



# SPEECH AI

AU SERVICE DE LA TRANSFORMATION  
MULTIMÉDIA MULTILINGUE



# PRÉSENTATION

# PRÉSENTATION

**Eléa Petton**

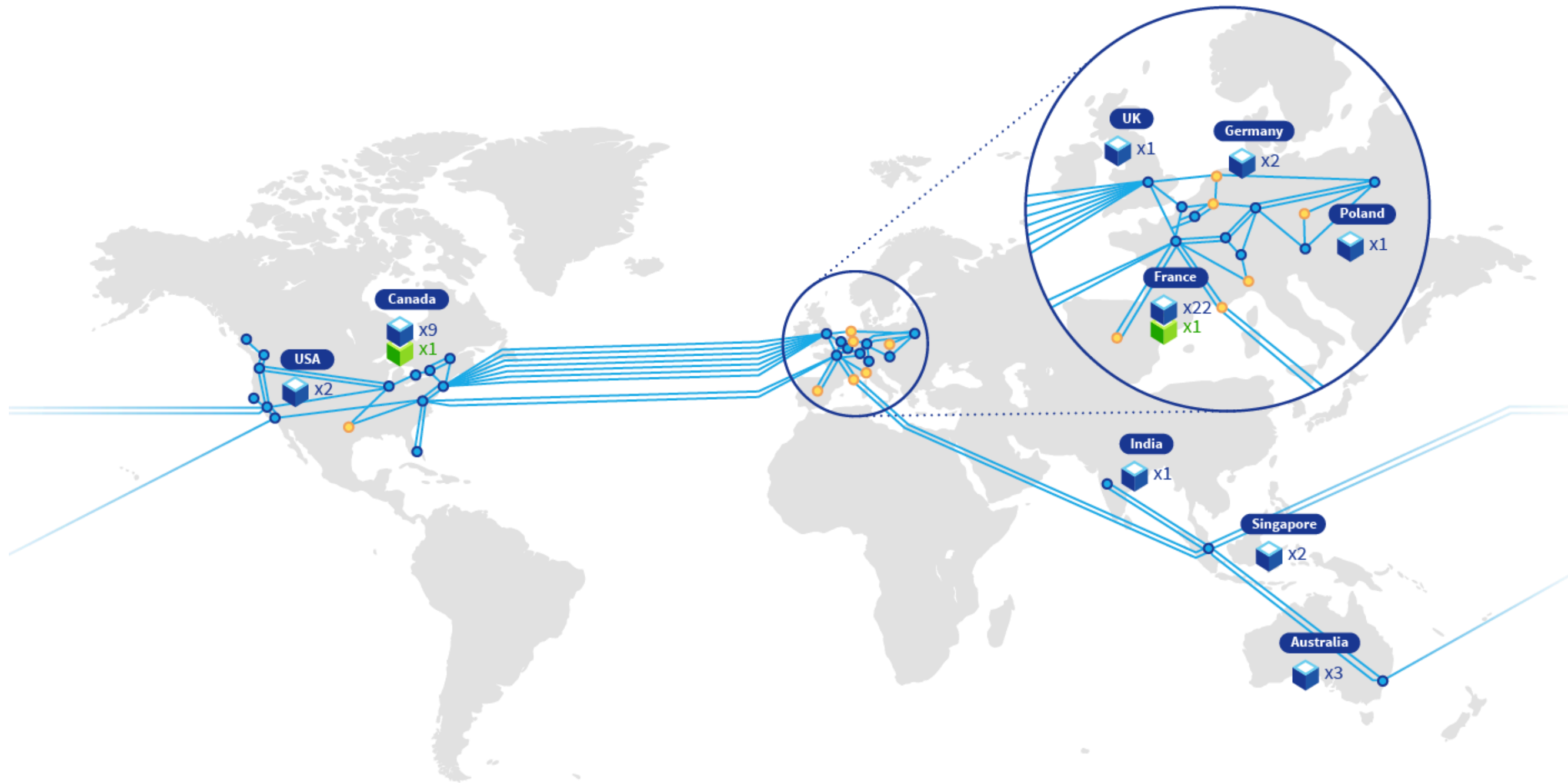
Machine Learning Engineer

OVHcloud

AI Solutions Team



# OVHcloud UNE PRÉSENCE MONDIALE



**43 datacenters** in 18 locations, across **9 countries**



**4 continents** connected through OVHcloud



**46 Points of Presence<sup>1</sup>** on a 100 Tbps<sup>2</sup> Bandwidth Network

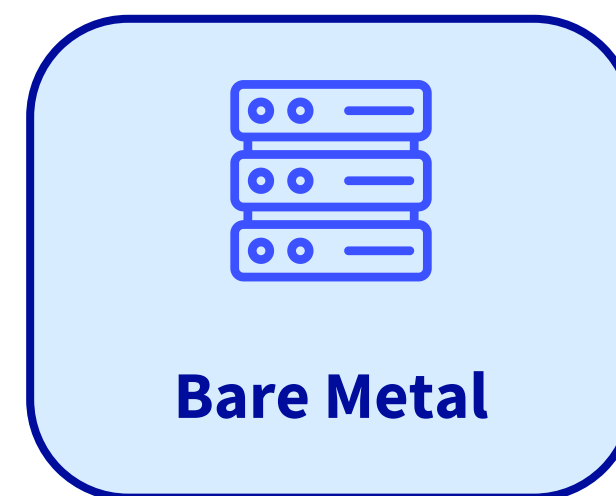
**Legend:** Active datacenters Planned datacenters for 2024 Point of presence Local zones OVHcloud optical fiber network

Source: OVHcloud

Notes: 1. A Point of Presence is a point at which the network establishes a connection with the Internet. 2. Tera bits per second

# À PROPOS D'OVHcloud

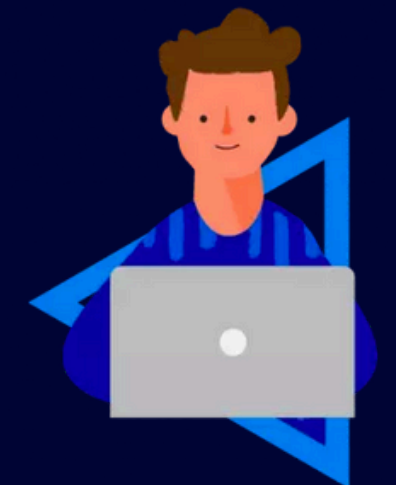
*OVHcloud est un fournisseur mondial de services de cloud computing spécialisé dans les solutions performantes et rentables pour mieux gérer, sécuriser et faire évoluer les données.*



# ÇA VOUS DIT ?



“ Embarquez dans le développement d’une solution de transcription temps réel de vos contenus multimédia...” ”



# INTRODUCTION

# INTRODUCTION

“ Quels sont les différents types de contenus multimédia et leur(s) public(s) ? ”



