



Luware  
**Nimbus**

# Create a Trained Speech Recognizer for Luware Nimbus

Step-by-Step Guide

# Contents

Introduction	2
How To Train a Microsoft Speech Recognizer Model	3
Creating Your Training Data	4
Train the Model	7
Test and Deploy Your Model	9
Add the Speech Recognizer to Luware Nimbus	12

# Introduction

This guide outlines the steps to create and integrate a custom speech recognizer model for Luware Nimbus using Microsoft Azure Speech Services.

## Benefits:

- Improved speech recognition accuracy for your specific use case within Luware Nimbus.
- Enhanced captions and transcriptions for virtual assistant interactions.

## Here's what you'll learn:

1. Training a Microsoft Speech Recognizer Model
2. Creating Your Training Data
3. Train the Model
4. Test and Deploy Your Model
5. Add the Speech Recognizer to Luware Nimbus

By following these steps, you can create a speech recognizer specifically tailored to your needs within Luware Nimbus, leading to improved accuracy and user experience.

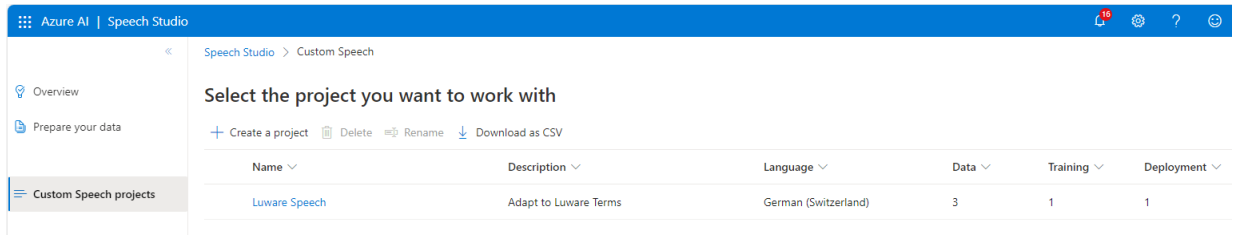
**Note:** This guide provides a high-level overview. Refer to the official Microsoft documentation for detailed instructions and code samples:

- **Training and testing datasets:** <https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-custom-speech-test-and-train>

# How To Train a Microsoft Speech Recognizer Model

Training a speech model starts in the **Custom Speech section in the Azure Speech Studio**.

To access the Custom Speech area, you need to create a Speech resource in the Azure portal and then sign in to Azure Speech Studio with your Azure credentials. There, you can create a new Custom speech project or use an existing one.



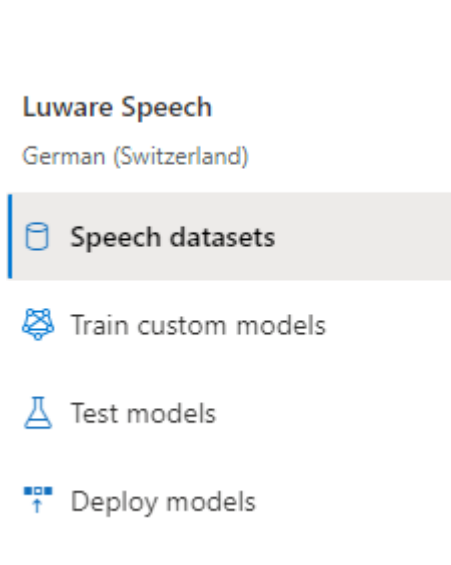
A Custom speech project consists of three components:

- data,
- models,
- and endpoints.

You need to upload your training and testing datasets to the data component, and then use them to train a custom model in the models' component.

You can also evaluate the performance of your model on the testing dataset and compare it with the baseline model.

Once you are satisfied with your model, you can deploy it to an endpoint and use it for speech recognition via the **Speech SDK** or **REST API**.



