



Lab 1 - Audiobox

Create an audio story with AI voices and sounds

Lab Objective

The goal of this lab is to equip students with the skills to use natural language prompts to generate various types of audio, including voices, sound effects, and soundscapes. Through hands-on experience with Meta's AudioBox, students will learn how AI models interpret and respond to prompts, ultimately creating a cohesive audio story.

Background

AudioBox is Meta's cutting-edge research model for audio generation. It allows users to generate custom audio by combining voice inputs with natural language text prompts. The AudioBox model family includes specialist models like AudioBox Speech and AudioBox Sound, all built upon the shared self-supervised model, AudioBox SSL. This lab will introduce you to the capabilities of AudioBox and guide you in creating your audio projects.

Part 1 Learn

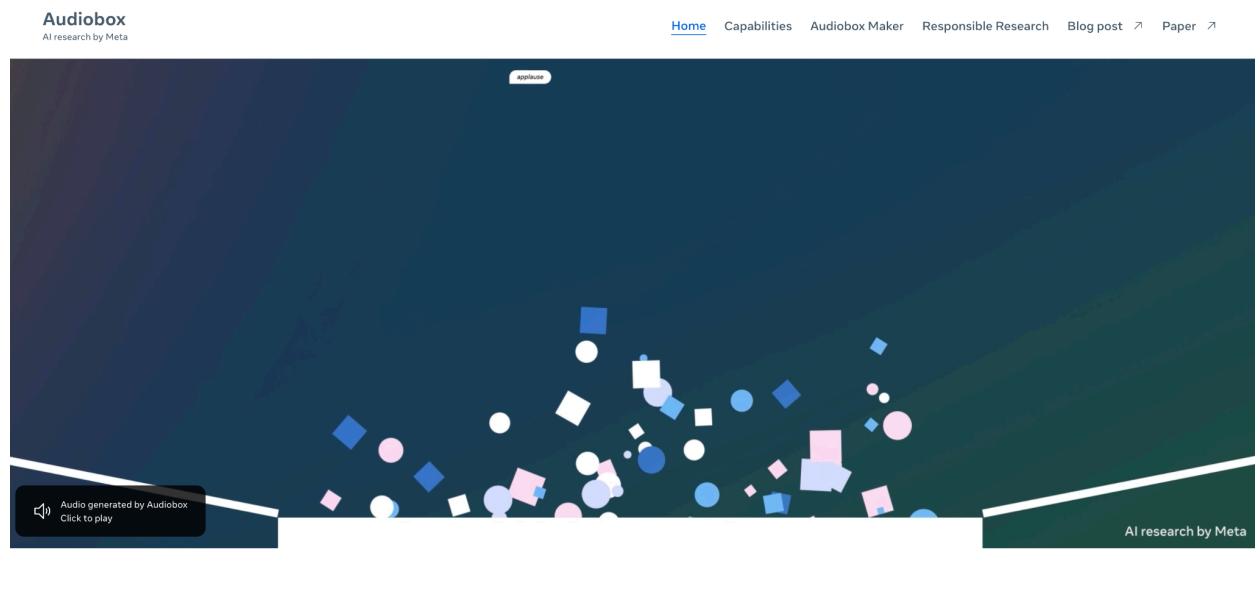
First, we are going to learn the effective prompting of AudioBox, go to the following website: <https://audiobox.metademolab.com/> At the site press accept and agree to the terms of use and Cookies Policy.

Allow the use of cookies from Meta on this browser? To find out more about the use of cookies, see our [Privacy Policy](#) and [Cookies Policy](#).

