DEVOXX™

Master Speech AI APIs to enhance your applications!

Hands-On-Lab

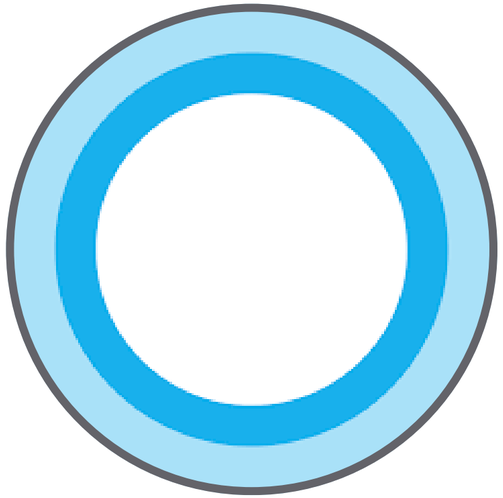
Mathieu Busquet & Eléa Petton

Thusday 08th Oct

OVHcloud

SPEECH AI IS EVERYWHERE TODAY

“Hey Siri”



“Hey Cortana”

“Alexa”

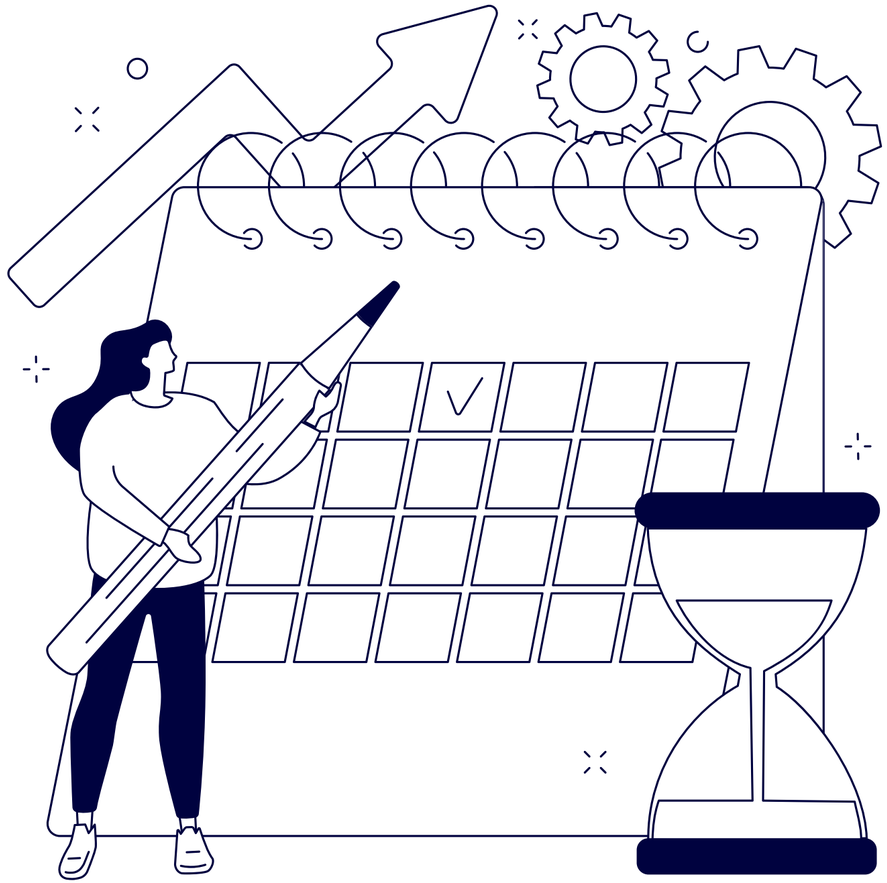
“Ok Google”

ABOUT US

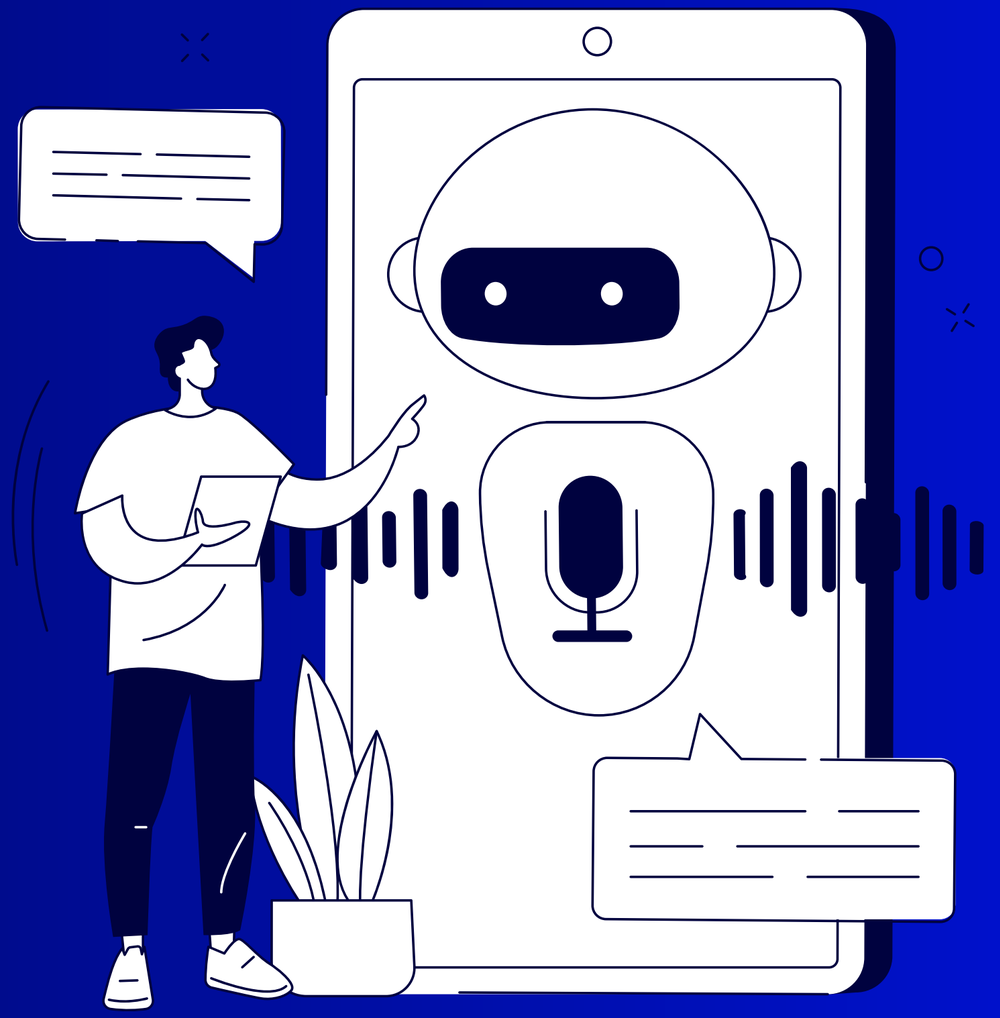
Éléa Petton
Machine Learning Engineer
AI Solutions Team
OVHcloud

Mathieu BUSQUET
Machine Learning Engineer
AI Solutions Team
OVHcloud

AGENDA

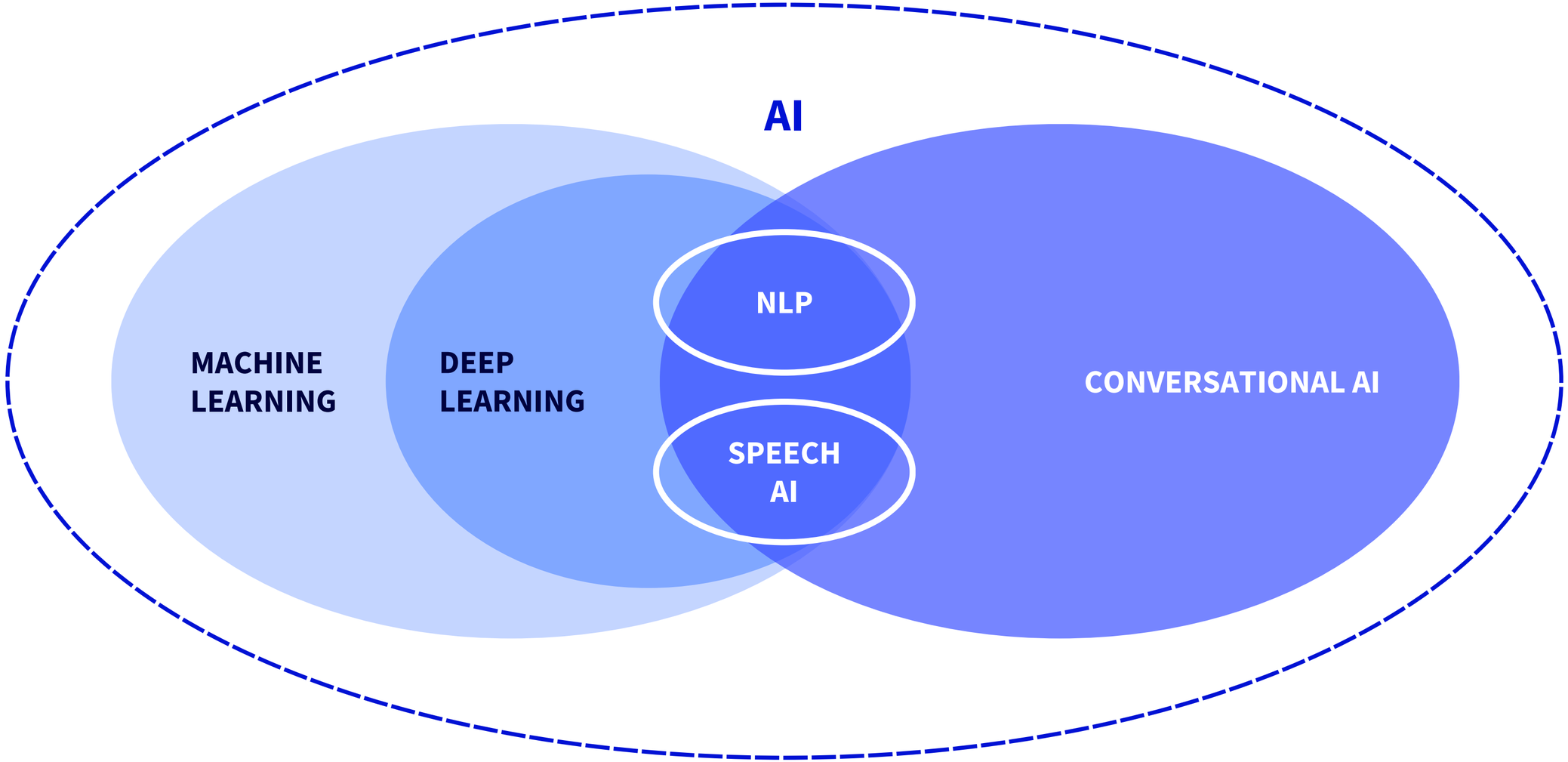


01 INTRODUCTION



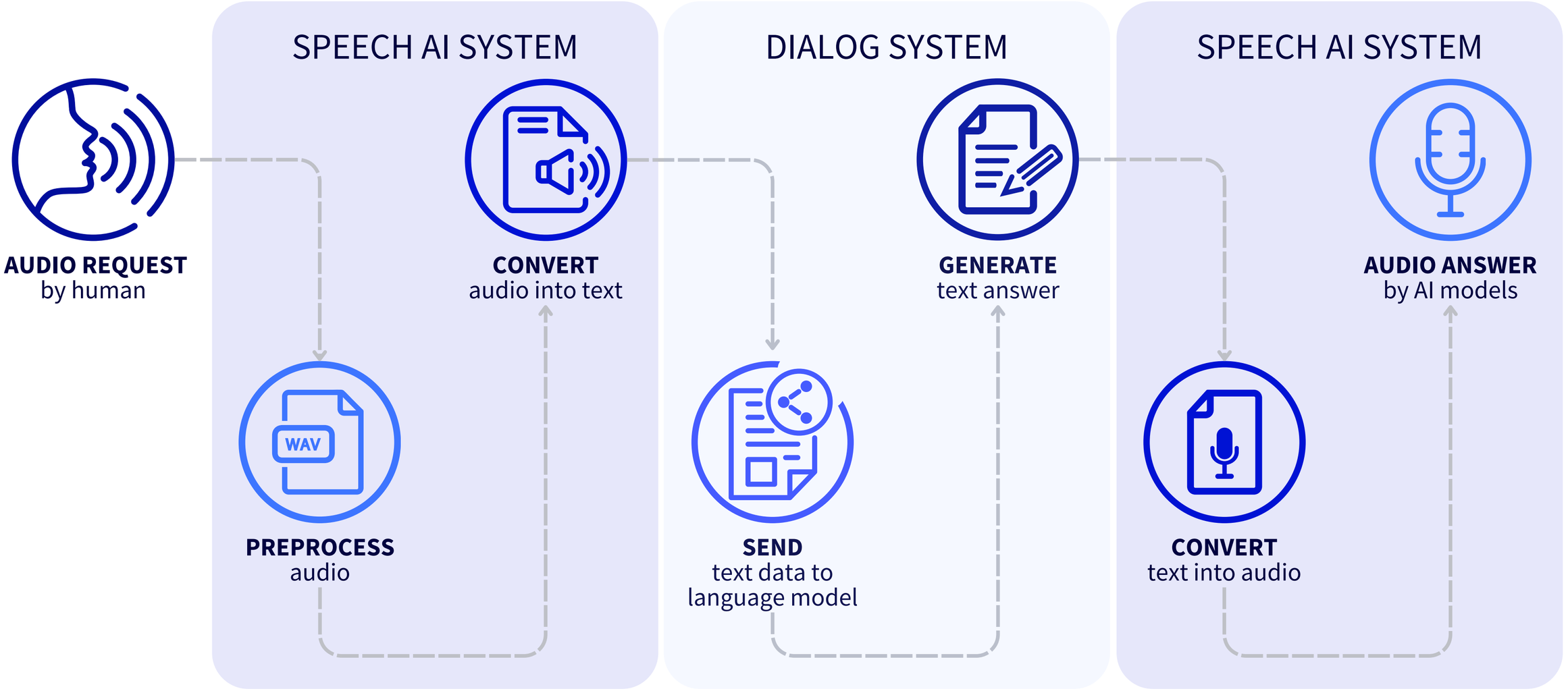
INTRODUCTION

How to consider Speech AI in AI ecosystem?