# Creating Your Training Data

The first step in creating your custom speech model is to prepare your training data. Training data consists of audio files and their corresponding transcripts that represent the speech patterns and vocabulary you want your model to recognize. **The quality and quantity of your training data directly affect the accuracy of your custom speech model.**

If you want to use the model for your Contact Center, we recommend using the **Audio + human-labelled transcript** data for training or testing. Such data allows you to provide both the original audio file and a manually corrected transcript for each utterance. The transcript should match the audio exactly and include punctuation, capitalization, and special symbols. This way, you can ensure that your model learns from the correct transcription and not from the errors of the automatic speech recognition system.

## To Create Your Training Data, Follow These Steps

- Collect audio files that match the characteristics of your target domain, such as **language, accent, channel, background noise**, etc. The audio files should be in **WAV format** with **16 kHz** sampling rate and **16-bit PCM** encoding.
- Transcribe each audio file using the best available method, such as the baseline speech recognition service, a third-party transcription service, or manual transcription. You can also use the transcription feature in the Speech Studio to generate transcripts automatically and edit them as needed.
- Create a TXT – File, that combines the WAV files with the according transcription. **The name of the WAV file must be separated by a tab (t)**. The transcription is only done with lower case letters. There is a guidance for different languages available.

```
nimbus.wav nimbus
luwarenimbus.wav luware nimbus
inputcustomer.wav input customer
attendantconsole.wav attendant console
advancedrouting.wav advanced routing
collectinformation.wav collect information erlaubt es input in parameter zu speichern
customersuccessmitarbeiter.wav customer success mitarbeiter
systemengineermitarbeiter.wav system engineer mitarbeiter
inputcustomeradvanced.wav input customer advanced element
de01cluster.wav de null eins cluster
unserendeutschencluster.wav unseren deutschen cluster
ch01cluster.wav ch null eins cluster
unserenschweizercluster.wav unseren schweizer cluster
instantmessagingueberinteract.wav instant messaging über interact
```

- Review and correct the transcripts to make sure they are accurate and consistent. You should also format the transcripts according to the guidelines provided in the documentation, such as **using brackets for non-speech events, using commas for pauses, and using quotation marks for direct speech.**
- Create a ZIP file, which contains the TXT file as well as the WAV files.
- Upload the ZIP file to the data component in the Speech Studio. You should also **provide some metadata for each file**, such as the locale, the speaker ID, and the transcription type.

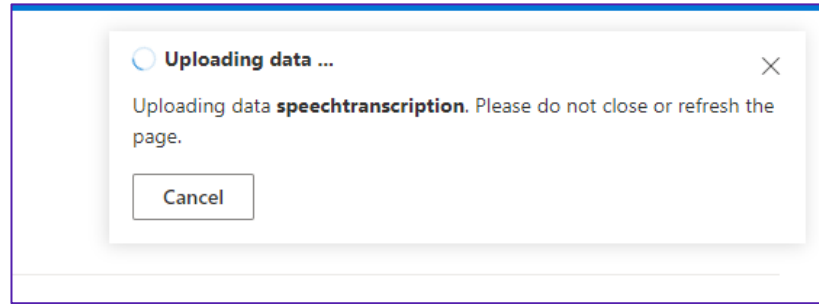
You can **use the import feature** in the Speech Studio to upload your data and fill in the metadata automatically.

The screenshot shows the 'Upload Speech Data' dialog box with the 'Choose data type' step selected. The left sidebar contains four steps: 'Choose data type' (selected), 'Upload data', 'Name and description', and 'Review and save'. The main area is titled 'What type of data do you want to upload' and lists five options, each with 'Training' and 'Testing' checkboxes:

