Lab 1 - Audiobox

Create an audio story with AI voices and sounds

Lab Objective

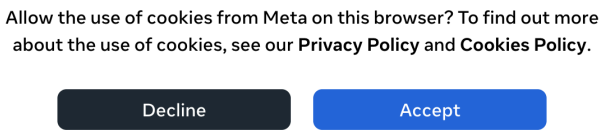
The goal of this lab is to equip students with the skills to use natural language prompts to generate various types of audio, including voices, sound effects, and soundscapes. Through hands-on experience with Meta's AudioBox, students will learn how AI models interpret and respond to prompts, ultimately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

Background

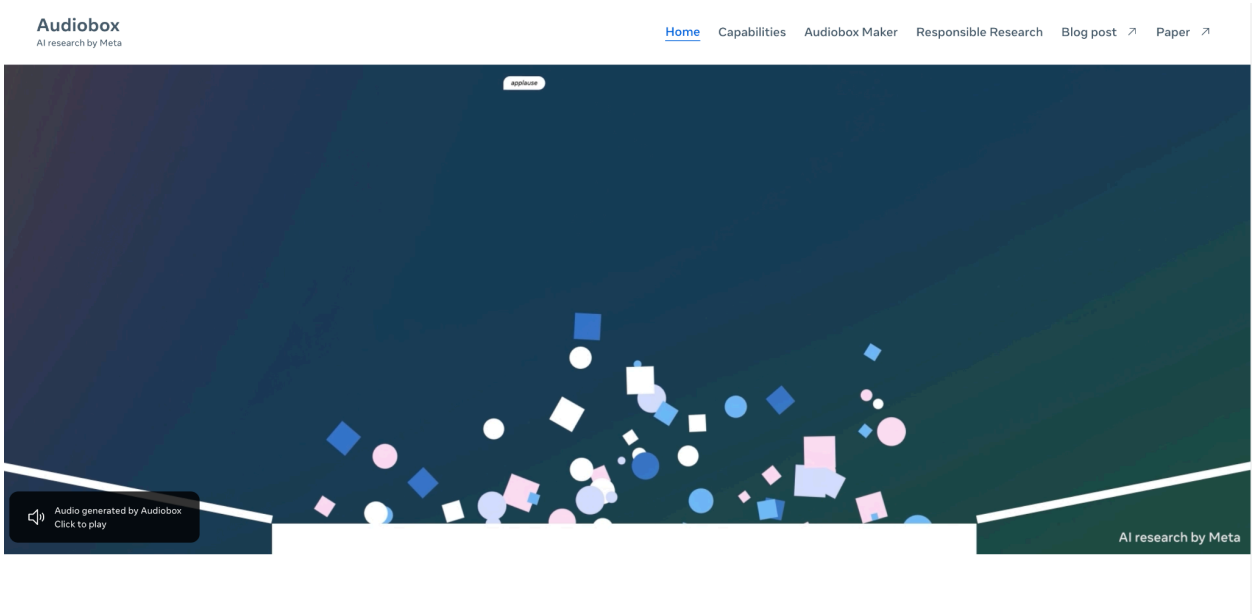
AudioBox is Meta’s cutting-edge research model for audio generation. It allows users to generate custom audio by combining voice inputs with natural language text prompts. The AudioBox model family includes specialist models like AudioBox Speech and AudioBox Sound, all built upon the shared self-supervised 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.

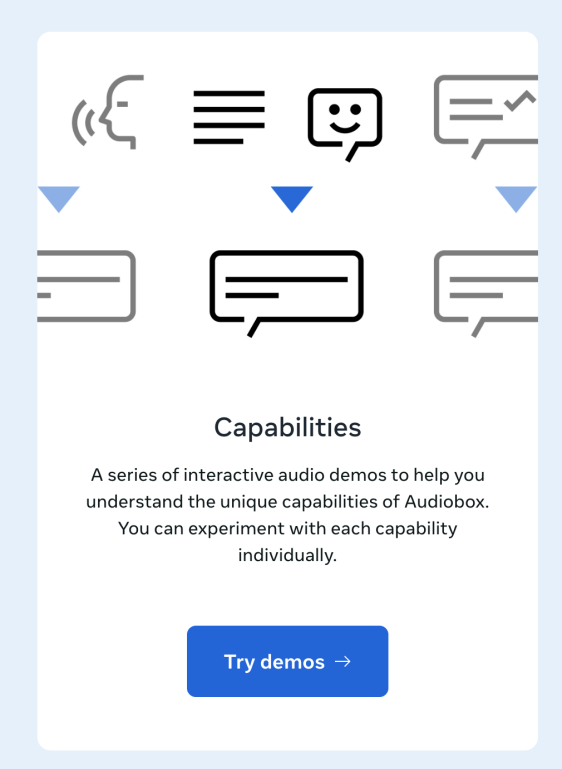


Explore the site, and watch the demo. Spend a few minutes exploring the site. Watch the demo video to understand the capabilities of AudioBox.