Decline

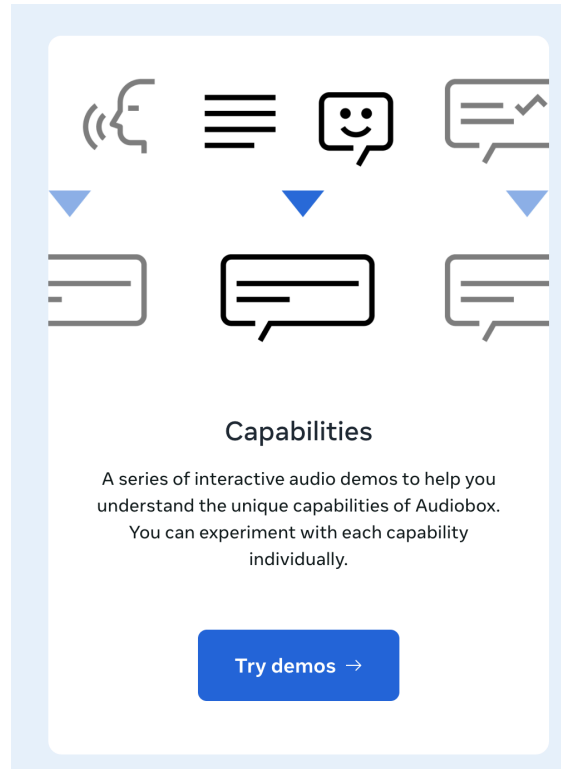
Accept

Explore the site, and watch the demo. Spend a few minutes exploring the site. Watch the demo video to understand the capabilities of AudioBox.

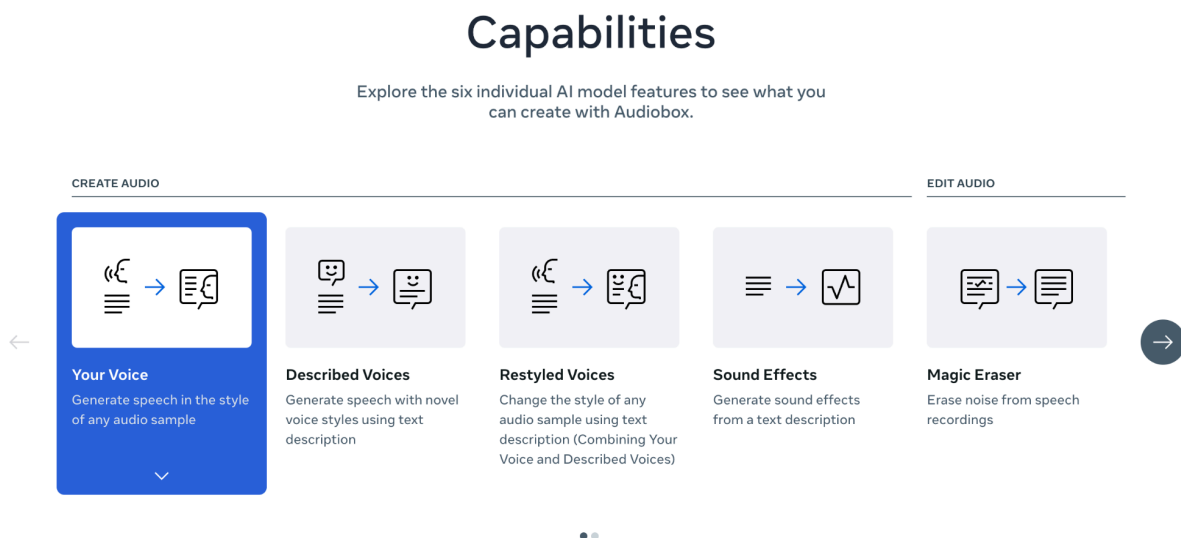


Look for the Try Demo button or go to this website.

<https://audiobox.metademolab.com/capabilities>



From there you will see some tools, we first explore the 'Your Voice' tool



Click the Record Your Voice Button

Your Voice

Generate speech in the style of your voice. Don't want to use your voice? Try a preset example.

Audiobox improves style similarity metric by more than 20% on average compared to Voicebox.

Record Your Voice

Record your own voice or use a sample recording. The AI model will use this to generate speech in the style of this reference vocal style.


Record your voice

Use a sample recording

Text To Speak

Add a short paragraph and the AI model will speak aloud the text.

Write here...

 0/125

Try an example

Generate →

View Results

Click the Record the reading and read the paragraph in the window.

Voice registration

To register your own voice, click **Record** and read the following prompt.