- Plain text**: A text (.txt) file with a large amount of unstructured sentences (with jargon) related to your target scenario. (Training: checked, Testing: unchecked)
- Audio + human-labeled transcript**: A collection (.zip) of audio files (.wav) as individual utterances, paired with a formatted transcript (.txt) denoting the audio file names, to improve the acoustic aspects like slight accents, speaking styles, and background noises; or enhance evaluate your model qualities with error rate numbers. (Training: checked, Testing: checked)
- Pronunciation**: A list of uncommon terms without standard pronunciations in a text (.txt) file, where each term should be designated with a displayed form and a spoken form. (Training: checked, Testing: unchecked)
- Audio**: A collection (.zip) of audio files to quickly inspect your model qualities without accuracy number. (Training: unchecked, Testing: checked)
- Transcript (automatic audio synthesis)** *Preview*: Don't even have any real audio? Upload a text (.txt) file with some testing sentences, and audio pair for each spoken sentence will be automatically synthesized, leveraging the state-of-the-art neural text to speech technology. (Training: checked, Testing: checked)

At the bottom, there are 'Next' and 'Cancel' buttons.

The screenshot shows the 'Upload Speech Data' dialog box with the 'Upload data' step selected. The left sidebar shows 'Upload data' as the active step. The main area is titled 'Upload your audio + transcript data' and includes the instruction: 'Upload your local files or import from Azure Blob or other shared web locations.' There are two tabs: 'Local file' (selected) and 'Azure blob or other shared web locations'. The 'Local file' tab shows a large grey area with a cloud icon and the text: 'Drag and drop. .zip (<2GB) including 16K/8KHz audio files and a human-labelled transcript text. [Browse for a file](#)'. At the bottom of this area, a green bar shows a file named 'speechtranscript.zip' with a checkmark icon and a close button. Below this bar is a link: '[Learn more about data requirements](#)'. At the bottom of the dialog, there are 'Back', 'Next', and 'Cancel' buttons.



Speech Studio > Custom Speech > Speech datasets

**Training and testing dataset** [Editor](#)

Upload speech data for testing or training a speech recognition model.

[Upload data](#)
[Train](#)
[Test](#)
[Export to Editor](#)
[Rename](#)
[Delete](#)
[Download as CSV](#)

Name	Description	Type	Created	Quantity	Status
speechtranscription	Luware specific terms	Audio + transcript	5/3/2024 1:22 PM	01:08s	Succeeded

- Label your data as training or testing data. You should use at least 80% of your data for training and the rest for testing. You can also create multiple datasets for different purposes, such as **tuning, validation, or evaluation**. Furthermore, you can use the dataset feature in the Speech Studio to create and manage your datasets.

Azure AI | Speech Studio

Speech Studio > Custom Speech > Speech datasets

**Training and testing dataset** [Editor](#)

Upload speech data for testing or training a speech recognition model.

[Upload data](#)
[Train](#)
[Test](#)
[Export to Editor](#)
[Rename](#)
[Delete](#)
[Download as CSV](#)