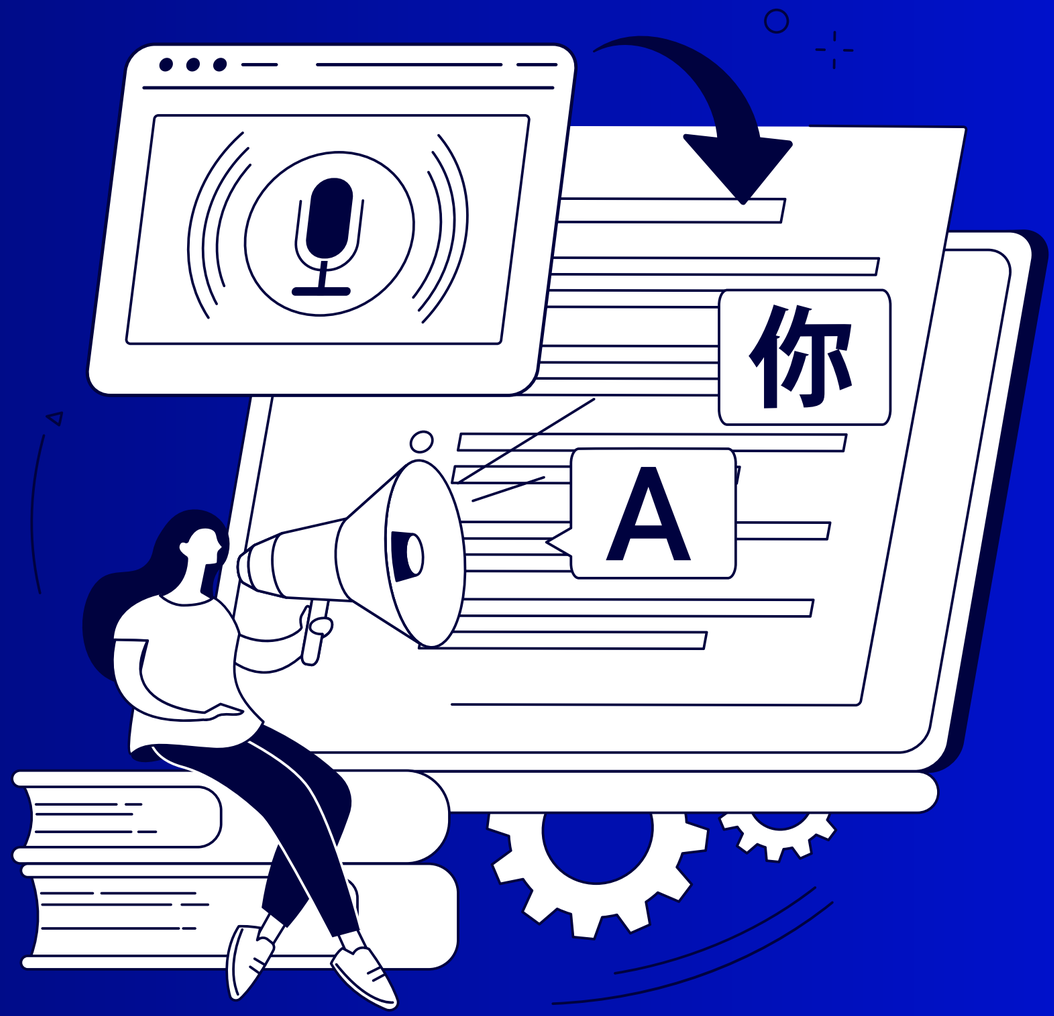


HOW DOES IT WORK?

Virtual Assistants in practice



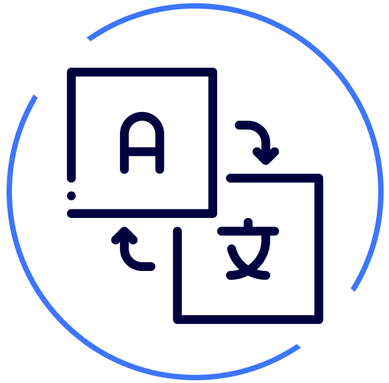
02 SPEECH AI CONCEPTS



SPEECH AI CONCEPTS

3 main tasks

ASR
Automatic Speech Recognition



NMT
Neural Machine Translation

TTS
Text To Speech

SPEECH AI CONCEPTS

3 main tasks

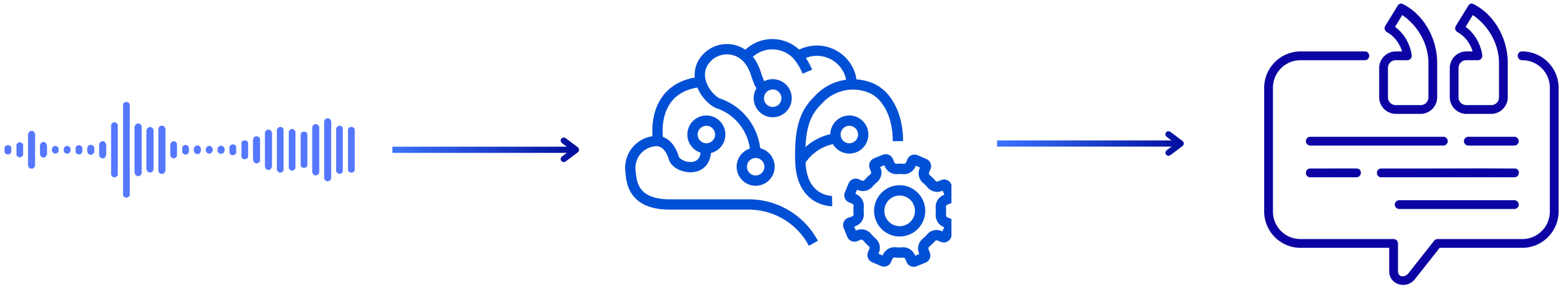
ASR
Automatic Speech Recognition

NMT
Neural Machine Translation

TTS
Text To Speech

AUTOMATIC SPEECH RECOGNITION

Transcribe human voice into text

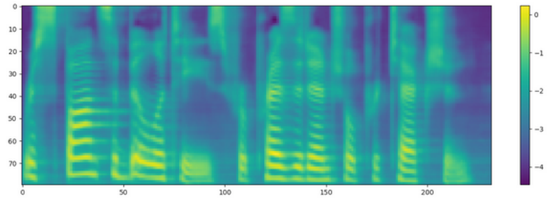


ASR PIPELINE

Transcribe human voice into text

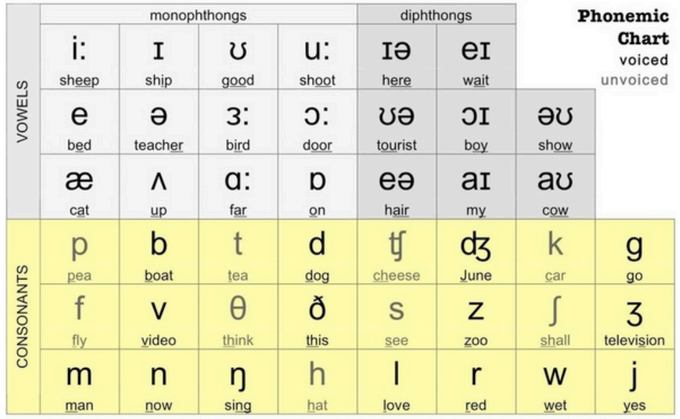
Speech

Feature Extraction



Spectrogram
(Freq. vs Time)

Acoustic Model



/h/ /ə/ /l/ /əʊ/ /h/ /aʊ/ /ɑː/ /j/ /uː/

Language Model

hello how are your
hello how are you
hello now are you

Decoder

hello how are you
~~hello, now are you~~

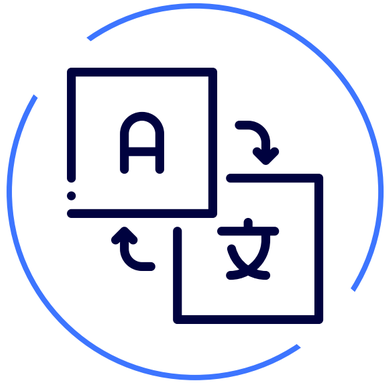
Punctuation & Capitalization Model

Hello, how are you**?**

SPEECH AI CONCEPTS

3 main tasks

ASR
Automatic Speech Recognition

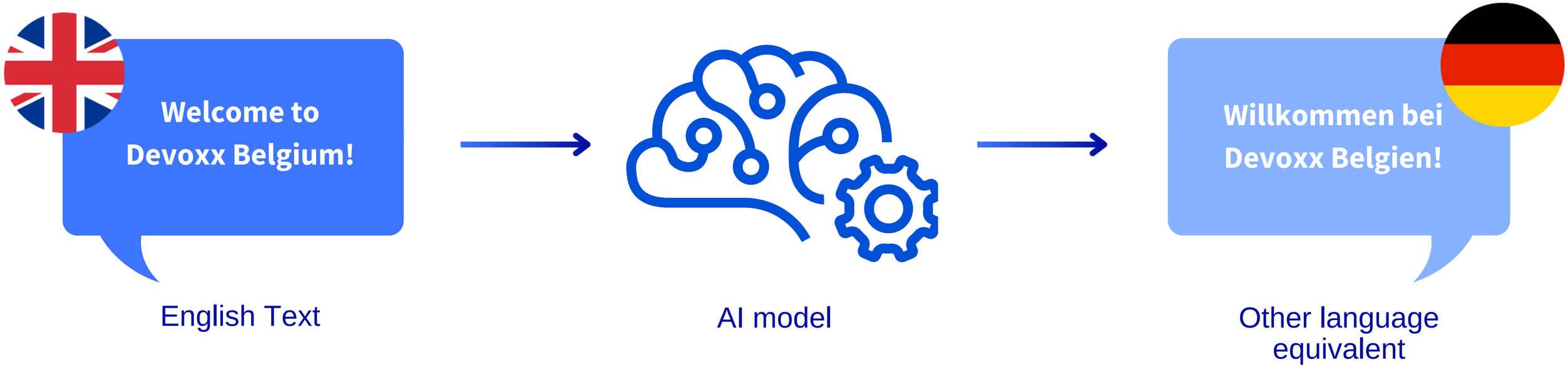


NMT
Neural Machine Translation

TTS
Text To Speech

NEURAL MACHINE TRANSLATION

Translate text into an other language



NMT PIPELINE

Translate text into an other language

English Text

Tokenization

“Welcome”, “to”, “Devoxx”, “Belgium”

NMT Encoder

[0,93
0,2
0,3]

Embeddings
(token representation in its specific context)

NMT Decoder

“Willkommen, “bei”, “Devoxx”, “Belgien”