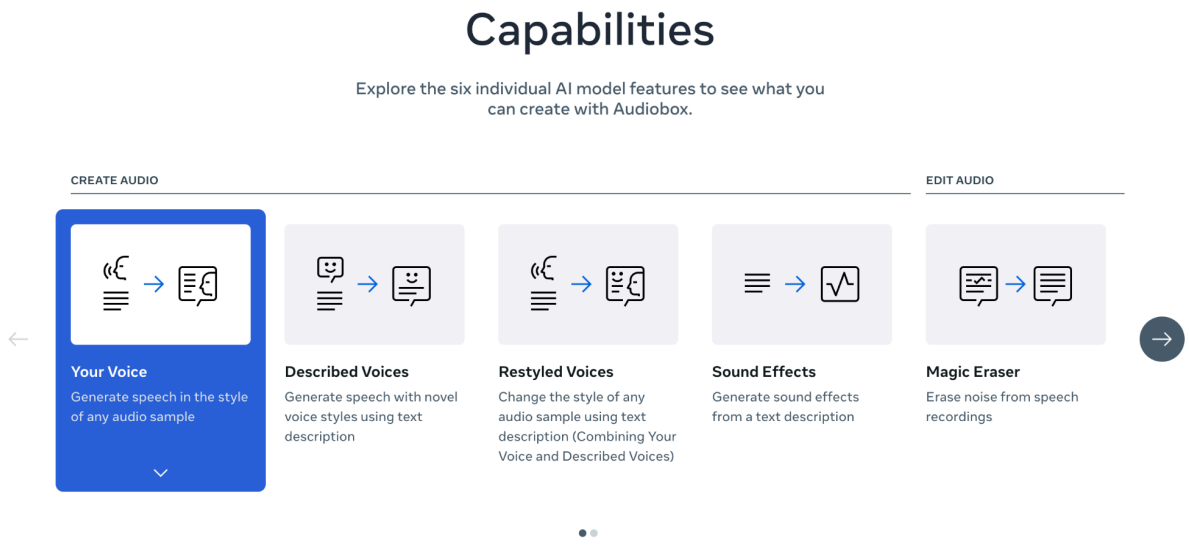


Look for the Try Demo button or go to this website.

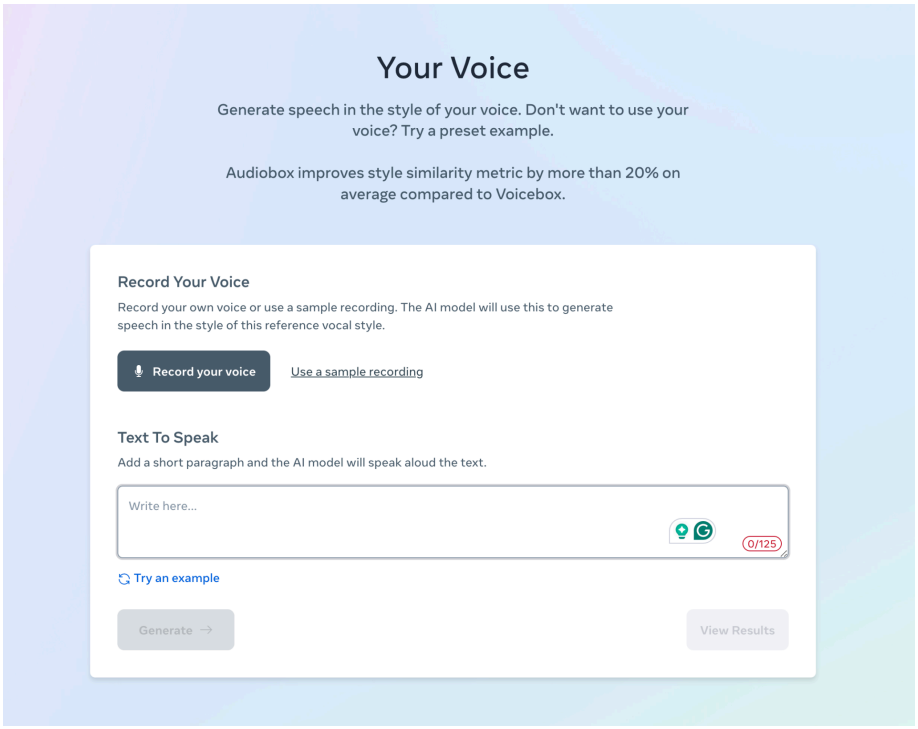
https://audiobox.metademolab.com/capabilities



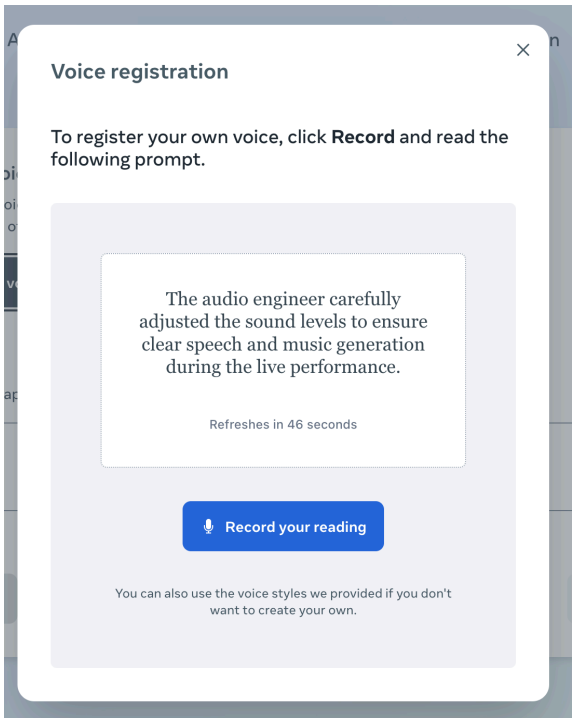
From there you will see some tools, we first explore the ‘Your Voice’ tool