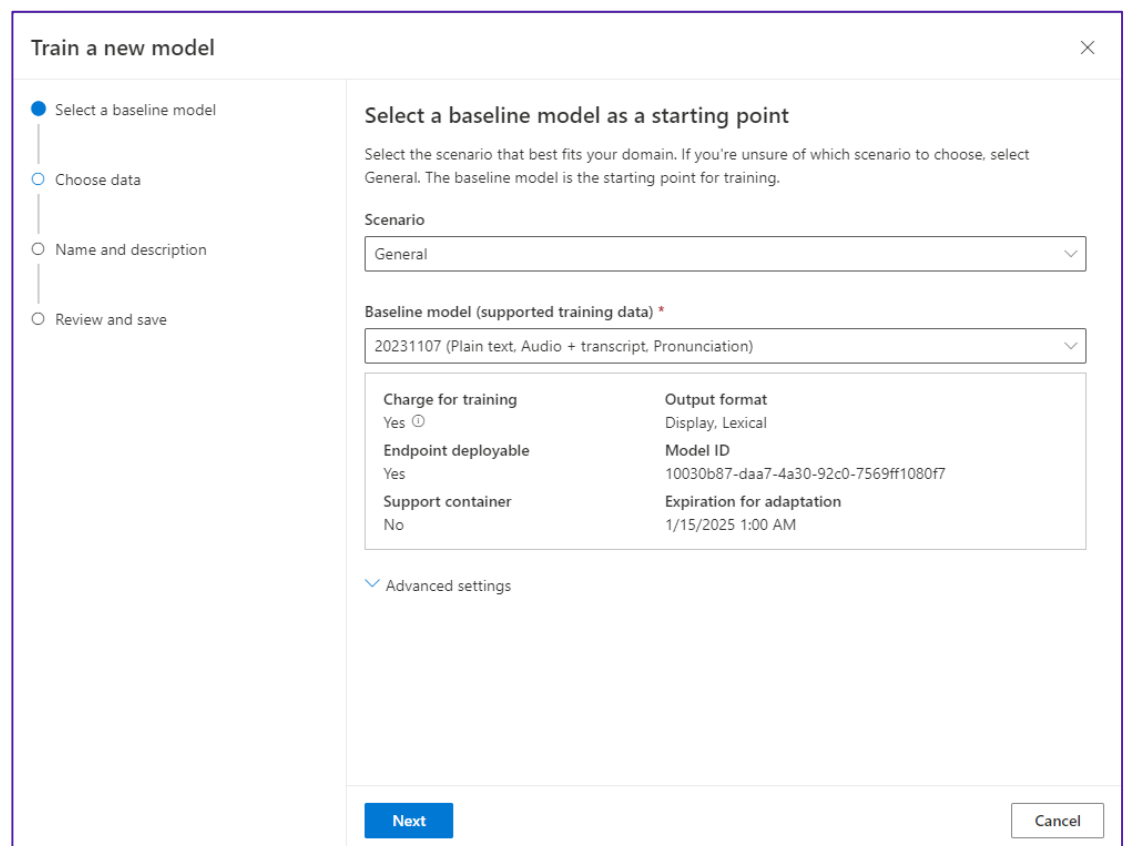
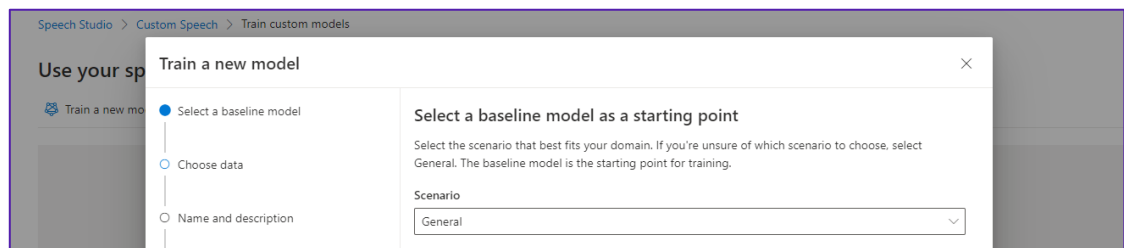
Name	Description	Type	Created	Quantity	Status
test_15_23		Audio	5/3/2024 3:31 PM	00:15s	Succeeded
Testdaten		Audio	5/3/2024 2:43 PM	01:14s	Succeeded
speechtranscription	Luware specific terms	Audio + transcript	5/3/2024 1:22 PM	01:08s	Succeeded