types de multimédia

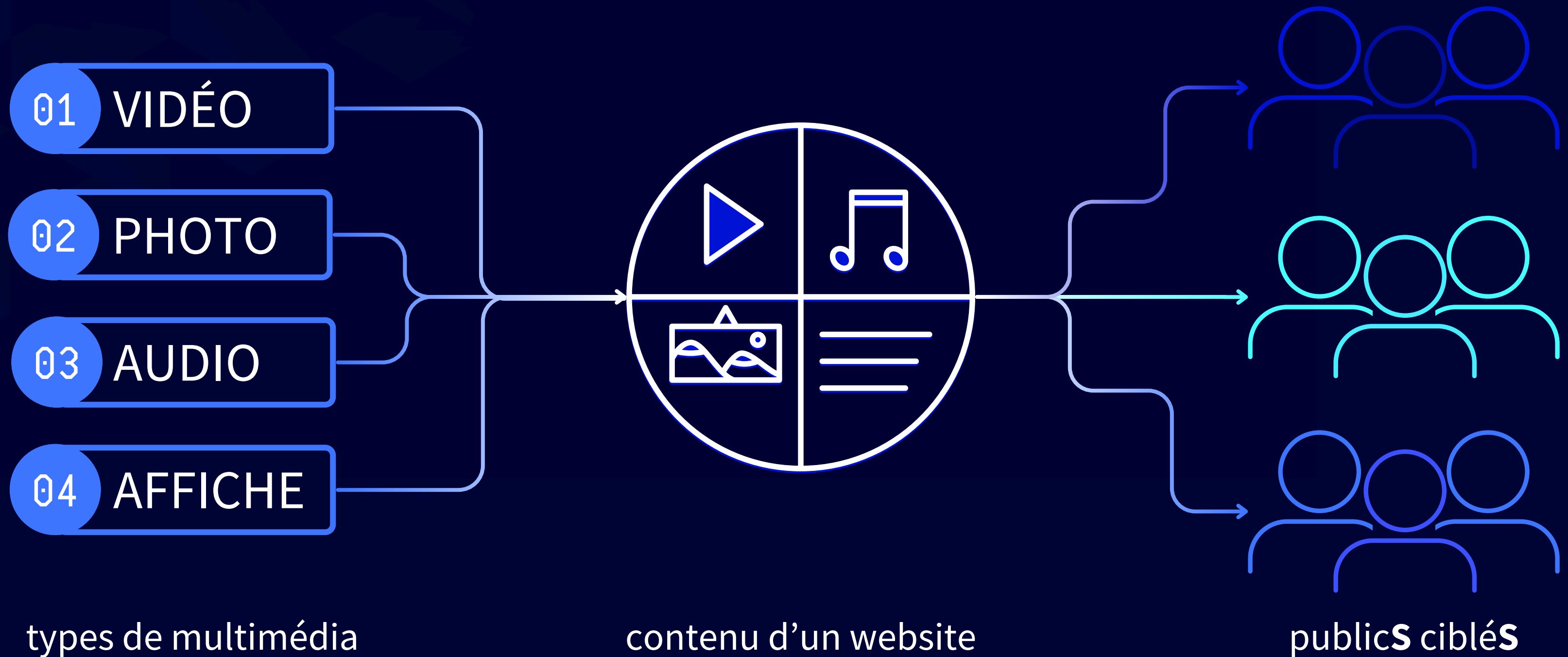
contenu d'un website

public ciblé



# INTRODUCTION

“ Quels sont les différents types de contenus multimédia et leur(s) public(s) ? ”



# TRANSFORMER SES CONTENUS MULTIMEDIA, QU'EST-CE QUE ÇA VEUT DIRE ?

## TRADUCTION



### Changer la langue de...

- sa page web
- son post Twitter, LinkedIn, ...
- ses slides

## TRANSCRIPTION



### Passer de l'oral à l'écrit pour...

- sous-titrer des vidéos, podcasts
- garder le contenu d'une réunion

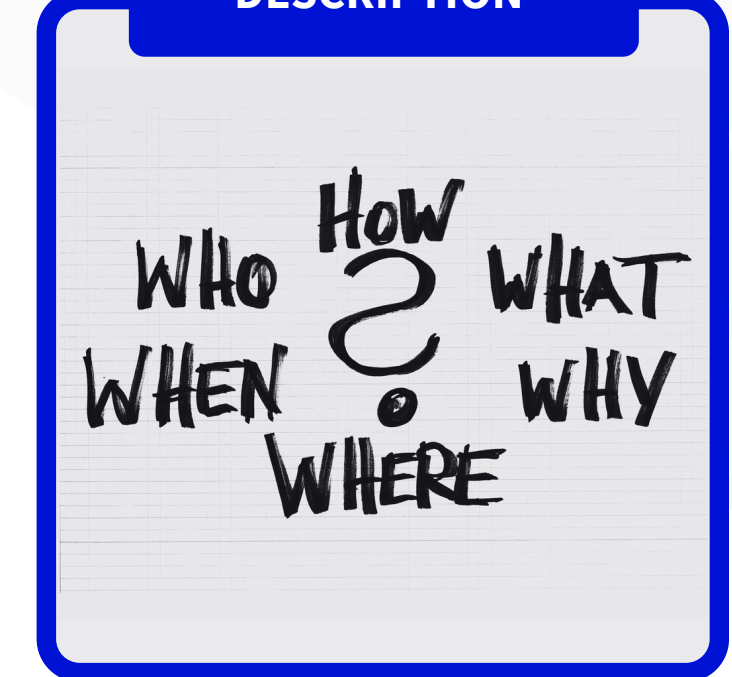
## SYNTHÈSE VOCALE



### Passer de l'écrit à l'oral pour...

- favoriser l'accessibilité
- doubler les voix

## DESCRIPTION



### Décrire ou résumer...

- une vidéo, un podcast
- le contenu d'une réunion
- une documentation

# EN PRATIQUE, ÇA DONNE QUOI ?



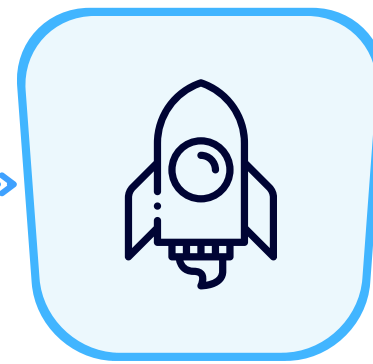
# NOS OBJECTIFS



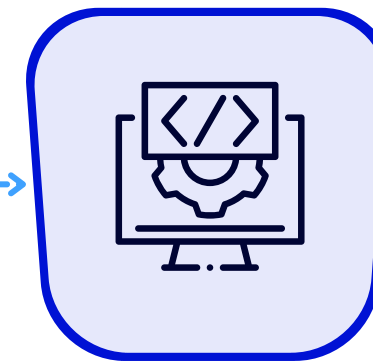
**DÉFINIR**  
les besoins



**SÉLECTIONNER**  
les modèles



**CHOISIR**  
ses AI Endpoints



**DÉVELOPPER**  
la partie client



**DÉPLOYER**  
l'app end to end



# DÉFINITION des besoins

# QUELS SONT NOS PERSONNAE ?



## SE DÉVELOPPER À L'INTERNATIONAL

- Traduire du contenu multimedia pour un usage multilingue
- Adapter les vidéos pour un public mondial



## ÉLARGIR SON PUBLIC...

- En tant que créateur de contenu, j'aimerais **élargir ma communauté**
- Créer et publier un **contenu plus attrayant**



## TRANSCRIRE LES MEETINGS

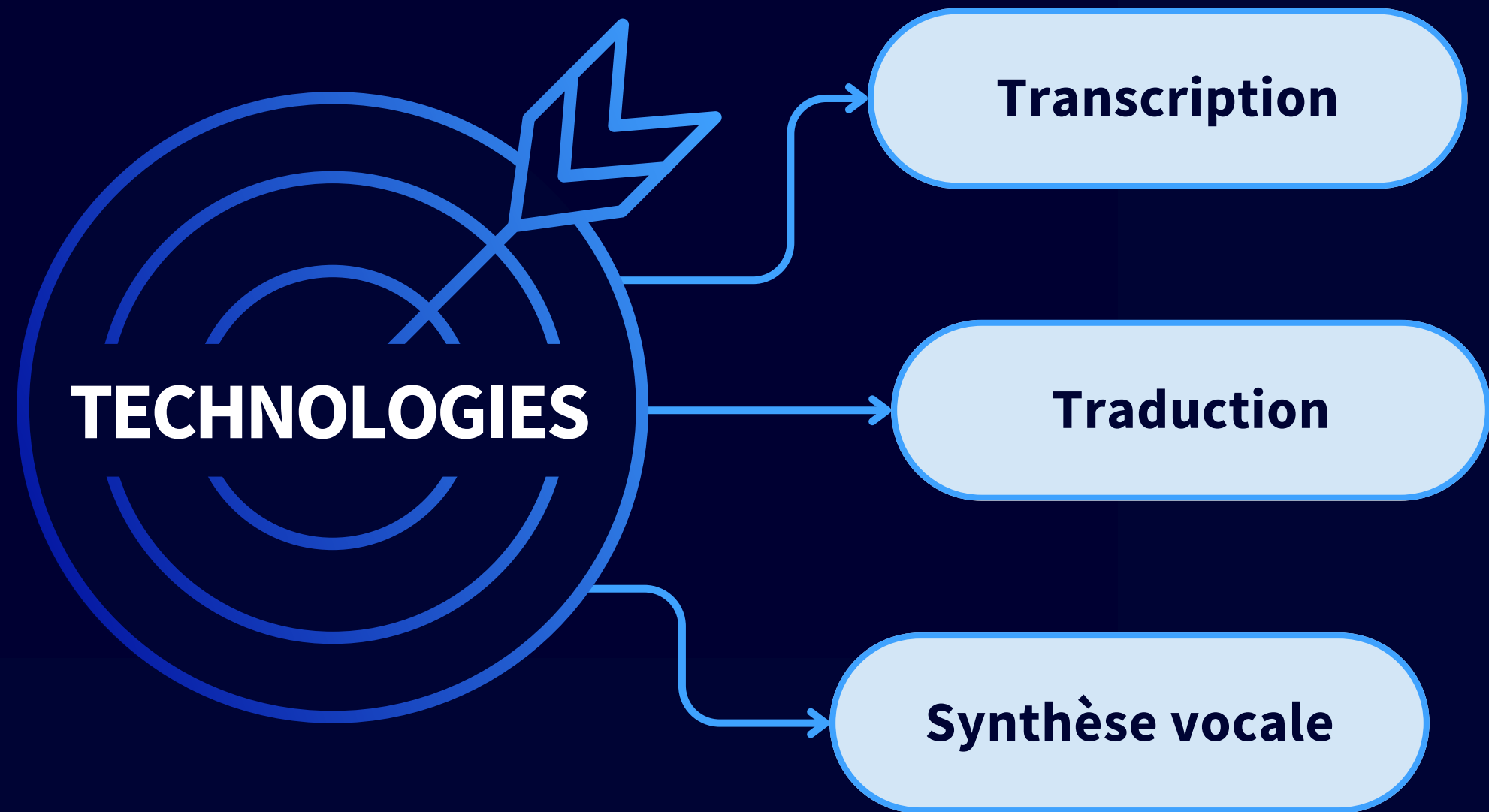
- **Retranscrire** les meetings à l'écrit
- **Transcrire** les meetings dans une autre langues
- **Résumer** les meetings