Detokenization

Other language equivalent

SPEECH AI CONCEPTS

3 main tasks

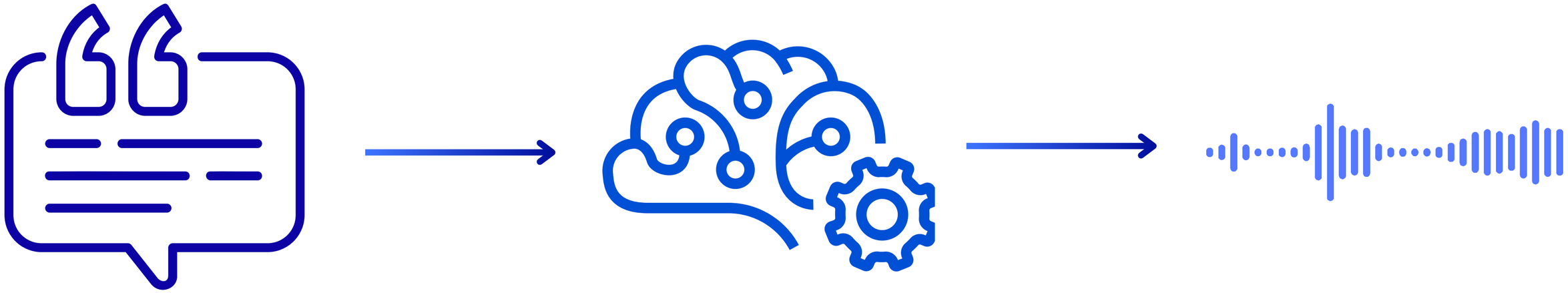
ASR
Automatic Speech Recognition

NMT
Neural Machine Translation

TTS
Text To Speech

TEXT TO SPEECH

Convert text into spoken words



TTS PIPELINE

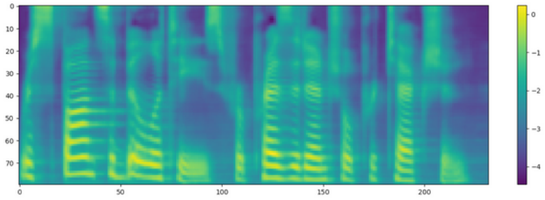
Convert text into spoken words

Text

Text Preprocessing

Text Encoder

Pitch and Duration Predictor for each phoneme



Spectrogram generator

VoCoder

Speech

03 HANDS-ON-LAB

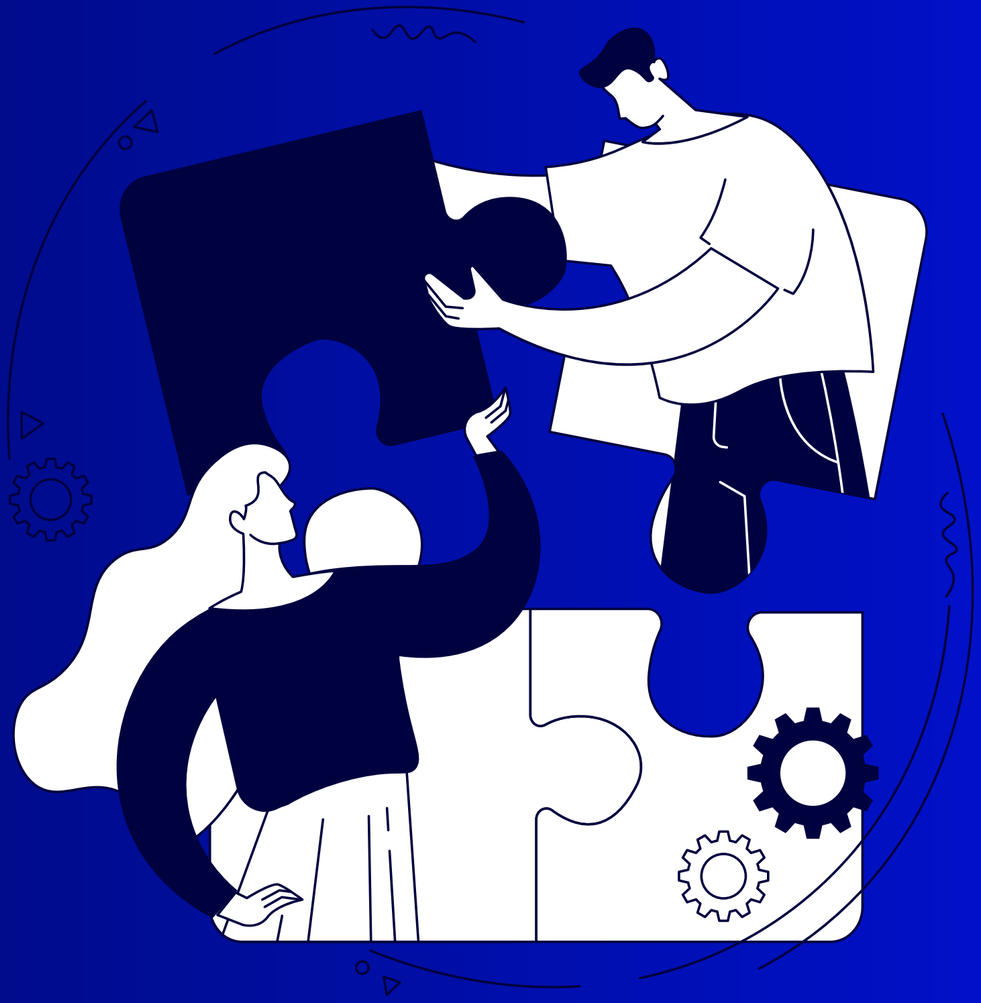


TABLE OF CONTENT

Hands-On-Lab schedule

Discover AI Endpoints

Get started with Speech AI

Focus on Speech AI key features

Develop Speech AI inference functions

Develop Gradio web app

Containerize web app

Deploy and test Video Translator app

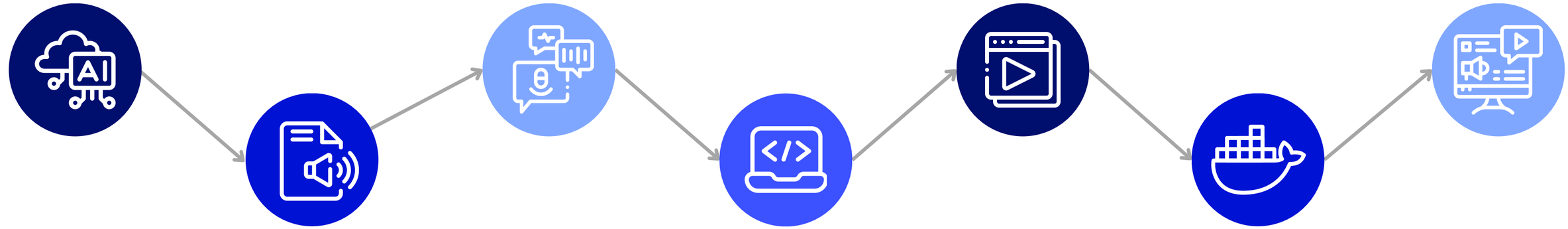


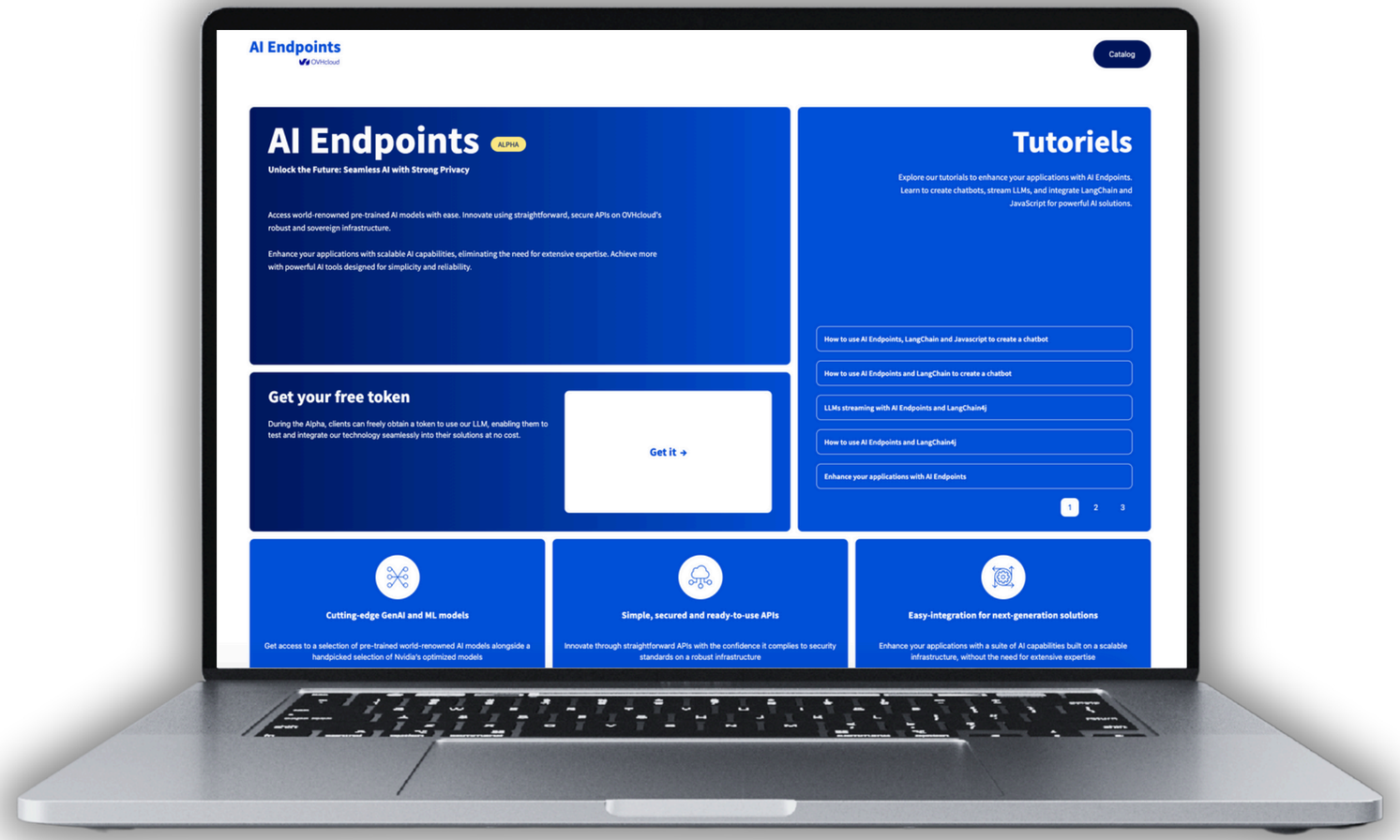
TABLE OF CONTENT

Hands-On-Lab schedule

Discover AI Endpoints

DISCOVER AI ENPOINTS

OVHcloud AI Endpoints website



DISCOVER AI ENPOINTS

Catalog of AI models

“ A **serverless platform** providing access to advanced AI models, such as **LLM,** NLP, translation, speech recognition, or image recognition. ”

ASSISTANT

AUDIO ANALYSIS

COMPUTER VISION

EMBEDDINGS

NLP

TRANSLATION

DISCOVER AI ENPOINTS

Democratizing AI

“ Improve your applications with **AI Endpoints** ”

DESIGNED FOR DEVELOPERS

with complete documentation, simple APIs, and code examples

COMMITTED TO PRIVACY

we do not store and do not share your data during or after the use of the model

CURATED LIST OF AI MODELS

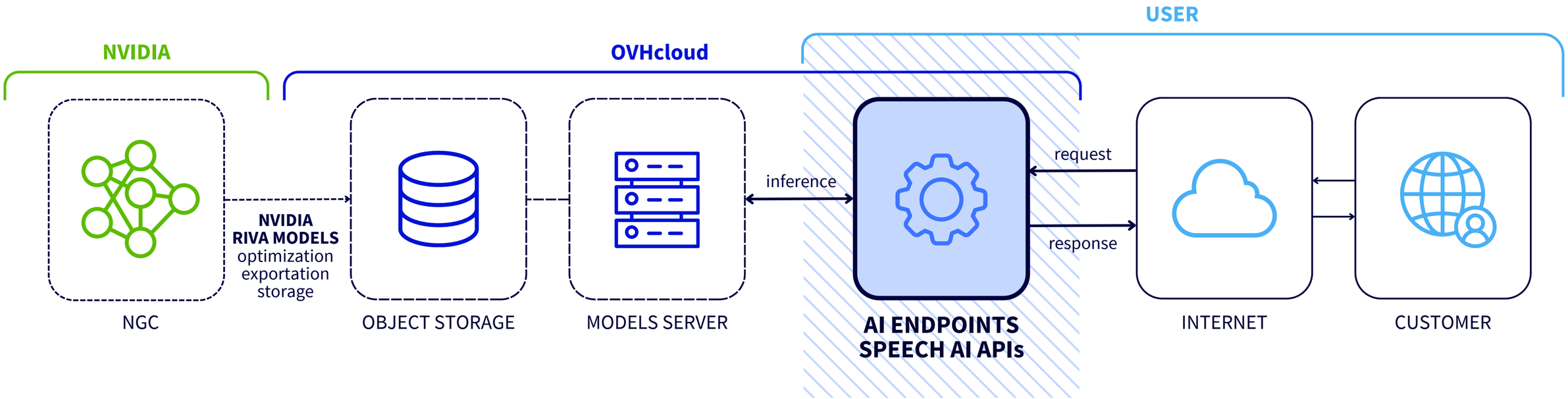
making available the latest models, optimized for maximum performance and accuracy

LOCK-IN FREE TECHNOLOGY

thanks to our transparency about the AI models used, clients can implement these models on their own infrastructure or on other cloud services

