



Speech-to-Text AI: Converting Audio to Readable Text

Discover how speech-to-text AI transforms spoken words into readable text. Learn applications and benefits of this innovative technology.

Presented by:

Akshita Gupta(202410116100017)

Akansha Tyagi(202410116100014)

Akansha Tomar(202410116100013)

Deepu Kumari(202410116100057)



Voice Recognition

Understanding Speech Recognition

What is Speech Recognition?

Technology that converts spoken language into text using AI models.

How It Works

Uses audio signal processing, feature extraction, and AI to decode speech.

History & Evolution

Originated in the 1950s, evolving to deep learning-based systems today.




```
280 world.width = isMobile ? window.innerWidth : DEFAU
281 world.height = isMobile ? window.innerHeight : DEF
282
283 // Center the player
284 player.position.x = world.width * 0.5;
285 player.position.y = world.height * 0.5;
```

Step-by-Step Implementation (Part 1)

Install Libraries

Run `!pip install SpeechRecognition` to add necessary tools.

Import Modules

Use `import speech_recognition as sr` and Google Colab files module.

```
In [216]: category_group = train[['wavfile', 'actuals', 'pred']].groupby(['actuals', 'pred']).count()  
category_group
```

Out[216]:

wavfile		
actuals	pred	
-1	0	271
	1	2304
1	0	93

Step-by-Step Implementation (Cont.)

1

Define Recognition Function

Create a function to process audio files with SpeechRecognition.

2

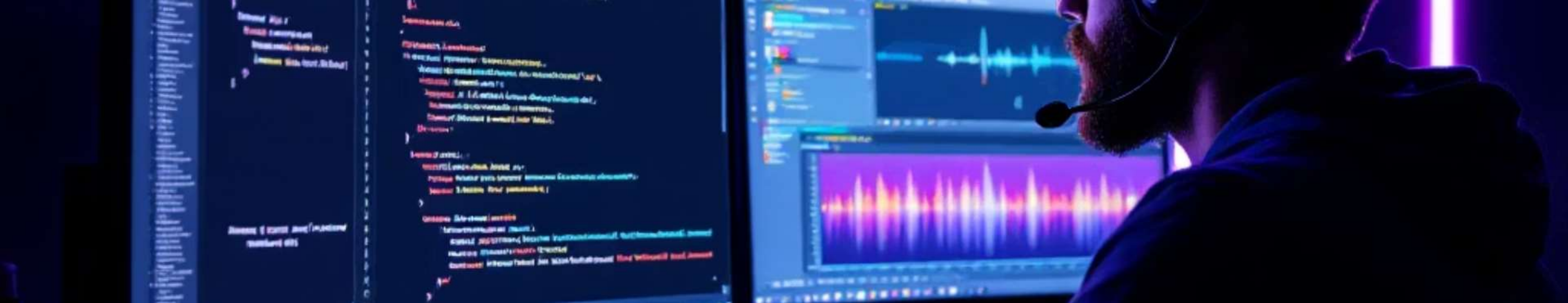
Initialize Recognizer

Use `sr.Recognizer()` to prepare speech decoding.

3

Read Audio File

Open audio with `sr.AudioFile(filename)` and record data.



Code Deep Dive: Speech Recognition in Action

- ``speech_recognition` Library`
Provides easy interfaces for converting audio to text.
- ``Recognizer()` Object`
Manages speech recognition operations and configurations.
- `Audio Processing`
`recognizer.record()` extracts audio segments for transcription.

Potential Applications of Speech-to-Text

Healthcare

- Medical transcription for records
- Improves doctor-patient communication

Education

- Automated lecture notes and summaries
- Accessibility for diverse learners

Business & Accessibility

- Meeting transcription and voice assistants
- Real-time captions for hearing impaired

Market size expected to reach \$26.5B by 2027

Challenges and Future Directions

Noise and Accent Handling

Improving recognition in noisy environments and diverse voices.

Low-Resource Languages

Enhancing accuracy where data is limited.

Future Enhancements

Real-time transcription, emotion recognition, and AI integrations.





Conclusion: The Future of Voice is Here



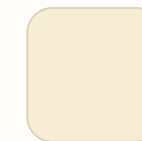
Transformation

Speech-to-text AI revolutionizes communication and accessibility.



Advancements

Continuous improvements drive diverse new applications.



Call to Action

Explore, innovate, and lead with speech technology today.