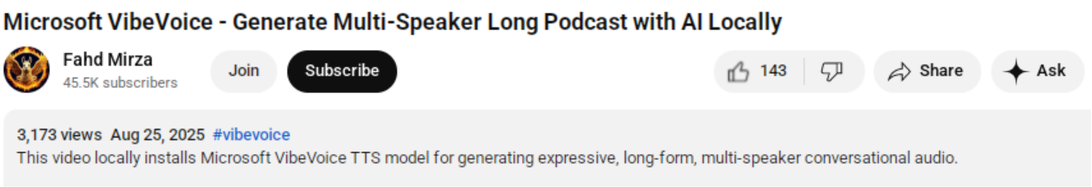
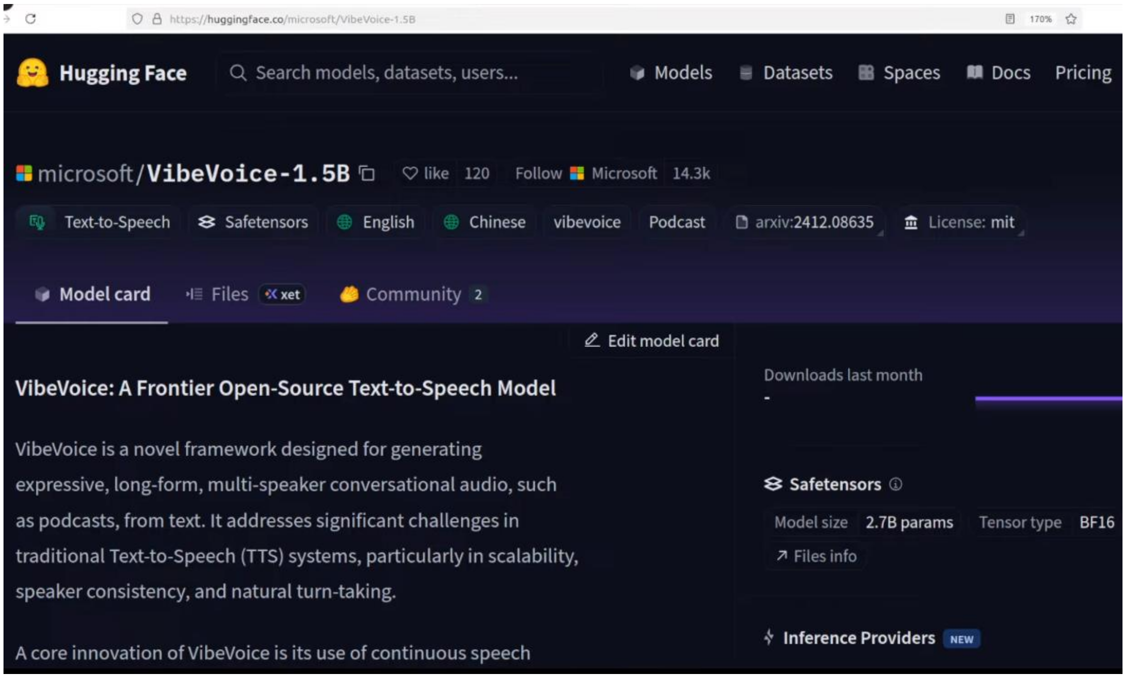


Microsoft VibeVoice - Generate Multi-Speaker Long Podcast with AI Locally

Fahd Mirza
45.5K subscribers

3,173 views Aug 25, 2025 #vibevoice
This video locally installs Microsoft VibeVoice TTS model for generating expressive, long-form, multi-speaker conversational audio.



https://huggingface.co/microsoft/VibeVoice-1.5B
170%
Hugging Face
Search models, datasets, users...
Models
Datasets
Spaces
Docs
Pricing

microsoft/VibeVoice-1.5B like 120 Follow Microsoft 14.3k

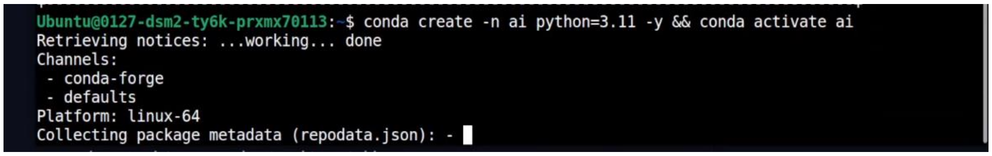
VibeVoice: A Frontier Open-Source Text-to-Speech Model

VibeVoice is a novel framework designed for generating expressive, long-form, multi-speaker conversational audio, such as podcasts, from text. It addresses significant challenges in traditional Text-to-Speech (TTS) systems, particularly in scalability, speaker consistency, and natural turn-taking.

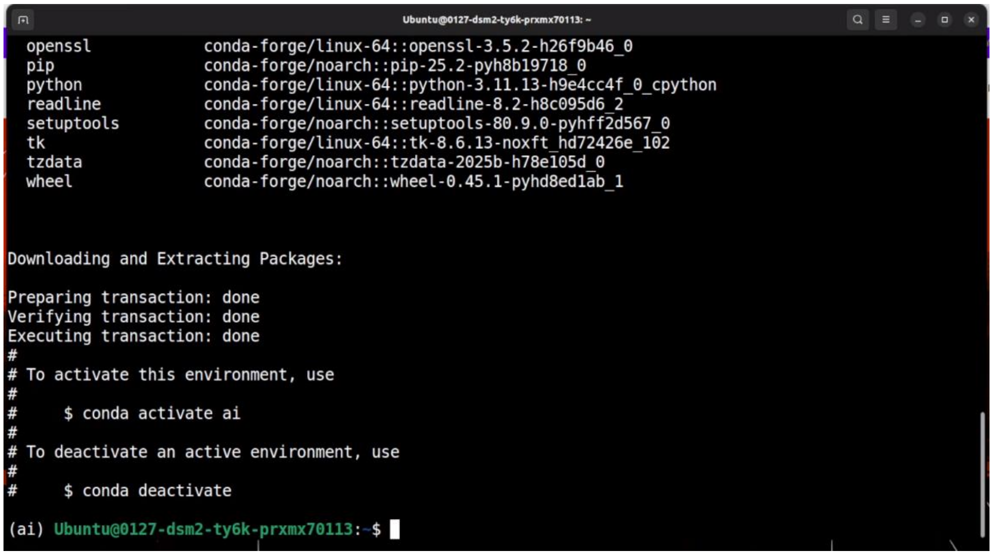
A core innovation of VibeVoice is its use of continuous speech

Downloads last month
-

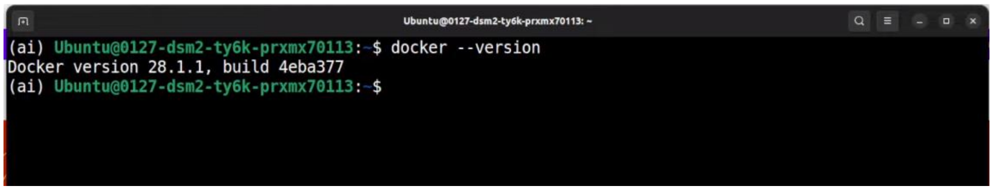
Safetensors
Model size 2.7B params
Tensor type BF16



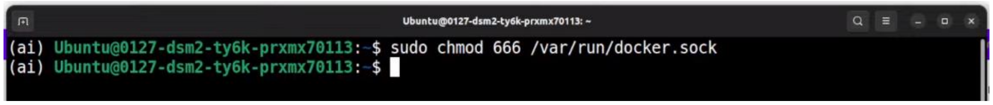
```
Ubuntu@0127-dsm2-ty6k-prxmx70113:~$ conda create -n ai python=3.11 -y && conda activate ai
Retrieving notices: ...working... done
Channels:
 - conda-forge
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): -
```



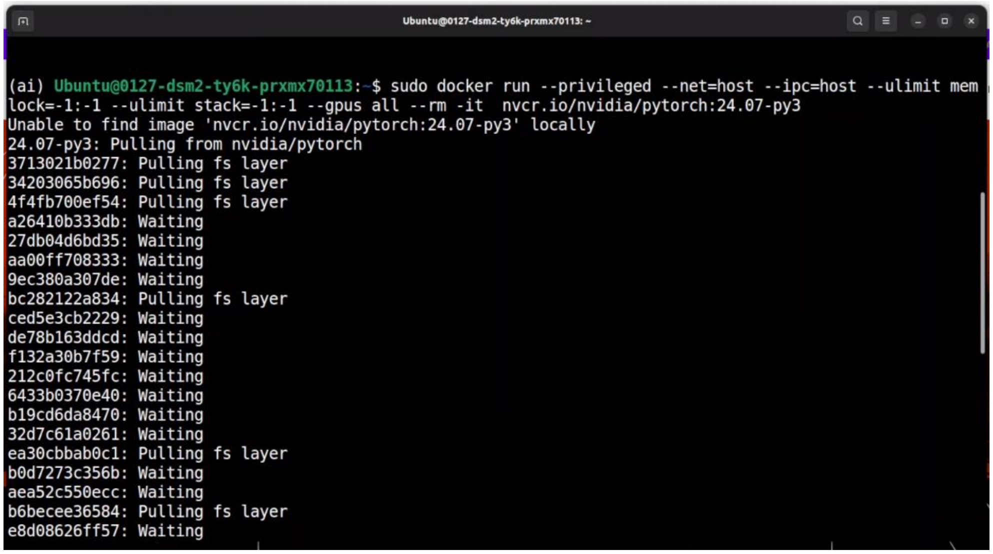
```
  openssl            conda-forge/linux-64::openssl-3.5.2-h26f9b46_0
  pip                conda-forge/noarch::pip-25.2-pyh8b19718_0
  python             conda-forge/linux-64::python-3.11.13-h9e4cc4f_0_cpython
  readline           conda-forge/linux-64::readline-8.2-h8c095d6_2
  setuptools         conda-forge/noarch::setuptools-80.9.0-pyhff2d567_0
  tk                 conda-forge/linux-64::tk-8.6.13-noxft_hd72426e_102
  tzdata             conda-forge/noarch::tzdata-2025b-h78e105d_0
  wheel              conda-forge/noarch::wheel-0.45.1-pyhd8ed1ab_1



Downloading and Extracting Packages:

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate ai
#
# To deactivate an active environment, use
#
#     $ conda deactivate

(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$
```