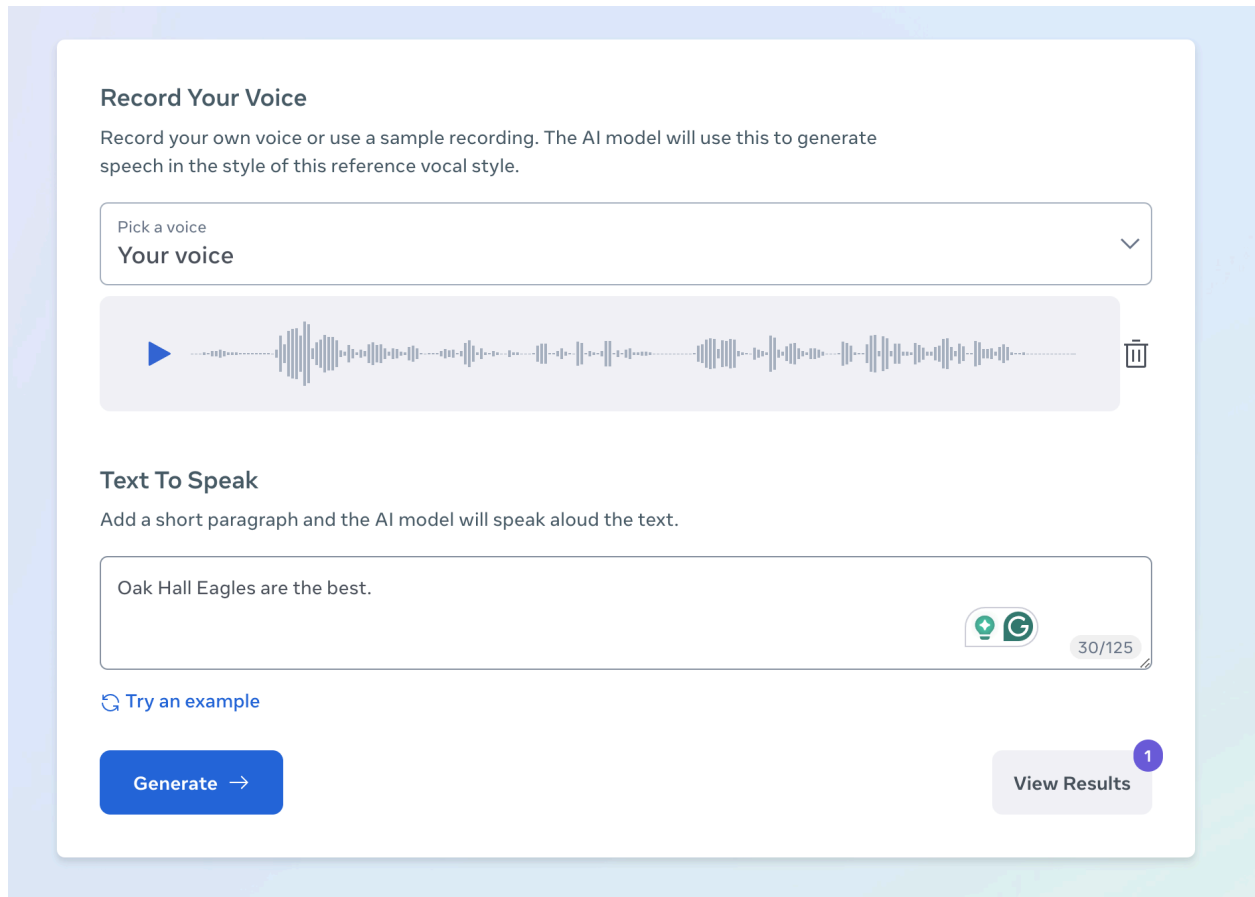
The audio engineer carefully adjusted the sound levels to ensure clear speech and music generation during the live performance.

Refreshes in 46 seconds

Record your reading

You can also use the voice styles we provided if you don't want to create your own.

Create your voice model. Not type a text in and generate the audio file.



The screenshot shows a user interface for creating a voice model. It is divided into two main sections: 'Record Your Voice' and 'Text To Speak'.

Record Your Voice
Record your own voice or use a sample recording. The AI model will use this to generate speech in the style of this reference vocal style.

Pick a voice
Your voice

Below the dropdown is a waveform visualization of a recorded audio sample, with a play button on the left and a trash icon on the right.

Text To Speak
Add a short paragraph and the AI model will speak aloud the text.

Oak Hall Eagles are the best.

On the right side of the text input field, there is a green icon with a plus sign and a 'G' logo, and a character count '30/125'.

Try an example

Generate →

View Results ¹

View the results. You will get two different files. Explore with different prompts.