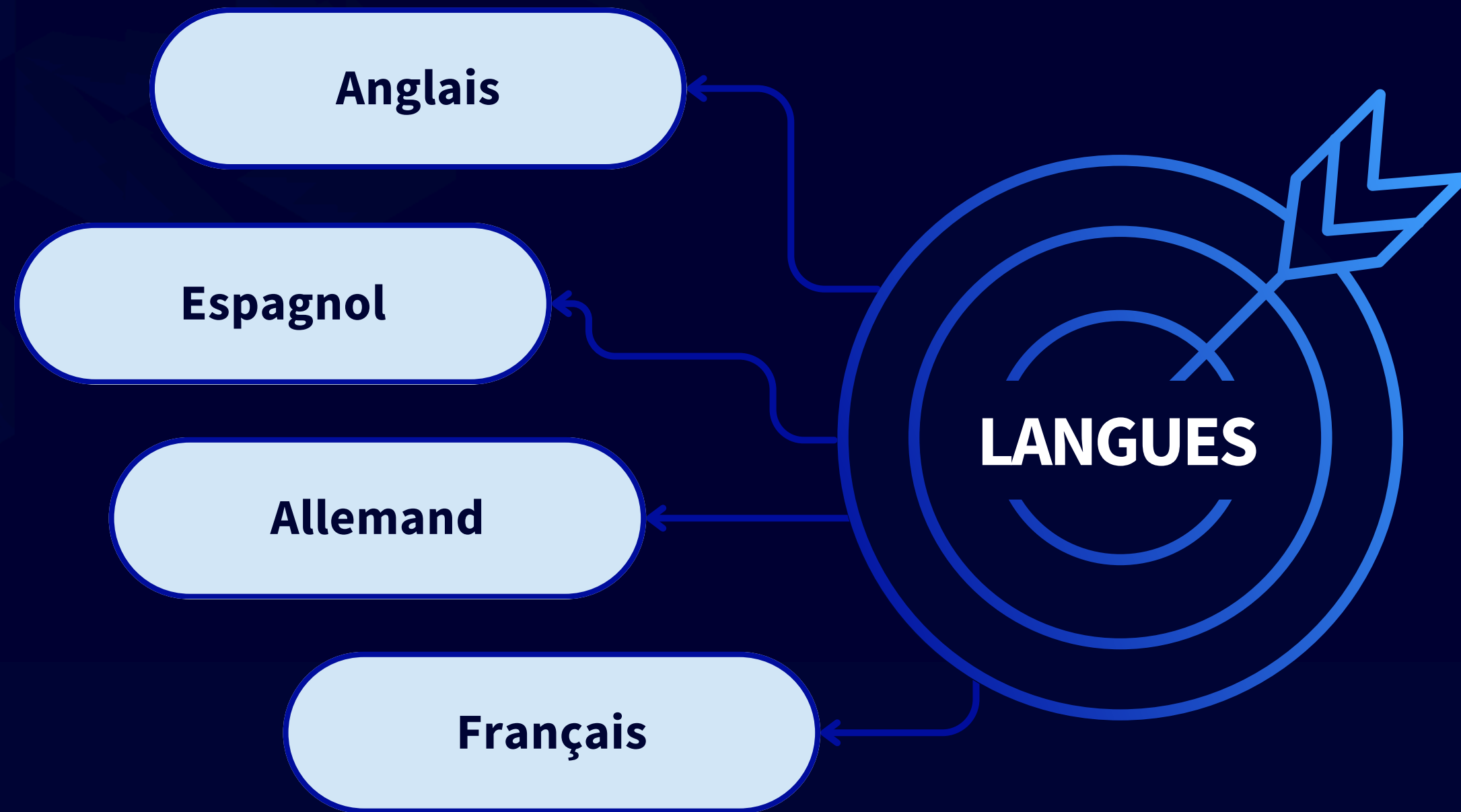
## ÊTRE PLUS INCLUSIF !

- **Sous-titrer les vidéos** pour les personnes malentendantes
- **Inclure les personnes** qui ne parlent pas la même langue

# QUELS SONT NOS BESOINS ?



# QUELS SONT NOS BESOINS ?





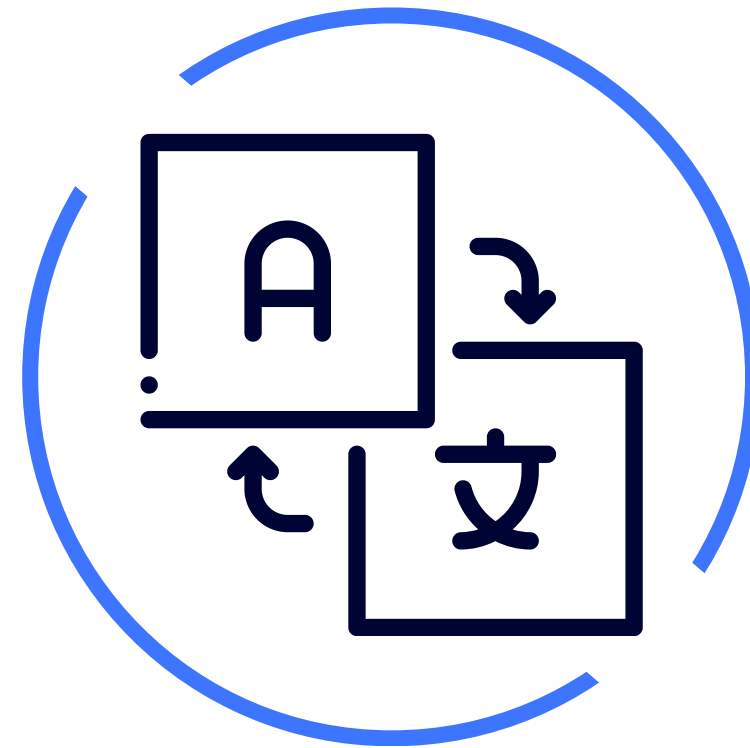
# QUELS SONT NOS BESOINS ?