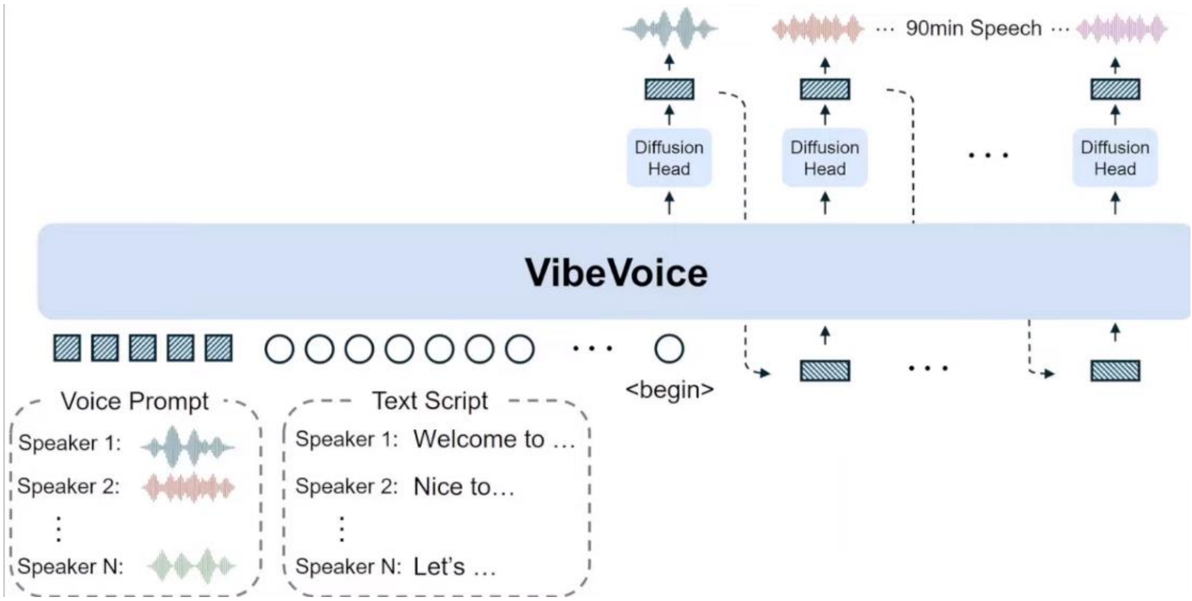
```
(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$ docker --version
Docker version 28.1.1, build 4eba377
(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$
```

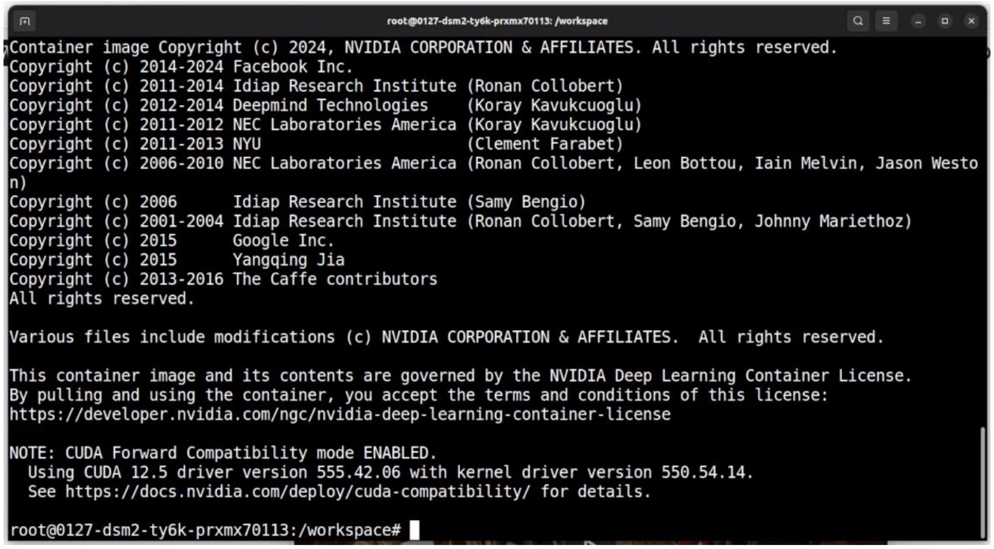


```
(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$ sudo chmod 666 /var/run/docker.sock
(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$ 
```

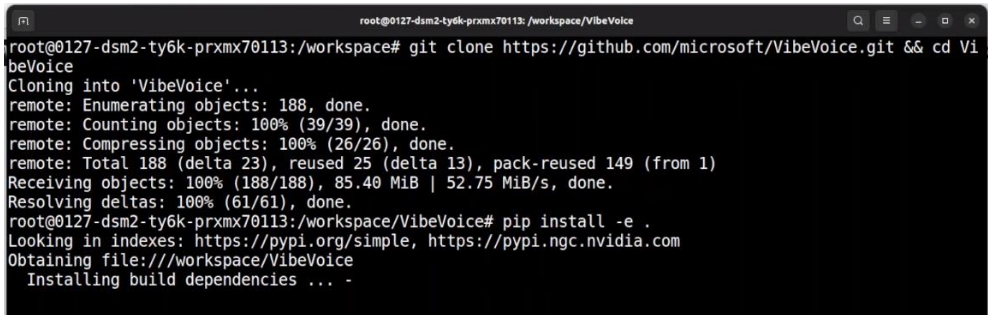


```
(ai) Ubuntu@0127-dsm2-ty6k-prxmx70113:~$ sudo docker run --privileged --net=host --ipc=host --ulimit mem
lock=-1:-1 --ulimit stack=-1:-1 --gpus all --rm -it  nvcr.io/nvidia/pytorch:24.07-py3
Unable to find image 'nvcr.io/nvidia/pytorch:24.07-py3' locally
24.07-py3: Pulling from nvidia/pytorch
3713021b0277: Pulling fs layer
34203065b696: Pulling fs layer
4f4fb700ef54: Pulling fs layer
a26410b333db: Waiting
27db04d6bd35: Waiting
aa00ff708333: Waiting
9ec380a307de: Waiting
bc282122a834: Pulling fs layer
ced5e3cb2229: Waiting
de78b163ddcd: Waiting
f132a30b7f59: Waiting
212c0fc745fc: Waiting
6433b0370e40: Waiting
b19cd6da8470: Waiting
32d7c61a0261: Waiting
ea30cbbab0c1: Pulling fs layer
b0d7273c356b: Waiting
aea52c550ecc: Waiting
b6becee36584: Pulling fs layer
e8d08626ff57: Waiting
```

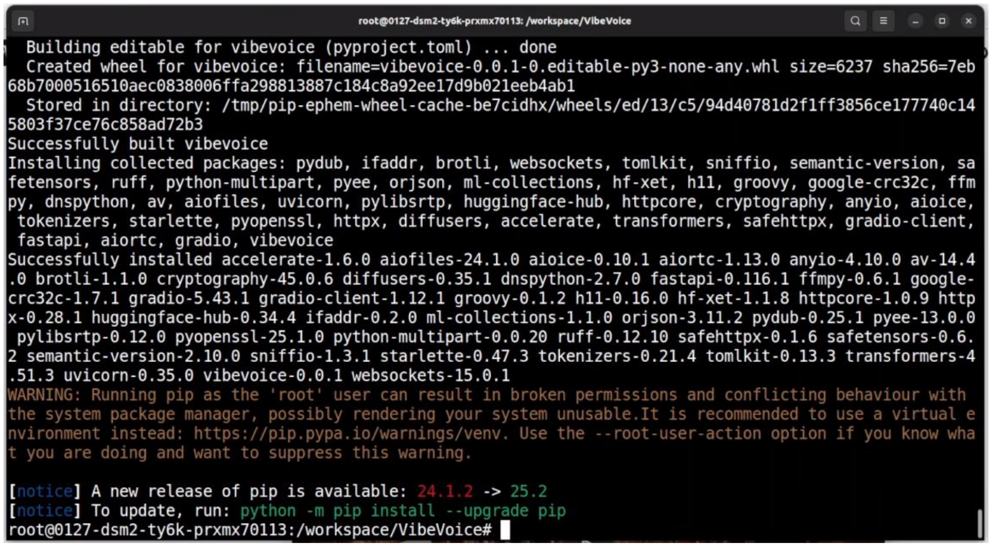




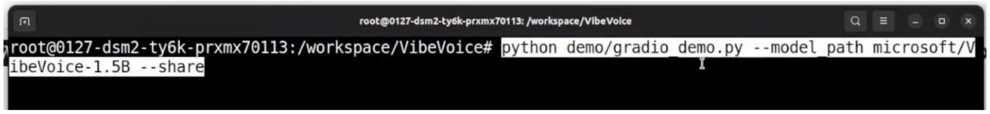
```
Container image Copyright (c) 2024, NVIDIA CORPORATION & AFFILIATES. All rights reserved.
Copyright (c) 2014-2024 Facebook Inc.
Copyright (c) 2011-2014 Idiap Research Institute (Ronan Collobert)
Copyright (c) 2012-2014 Deepmind Technologies    (Koray Kavukcuoglu)
Copyright (c) 2011-2012 NEC Laboratories America (Koray Kavukcuoglu)
Copyright (c) 2011-2013 NYU                      (Clement Farabet)
Copyright (c) 2006-2010 NEC Laboratories America (Ronan Collobert, Leon Bottou, Iain Melvin, Jason Westo
n)
Copyright (c) 2006      Idiap Research Institute (Samy Bengio)
Copyright (c) 2001-2004 Idiap Research Institute (Ronan Collobert, Samy Bengio, Johnny Mariethoz)
Copyright (c) 2015      Google Inc.
Copyright (c) 2015      Yangqing Jia
Copyright (c) 2013-2016 The Caffe contributors
All rights reserved.

Various files include modifications (c) NVIDIA CORPORATION & AFFILIATES.  All rights reserved.

This container image and its contents are governed by the NVIDIA Deep Learning Container License.
By pulling and using the container, you accept the terms and conditions of this license:
https://developer.nvidia.com/ngc/nvidia-deep-learning-container-license

NOTE: CUDA Forward Compatibility mode ENABLED.
  Using CUDA 12.5 driver version 555.42.06 with kernel driver version 550.54.14.
  See https://docs.nvidia.com/deploy/cuda-compatibility/ for details.

root@0127-dsm2-ty6k-prxmx70113:/workspace# 
```



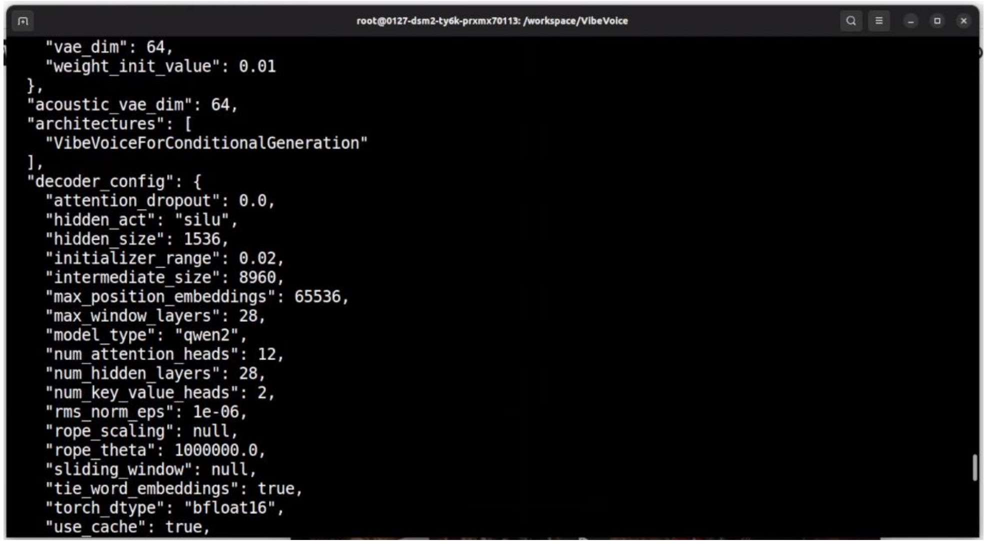
```
root@0127-dsm2-ty6k-prxmx70113:/workspace# git clone https://github.com/microsoft/VibeVoice.git && cd Vi
beVoice
Cloning into 'VibeVoice'...
remote: Enumerating objects: 188, done.
remote: Counting objects: 100% (39/39), done.
remote: Compressing objects: 100% (26/26), done.
remote: Total 188 (delta 23), reused 25 (delta 13), pack-reused 149 (from 1)
Receiving objects: 100% (188/188), 85.40 MiB | 52.75 MiB/s, done.
Resolving deltas: 100% (61/61), done.
root@0127-dsm2-ty6k-prxmx70113:/workspace/VibeVoice# pip install -e .
Looking in indexes: https://pypi.org/simple, https://pypi.ngc.nvidia.com
Obtaining file:///workspace/VibeVoice
  Installing build dependencies ... -
```