Questions to Consider:

- What differences do you notice between the generated files?
- Can you tell it's an AI-generated voice?
- How does the prompt affect the output?

Your Voice Results



The AudioBox model is a generative flow matching model that involves random sampling. That means its generations will vary from one generation to the next, even given the same input. In this research demo, we generate multiple outputs to demonstrate the diversity of potential outputs of the model. See more information [here](#) about our efforts to engage in Responsible AI Research for Speech.

...

New

Oak Hall Eagles are the best.

Audio result (Option 1)

Audio result (Option 2)

Generated by: Audiobox Speech

Now click on Restylised Voice

Capabilities

Explore the six individual AI model features to see what you can create with Audiobox.

CREATE AUDIO

EDIT AUDIO

Your Voice
Generate speech in the style of any audio sample

Described Voices
Generate speech with novel voice styles using text description

Restyled Voices
Change the style of any audio sample using text description (Combining Your Voice and Described Voices)
▼

Sound Effects
Generate sound effects from a text description

Magic Eraser
Erase noise from speech recordings

● ●

Now we will take the same prompt from before, but now add emotion to it. Prompt it on how you want it to sound.

Record Your Voice

Record your own voice or use a sample recording. The AI model will use this to generate speech in the style of this reference vocal style.

Pick a voice

Your voice



Text To Speak

Add a short paragraph and the AI model will speak aloud the text.

Oak Hall Eagles are the best.



30/125

[Try an example](#)

Describe the new style

Describe how you want to modify the style of the voice. The model will use this new voice style to narrate your script.

excited guy at a football game



30/250

[Try an example](#)

Generating...


View Results






View the results.




Restyled Voices Results

The AudioBox model is a generative flow matching model that involves random sampling. That means its generations will vary from one generation to the next, even given the same input. In this research demo, we generate multiple outputs to demonstrate the diversity of potential outputs of the model. See more information [here](#) about our efforts to engage in Responsible AI Research for Speech.


... New Oak Hall Eagles are the best. 






Audio result (Option 1)






Audio result (Option 2)

... Oak Hall Eagles are the best. 



Audio result (Option 1)



Audio result (Option 2)

Generated by: Audiobox

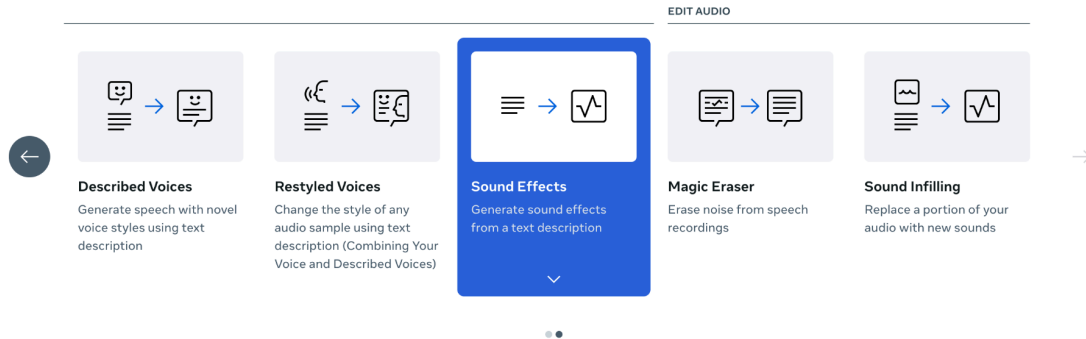
Questions to Consider:

- Did the voice capture the intended emotion?
- How could you refine your prompt for better results?
- What patterns do you observe in how emotions are expressed?

Now we will try sound effects. Click on Sound Effects.

Capabilities

Explore the six individual AI model features to see what you can create with Audiobox.



Type in a sound effect that you are interested in. A simple sound effect prompt (e.g., “thunderstorm,” “car honking”).


Sound Effects

Generate sound effects from a text description.

Audiobox reduces FAD by 50% compared to prior state-of-the-art, matching real audio in quality and faithfulness.

Describe A Sound Effect

Describe the characteristics of the sound you would like to create. The model will use this to generate your sound effect.