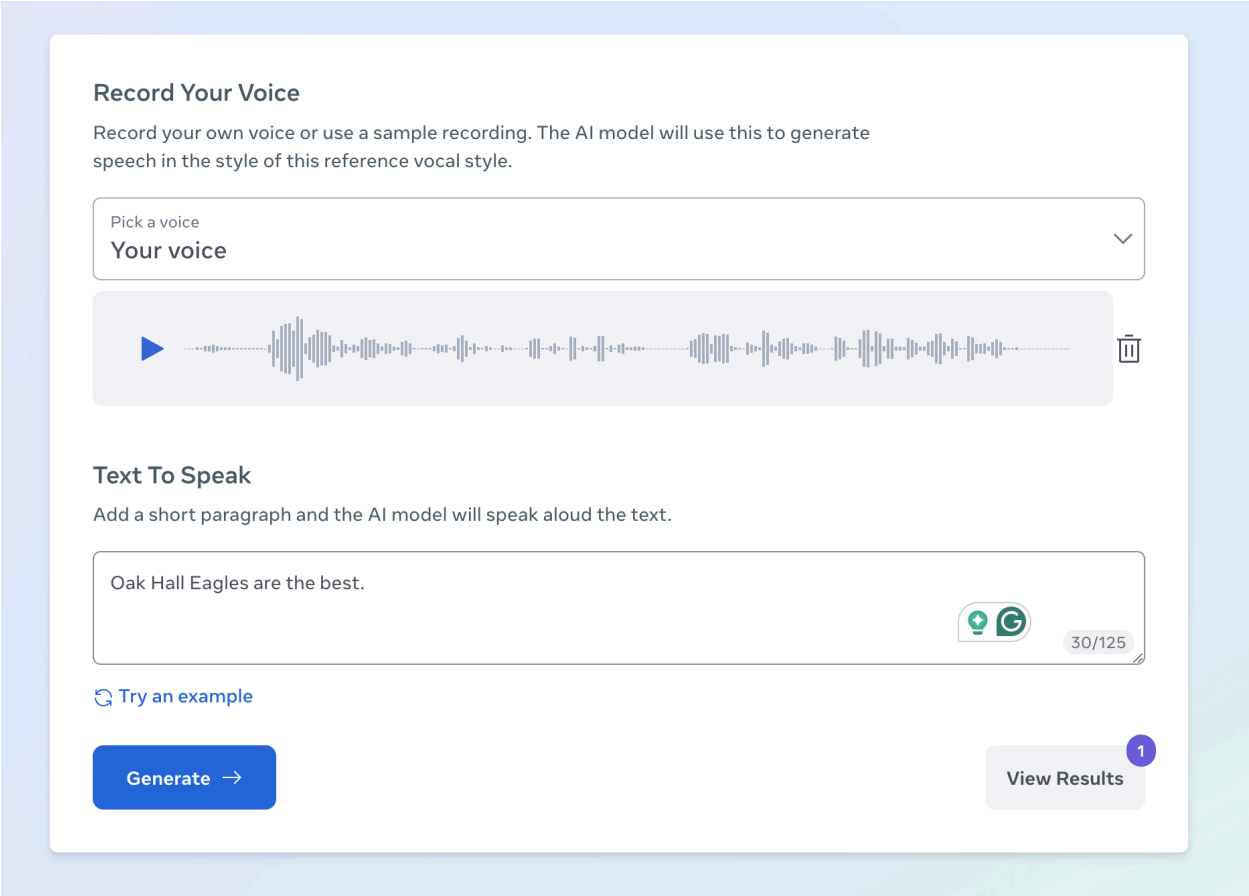
Click the Record Your Voice Button



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



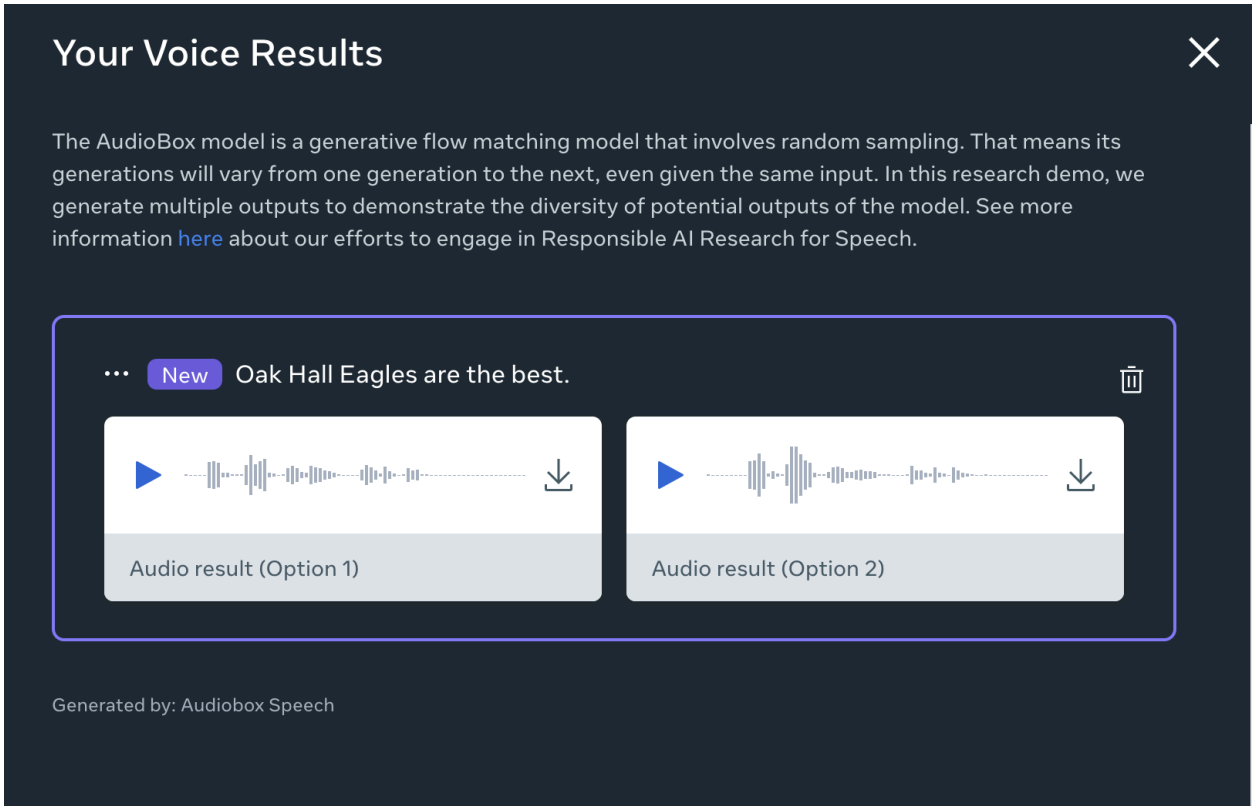
Create your voice model. Not type a text in and generate the audio file.