# 3 TÂCHES CIBLÉES



**AUDIO**  
recognition



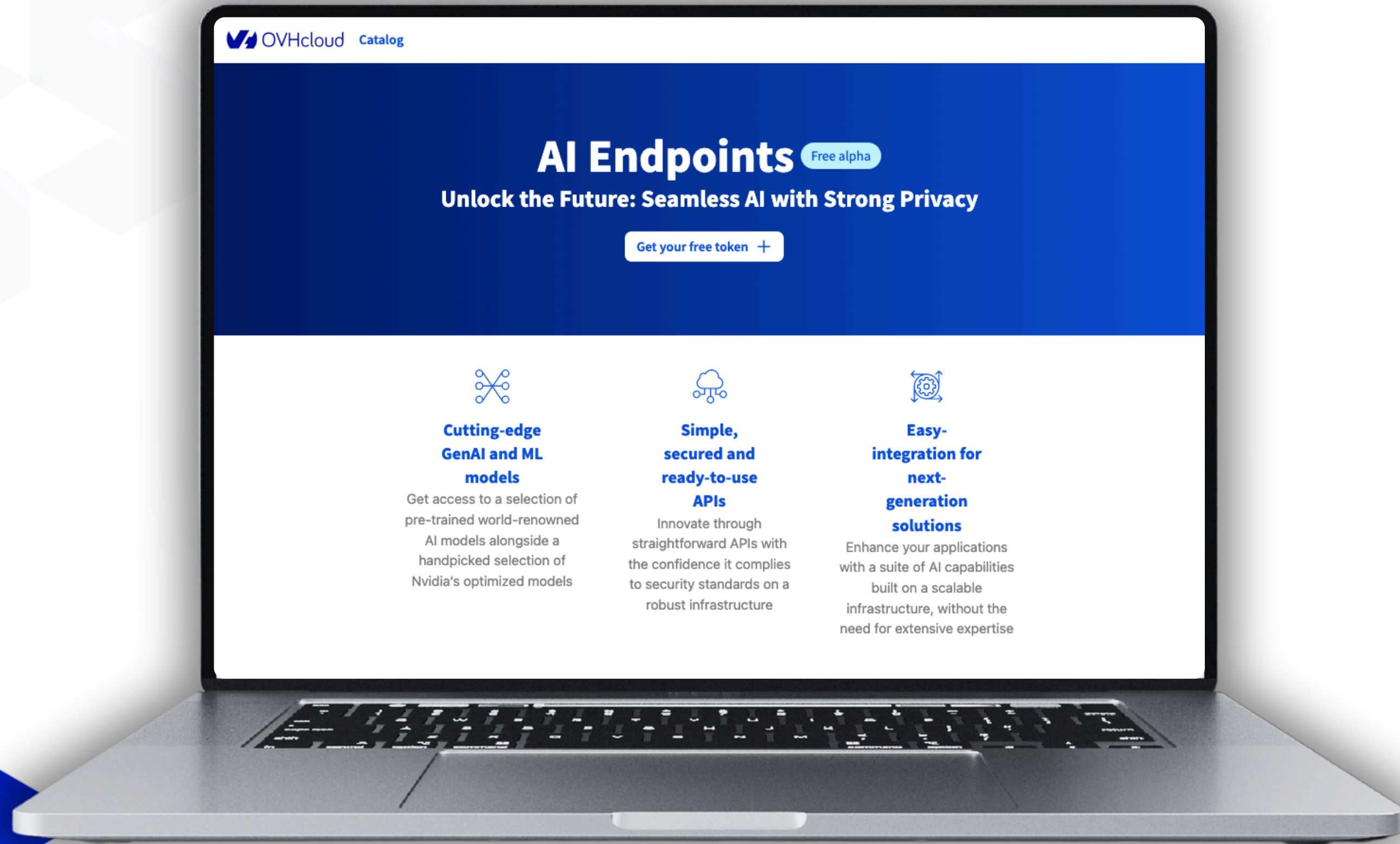
**TEXT**  
translation



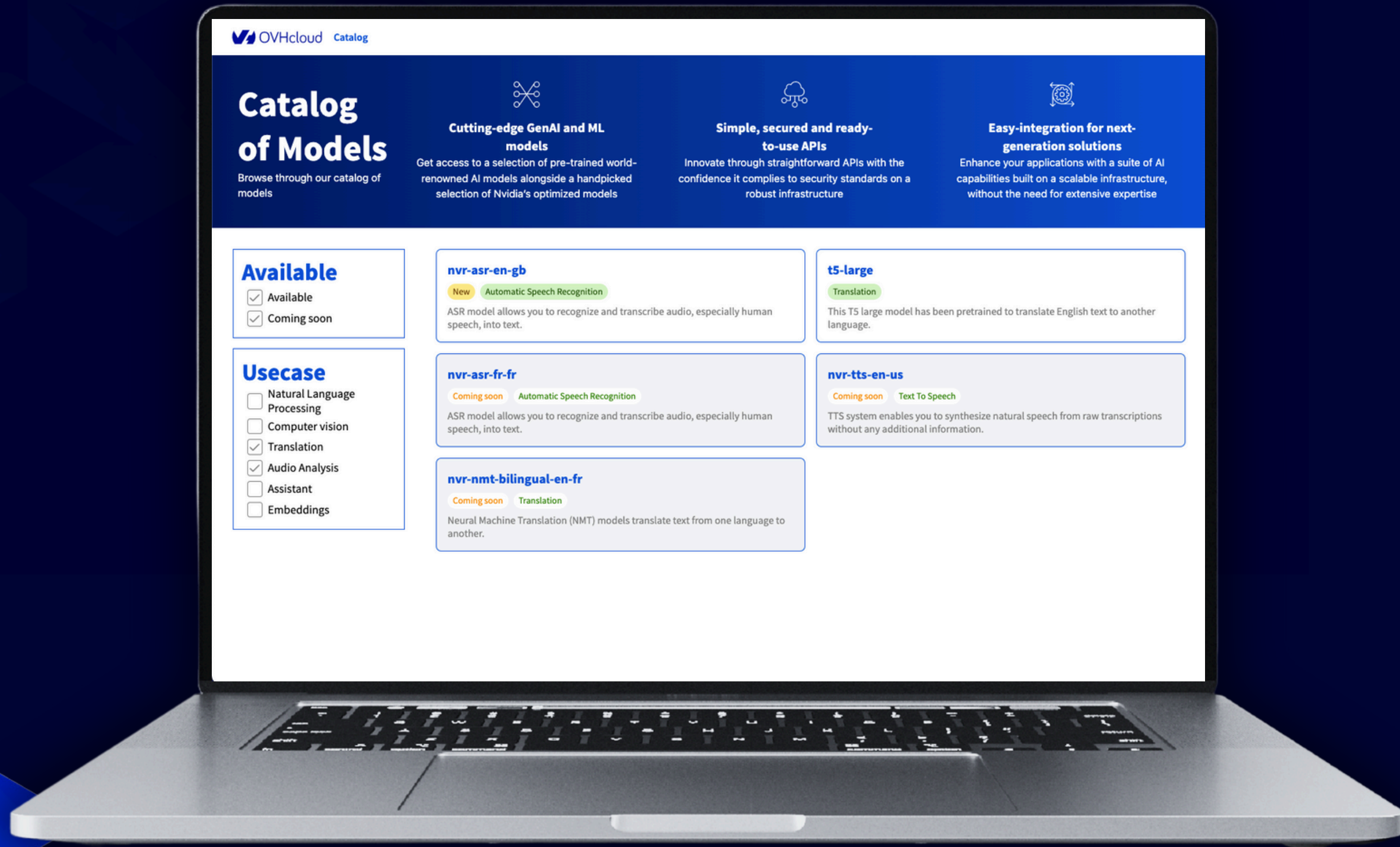
**SPEECH**  
synthesis

# SÉLECTION des modèles

# OÙ TROUVER CES MODÈLES ?



# COMMENT CHOISIR LES MODÈLES ?



# ASR ENDPOINT

The screenshot shows the OVHcloud AI Endpoints interface for the 'nvr-asr-en-gb' model. The interface is divided into a left sidebar and a main content area. The sidebar, titled 'AI Endpoints', features a dark blue background with white text and icons. It includes a 'Get your free token +' button, a 'Cutting-edge GenAI and ML models' section with a neural network icon, a 'Simple, secured and ready-to-use APIs' section with a brain icon, and an 'Easy-integration' section with a gear icon. The main content area has a white background and contains the following elements: a breadcrumb 'Audio Analysis > nvr-asr-en-gb', the model name 'nvr-asr-en-gb', a 'Automatic Speech Recognition' tag, a description 'ASR model allows you to recognize and transcribe audio, especially human speech, into text.', navigation links for 'Playground', 'Documentation', and 'API', and a 'Try out the model by playing with it.' instruction. Below this is a dark-themed playground interface with a file upload area (containing 'Drop Audio Here - or - Click to Upload'), a 'Submit' button, a 'Clear' button, and an 'Examples' section with a file named 'audio\_en\_gb.mp3'. A 'Transcription' output field is also visible on the right side of the playground.