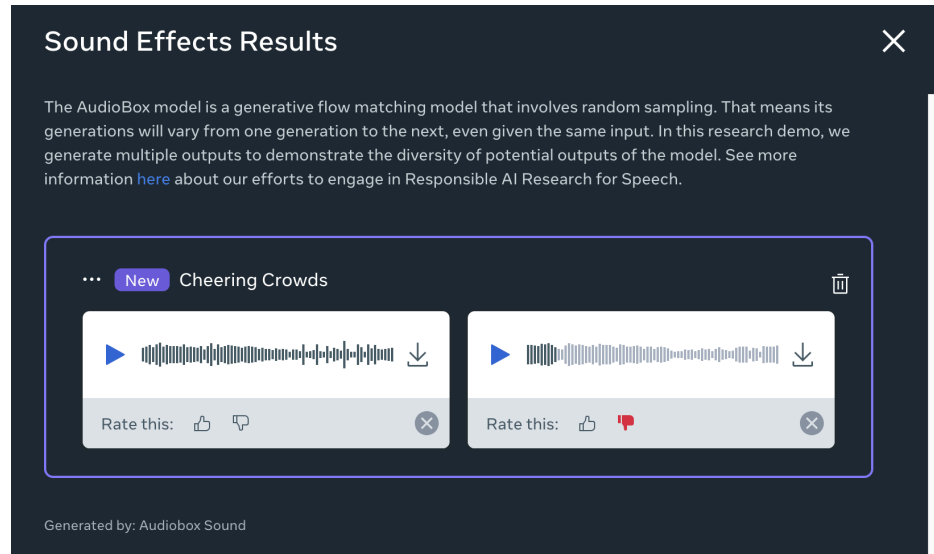
15/250

[Try an example](#)

Generate →

View Results

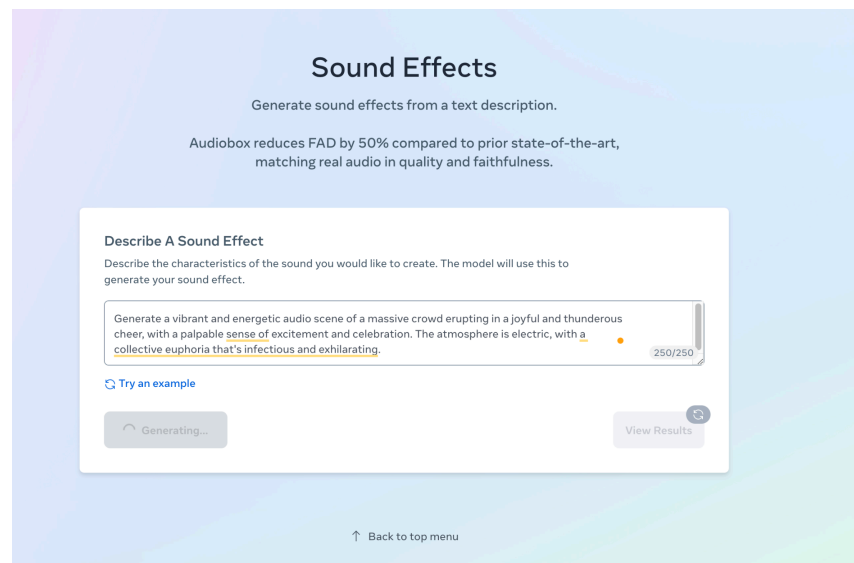
View the results



Questions to Consider:

- How accurate was the sound effect?
- What elements of the sound did the AI capture well, and what could be improved?

Try out a more detailed Prompt:








View the new results. Compare and iterate. Try out new versions.

Sound Effects Results


The AudioBox model is a generative flow matching model that involves random sampling. That means its generations will vary from one generation to the next, even given the same input. In this research demo, we generate multiple outputs to demonstrate the diversity of potential outputs of the model. See more information [here](#) about our efforts to engage in Responsible AI Research for Speech.


... **New** Generate a vibrant and energetic audio scene of a massive crowd erupting in a joyful and thunderous cheer, with a palpable sense of...