DISCOVER AI ENPOINTS

Concerning Speech AI Endpoints



DISCOVER AI ENPOINTS

Concerning Speech AI Endpoints

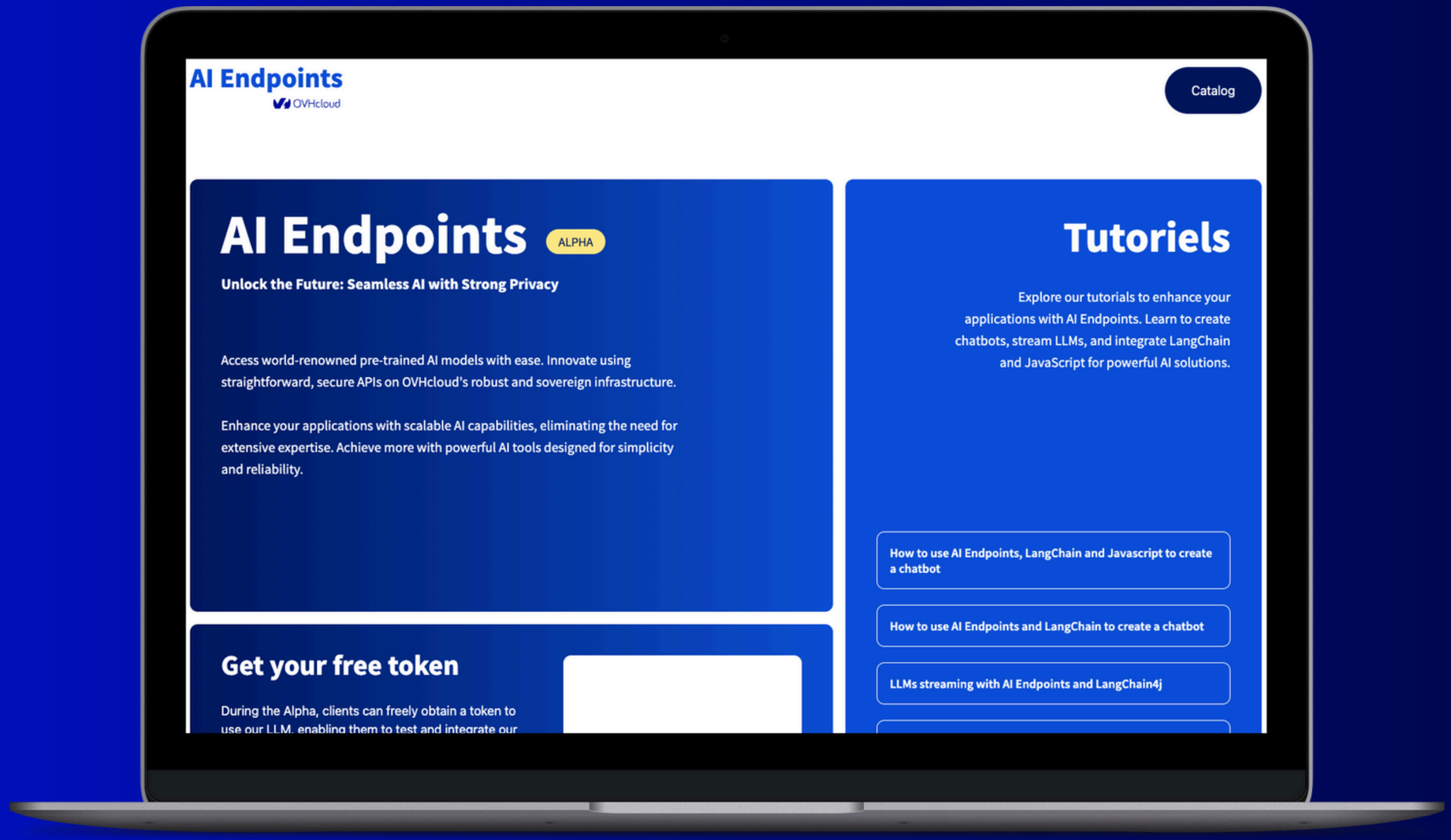
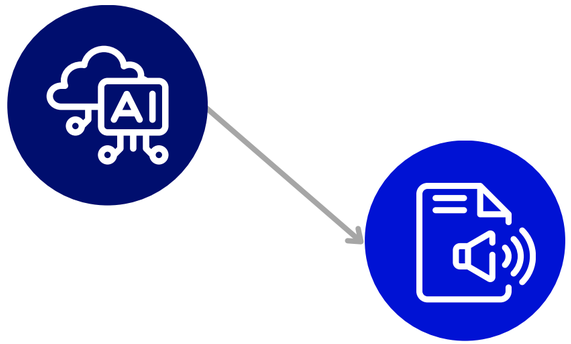


TABLE OF CONTENT

Hands-On-Lab schedule

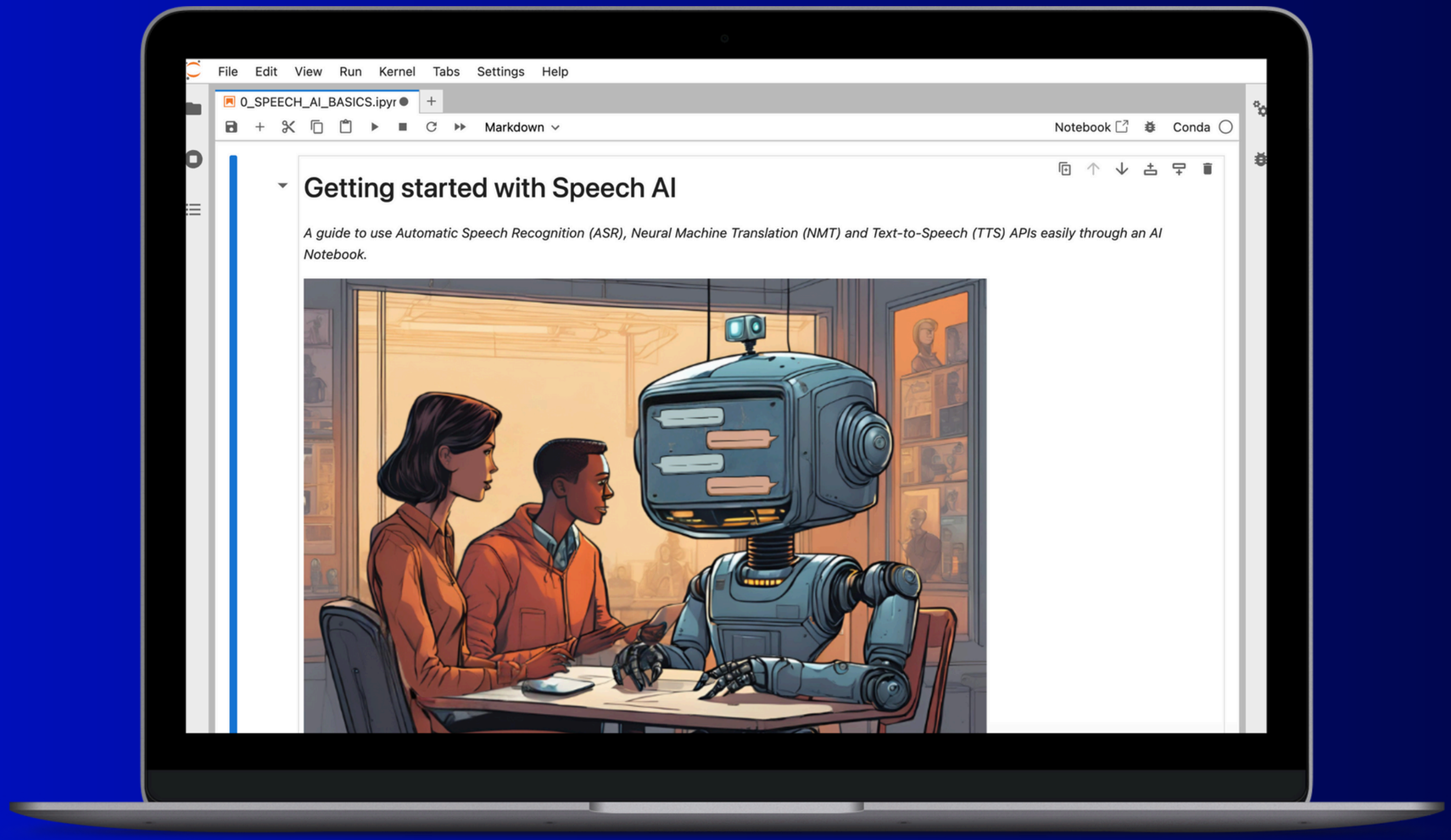
Discover AI Endpoints



Get started with Speech AI

GET STARTED WITH SPEECH AI

How to use ASR, NMT and TTS APIs easily?



GET STARTED WITH SPEECH AI

Connect your Speech AI Endpoints to each other!

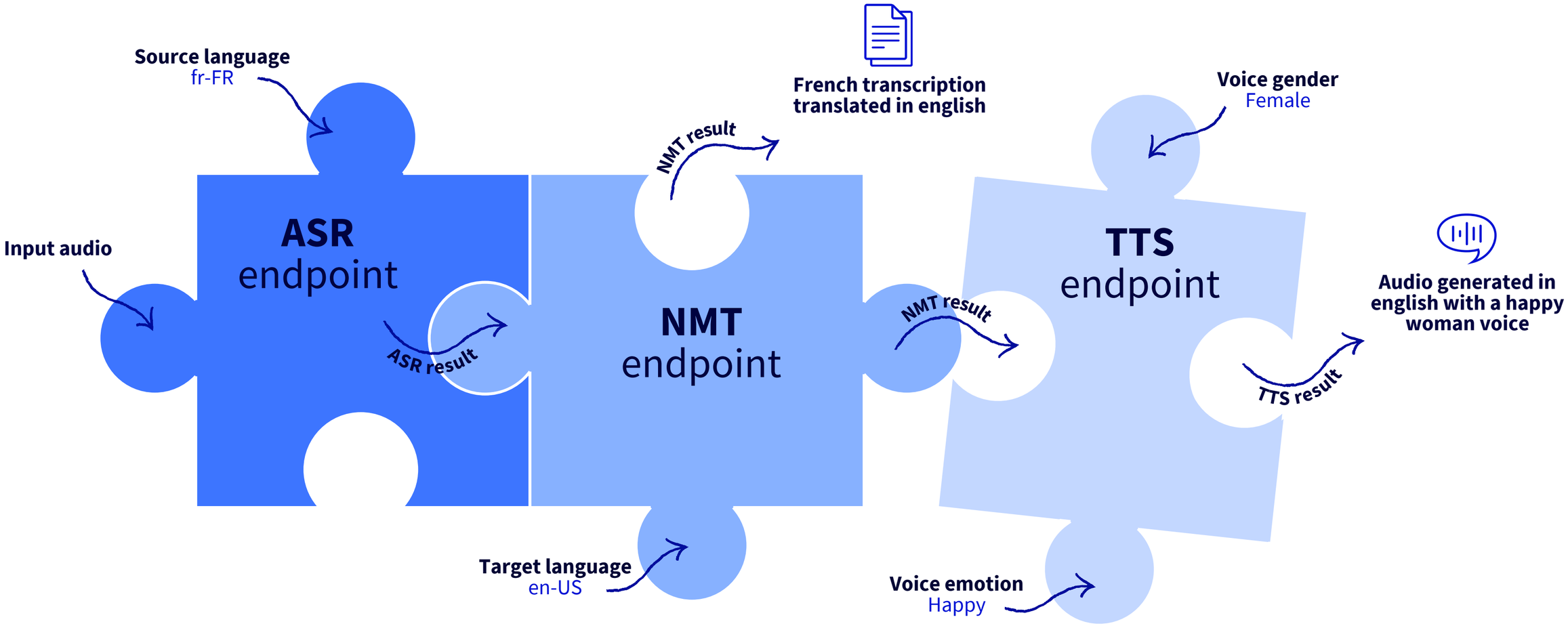
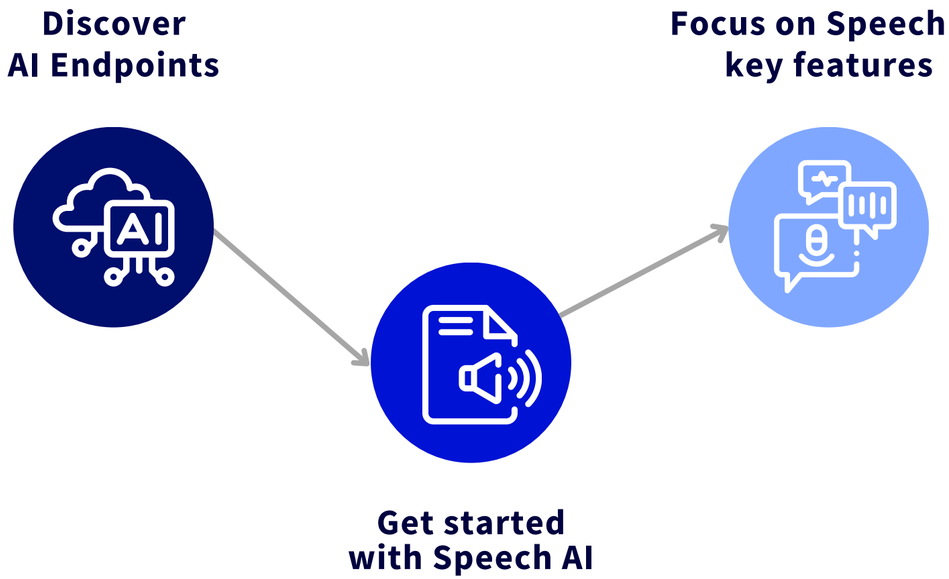