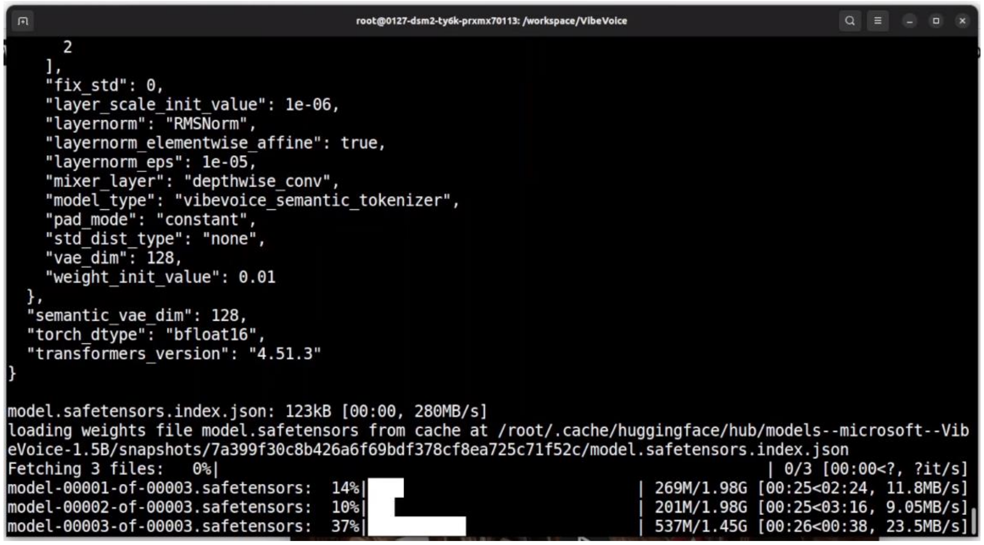
```
  Building editable for vibevoice (pyproject.toml) ... done
  Created wheel for vibevoice: filename=vibevoice-0.0.1-0.editable-py3-none-any.whl size=6237 sha256=7eb
68b7000516510aec0838006ffa298813887c184c8a92ee17d9b021eeb4ab1
  Stored in directory: /tmp/pip-ephem-wheel-cache-be7cidhx/wheels/ed/13/c5/94d40781d2f1ff3856ce177740c14
5803f37ce76c858ad72b3
Successfully built vibevoice
Installing collected packages: pydub, ifaddr, brotli, websockets, tomlkit, sniffio, semantic-version, sa
fetensors, ruff, python-multipart, pyee, orjson, ml-collections, hf-xet, h11, groovy, google-crc32c, ffm
py, dnspython, av, aiofiles, uvicorn, pylibsrtp, huggingface-hub, httpcore, cryptography, anyio, aioice,
 tokenizers, starlette, pyopenssl, httpx, diffusers, accelerate, transformers, safehttpx, gradio-client,
 fastapi, aiortc, gradio, vibevoice
Successfully installed accelerate-1.6.0 aiofiles-24.1.0 aioice-0.10.1 aiortc-1.13.0 anyio-4.10.0 av-14.4
.0 brotli-1.1.0 cryptography-45.0.6 diffusers-0.35.1 dnspython-2.7.0 fastapi-0.116.1 ffmpy-0.6.1 google-
crc32c-1.7.1 gradio-5.43.1 gradio-client-1.12.1 groovy-0.1.2 h11-0.16.0 hf-xet-1.1.8 httpcore-1.0.9 http
x-0.28.1 huggingface-hub-0.34.4 ifaddr-0.2.0 ml-collections-1.1.0 orjson-3.11.2 pydub-0.25.1 pyee-13.0.0
 pylibsrtp-0.12.0 pyopenssl-25.1.0 python-multipart-0.0.20 ruff-0.12.10 safehttpx-0.1.6 safetensors-0.6.
2 semantic-version-2.10.0 sniffio-1.3.1 starlette-0.47.3 tokenizers-0.21.4 tomlkit-0.13.3 transformers-4
.51.3 uvicorn-0.35.0 vibevoice-0.0.1 websockets-15.0.1
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with
the system package manager, possibly rendering your system unusable.It is recommended to use a virtual e
nvironment instead: https://pip.pypa.io/warnings/venv. Use the --root-user-action option if you know wha
t you are doing and want to suppress this warning.

[notice] A new release of pip is available: 24.1.2 -> 25.2
[notice] To update, run: python -m pip install --upgrade pip
root@0127-dsm2-ty6k-prxmx70113:/workspace/VibeVoice# 
```



```
root@0127-dsm2-ty6k-prxmx70113:/workspace/VibeVoice# python demo/gradio_demo.py --model_path microsoft/V
ibeVoice-1.5B --share
```

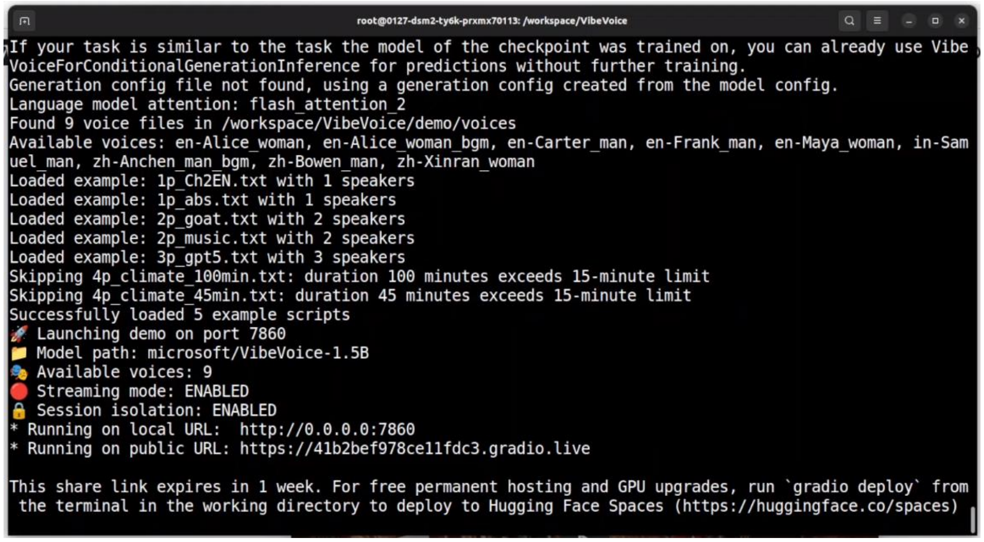


```
    "vae_dim": 64,
    "weight_init_value": 0.01
  },
  "acoustic_vae_dim": 64,
  "architectures": [
    "VibeVoiceForConditionalGeneration"
  ],
  "decoder_config": {
    "attention_dropout": 0.0,
    "hidden_act": "silu",
    "hidden_size": 1536,
    "initializer_range": 0.02,
    "intermediate_size": 8960,
    "max_position_embeddings": 65536,
    "max_window_layers": 28,
    "model_type": "qwen2",
    "num_attention_heads": 12,
    "num_hidden_layers": 28,
    "num_key_value_heads": 2,
    "rms_norm_eps": 1e-06,
    "rope_scaling": null,
    "rope_theta": 1000000.0,
    "sliding_window": null,
    "tie_word_embeddings": true,
    "torch_dtype": "bfloat16",
    "use_cache": true,
```



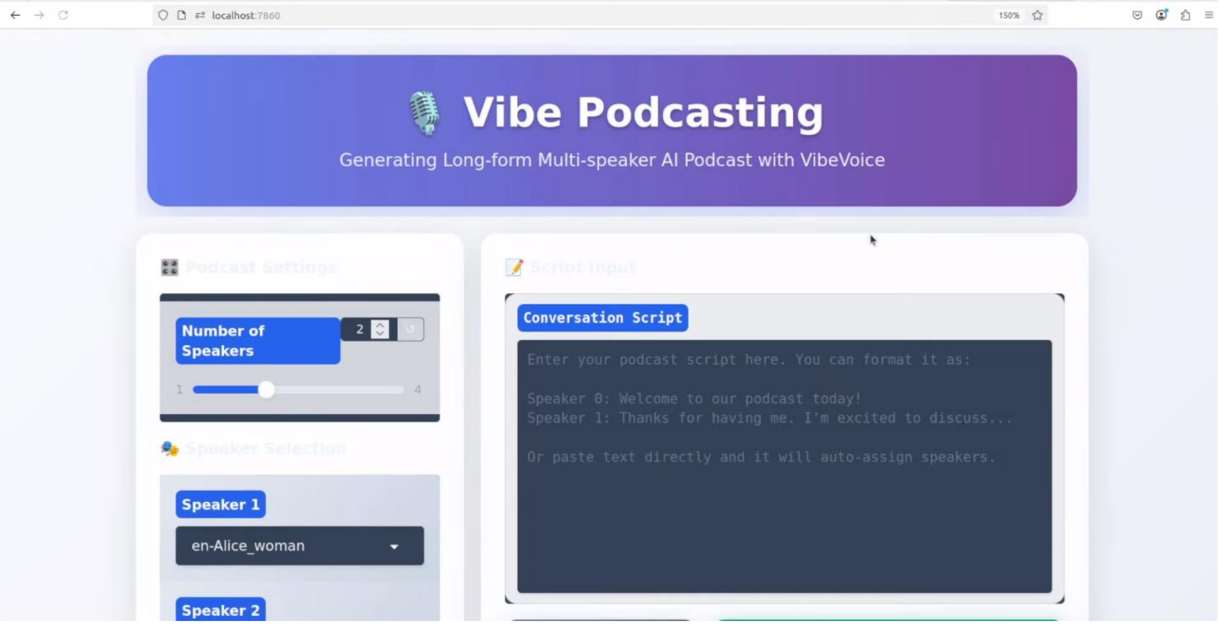
root@0127-dsm2-ty6k-prxmx70113: /workspace/VibeVoice

```
      2
    ],
    "fix_std": 0,
    "layer_scale_init_value": 1e-06,
    "layernorm": "RMSNorm",
    "layernorm_elementwise_affine": true,
    "layernorm_eps": 1e-05,
    "mixer_layer": "depthwise_conv",
    "model_type": "vibevoice_semantic_tokenizer",
    "pad_mode": "constant",
    "std_dist_type": "none",
    "vae_dim": 128,
    "weight_init_value": 0.01
  },
  "semantic_vae_dim": 128,
  "torch_dtype": "bfloat16",
  "transformers_version": "4.51.3"
}

model.safetensors.index.json: 123kB [00:00, 280MB/s]
loading weights file model.safetensors from cache at /root/.cache/huggingface/hub/models--microsoft--Vib
eVoice-1.5B/snapshots/7a399f30c8b426a6f69bdf378cf8ea725c71f52c/model.safetensors.index.json
Fetching 3 files:   0%|                                                       | 0/3 [00:00<?, ?it/s]
model-00001-of-00003.safetensors:  14%|██                                | 269M/1.98G [00:25<02:24, 11.8MB/s]
model-00002-of-00003.safetensors:  10%|█                                 | 201M/1.98G [00:25<03:16, 9.05MB/s]
model-00003-of-00003.safetensors:  37%|██████                            | 537M/1.45G [00:26<00:38, 23.5MB/s]
```