Rate this:   


Audio result (Option 2)

... Cheering Crowds


Audio result (Option 1)


Audio result (Option 2)

Generated by: Audiobox Sound

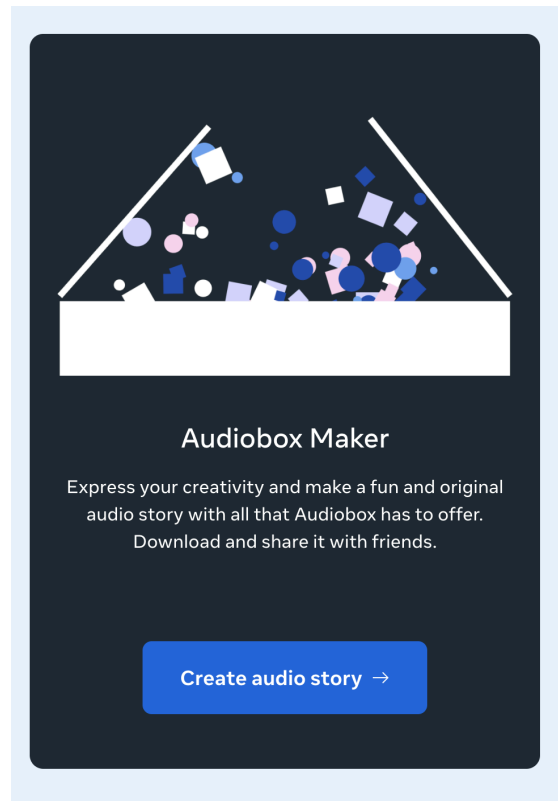
Questions to Consider:

- How did the added details in the prompt change the output?
- What new elements were introduced, and how did they enhance the sound?

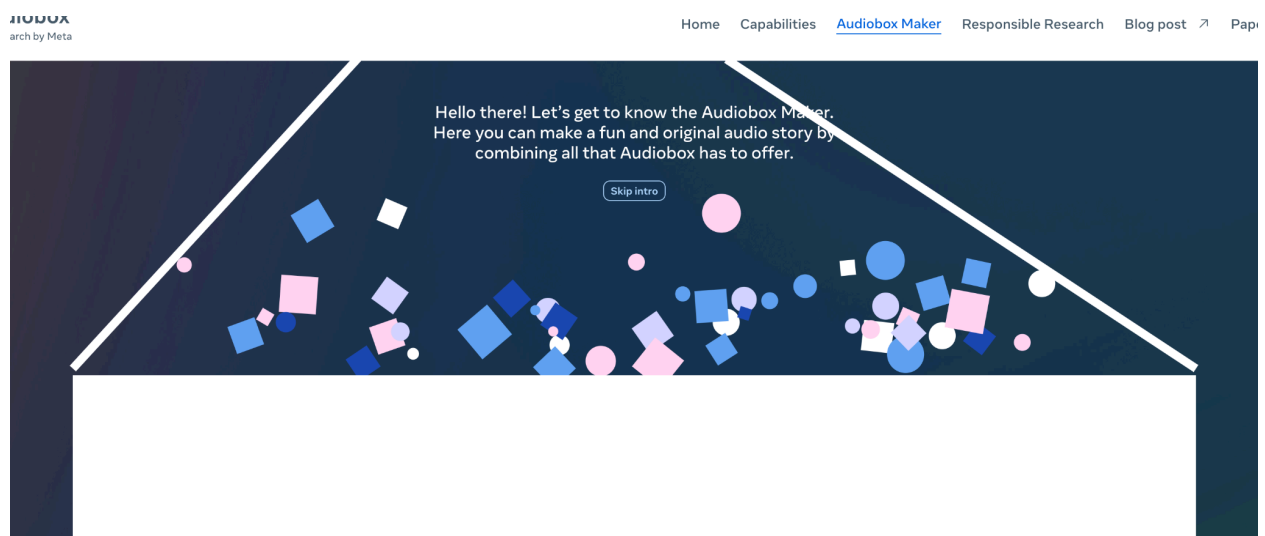
Create

Now that we have learned about the many tools let's look at they can be used to make a story come to life with sounds. Go to the Audiobox Maker tool or visit this website -