TABLE OF CONTENT

Hands-On-Lab schedule

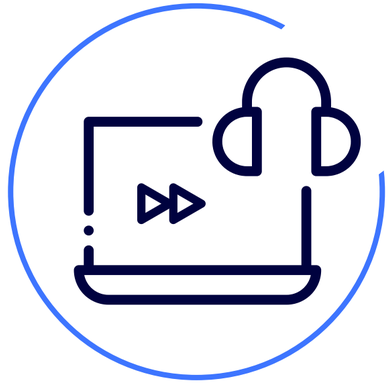


FOCUS ON SPEECH AI KEY FEATURES

Enhance your Speech AI Endpoints with cutting-edge features

GENERATE
SRT file

KEEP SILENCE
during translation



SUPERIMPOSE
audio on video

FOCUS ON SPEECH AI KEY FEATURES

Master Speech AI API endpoints by developing key features

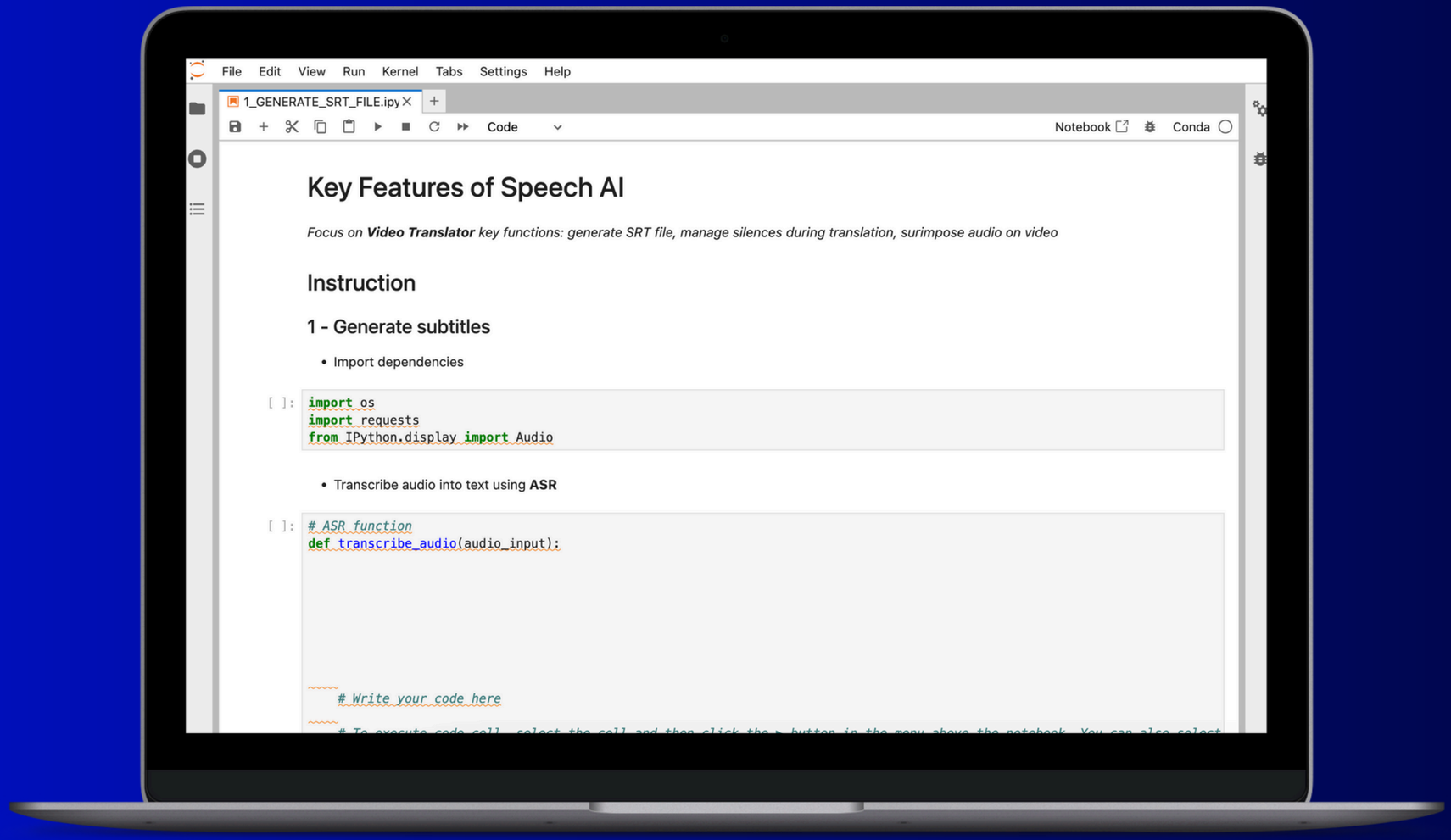
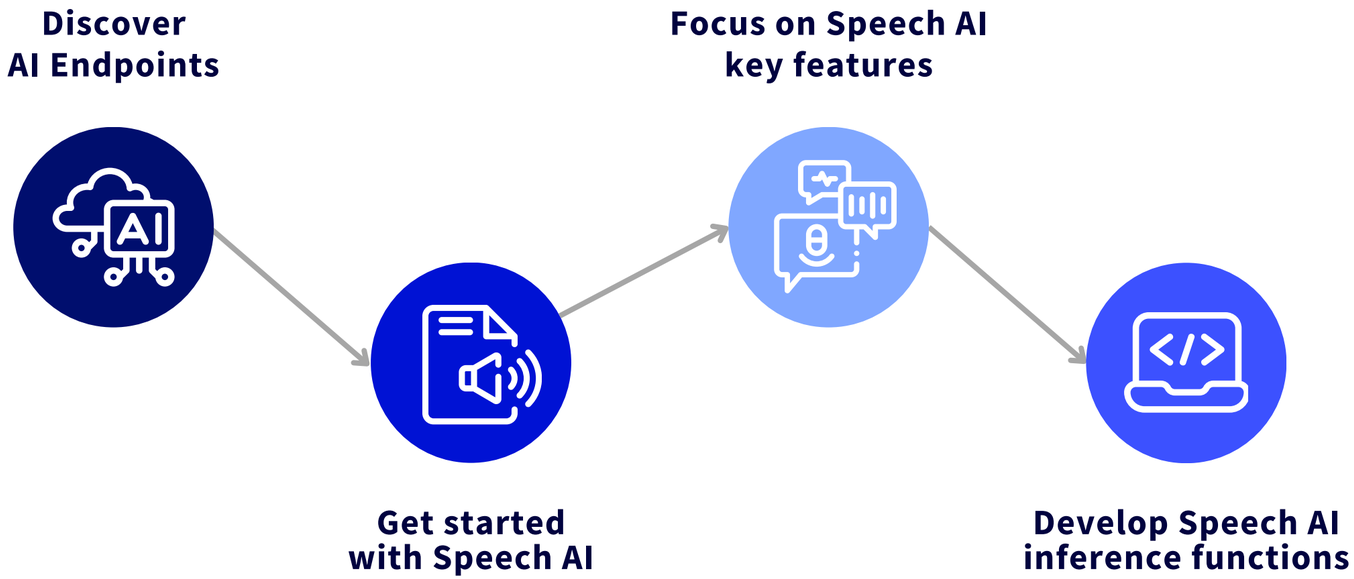