root@0127-dsm2-ty6k-prxmx70113: /workspace/VibeVoice

```
If your task is similar to the task the model of the checkpoint was trained on, you can already use Vibe
VoiceForConditionalGenerationInference for predictions without further training.
Generation config file not found, using a generation config created from the model config.
Language model attention: flash_attention_2
Found 9 voice files in /workspace/VibeVoice/demo/voices
Available voices: en-Alice_woman, en-Alice_woman_bgm, en-Carter_man, en-Frank_man, en-Maya_woman, in-Sam
uel_man, zh-Anchen_man_bgm, zh-Bowen_man, zh-Xinran_woman
Loaded example: 1p_Ch2EN.txt with 1 speakers
Loaded example: 1p_abs.txt with 1 speakers
Loaded example: 2p_goat.txt with 2 speakers
Loaded example: 2p_music.txt with 2 speakers
Loaded example: 3p_gpt5.txt with 3 speakers
Skipping 4p_climate_100min.txt: duration 100 minutes exceeds 15-minute limit
Skipping 4p_climate_45min.txt: duration 45 minutes exceeds 15-minute limit
Successfully loaded 5 example scripts
🚀 Launching demo on port 7860
📁 Model path: microsoft/VibeVoice-1.5B
🎭 Available voices: 9
🔴 Streaming mode: ENABLED
🔒 Session isolation: ENABLED
* Running on local URL:  http://0.0.0.0:7860
* Running on public URL: https://41b2bef978ce11fdc3.gradio.live

This share link expires in 1 week. For free permanent hosting and GPU upgrades, run `gradio deploy` from
 the terminal in the working directory to deploy to Hugging Face Spaces (https://huggingface.co/spaces)
```



localhost:7860

🎙️ Vibe Podcasting

Generating Long-form Multi-speaker AI Podcast with VibeVoice

🎛️ Podcast Settings

Number of Speakers: 2

📝 Script Input

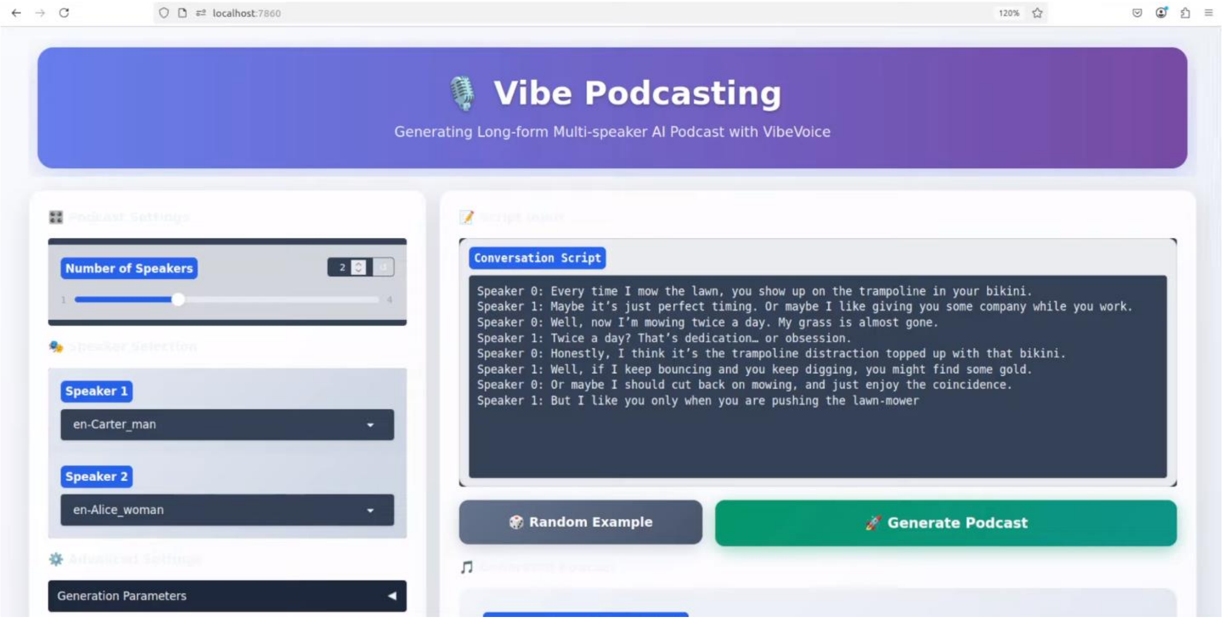
Conversation Script

```
Enter your podcast script here. You can format it as:

Speaker 0: Welcome to our podcast today!
Speaker 1: Thanks for having me. I'm excited to discuss...

Or paste text directly and it will auto-assign speakers.
```

🎭 Speaker Selection

Speaker 1: en-Alice_woman

Speaker 2



🎙️ Vibe Podcasting

Generating Long-form Multi-speaker AI Podcast with VibeVoice

Podcast Settings

Number of Speakers: 2

Speaker Selection

Speaker 1: en-Carter_man

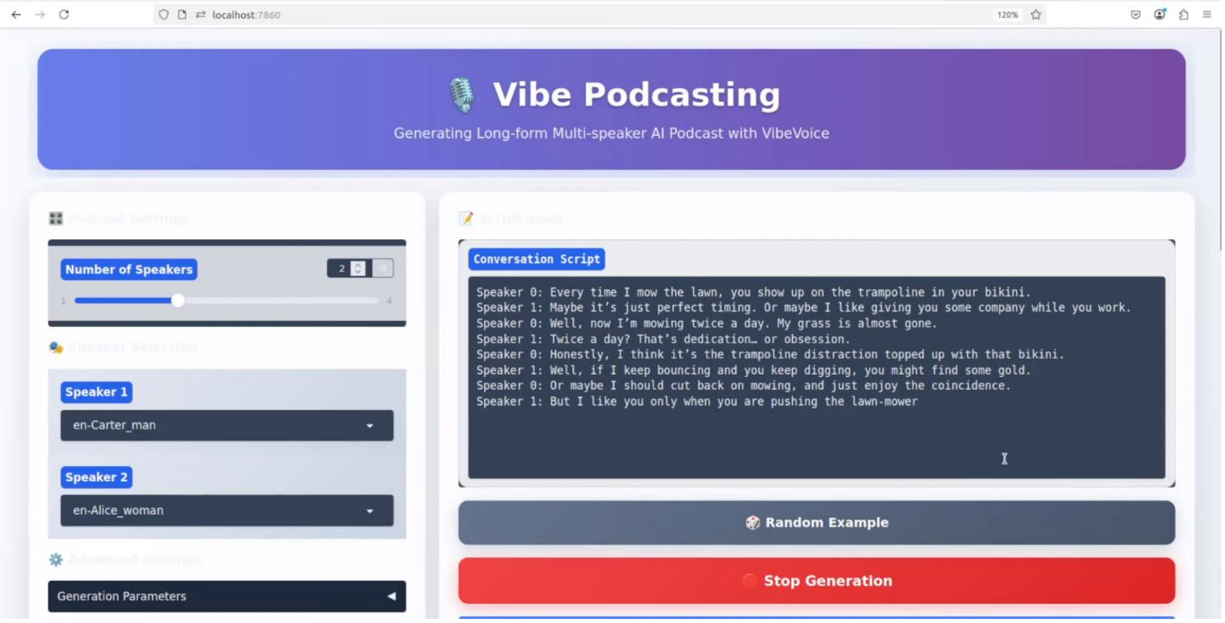
Speaker 2: en-Alice_woman

Advanced Settings

📝 Script Input

Conversation Script

```
Speaker 0: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 1: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 0: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 1: Twice a day? That’s dedication… or obsession.
Speaker 0: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 1: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 0: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 1: But I like you only when you are pushing the lawn-mower
```



🎙️ Vibe Podcasting

Generating Long-form Multi-speaker AI Podcast with VibeVoice

Podcast Settings

Number of Speakers: 2

Speaker Selection

Speaker 1: en-Carter_man

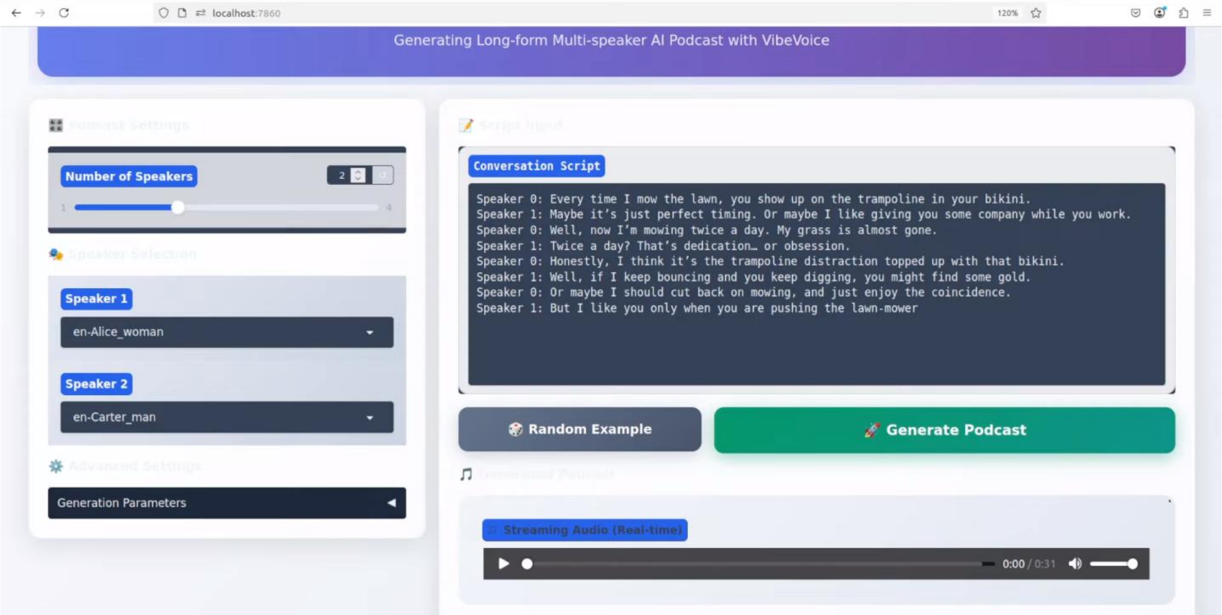
Speaker 2: en-Alice_woman

Advanced Settings

📝 Script Input

Conversation Script

```
Speaker 0: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 1: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 0: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 1: Twice a day? That’s dedication… or obsession.
Speaker 0: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 1: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 0: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 1: But I like you only when you are pushing the lawn-mower
```