# Train the Model

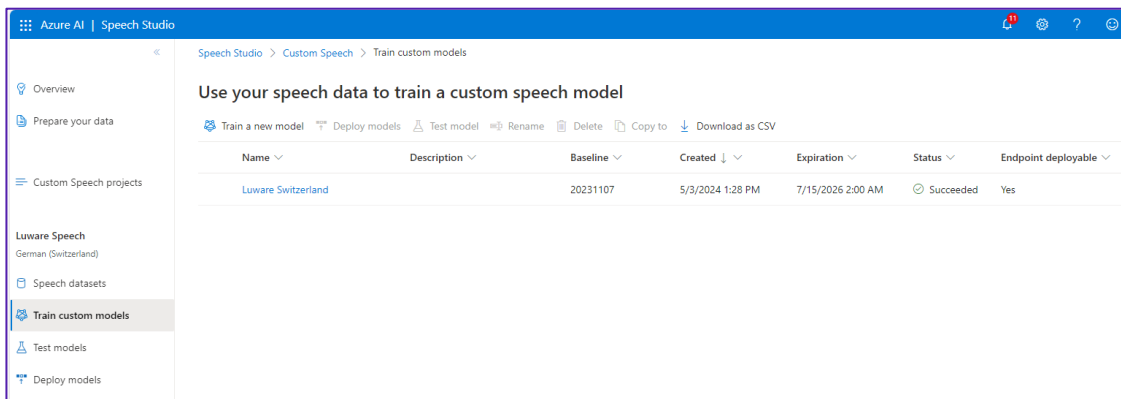
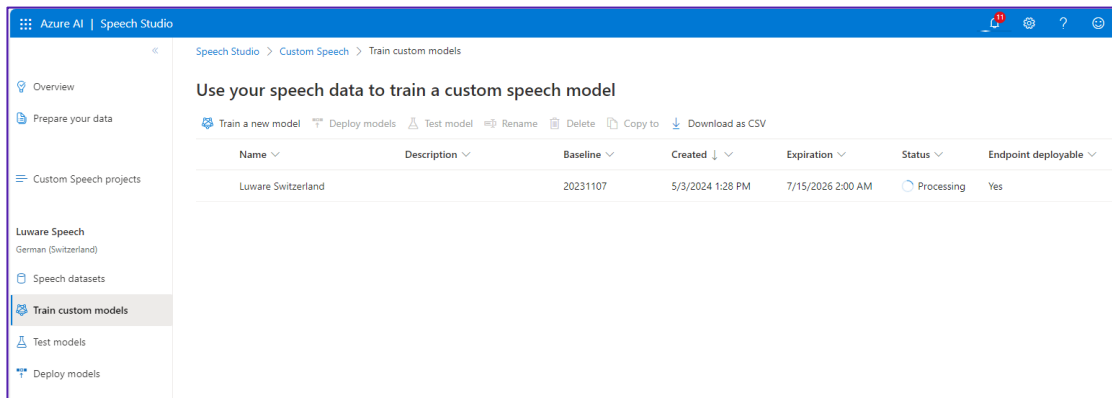
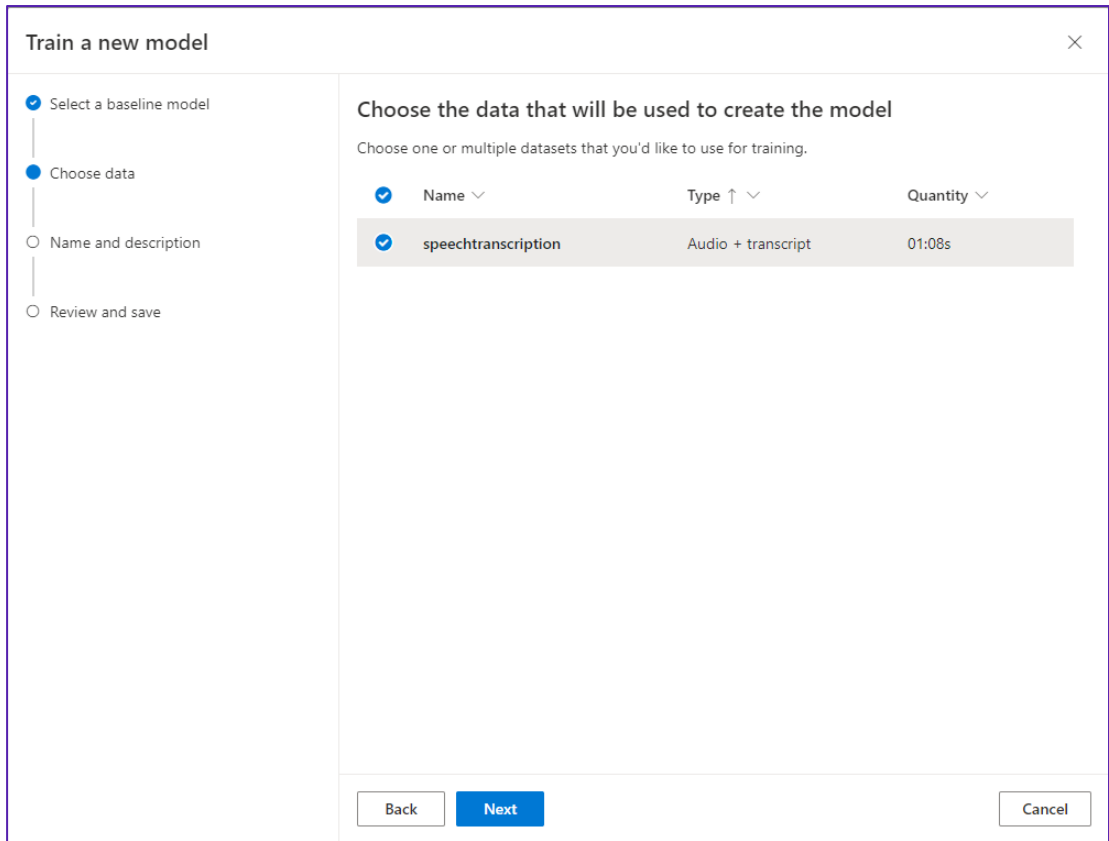
Switch to the **train custom models** tab to train the model.