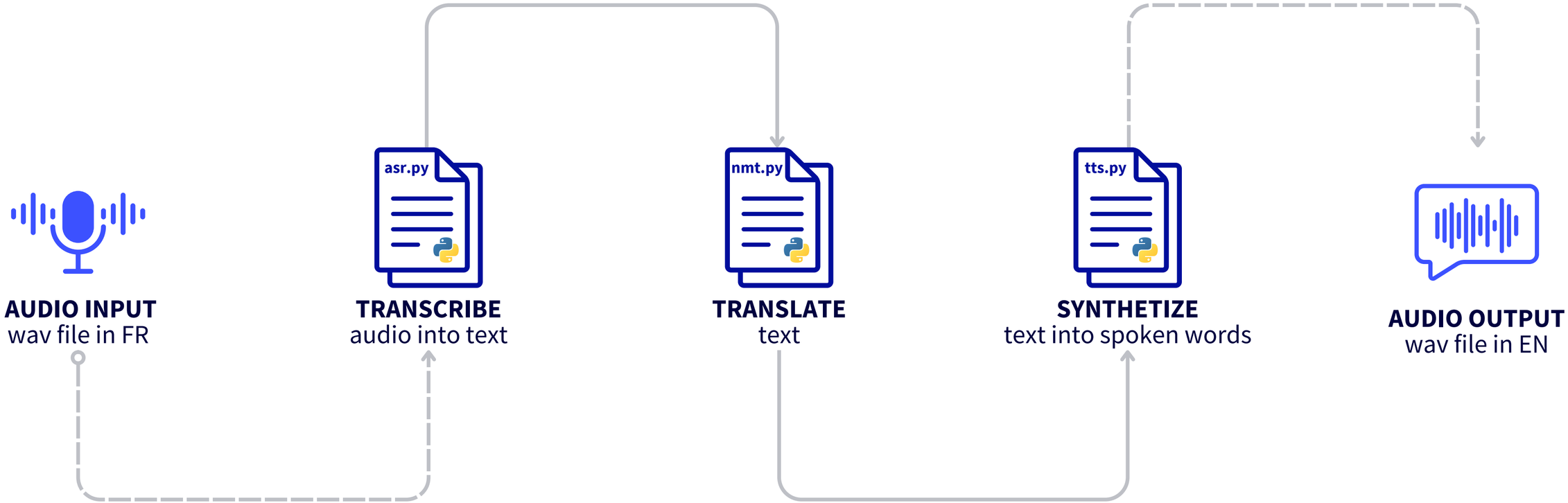
TABLE OF CONTENT

Hands-On-Lab schedule



DEVELOP SPEECH AI INFERENCE FUNCTIONS

Develop ASR, NMT and TTS scripts



DEVELOP SPEECH AI INFERENCE FUNCTIONS

Use Speech AI Endpoints in Python scripts

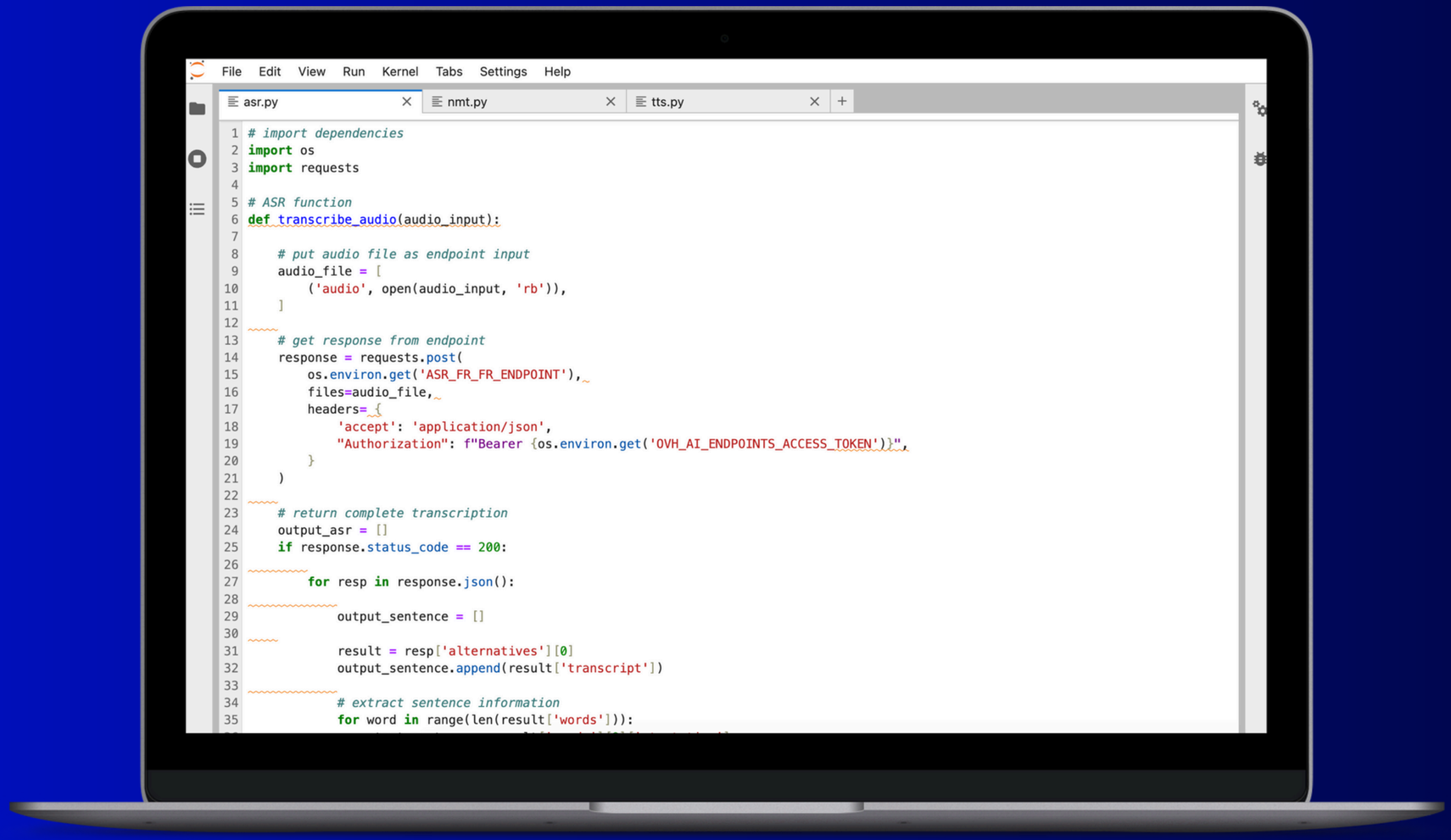
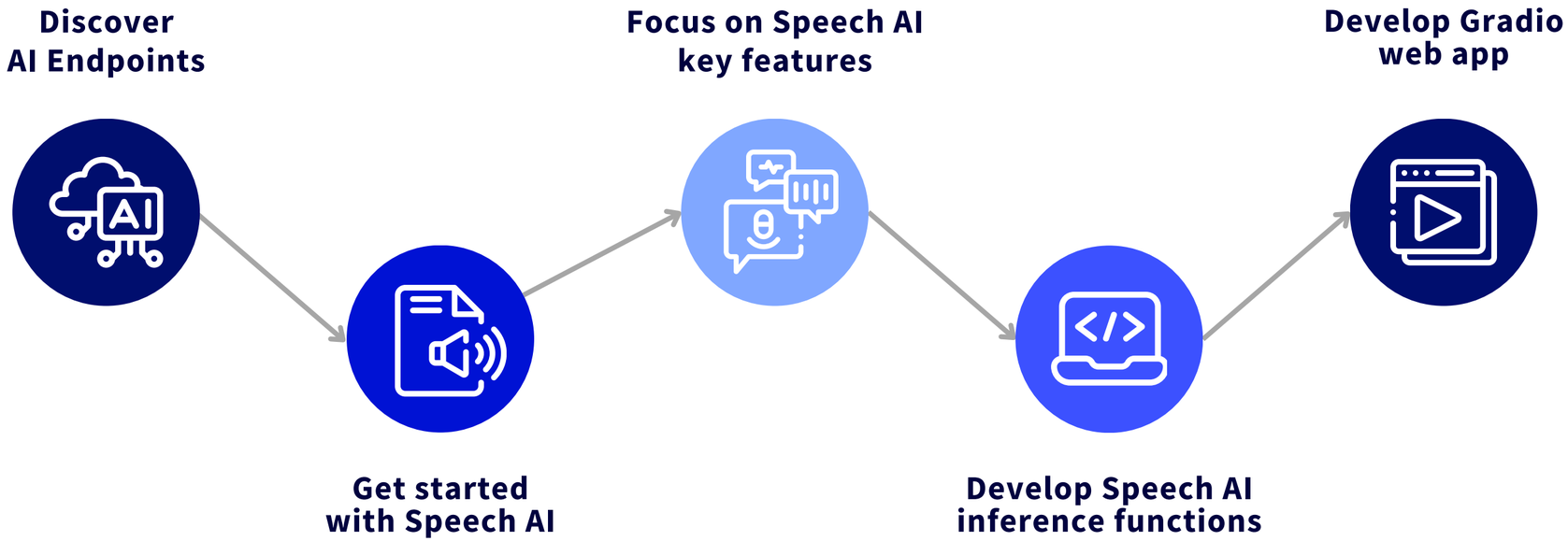


TABLE OF CONTENT

Hands-On-Lab schedule



DEVELOP GRADIO WEB APP

Gradio web app overview

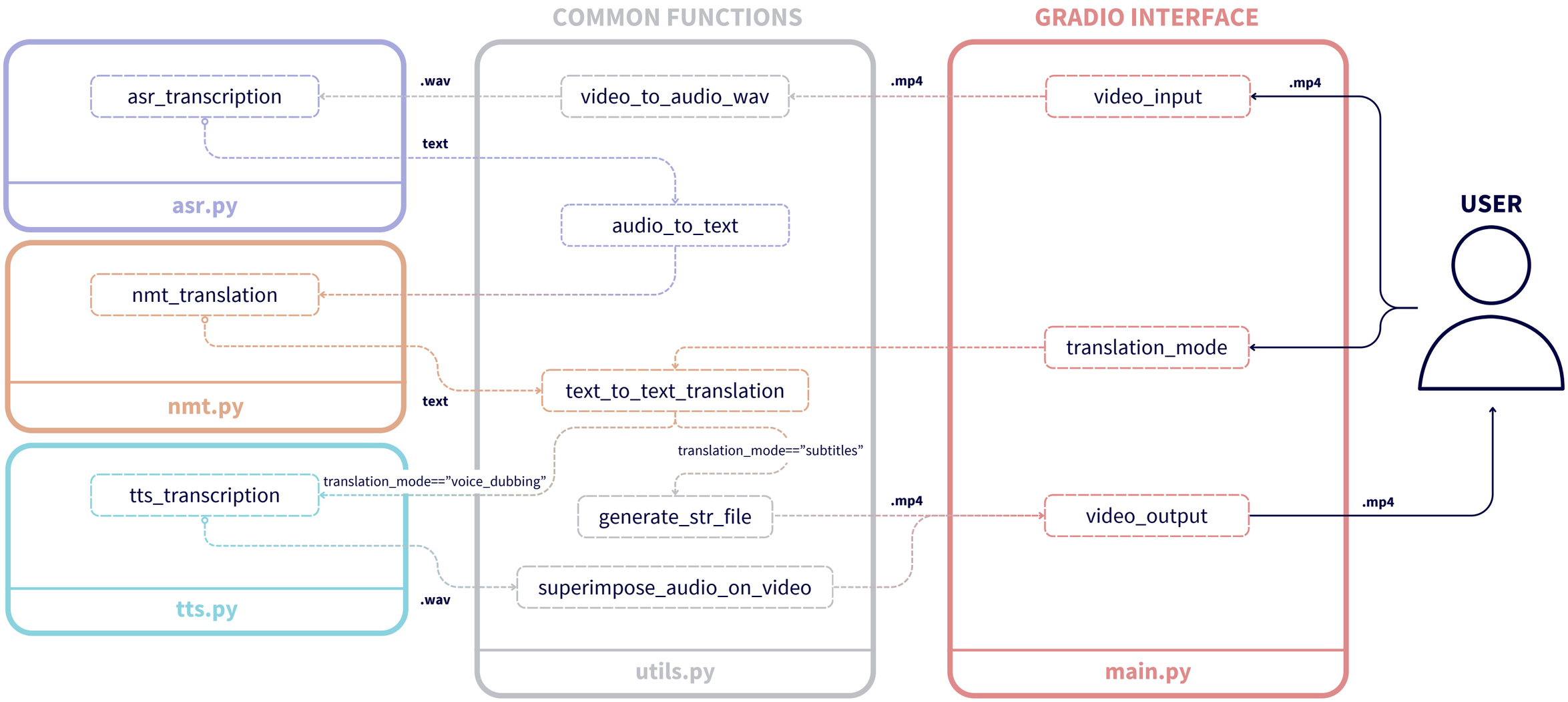
DEVELOP GRADIO WEB APP

Define Gradio app features

- **Upload** a video
- **Transcribe** the audio part into text
- **Subtitle** video in any language
- **Dub** the video in another language
- **Choose** the gender of the dubbing voice
- **Download** resulting video

DEVELOP GRADIO WEB APP

Connect ASR, NMT and TTS inside a Gradio app



DEVELOP GRADIO WEB APP

Develop your Gradio app in Python

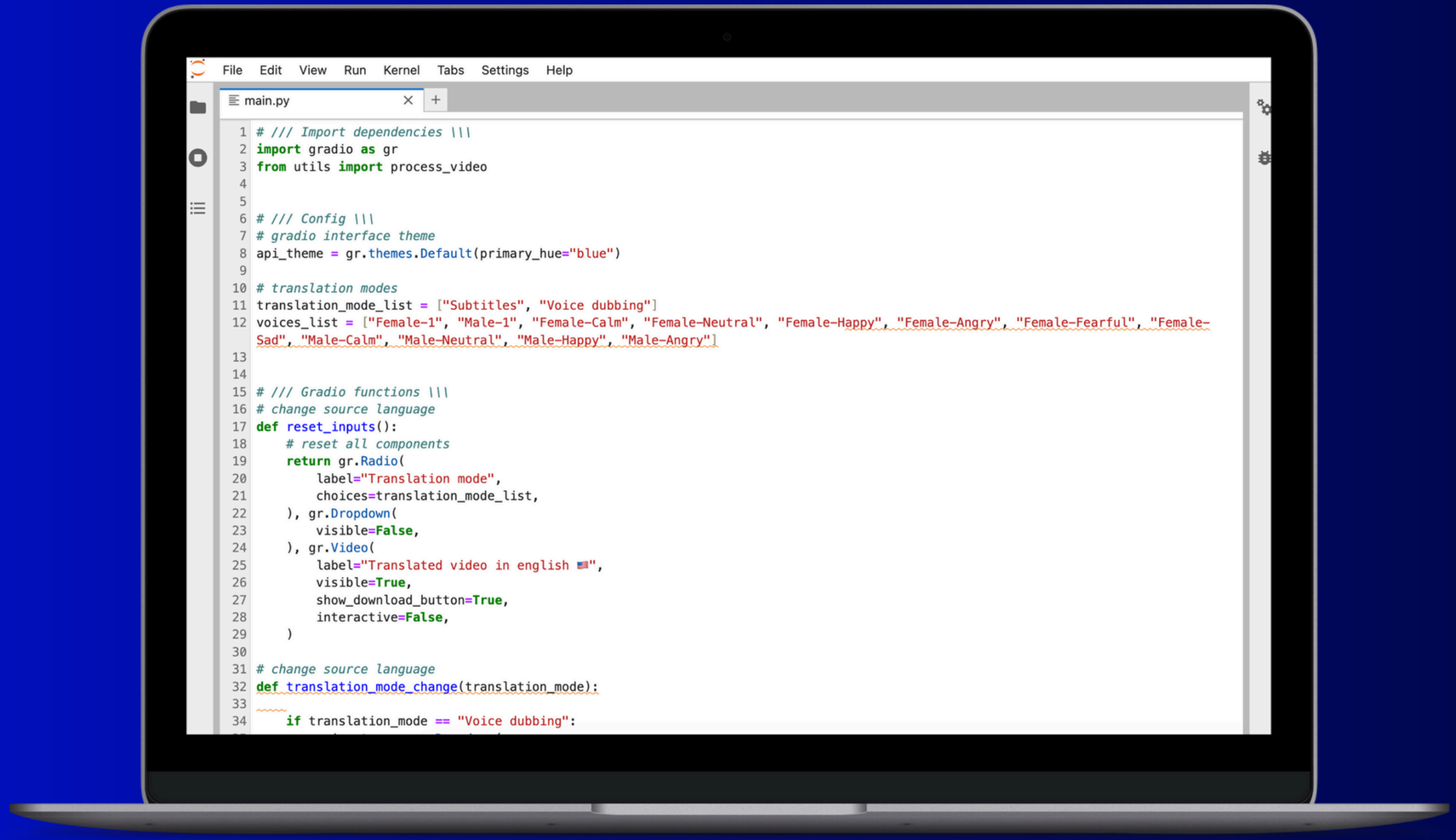
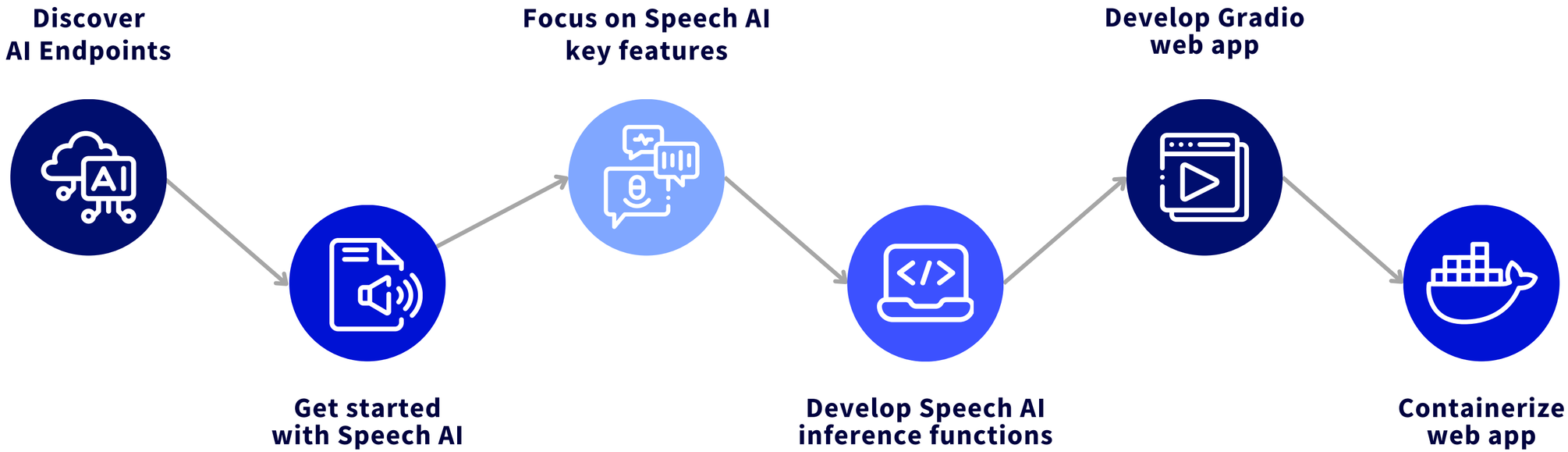