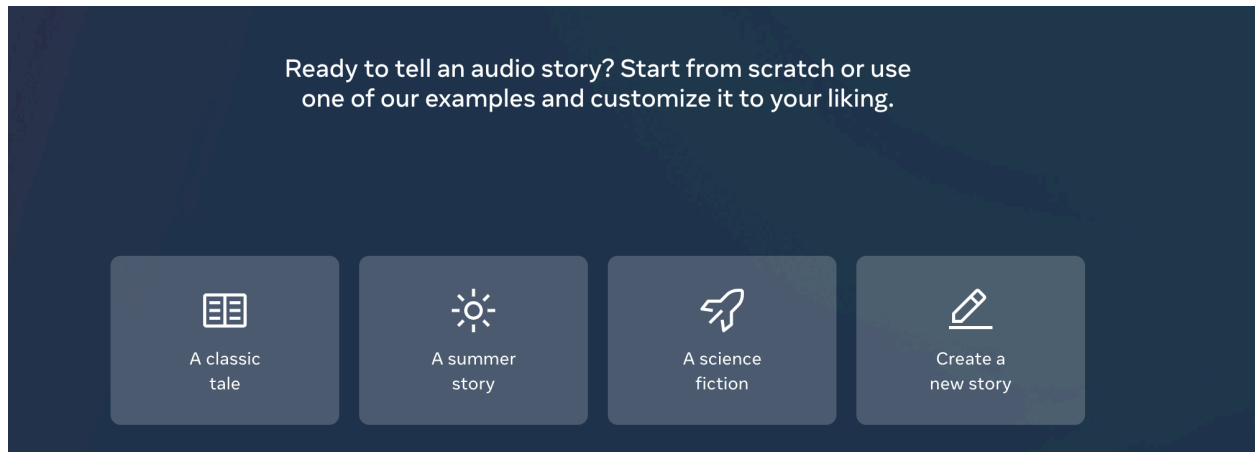
<https://audiobox.metademolab.com/maker>



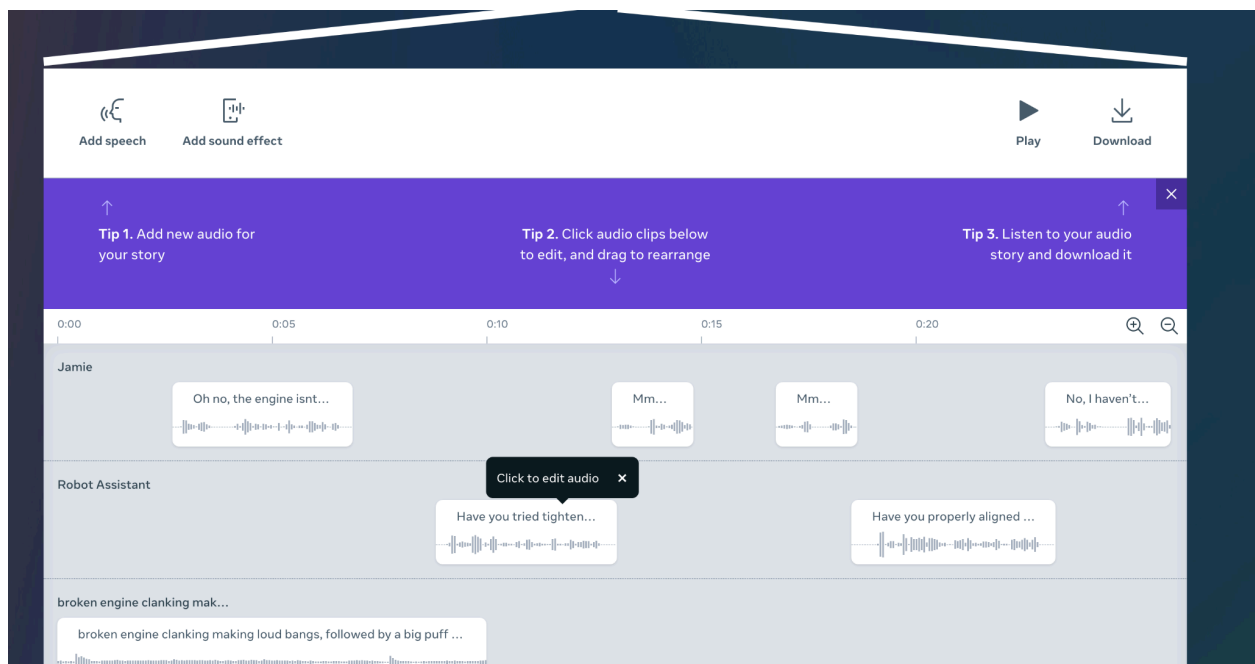
Click anywhere to start and listen to the intro.