Generating Long-form Multi-speaker AI Podcast with VibeVoice

Podcast Settings

Number of Speakers: 2

Speaker Selection

Speaker 1: en-Alice_woman

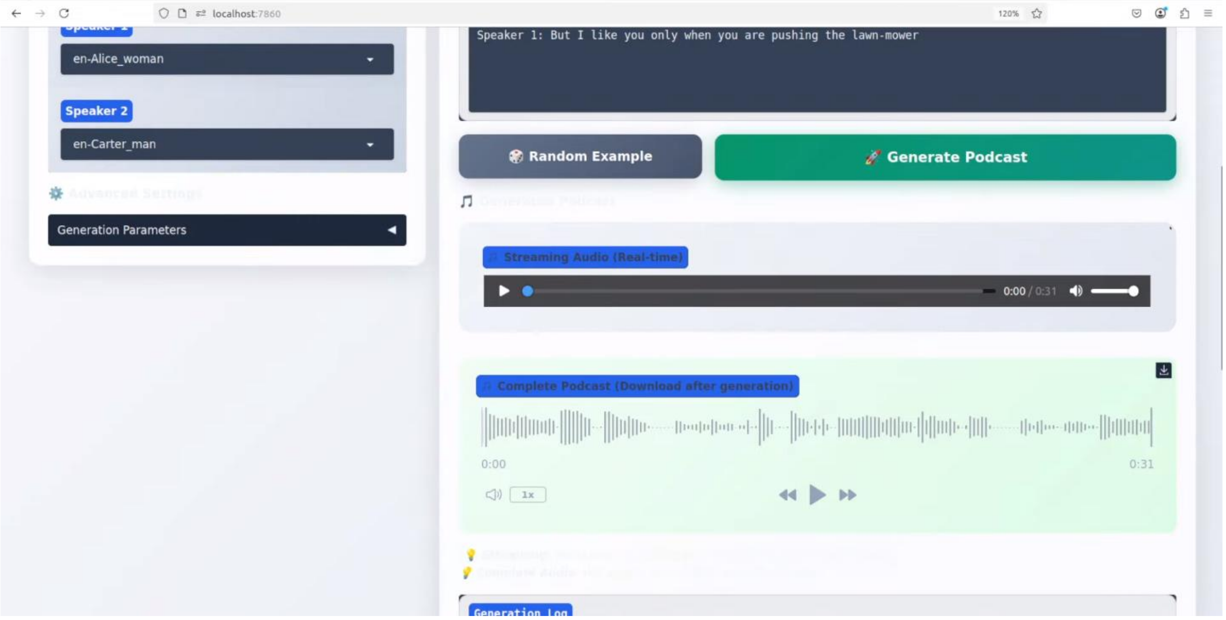
Speaker 2: en-Carter_man

Advanced Settings

📝 Script Input

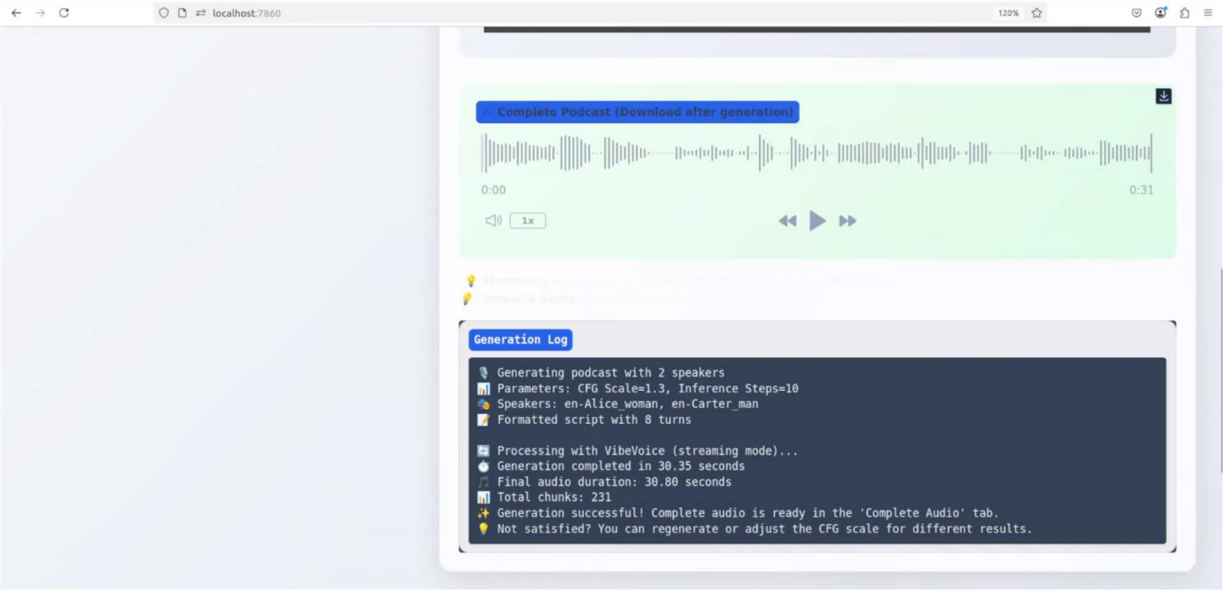
Conversation Script

```
Speaker 0: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 1: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 0: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 1: Twice a day? That’s dedication… or obsession.
Speaker 0: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 1: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 0: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 1: But I like you only when you are pushing the lawn-mower
```

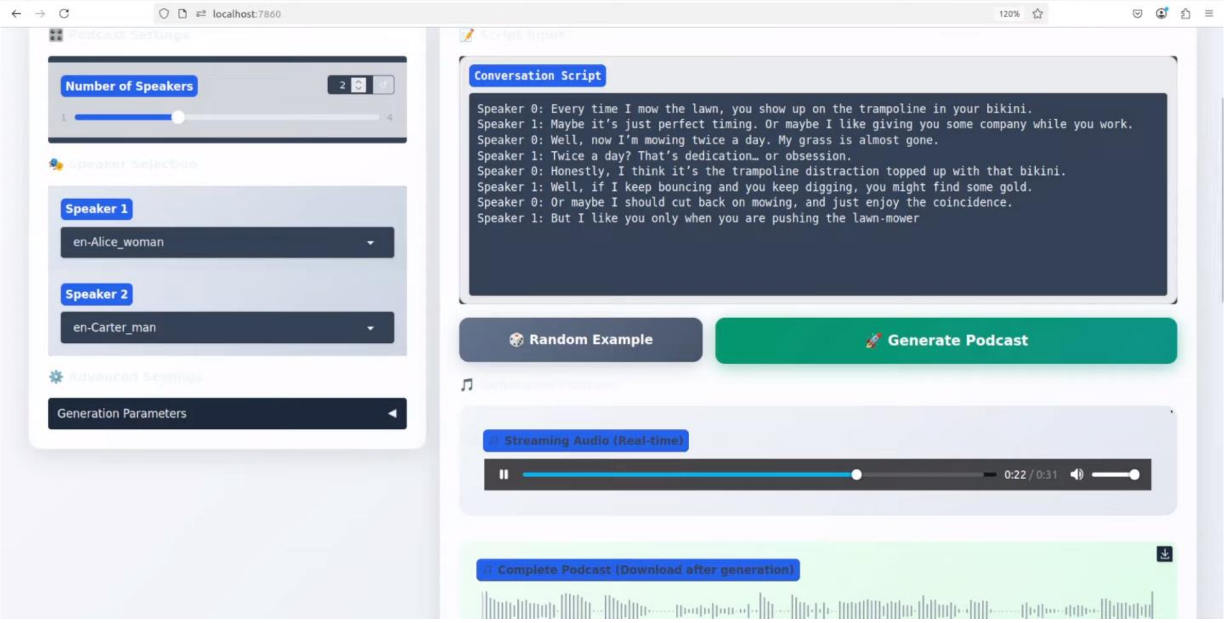


Speaker 2: en-Carter_man

```
Speaker 1: But I like you only when you are pushing the lawn-mower
```



```
🎙 Generating podcast with 2 speakers
📊 Parameters: CFG Scale=1.3, Inference Steps=10
🎭 Speakers: en-Alice_woman, en-Carter_man
📝 Formatted script with 8 turns

🔄 Processing with VibeVoice (streaming mode)...
⏱ Generation completed in 30.35 seconds
🎵 Final audio duration: 30.80 seconds
📊 Total chunks: 231
✨ Generation successful! Complete audio is ready in the 'Complete Audio' tab.
💡 Not satisfied? You can regenerate or adjust the CFG scale for different results.
```