TABLE OF CONTENT

Hands-On-Lab schedule



CONTAINERIZE WEB APP

Create the Dockerfile

```
FROM python:3.10


WORKDIR /workspace

ADD . /workspace


RUN apt-get update && apt-get install -y ffmpeg libsndfile1-dev

RUN pip install -r requirements.txt


RUN chown -R 42420:42420 /workspace

ENV HOME=/workspace

CMD [ "python3" , "/workspace/main.py" ]
```

Dockerfile

TABLE OF CONTENT

Hands-On-Lab schedule

Discover AI Endpoints

Get started with Speech AI

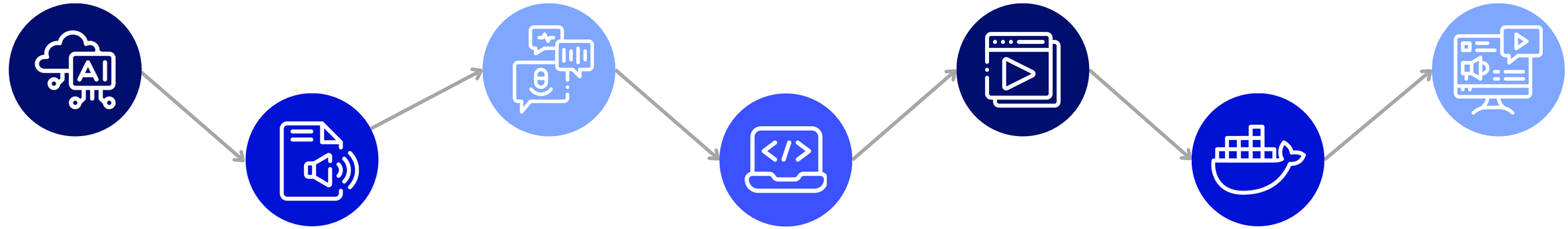
Focus on Speech AI key features

Develop Speech AI inference functions

Develop Gradio web app

Containerize web app

Deploy and test Video Translator app



CONTAINERIZE WEB APP

Define your infrastructure requirements

COMPUTE RESOURCES

- 4 CPUs

HIGH AVAILABILITY

- API scalable on the fly
- Custom number of replicas

SECURE ACCESS

- Private mode
- Personal token access

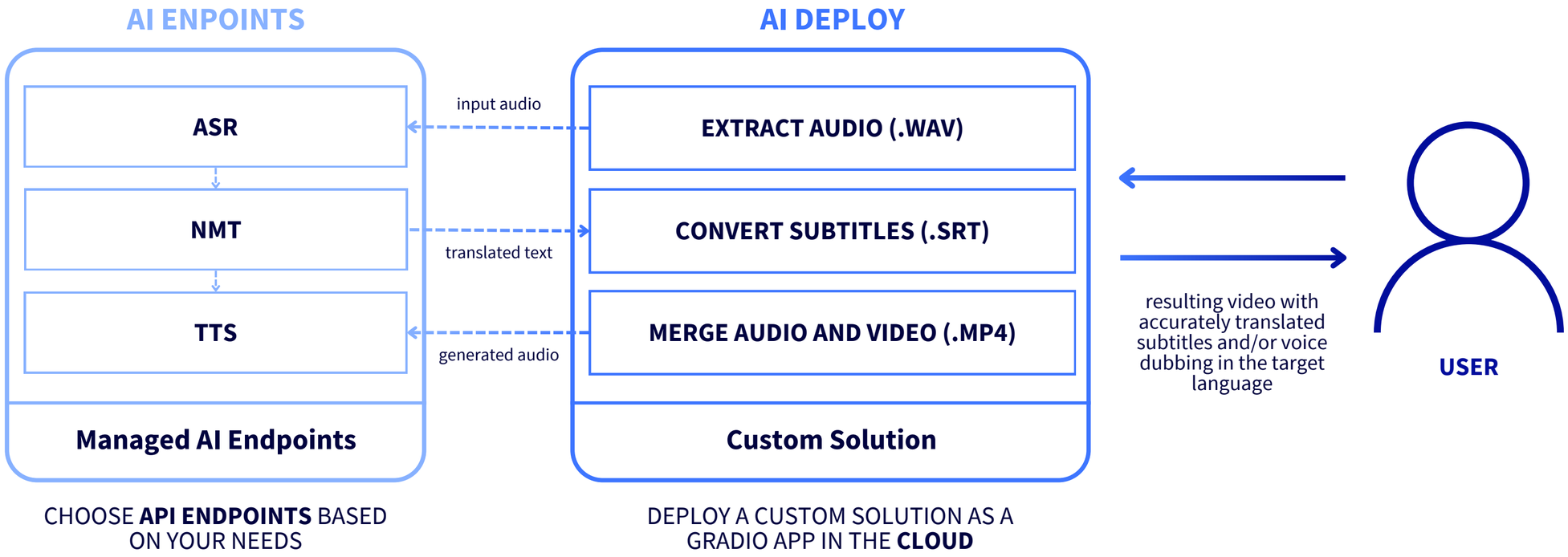
DEPLOY AND TEST VIDEO TRANSLATOR APP

Introduction to OVHcloud AI Deploy solution

01	**CONTAINER AS A SERVICE**	Customer provides a docker container through Docker registry
02	**COMPUTE RESOURCES**	Container is running in the cloud over GPU (or CPU)
03	**BILLING METHOD**	Customer is billed per minute used
04	**API/APP DEPLOYMENT**	Industrial way of deploying stateless API(s)
05	**SCALING STRATEGY**	Scalable on the fly