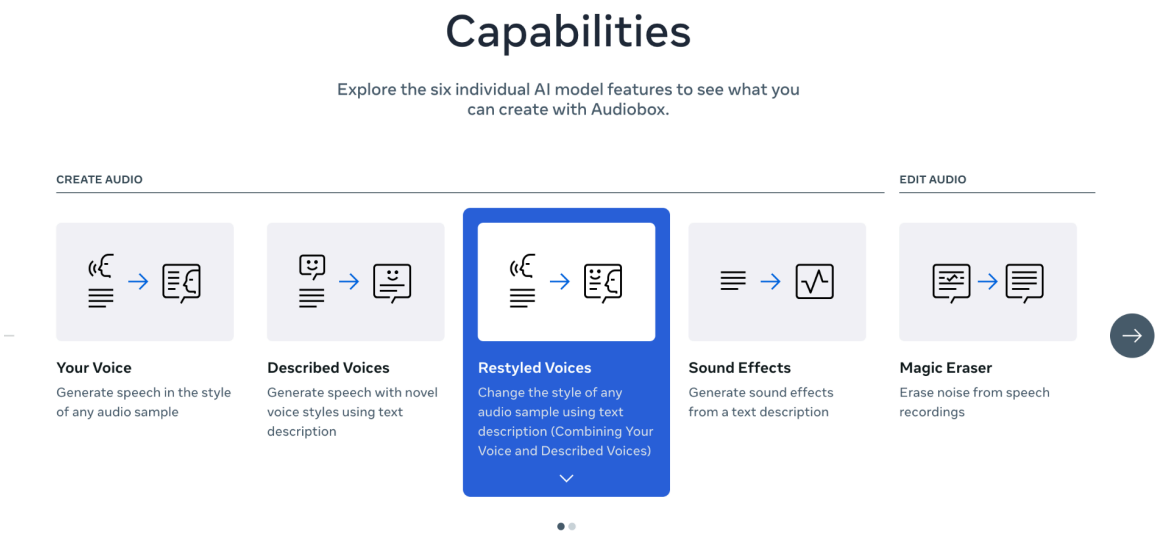
View the results. You will get two different files. Explore with different prompts.

Questions to Consider:

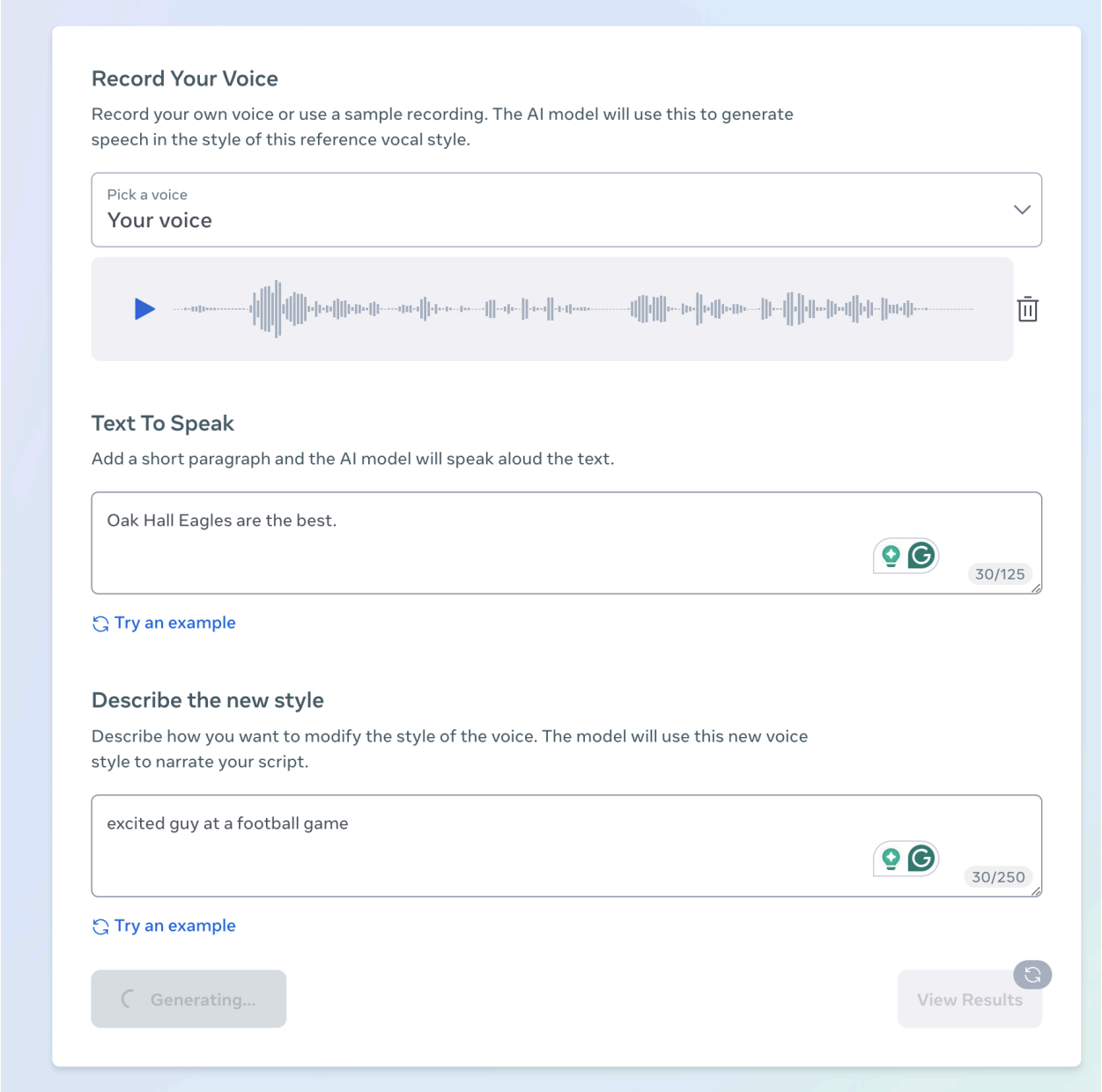
- What differences do you notice between the generated files?
- Can you tell it's an AI-generated voice?
- How does the prompt affect the output?