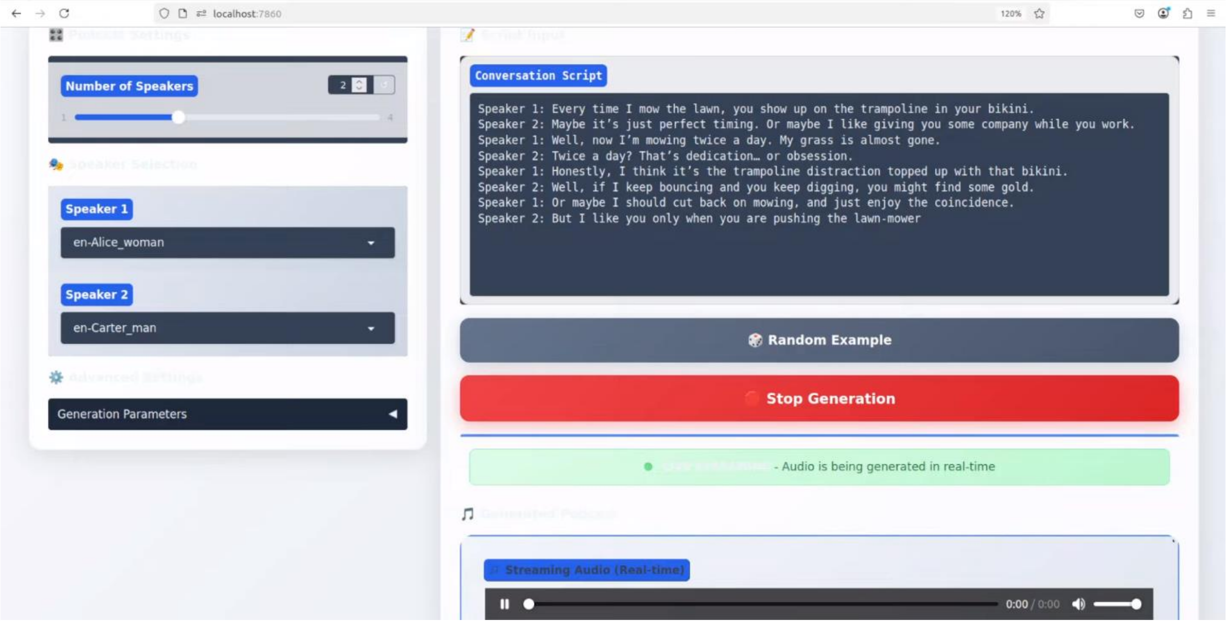


Number of Speakers: 2

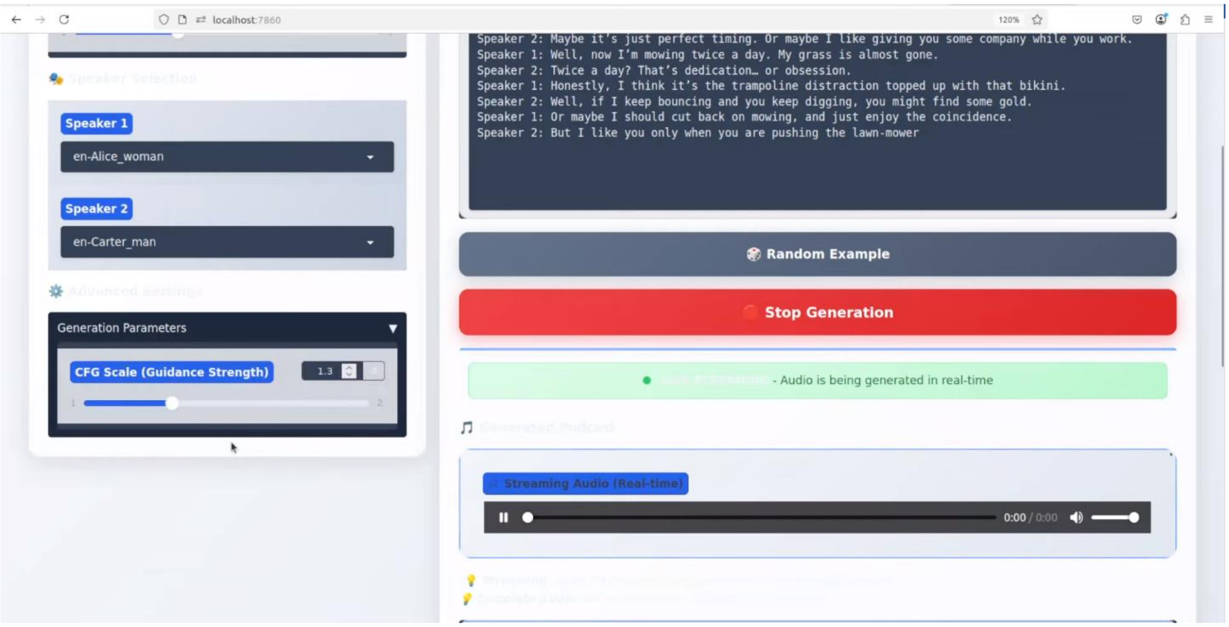
Speaker 1: en-Alice_woman

Speaker 2: en-Carter_man

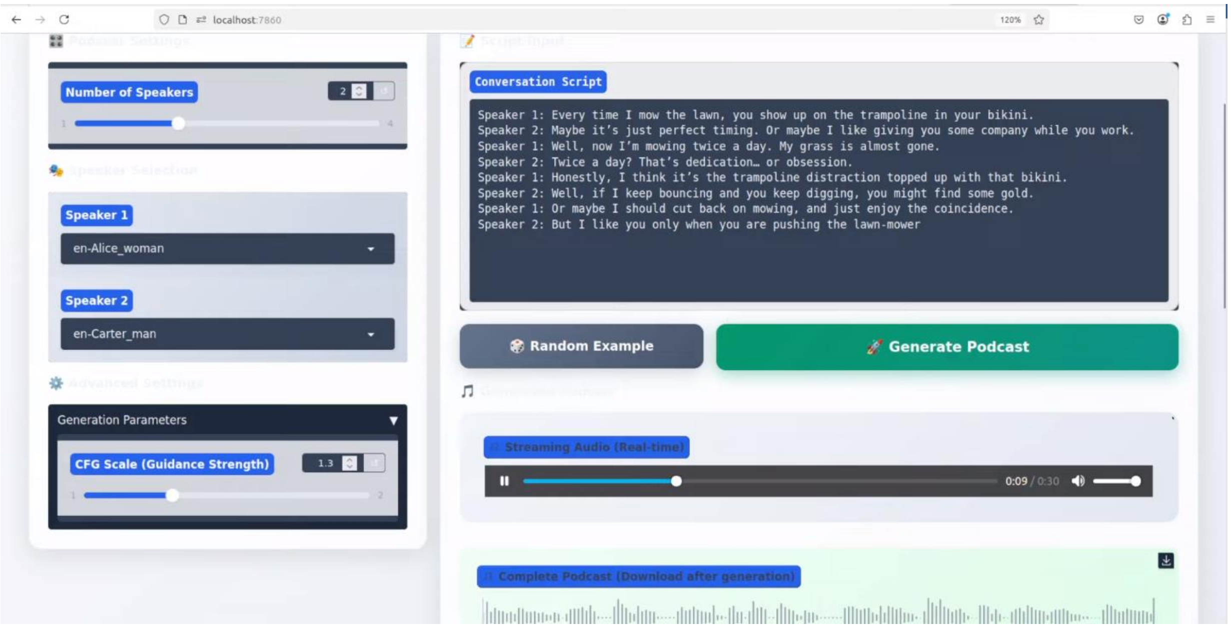
```
Speaker 0: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 1: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 0: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 1: Twice a day? That's dedication… or obsession.
Speaker 0: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 1: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 0: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 1: But I like you only when you are pushing the lawn-mower
```



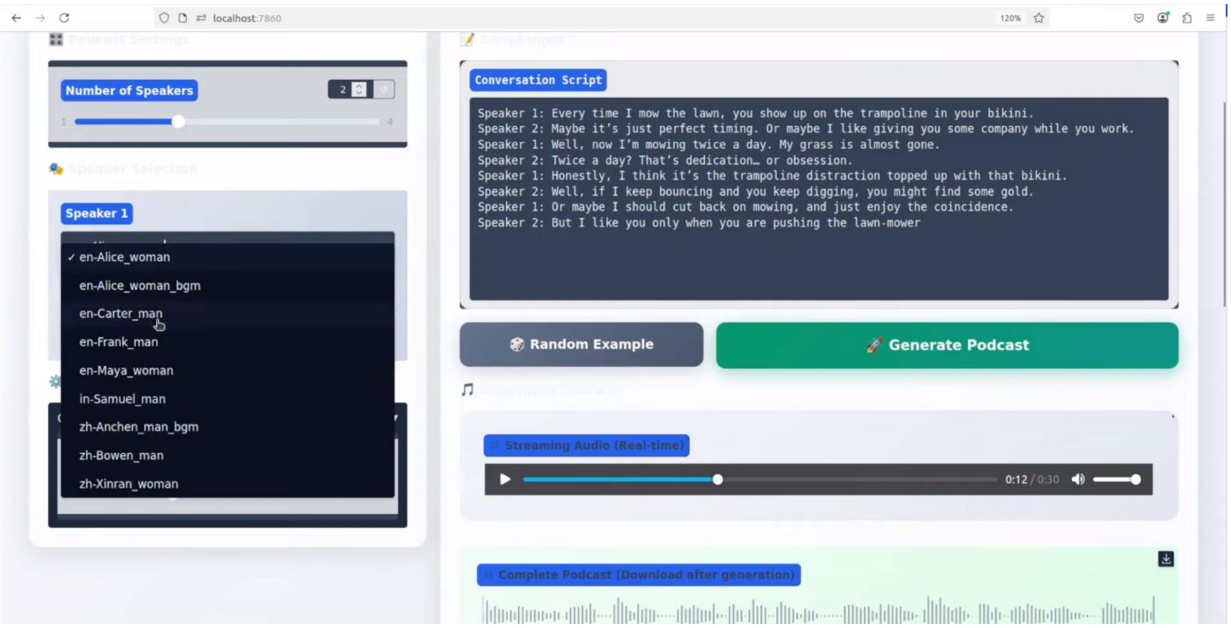
```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That’s dedication… or obsession.
Speaker 1: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```



```
Speaker 2: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That’s dedication… or obsession.
Speaker 1: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```



```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it’s just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I’m mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That’s dedication… or obsession.
Speaker 1: Honestly, I think it’s the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```

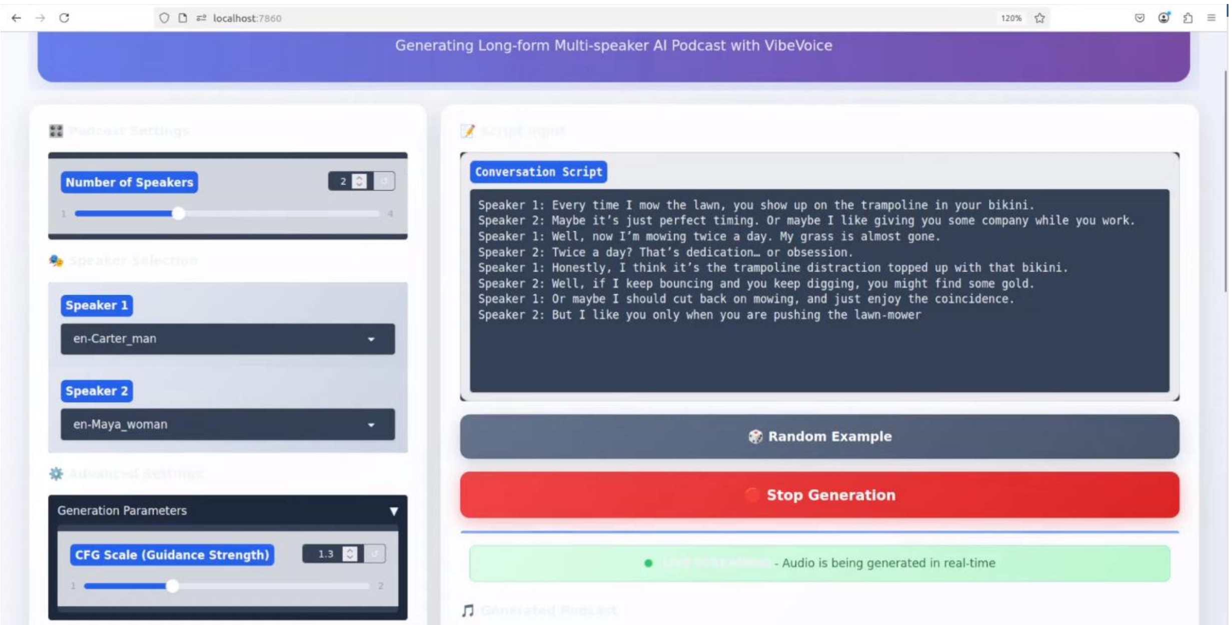


Number of Speakers: 2

Speaker 1

Conversation Script

```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That's dedication… or obsession.
Speaker 1: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```