To train your custom model, follow these steps:

- Select the dataset you want to use for training from the dropdown menu.
- Choose a base model that matches the language and domain of your data. You can also select a custom model that you have previously trained as a base model.
- Specify a name and a description for your new custom model.
- Optionally, adjust the training settings, such as the number of epochs, the learning rate, and the dropout rate.
- Click on the Train button and wait for the training process to complete. You can monitor the progress and the performance metrics on the dashboard.
- Once the training is done, you can download or deploy your custom model for testing or production purposes. You can also compare your custom model with the base model or other custom models on the Evaluate tab.







# Test and Deploy Your Model

Once the model has been trained, you can now test the model with test data.

### Create new test

- Choose testing type
- Choose testing data
- Choose models to evaluate
- Name and description
- Review and evaluate

#### Choose a test type

You can inspect model quality on selected audio (**Inspect quality**) or evaluate and compare the accuracy of a selected model with another custom model or baseline model using selected data (**Evaluate accuracy**).

**Inspect quality (Audio-only data)**  
Visually inspect the recognition quality of audio data using the model.

**Evaluate accuracy (Audio + transcript data)**  
Accurately measure model performance by leveraging human-labeled transcriptions from Audio + transcript data.

[Learn more about inspecting quality with audio data](#)

**Next** Cancel



### Create new test

- Choose testing type
- Choose testing data
- Choose models to evaluate
- Name and description
- Review and evaluate

#### Choose which testing datasets to use

Choose one audio dataset that you'd like to use for testing. Or upload additional from the Speech dataset page.

Name ↓ ▾	Type ▾	Quantity ▾
Testdaten	Audio	01:14s
<input checked="" type="checkbox"/> test_15_23	Audio	00:15s

**Back** **Next** Cancel



### Create new test

- Choose testing type
- Choose testing data
- Choose models to evaluate
- Name and description
- Review and evaluate

#### Choose models to evaluate

Select up to two model to evaluate and compare accuracy.

**Model 1 \***

Luware Switzerland

**Model 2**

[None]



### Create new test

- Choose testing type
- Choose testing data
- Choose models to evaluate
- Name and description
- Review and evaluate

#### Name and description

Create a test to evaluate the quality of Microsoft's speech to text baseline model, you can compare two models side by side to evaluate accuracy.