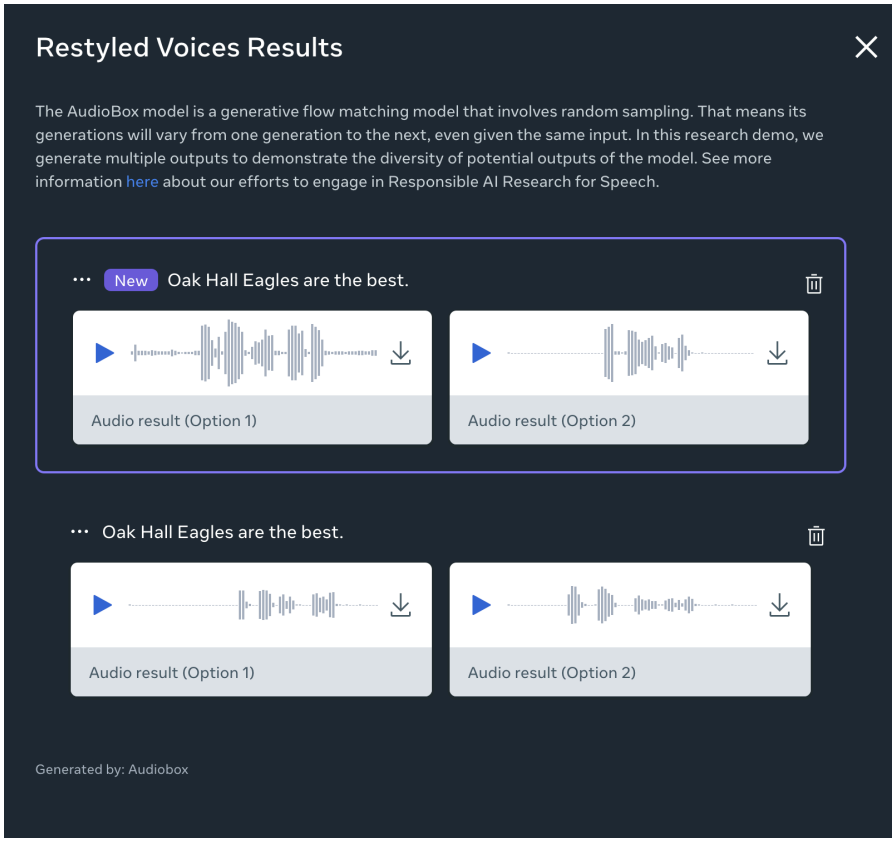
Now click on Restylised Voice



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



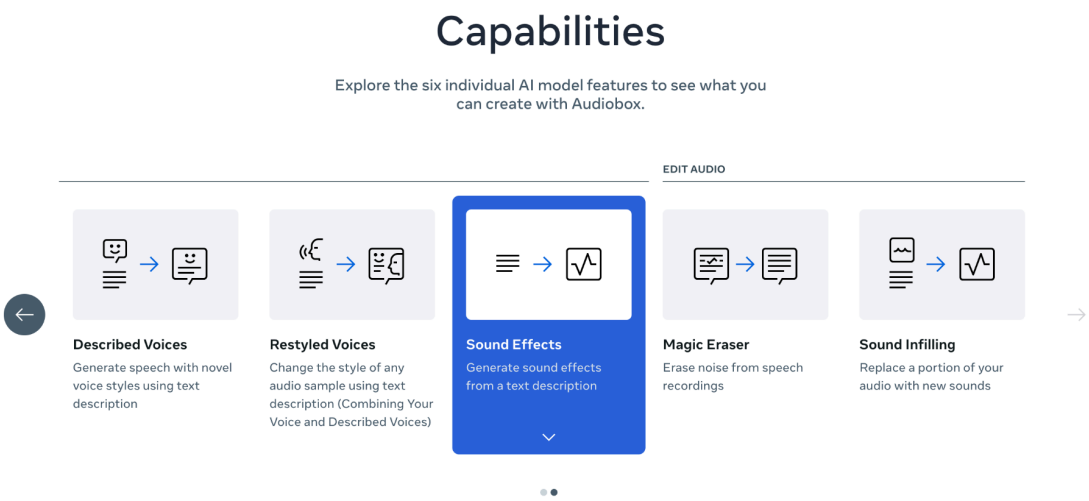
View the results.