# TTS ENDPOINT

## nvr-tts-en-us

Coming soon

Text To Speech

TTS system enables you to synthesize natural speech from raw transcriptions without any additional information.

↳ Coming Soon

# NMT ENDPOINT

## nvr-nmt-bilingual-en-fr

Coming soon

Translation

Neural Machine Translation (NMT) models translate text from one language to another.

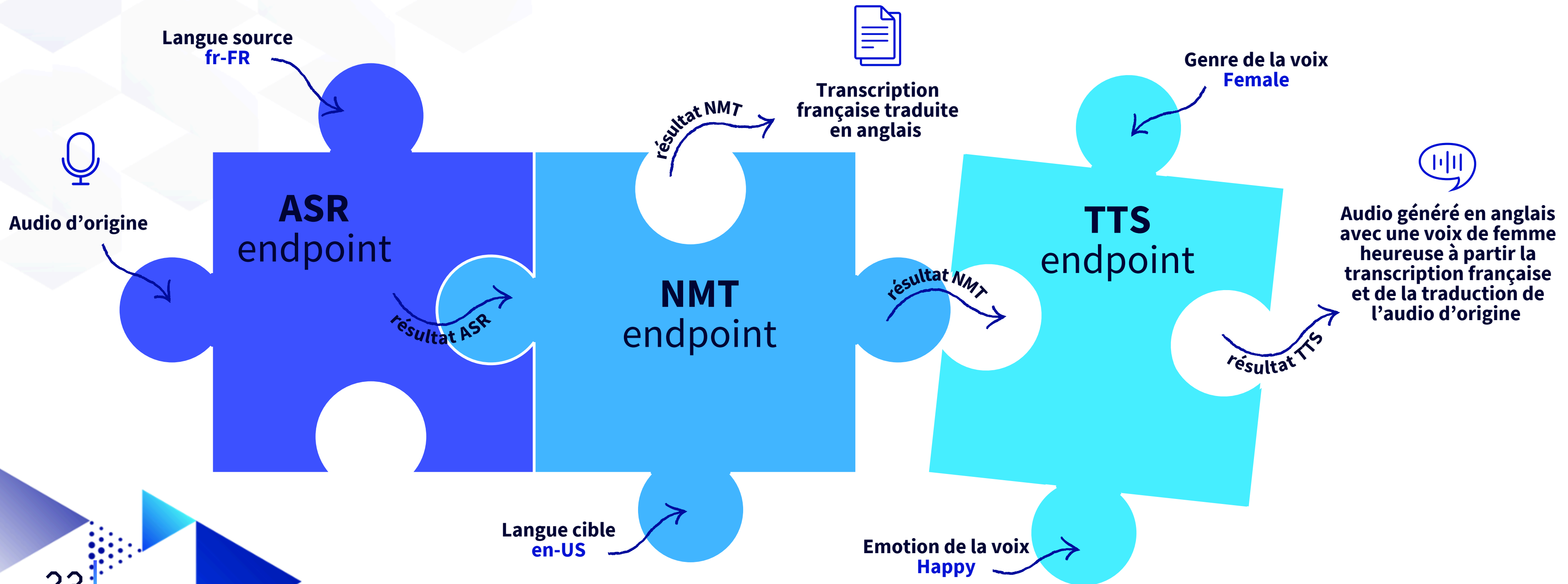
↳ **Coming Soon**



# COMMENT CONNECTER CES AI ENDPOINTS ENTRE EUX ?

# CONNECTER LES ENDPOINTS API

“ Connectez vos AI Endpoints selon vos besoins de Speech AI ! ”



# CONSTRUIRE sa solution


# CONSTRUIRE SA SOLUTION

## Multimedia translator

Transcribe your video:

<https://www.youtube.com/...>

 French

 English

Voice

Subtitles

 Male

 .mp4

 Download



# CONSTRUIRE SA SOLUTION

- ↳ **Entrer** un lien de vidéo YouTube
- ↳ **Transcrire** la partie audio de la vidéo en texte
- ↳ **Sous-titrer** la vidéo dans n'importe quelle langue
- ↳ **Doubler** la voix de le vidéo dans une autre langue
- ↳ **Choisir** le genre de la voix de doublage
- ↳ **Télécharger** la vidéo résultante