Generating Long-form Multi-speaker AI Podcast with VibeVoice

Number of Speakers: 2

Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

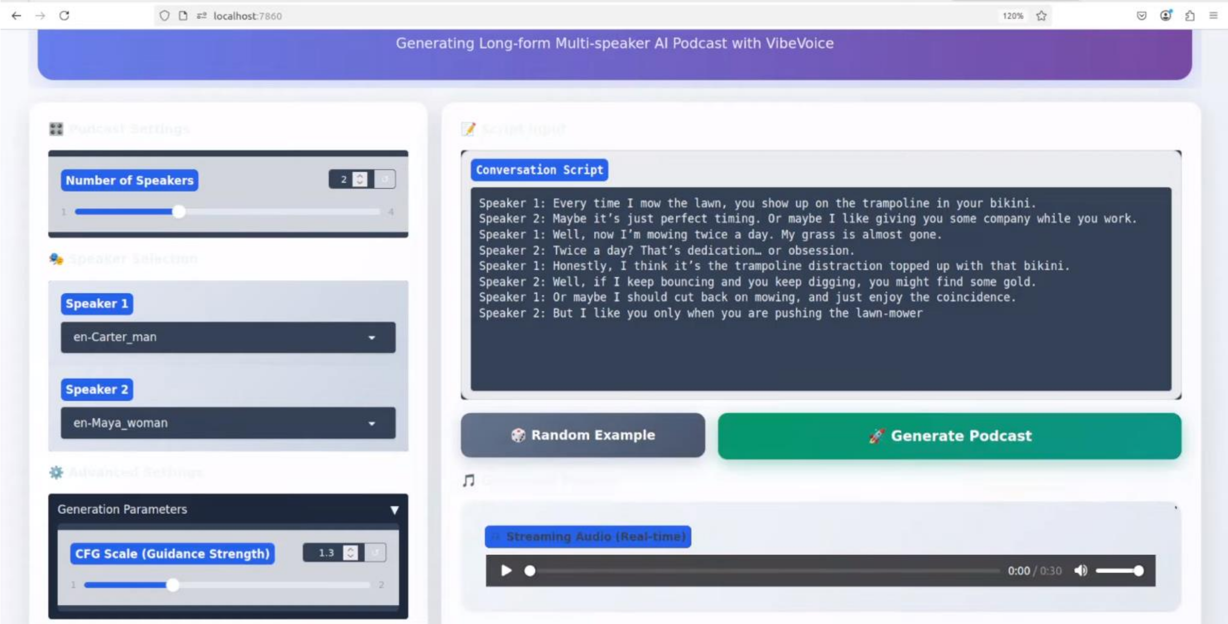
Generation Parameters

CFG Scale (Guidance Strength): 1.3

Conversation Script

```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That's dedication… or obsession.
Speaker 1: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```

● Live Streaming - Audio is being generated in real-time



Generating Long-form Multi-speaker AI Podcast with VibeVoice

Number of Speakers: 2

Speaker 1: en-Carter_man

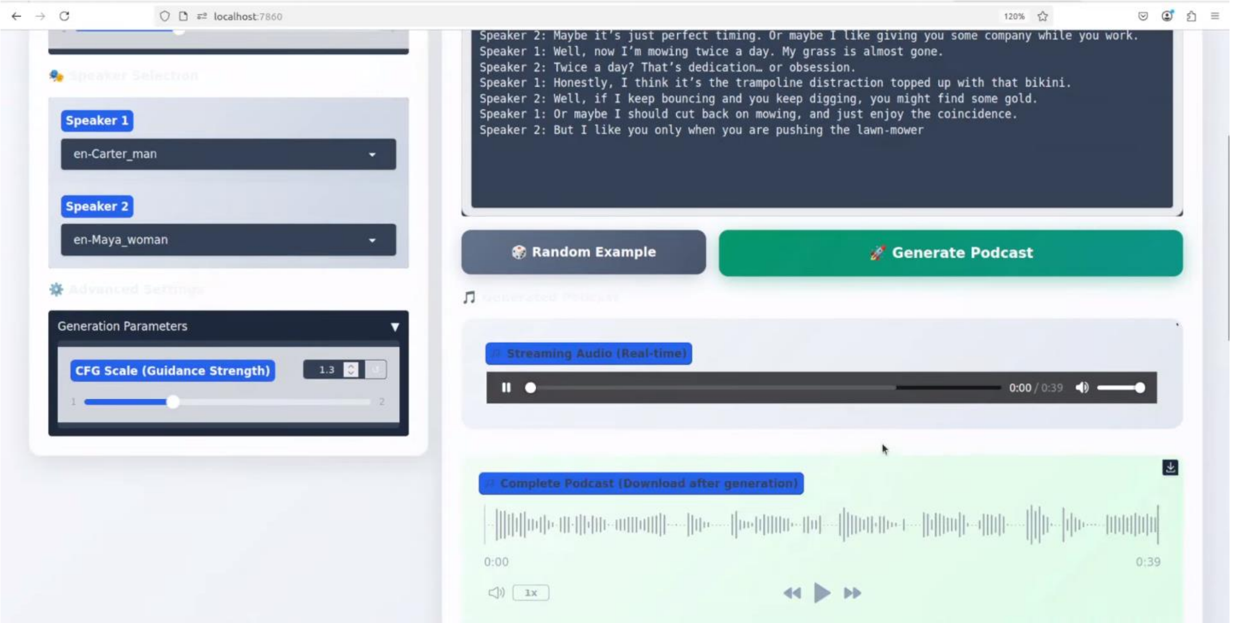
Speaker 2: en-Maya_woman

Generation Parameters

CFG Scale (Guidance Strength): 1.3

Conversation Script

```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That's dedication… or obsession.
Speaker 1: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```

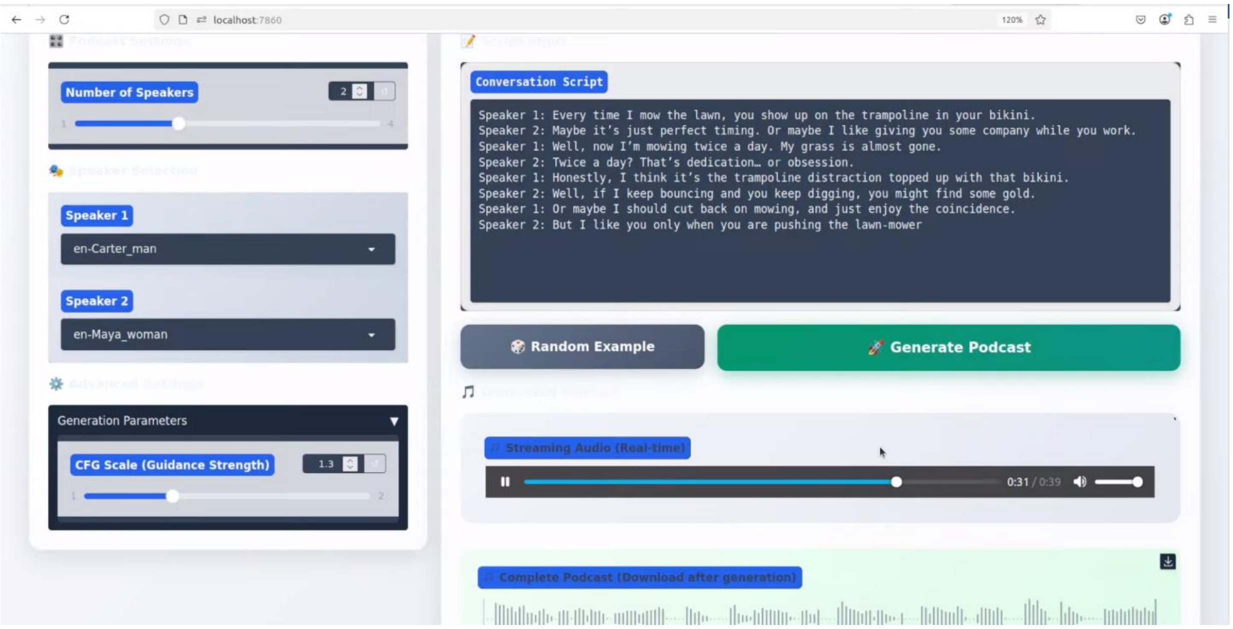


Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

CFG Scale (Guidance Strength): 1.3

```
Speaker 2: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That's dedication… or obsession.
Speaker 1: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```



Number of Speakers: 2

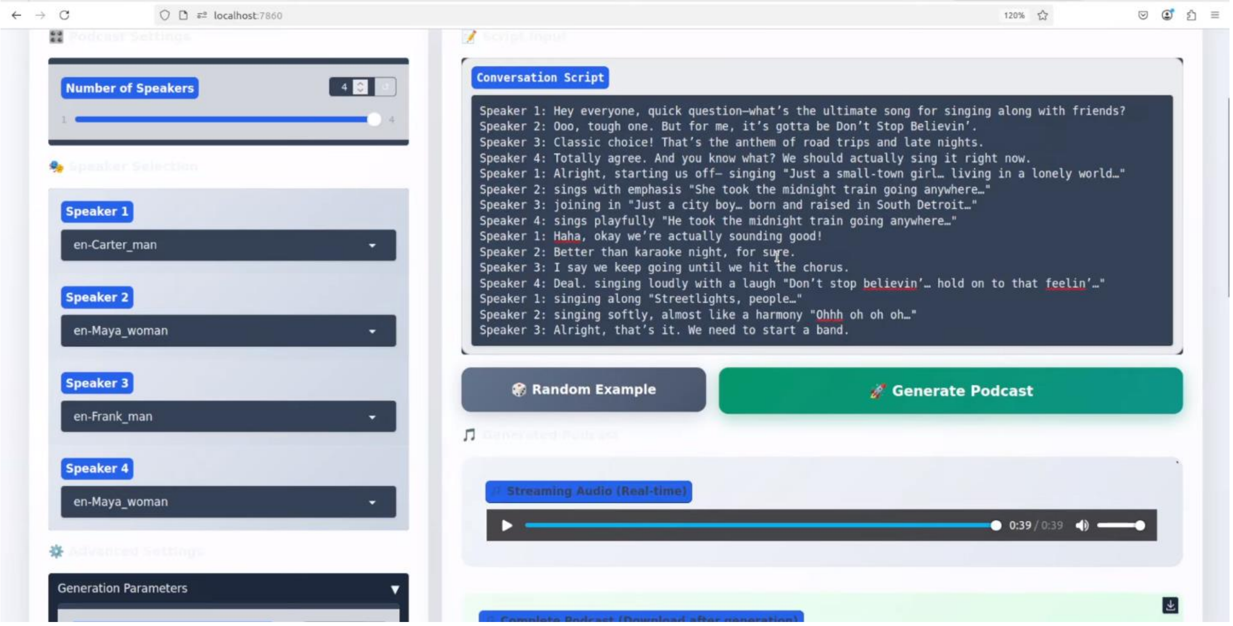
Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

CFG Scale (Guidance Strength): 1.3

Conversation Script