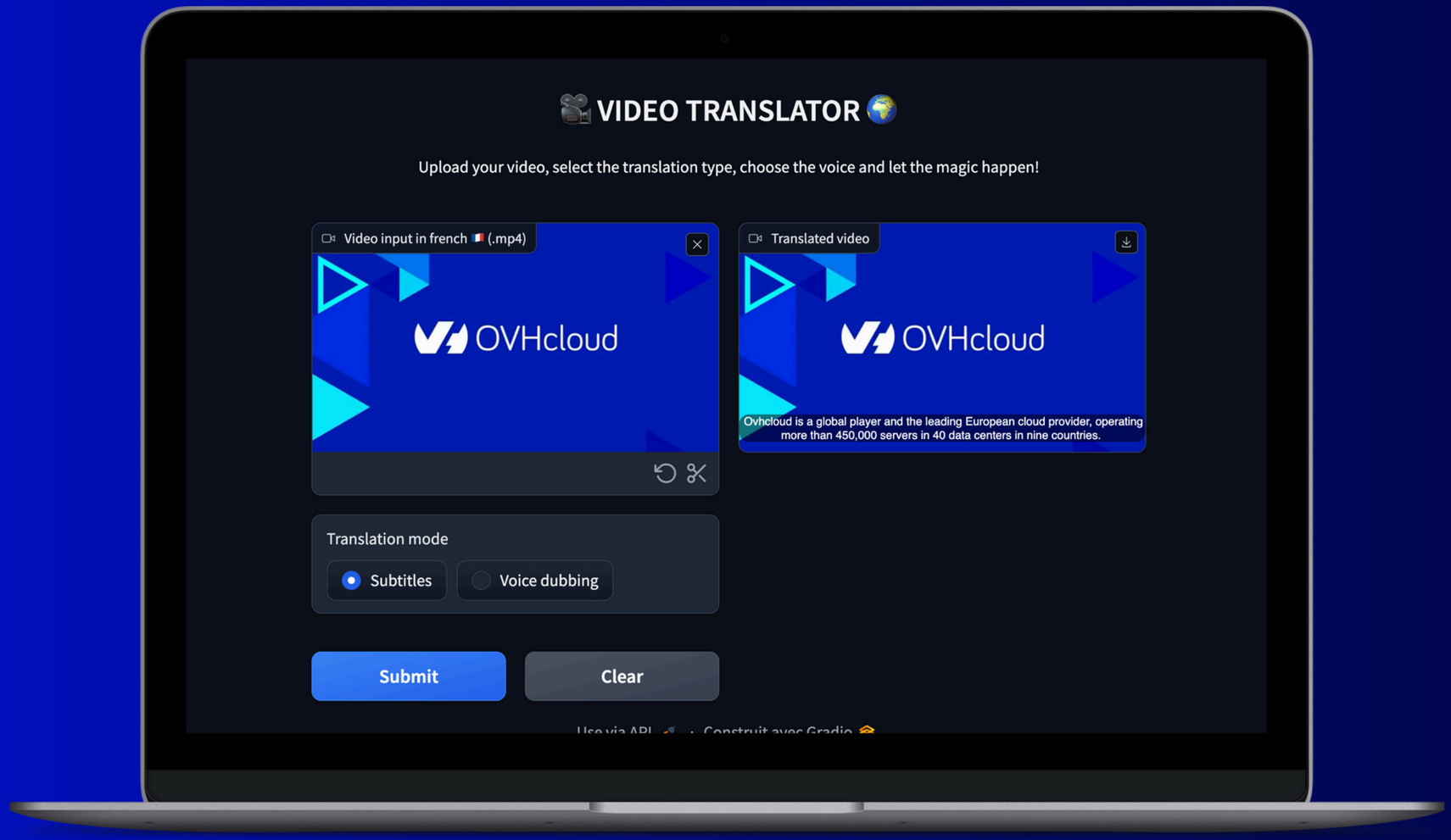
DEPLOY AND TEST VIDEO TRANSLATOR APP

Choose OVHcloud AI Deploy to deploy the solution



DEPLOY AND TEST VIDEO TRANSLATOR APP

Launch the app deployment in the cloud

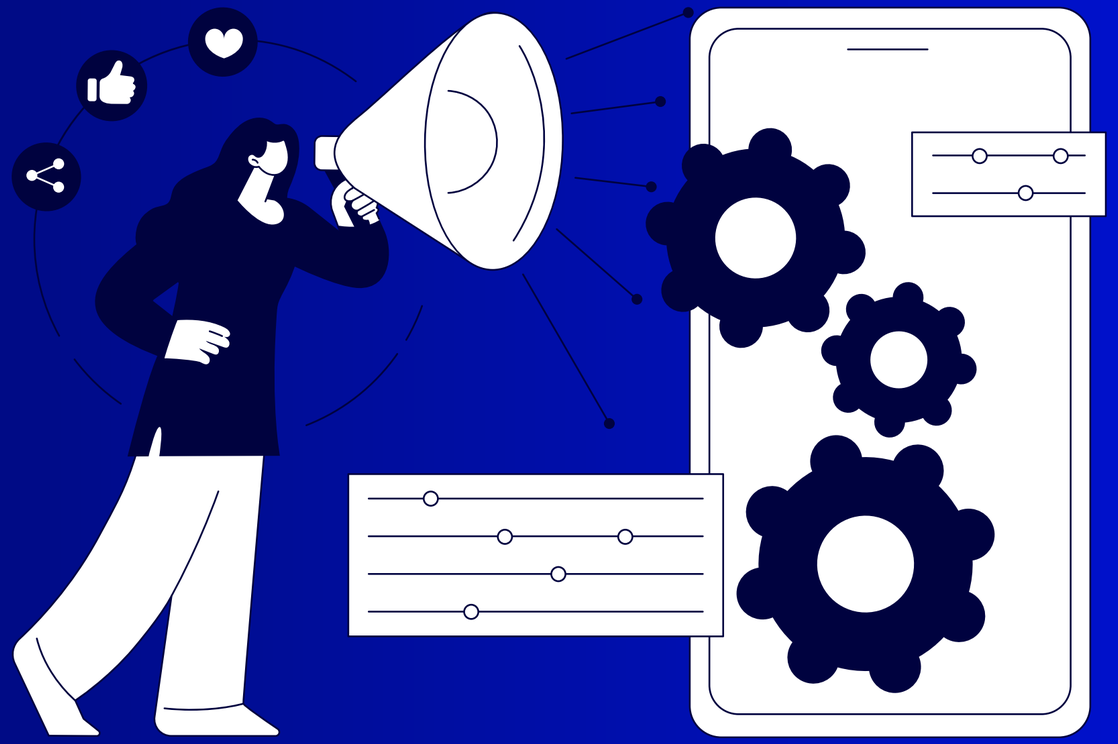


DEPLOY AND TEST VIDEO TRANSLATOR APP

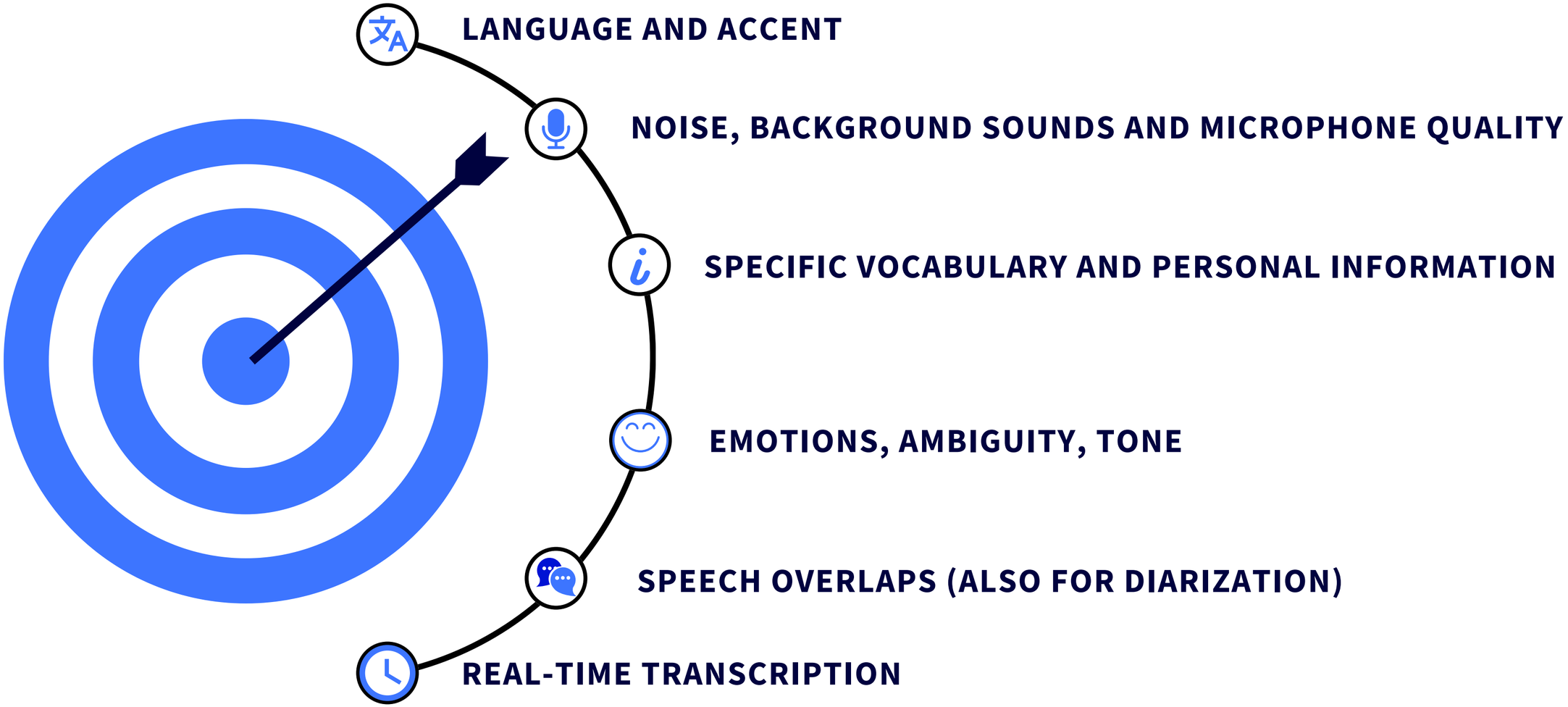
Live testing time!

https://bit.ly/video-translator-devoxx-be

04 CHALLENGES OF SPEECH AI



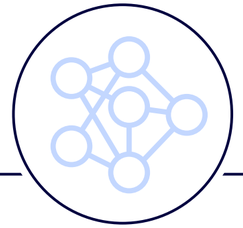
CHALLENGES OF SPEECH AI



05 CONCLUSION

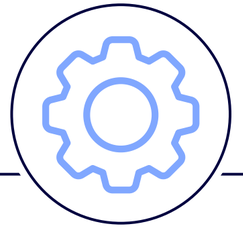
CONCLUSION

AI Endpoints and the Video Translator in summary



Speech AI models

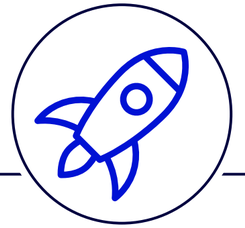
- **speech** and **translation** AI models can work together
- **real-time** conversational AI
- **accurate models** in several languages
- **transcribe** any audio into text
- transform text into **spoken words**



OVHcloud AI Endpoints

- **cutting-edge** GenAI and ML models
- **simple**, **secured** and **ready-to-use** AI APIs
- **easy-integration** for next-generation solutions

OVHcloud



OVHcloud AI Deploy

- turnkey **serverless** solution
- **Container As A Service** platform
- the best of **GPUs** (H100, A100, L4, L40S, V100S)
- **per-minute** billing

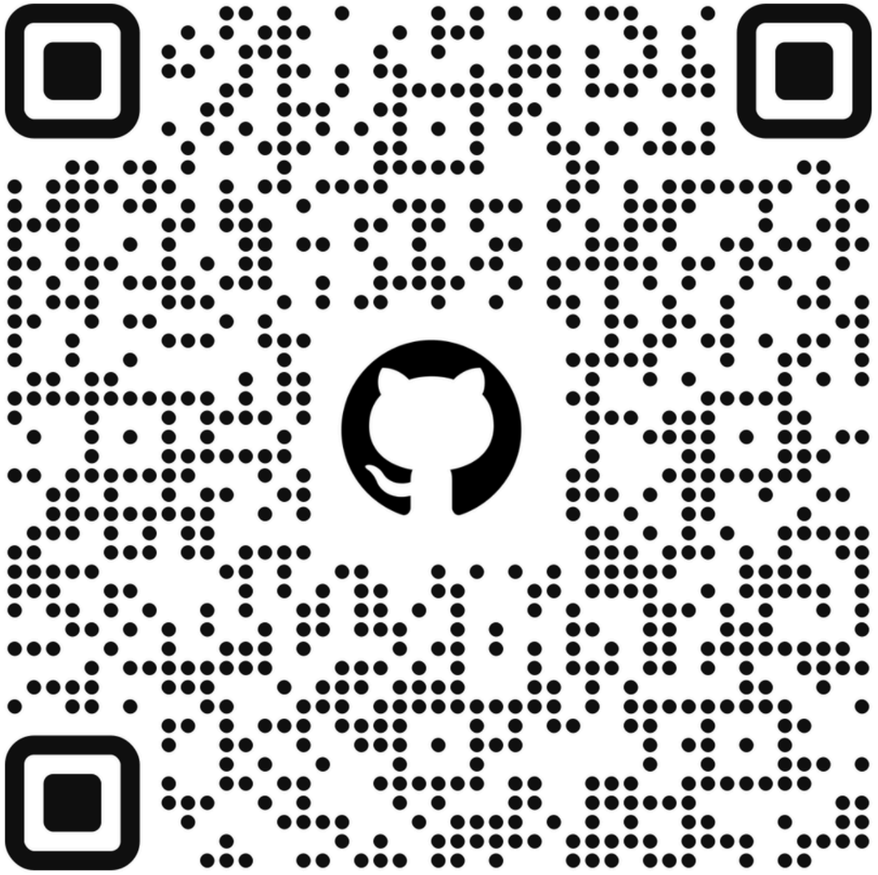
OVHcloud

Going further

- build your own solution based on **Open-Source models**
- create an **Audio Virtual Assistant** in less than 100 lines of code
- **transcribe** and **summarize** any meetings using **diarization**
- **Be creative!**

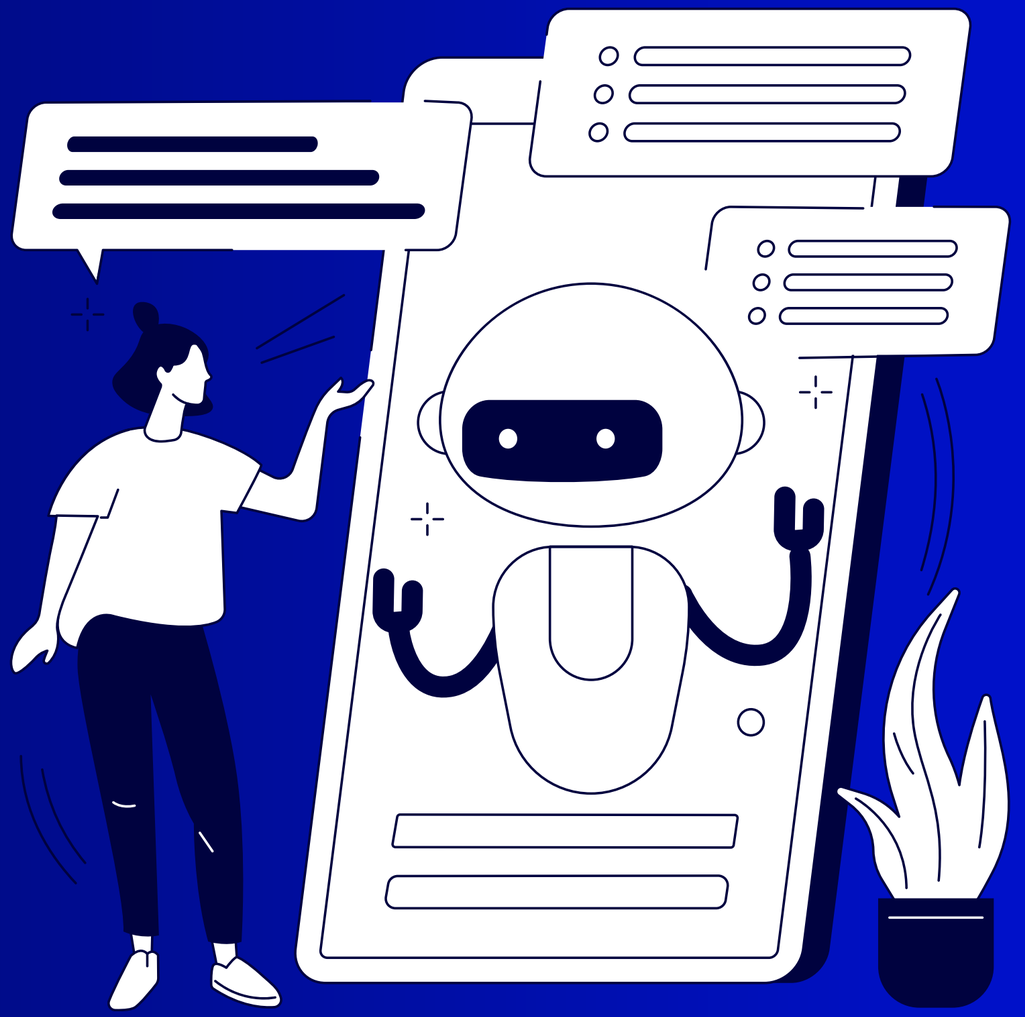
CONCLUSION

Find out Hands-On-Lab resources



https://github.com/eleapttn/workshop-mastering-speech-ai.git

06 REFERENCES



REFERENCES

AI Endpoints and the Video Translator in summary

AI Endpoints

- https://endpoints.ai.cloud.ovh.net/

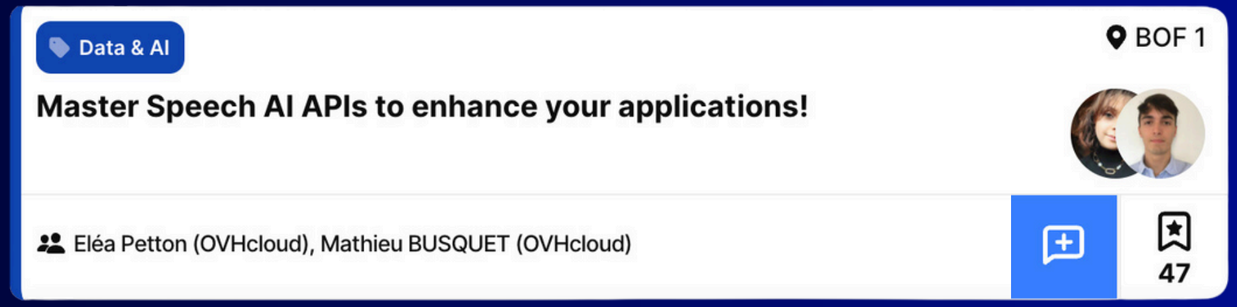
GitHub repositories

- https://github.com/eleapttn/workshop-mastering-speech-ai.git
- https://github.com/ovh/public-cloud-examples/tree/main/ai/ai-endpoints

Blog articles

- https://blog.ovhcloud.com/master-speech-ai-and-build-your-own-video-translator-app-with-ai-endpoints/
- https://blog.ovhcloud.com/how-to-build-a-speech-to-text-application-with-python-1-3/
- https://blog.ovhcloud.com/build-a-powerful-audio-virtual-assistant-with-ai-endpoints/
- https://blog.ovhcloud.com/create-audio-summarizer-assistant-with-ai-endpoints/

THANK YOU!


Data & AI
BOF 1
Master Speech AI APIs to enhance your applications!
Eléa Petton (OVHcloud), Mathieu BUSQUET (OVHcloud)
47