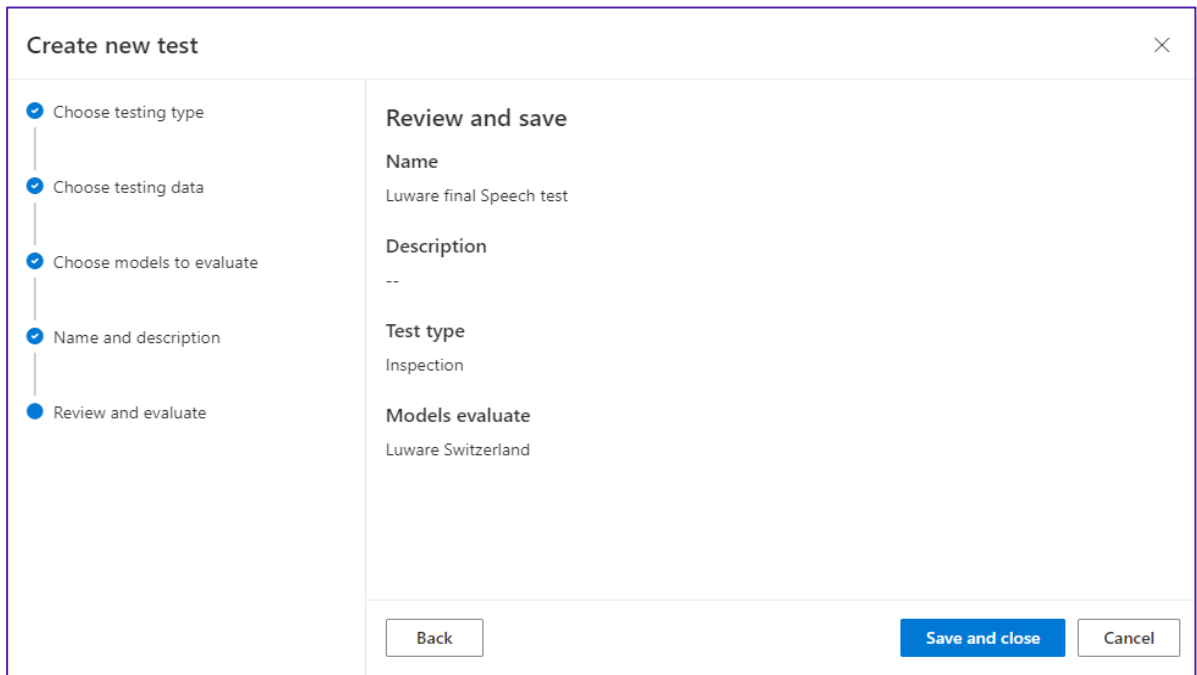
**Name \***

Luware final Speech test

**Description**

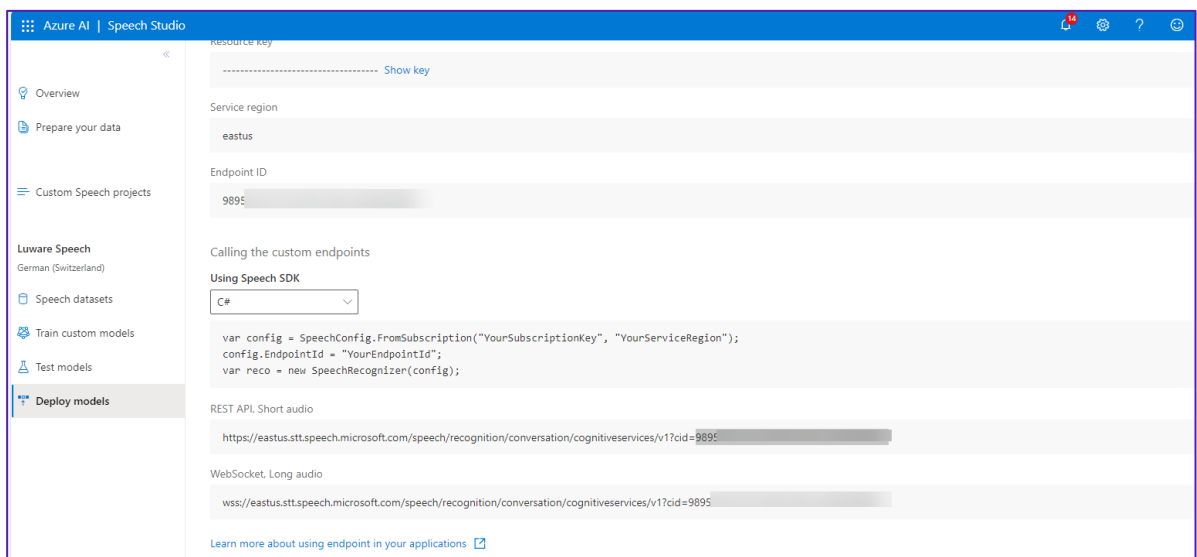
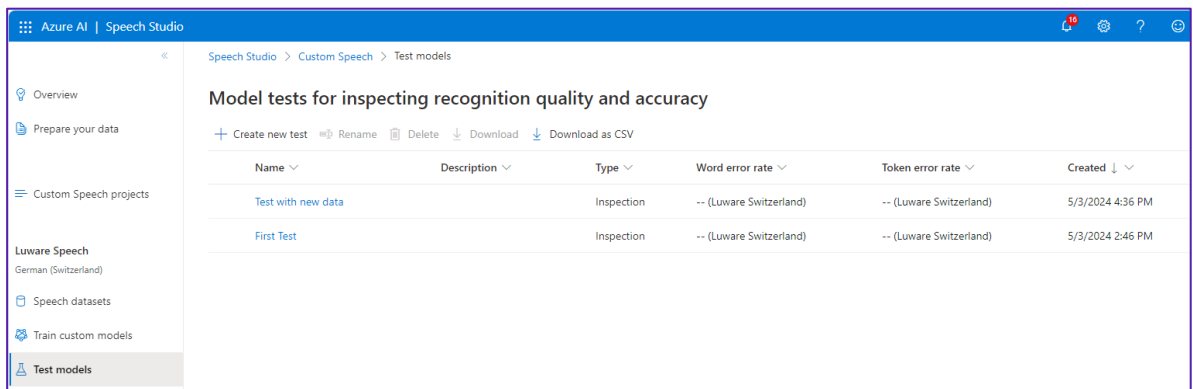
Enter test description (max 500 characters)





Now you can switch to the Tab **deploy models** and deploy the model. As a result, you will get the new model.

To enable the new model in Luware Nimubs, you need to **copy the Key** as well as the **region**, the model is created in.



# Add the Speech Recognizer to Luware Nimbus

You have successfully created and deployed your custom speech model.

## To use it in Luware Nimbus, follow these steps:

- Go to the Azure portal and select your Custom Speech resource.
- On the left pane, click on Keys and Endpoint. You will see two keys and an endpoint URL. Copy one of the keys and the region name from the URL. For example, if the URL is [URL], then the region name is westeurope.
- Open the Luware Nimbus Admin Portal and navigate to Settings > Virtual Assistants.
- Click on the + button to create a new Speech Recognizer. Give it a name and select the organizational unit that will use it.
- Paste the key and the region name that you copied from Azure into the corresponding fields.
- Click on Save to create the Speech Recognizer.
- Now go to Service Settings > Virtual Assistant and check the boxes for Captions and/or Transcription, depending on your preference.
- Select the Speech Recognizer that you just created from the drop-down menu.
- Click on Save to apply the changes.

**Congratulations!** You've successfully integrated your custom speech model into Luware Nimbus.

## To test its performance:

- Initiate a call with a virtual assistant.
- Observe the generated captions and/or transcriptions.

## For ongoing monitoring:

- Visit the Azure portal to track the model's quality and accuracy metrics.

By following these steps, you can assess the effectiveness of your custom speech model in real-world scenarios and make necessary adjustments.



[solutions@luware.com](mailto:solutions@luware.com)

+41 58 404 28 00

[www.luware.com](http://www.luware.com)