```
Speaker 1: Every time I mow the lawn, you show up on the trampoline in your bikini.
Speaker 2: Maybe it's just perfect timing. Or maybe I like giving you some company while you work.
Speaker 1: Well, now I'm mowing twice a day. My grass is almost gone.
Speaker 2: Twice a day? That's dedication… or obsession.
Speaker 1: Honestly, I think it's the trampoline distraction topped up with that bikini.
Speaker 2: Well, if I keep bouncing and you keep digging, you might find some gold.
Speaker 1: Or maybe I should cut back on mowing, and just enjoy the coincidence.
Speaker 2: But I like you only when you are pushing the lawn-mower
```



Number of Speakers: 4

Speaker 1: en-Carter_man

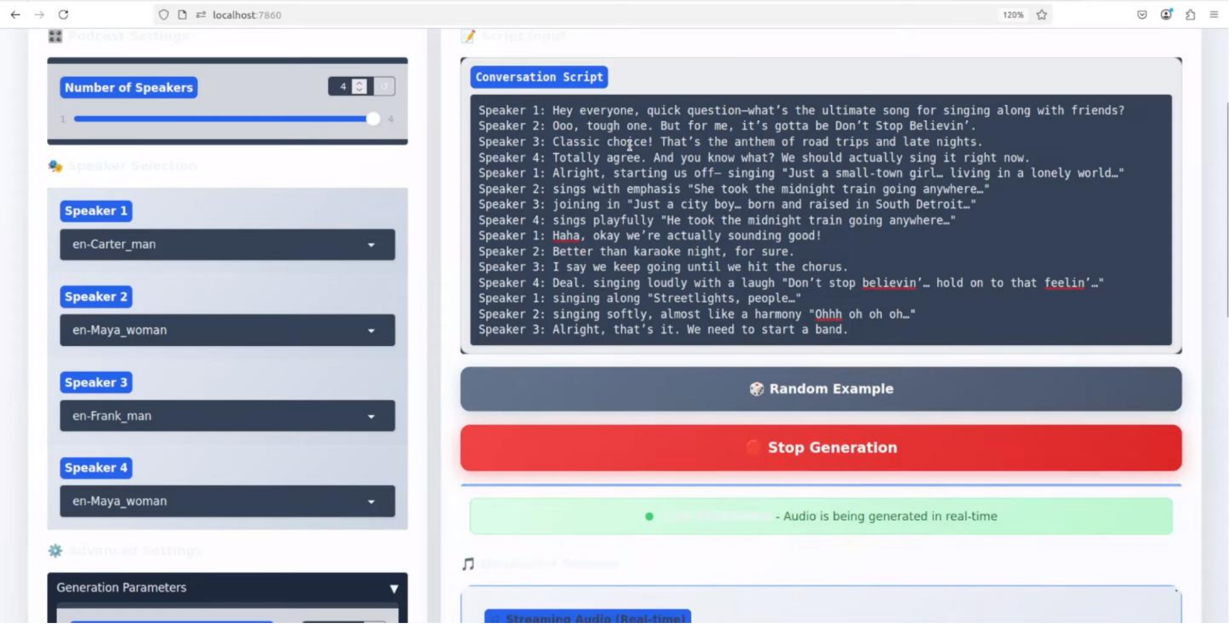
Speaker 2: en-Maya_woman

Speaker 3: en-Frank_man

Speaker 4: en-Maya_woman

Conversation Script

```
Speaker 1: Hey everyone, quick question—what's the ultimate song for singing along with friends?
Speaker 2: Ooo, tough one. But for me, it's gotta be Don't Stop Believin'.
Speaker 3: Classic choice! That's the anthem of road trips and late nights.
Speaker 4: Totally agree. And you know what? We should actually sing it right now.
Speaker 1: Alright, starting us off— singing "Just a small-town girl… living in a lonely world…"
Speaker 2: sings with emphasis "She took the midnight train going anywhere…"
Speaker 3: joining in "Just a city boy… born and raised in South Detroit…"
Speaker 4: sings playfully "He took the midnight train going anywhere…"
Speaker 1: Haha, okay we're actually sounding good!
Speaker 2: Better than karaoke night, for sure.
Speaker 3: I say we keep going until we hit the chorus.
Speaker 4: Deal. singing loudly with a laugh "Don't stop believin'… hold on to that feelin'…"
Speaker 1: singing along "Streetlights, people…"
Speaker 2: singing softly, almost like a harmony "Ohhh oh oh oh…"
Speaker 3: Alright, that's it. We need to start a band.
```



Number of Speakers: 4

Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

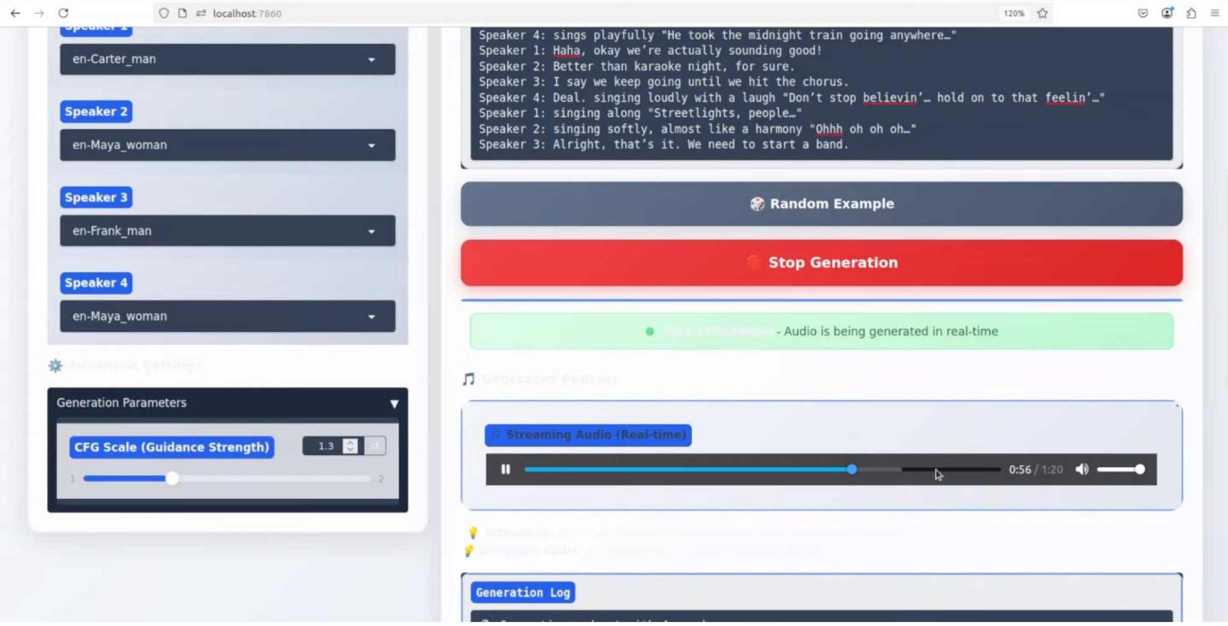
Speaker 3: en-Frank_man

Speaker 4: en-Maya_woman

Conversation Script

```
Speaker 1: Hey everyone, quick question—what’s the ultimate song for singing along with friends?
Speaker 2: Ooo, tough one. But for me, it’s gotta be Don’t Stop Believin’.
Speaker 3: Classic choice! That’s the anthem of road trips and late nights.
Speaker 4: Totally agree. And you know what? We should actually sing it right now.
Speaker 1: Alright, starting us off— singing "Just a small-town girl… living in a lonely world…"
Speaker 2: sings with emphasis "She took the midnight train going anywhere…"
Speaker 3: joining in "Just a city boy… born and raised in South Detroit…"
Speaker 4: sings playfully "He took the midnight train going anywhere…"
Speaker 1: Haha, okay we’re actually sounding good!
Speaker 2: Better than karaoke night, for sure.
Speaker 3: I say we keep going until we hit the chorus.
Speaker 4: Deal. singing loudly with a laugh "Don’t stop believin’… hold on to that feelin’…"
Speaker 1: singing along "Streetlights, people…"
Speaker 2: singing softly, almost like a harmony "Ohhh oh oh oh…"
Speaker 3: Alright, that’s it. We need to start a band.
```

- Audio is being generated in real-time



Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

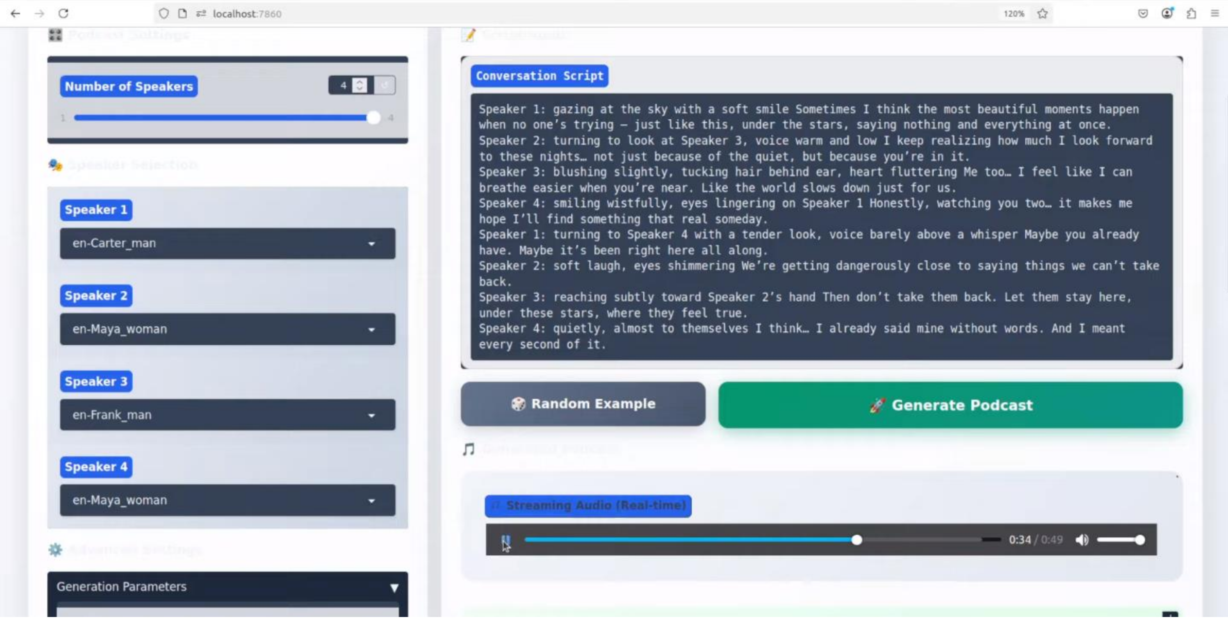
Speaker 3: en-Frank_man

Speaker 4: en-Maya_woman

CFG Scale (Guidance Strength): 1.3

```
Speaker 4: sings playfully "He took the midnight train going anywhere…"
Speaker 1: Haha, okay we’re actually sounding good!
Speaker 2: Better than karaoke night, for sure.
Speaker 3: I say we keep going until we hit the chorus.
Speaker 4: Deal. singing loudly with a laugh "Don’t stop believin’… hold on to that feelin’…"
Speaker 1: singing along "Streetlights, people…"
Speaker 2: singing softly, almost like a harmony "Ohhh oh oh oh…"
Speaker 3: Alright, that’s it. We need to start a band.
```

- Audio is being generated in real-time



Number of Speakers: 4

Speaker 1: en-Carter_man

Speaker 2: en-Maya_woman

Speaker 3: en-Frank_man

Speaker 4: en-Maya_woman

Conversation Script

```
Speaker 1: gazing at the sky with a soft smile Sometimes I think the most beautiful moments happen
when no one’s trying — just like this, under the stars, saying nothing and everything at once.
Speaker 2: turning to look at Speaker 3, voice warm and low I keep realizing how much I look forward
to these nights… not just because of the quiet, but because you’re in it.
Speaker 3: blushing slightly, tucking hair behind ear, heart fluttering Me too… I feel like I can
breathe easier when you’re near. Like the world slows down just for us.
Speaker 4: smiling wistfully, eyes lingering on Speaker 1 Honestly, watching you two… it makes me
hope I’ll find something that real someday.
Speaker 1: turning to Speaker 4 with a tender look, voice barely above a whisper Maybe you already
have. Maybe it’s been right here all along.
Speaker 2: soft laugh, eyes shimmering We’re getting dangerously close to saying things we can’t take
back.
Speaker 3: reaching subtly toward Speaker 2’s hand Then don’t take them back. Let them stay here,
under these stars, where they feel true.
Speaker 4: quietly, almost to themselves I think… I already said mine without words. And I meant
every second of it.
```