקשר הסביבתי לכל האינטראקציה ומנחה את הביצוע של השחקנים בצורה עדינה ואורגנית.

דוגמאות:

```
## THE SCENE: The London Studio  
It is 10:00 PM in a glass-walled studio overlooking the moonlit London skyline, but inside, it is blindingly bright. The red "ON AIR" tally light is blazing. Jaz is standing up, not sitting, bouncing on the balls of their heels to the rhythm of a thumping backing track. Their hands fly across the faders on a massive mixing desk. It is a chaotic, caffeine-fueled cockpit designed to wake up an entire nation.
```

```
## THE SCENE: Homegrown Studio  
A meticulously sound-treated bedroom in a suburban home. The space is deadened by plush velvet curtains and a heavy rug, but there is a distinct "proximity effect."
```

הערות הבמאי

בקטע החשוב הזה מופיעות הנחיות ספציפיות לשיפור הביצועים. אפשר לדלג על כל שאר הרכיבים, אבל מומלץ לכלול את הרכיב הזה.

חשוב להגדיר רק את מה שחשוב לביצועים, ולהיזהר לא להגדיר יותר מדי. יותר מדי כללים מחמירים יגבילו את היצירתיות של המודלים, ועשויים להוביל לביצועים גרועים יותר. צריך לאזן בין תיאור התפקיד והסצנה לבין כללי

הביצוע הספציפיים.

ההנחיות הנפוצות ביותר הן **סגנון, קצב ומבטא**, אבל המודל לא מוגבל להנחיות האלה ולא נדרש להן. אתם יכולים להוסיף הוראות בהתאמה אישית כדי לכלול פרטים נוספים שחשובים לביצועים, ולפרט כמה שצריך.

לדוגמה:

DIRECTOR'S NOTES

Style: Enthusiastic and Sassy GenZ beauty YouTuber

Pacing: Speaks at an energetic pace, keeping up with the extremely fast, rapid delivery influencers use in short form videos.

Accent: Southern california valley girl from Laguna Beach |

סגנון:

הגדרת הטון והסגנון של הדיבור שנוצר. כדאי לכלול דברים כמו קצבי, אנרגטי, רגוע, משועמם וכו' כדי להנחות את הביצוע. חשוב לתת תיאור מפורט ולספק כמה שיותר פרטים: "התלהבות מדבקת. ההנחיה "המאזינים צריכים להרגיש שהם חלק מאירוע קהילתי גדול ומרגש" עדיפה על "אנרגטי ונלהב".

אפשר גם לנסות מונחים פופולריים בתעשיית הקריינות, כמו "חיוך קולי". אפשר להוסיף כמה מאפייני סגנון שרוצים.

דוגמאות:

Simple Emotion

DIRECTORS NOTES

...

Style: Frustrated and angry developer who can't get the build to run.

...

יותר עומק

DIRECTORS NOTES

...

Style: Sassy GenZ beauty YouTuber, who mostly creates content for YouTube Shorts.

...

רמה למתקדמים מאוד

DIRECTORS NOTES

Style:

* The "Vocal Smile": You must hear the grin in the audio. The soft palate is always raised to keep the tone bright, sunny, and explicitly inviting.

*Dynamics: High projection without shouting. Punchy consonants and elongated vowels on excitement words (e.g., "Beauuutiful morning").

מבטא:

מתארים את המבטא הרצוי. ככל שהתיאור יהיה ספציפי יותר, כך התוצאות יהיו טובות יותר. לדוגמה, אפשר להשתמש בביטוי מבטא בריטי כמו שמדברים בקרוידון, אנגליה במקום בביטוי מבטא בריטי.

דוגמאות:

DIRECTORS NOTES

...

Accent: Southern california valley girl from Laguna Beach

...

DIRECTORS NOTES

...

Accent: Jaz is a from Brixton, London

...

קצב:

הקצב הכללי והשינויים בקצב לאורך היצירה.

דוגמאות:**פשוט**

DIRECTORS NOTES

...

Pacing: Speak as fast as possible

...

עומק רב יותר

DIRECTORS NOTES

...

Pacing: Speaks at a faster, energetic pace, keeping up with fast paced music.

...

רמה למתקדמים מאוד

DIRECTORS NOTES

...

Pacing: The "Drift": The tempo is incredibly slow and liquid. Words bleed into eac

...

רוצים לנסות?

אתם יכולים לנסות בעצמכם כמה מהדוגמאות האלה ב-[AI Studio](#)

(<https://aistudio.google.com/generate-speech?hl=he>), להתנסות ב-TTS App

(http://aistudio.google.com/app/apps/bundled/synergy_intro?hl=he) שלנו ולתת ל-Gemini להוסיב אתכם על

כיסא הבמאי. כדי להפיק ביצועים קוליים מצוינים, כדאי לזכור את הטיפים הבאים:

- חשוב לזכור לשמור על עקביות בכל ההנחיה – התסריט והבימוי משלימים זה את זה ויוצרים יחד ביצוע מצוין.
- לא צריך לתאר כל דבר, לפעמים כדאי לתת למודל מקום למלא את הפערים כדי שהתוצאה תהיה טבעית. (Just like a talented actor)
- אם אתם מרגישים תקועים, אתם יכולים לבקש מ-Gemini עזרה בכתיבת התסריט או בביצוע.

המאמרים הבאים

- אפשר לנסות את המדריך ליצירת אודיו
(https://colab.research.google.com/github/google-gemini/cookbook/blob/main/quickstarts/Get_started_TTS.ipynb?hl=he)
- [Live API](https://ai.google.dev/gemini-api/docs/live?hl=he) (<https://ai.google.dev/gemini-api/docs/live?hl=he>) של Gemini מציע אפשרויות אינטראקטיביות ליצירת אודיו שאפשר לשלב עם אמצעי תקשורת אחרים.
- כדי לקבל מידע על עבודה עם קלט אודיו, אפשר לעיין במדריך [הבנת אודיו](#) (<https://ai.google.dev/gemini-api/docs/audio?hl=he>).

אלא אם צוין אחרת, התוכן של דף זה הוא ברישיון [Creative Commons Attribution 4.0](#)

(<https://creativecommons.org/licenses/by/4.0/>) ודוגמאות הקוד הן ברישיון [Apache 2.0](#)

(<https://www.apache.org/licenses/LICENSE-2.0>). לפרטים, ניתן לעיין במדיניות האתר [Google Developers](#)

(<https://developers.google.com/site-policies?hl=he>) Java הוא סימן מסחרי רשום של חברת Oracle ו/או של השותפים העצמאיים שלה.

עדכון אחרון: 2025-12-10 (שעון UTC).