# DÉVELOPPER LES FEATURES CLÉS



**GÉNÉRER**  
un fichier SRT  
de sous-titres



**CONSERVER**  
les silences pendant  
la traduction



**DOUBLER**  
l'audio d'une vidéo  
dans une autre langue

# DÉVELOPPER LES FEATURES CLÉS



**GÉNÉRER**  
un fichier SRT  
de sous-titres

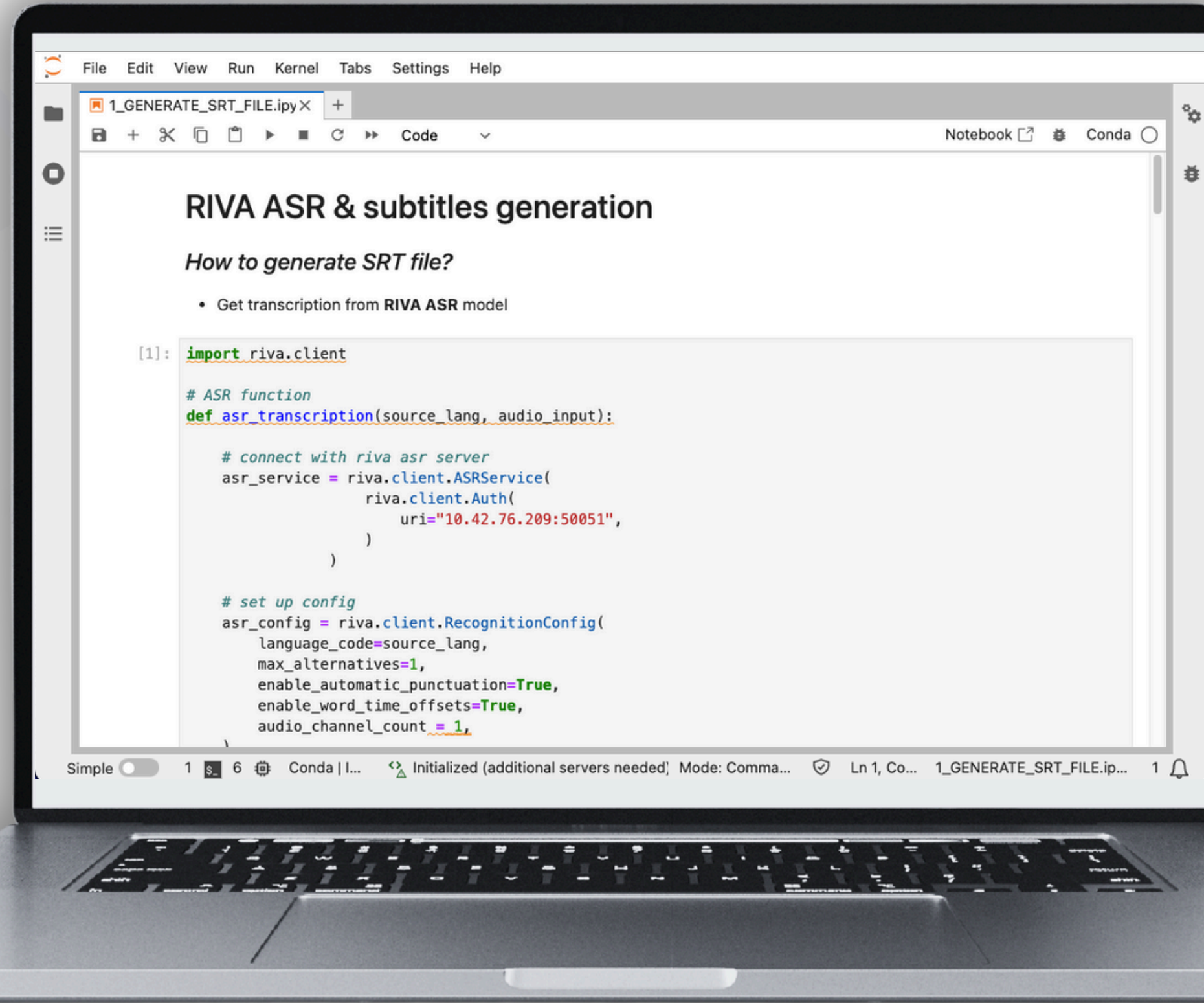


**CONSERVER**  
les silences pendant  
la traduction



**DOUBLER**  
l'audio d'une vidéo  
dans une autre langue

# GÉNÉRER UN FICHER SRT



```
File Edit View Run Kernel Tabs Settings Help
1_GENERATE_SRT_FILE.ipynx +
+ ✂ 📄 📄 ▶ ⏪ ⏩ Code Notebook Conda
RIVA ASR & subtitles generation
How to generate SRT file?
• Get transcription from RIVA ASR model
[1]: import riva.client

# ASR function
def asr_transcription(source_lang, audio_input):

    # connect with riva asr server
    asr_service = riva.client.ASRService(
        riva.client.Auth(
            uri="10.42.76.209:50051",
        )
    )

    # set up config
    asr_config = riva.client.RecognitionConfig(
        language_code=source_lang,
        max_alternatives=1,
        enable_automatic_punctuation=True,
        enable_word_time_offsets=True,
        audio_channel_count = 1,
```

Simple 1 6 Conda | l... Initialized (additional servers needed) Mode: Comma... Ln 1, Co... 1\_GENERATE\_SRT\_FILE.ip... 1



# DÉVELOPPER LES FEATURES CLÉS



**GÉNÉRER**  
un fichier SRT  
de sous-titres

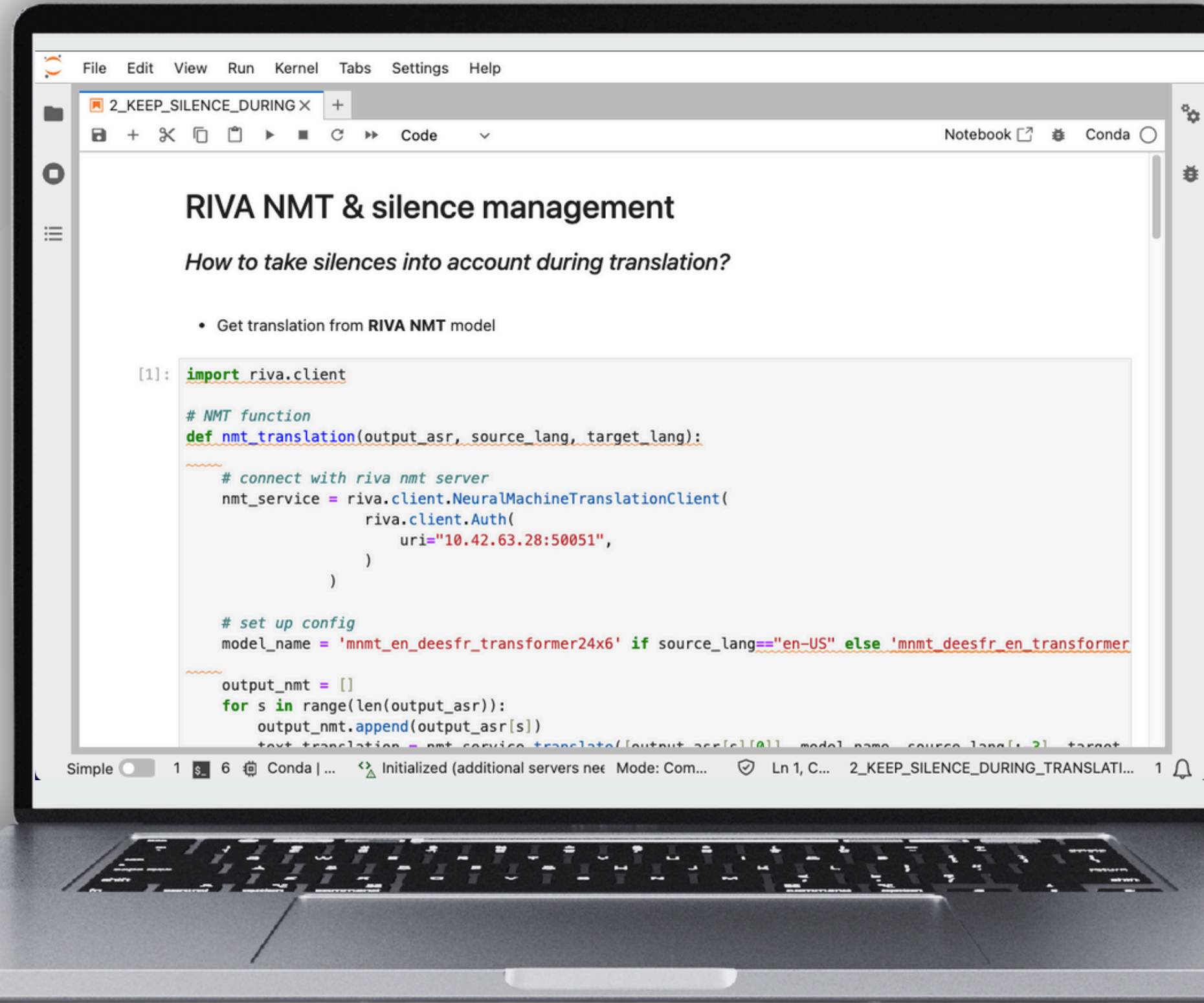


**CONSERVER**  
les silences pendant  
la traduction



**DOUBLER**  
l'audio d'une vidéo  
dans une autre langue

# CONSERVER LES SILENCES



The image shows a laptop screen displaying a Jupyter Notebook. The notebook is titled "RIVA NMT & silence management" and contains a code cell with the following Python code:

```
[1]: import riva.client

# NMT function
def nmt_translation(output_asr, source_lang, target_lang):

    # connect with riva nmt server
    nmt_service = riva.client.NeuralMachineTranslationClient(
        riva.client.Auth(
            uri="10.42.63.28:50051",
        )
    )

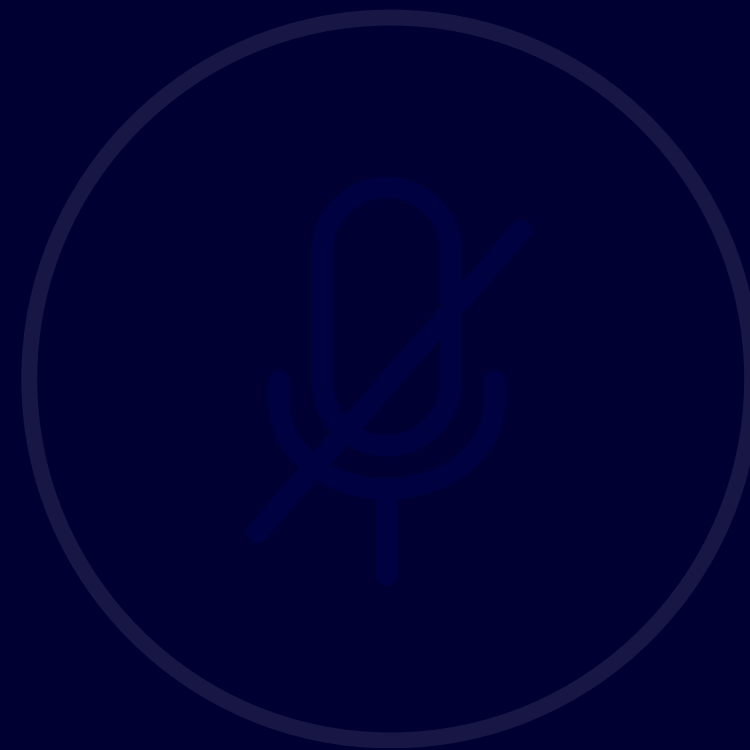
    # set up config
    model_name = 'mnmt_en_deesfr_transformer24x6' if source_lang=="en-US" else 'mnmt_deesfr_en_transformer'

    output_nmt = []
    for s in range(len(output_asr)):
        output_nmt.append(output_asr[s])
        text_translation = nmt_service.translate([output_asr[s][0]], model_name, source_lang[s], target
```