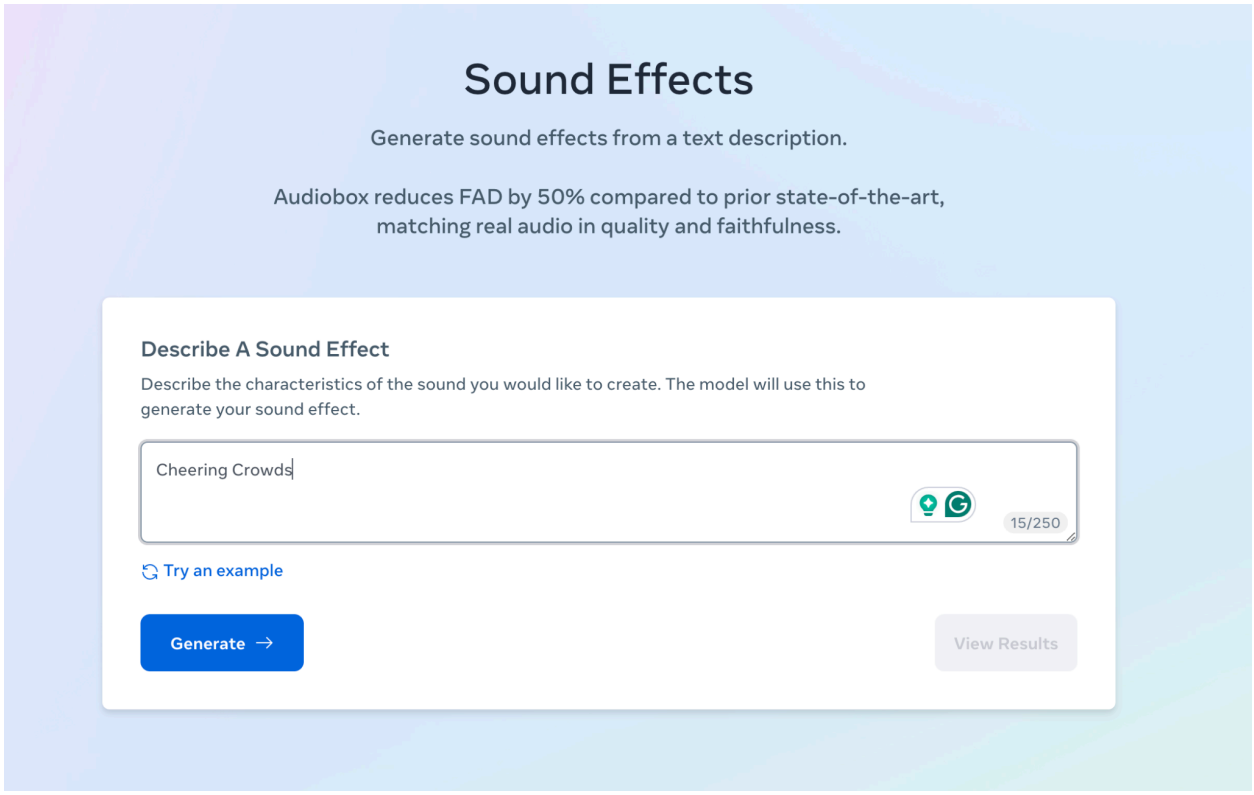
Questions to Consider:

- 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?

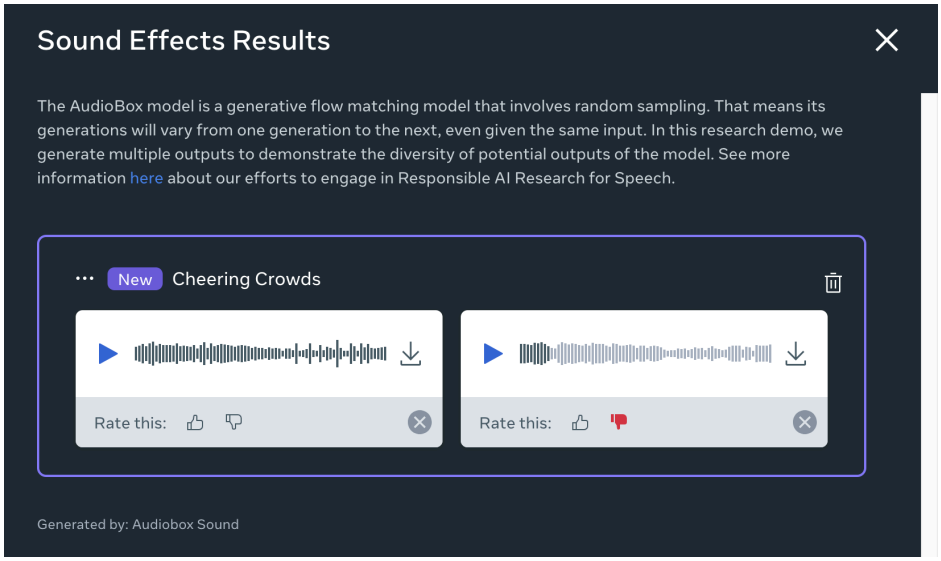
Now we will try sound effects. Click on Sound Effects.