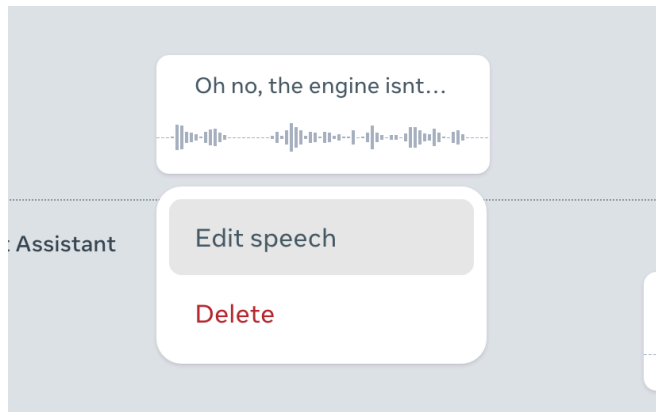
We will first learn by looking at the examples. Click on one of the examples.



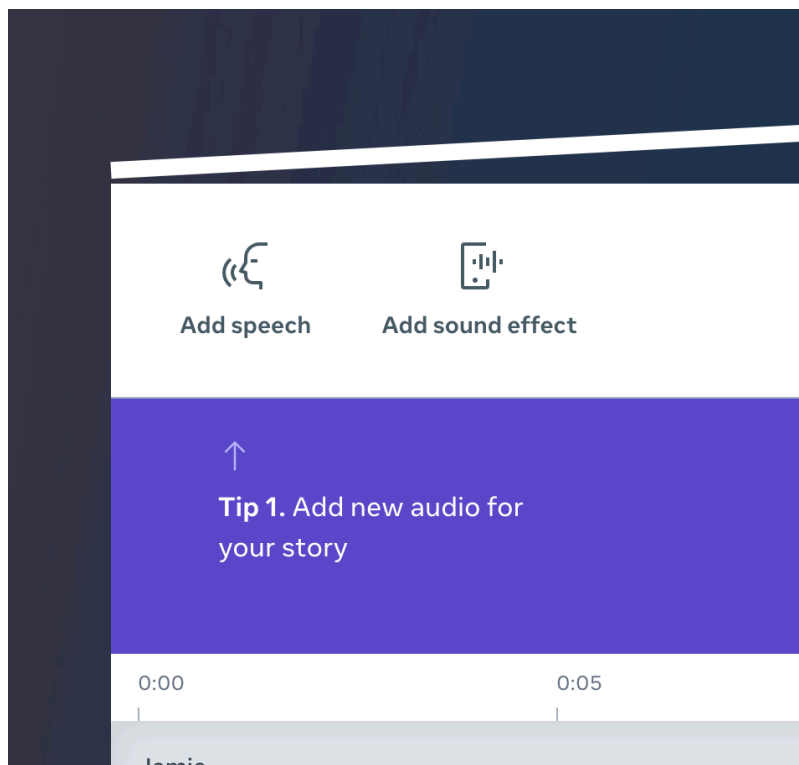
This will prepopulate the interface for you. Explore how it works. Press the play button. Remember none of these sounds are recorded, they are all AI-generated.



Now try editing the story. Change what is said or how it is said. Change sounds. See how it changes the story..



Try adding yourself to the story, or add in sound effects.



Keep iterating and experimenting.

Homework

Now that you have learned the tools I want you to challenge yourself and make a new story from scratch. Use the Audiobox Maker - <https://audiobox.metademolab.com/maker> and make an interesting Audio Story. Once done download it and email / share it to me - sounny@gmail.com The most interesting story will be talked about in class next week.

Learning Outcome

Rate yourself on how well you have achieved the following learning outcomes:

1. I have a basic understanding of prompt engineering.

- ☐ Not at all
- ☐ Somewhat
- ☐ Very well

2. I understand how natural language can be used to generate and control audio outputs in an AI system.

- ☐ Not at all
- ☐ Somewhat
- ☐ Very well

3. I am confident experimenting with simple commands to create desired outcomes.

- ☐ Not at all
- ☐ Somewhat
- ☐ Very well

Useful URLs

AudioBox StoryMaker - <https://audiobox.metademolab.com/storymaker/>

Research Paper about Makign AudioBox

<https://ai.meta.com/research/publications/audiobox-unified-audio-generation-with-natural-language-prompts/>