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



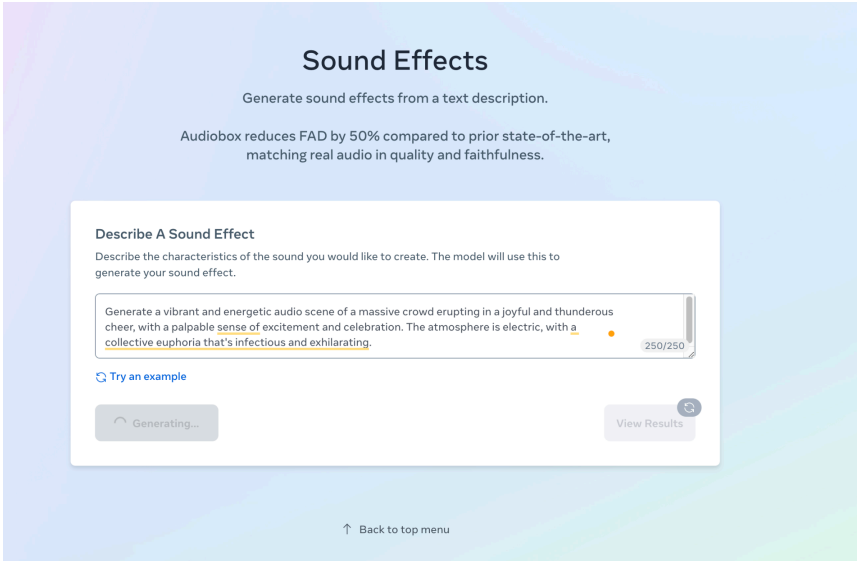
View the results



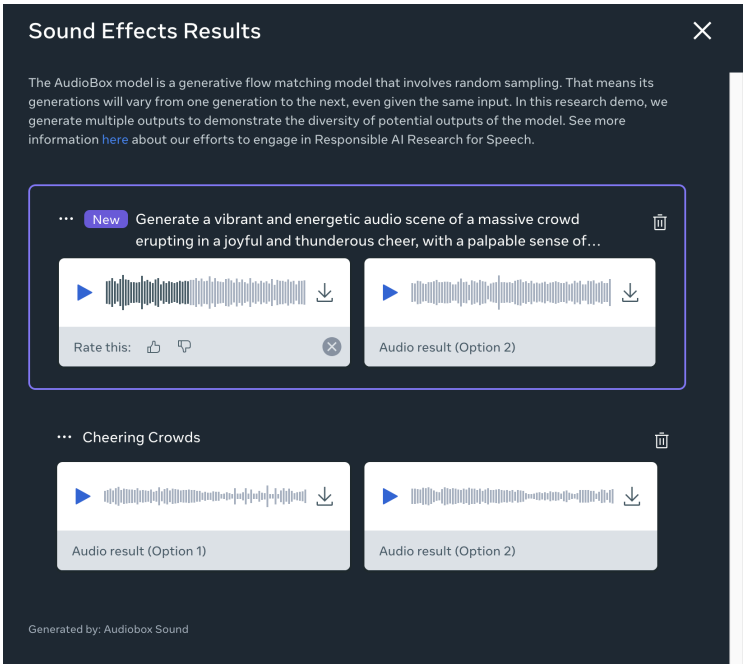
Questions to Consider:

- 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.

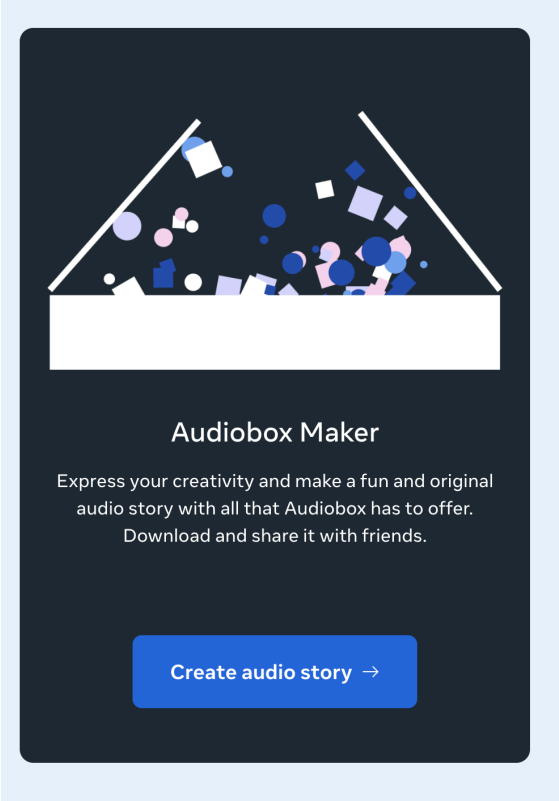


Questions to Consider:

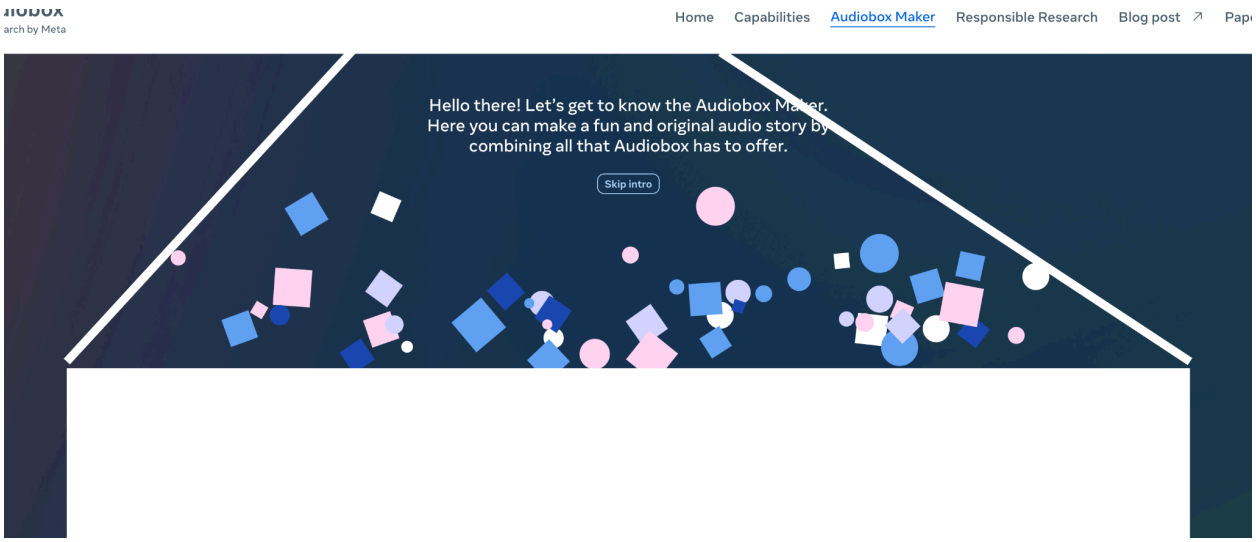
- How did the added details in the prompt change the output?
- What new elements were introduced, and how did they enhance the sound?

Create

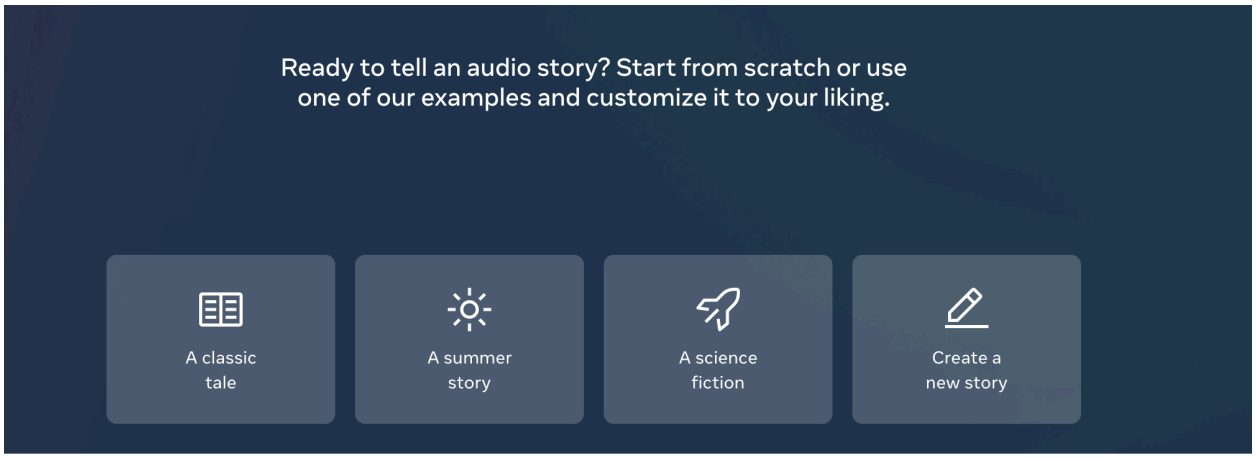
Now that we have learned about the many tools let's look at they can be used to make a story come to life with sounds. Go to the Audiobox Maker tool or visit this website - https://audiobox.metademolab.com/maker



Click anywhere to start and listen to the intro.

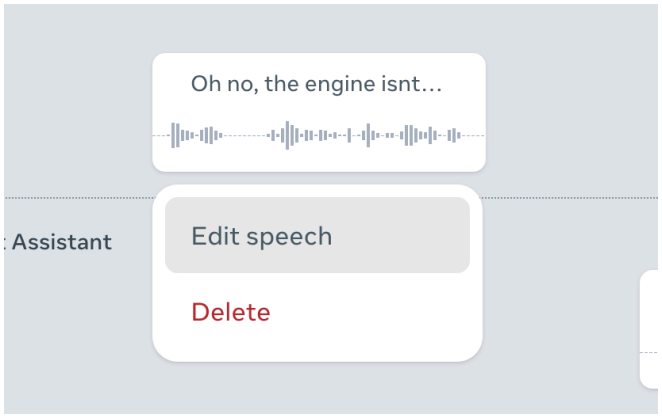


We will first learn by looking at the examples. Click on one of the examples.

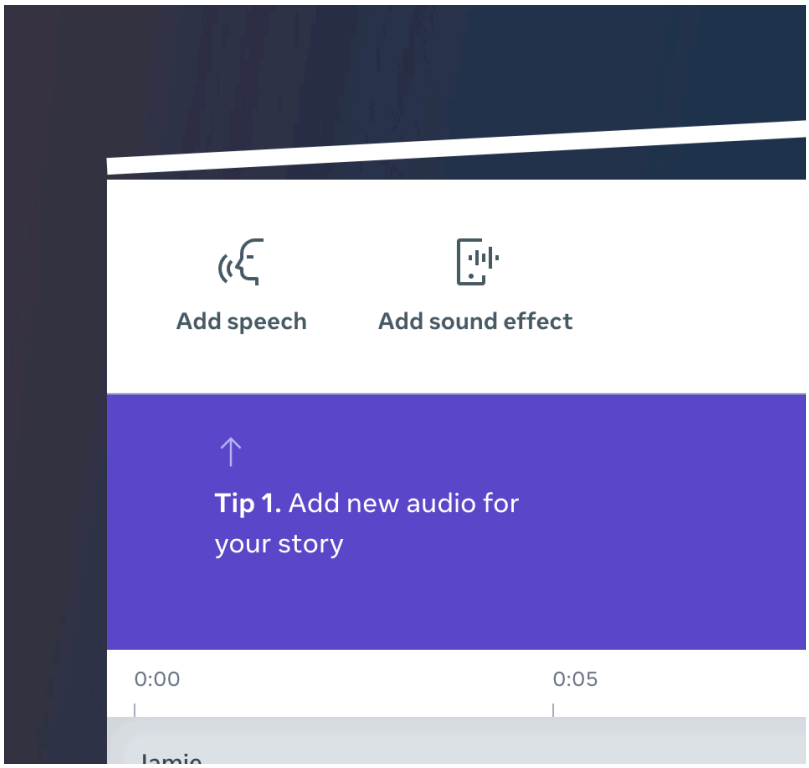


This will prepopulate the interface for you. Explore how it works. Press the play button. Remember none of these sounds are recorded, they are all AI-generated.

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/