# DÉVELOPPER LES FEATURES CLÉS



**GÉNÉRER**  
un fichier SRT  
de sous-titres

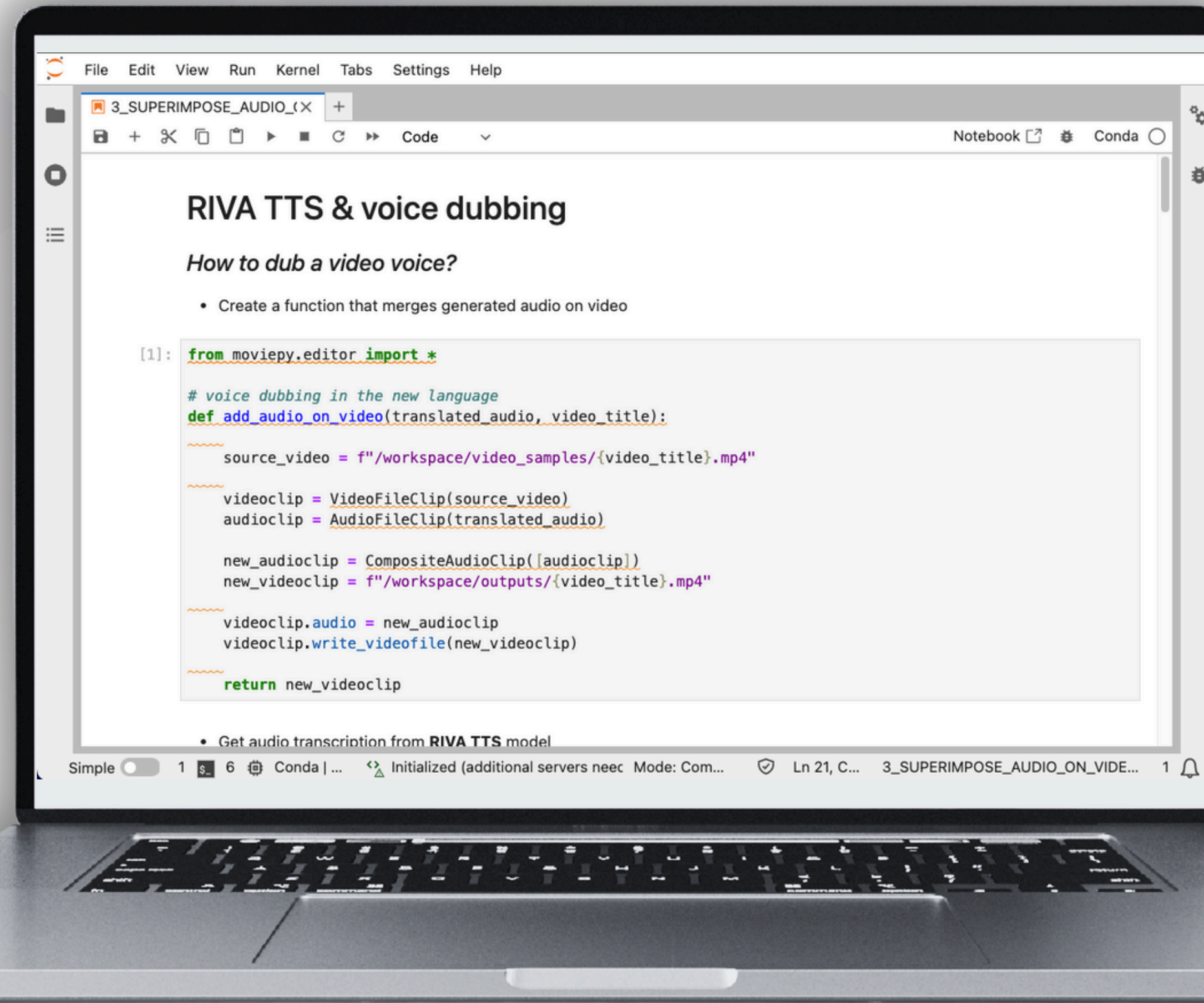


**CONSERVER**  
les silences pendant  
la traduction



**DOUBLER**  
l'audio d'une vidéo  
dans une autre langue

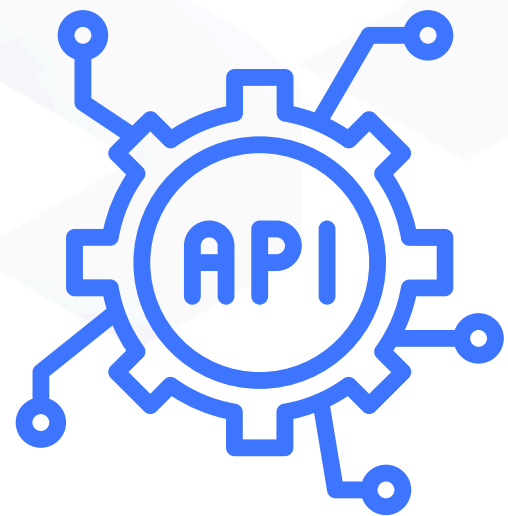
# DOUBLER L'AUDIO D'UNE VIDÉO



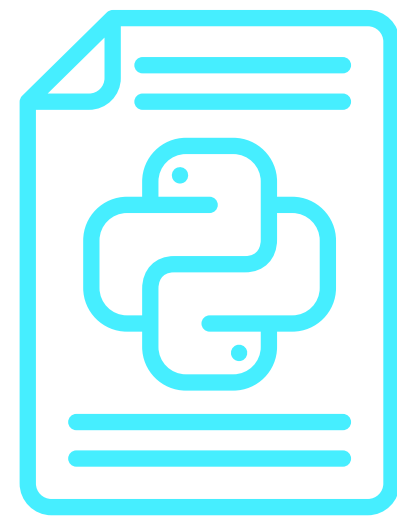
# DÉPLOIEMENT

## l'app end-to-end

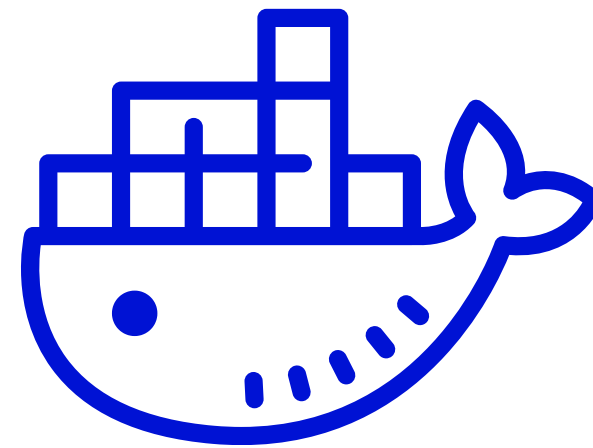
# DÉPLOIEMENT L'APP END-TO-END



**SÉLECTION**  
des endpoints



**DÉVELOPPEMENT**  
de sa solution



**CONTENEURISATION**  
de sa solution

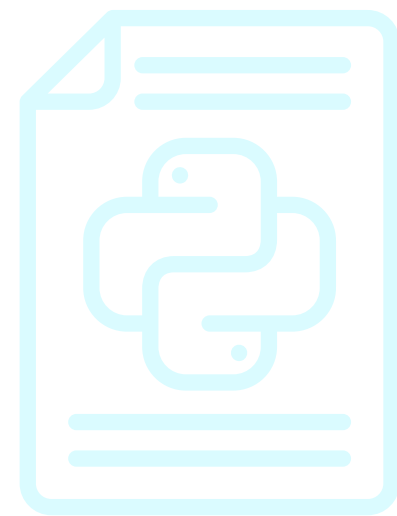


**DÉPLOIEMENT**  
de l'app end-to-end

# DÉPLOIEMENT L'APP END-TO-END



**SÉLECTION**  
des endpoints



**DÉVELOPPEMENT**  
de sa solution



**CONTENEURISATION**  
de sa solution



**DÉPLOIEMENT**  
de l'app end-to-end

# SÉLECTION DES ENDPOINTS

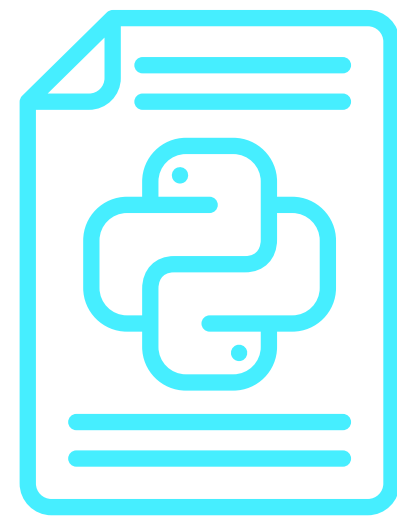




# DÉPLOIEMENT L'APP END-TO-END



**SÉLECTION**  
des endpoints



**DÉVELOPPEMENT**  
de sa solution



**CONTENEURISATION**  
de sa solution



**DÉPLOIEMENT**  
de l'app end-to-end

# DÉVELOPPEMENT DE SA SOLUTION

## Multimedia translator

Transcribe your video:

<https://www.youtube.com/...>

 French

 English

Voice

Subtitles

 Male

 .mp4

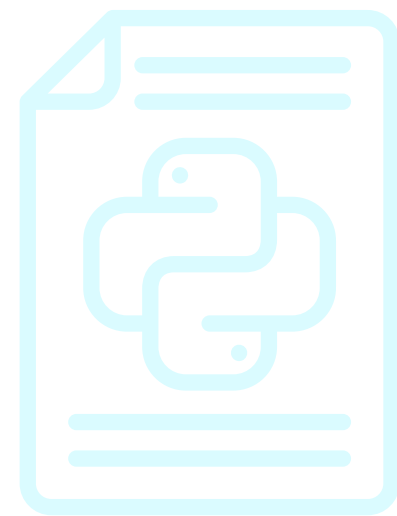
 Download



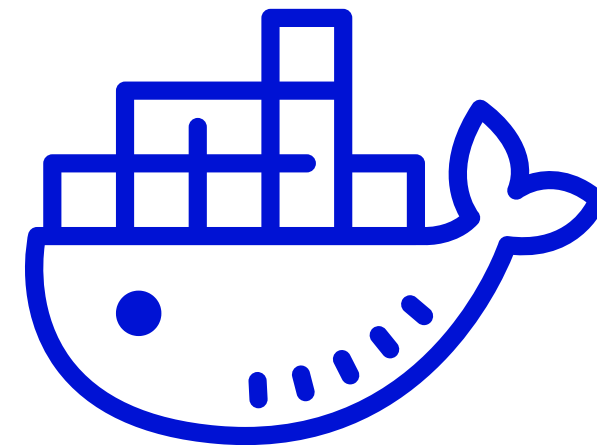
# DÉPLOIEMENT L'APP END-TO-END



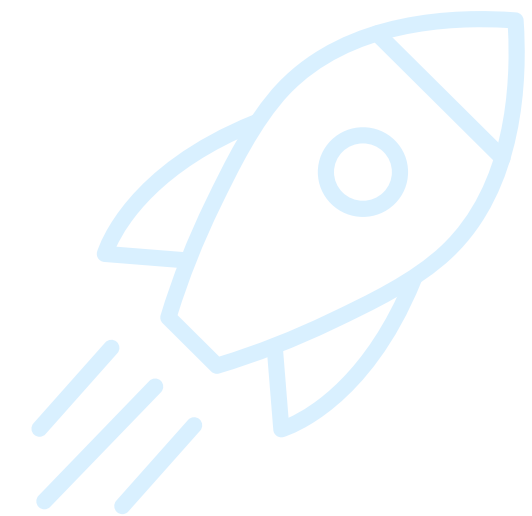
**SÉLECTION**  
des endpoints



**DÉVELOPPEMENT**  
de sa solution




**CONTENEURISATION**  
de sa solution



**DÉPLOIEMENT**  
de l'app end-to-end

# CONTENEURISATION DE SA SOLUTION



```
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
```

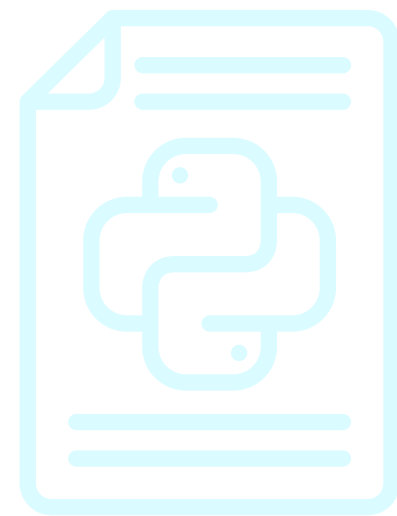
```
CMD [ "python3" , "/workspace/main.py" ]
```

Dockerfile

# DÉPLOIEMENT L'APP END-TO-END



**SÉLECTION**  
des endpoints



**DÉVELOPPEMENT**  
de sa solution



**CONTENEURISATION**  
de sa solution



**DÉPLOIEMENT**  
de l'app end-to-end

# SOLUTION DE DÉPLOIEMENT



## COMPUTE RESOURCES

- 1 GPU
- H100



## HIGH AVAILABILITY

- API scalable on the fly
- Custom number of replicas



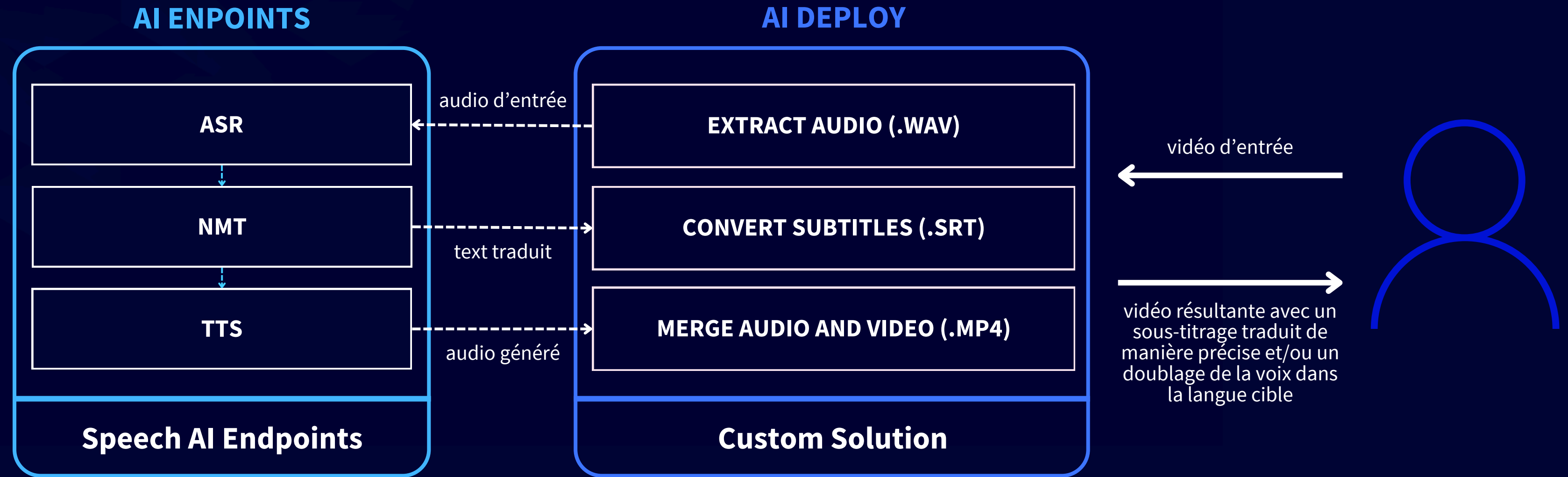
## SECURE ACCESS

- Private mode
- Personal token access

# OVHcloud AI DEPLOY

01	<b>CONTENEUR AS A SERVICE</b>	Le client fournit un conteneur Docker via une registry Docker
02	<b>RESSOURCES DE CALCUL</b>	Le conteneur s'exécute dans le cloud sur GPU (ou CPU)
03	<b>MODE DE FACTURATION</b>	Le client est facturé à la minute utilisée
04	<b>DÉPLOIEMENT API/APP</b>	Manière industrielle de déployer des API(s) stateless
05	<b>STRATÉGIE D'AUTOSCALING</b>	Scalable à la volé

# DÉPLOIEMENT L'APP END-TO-END



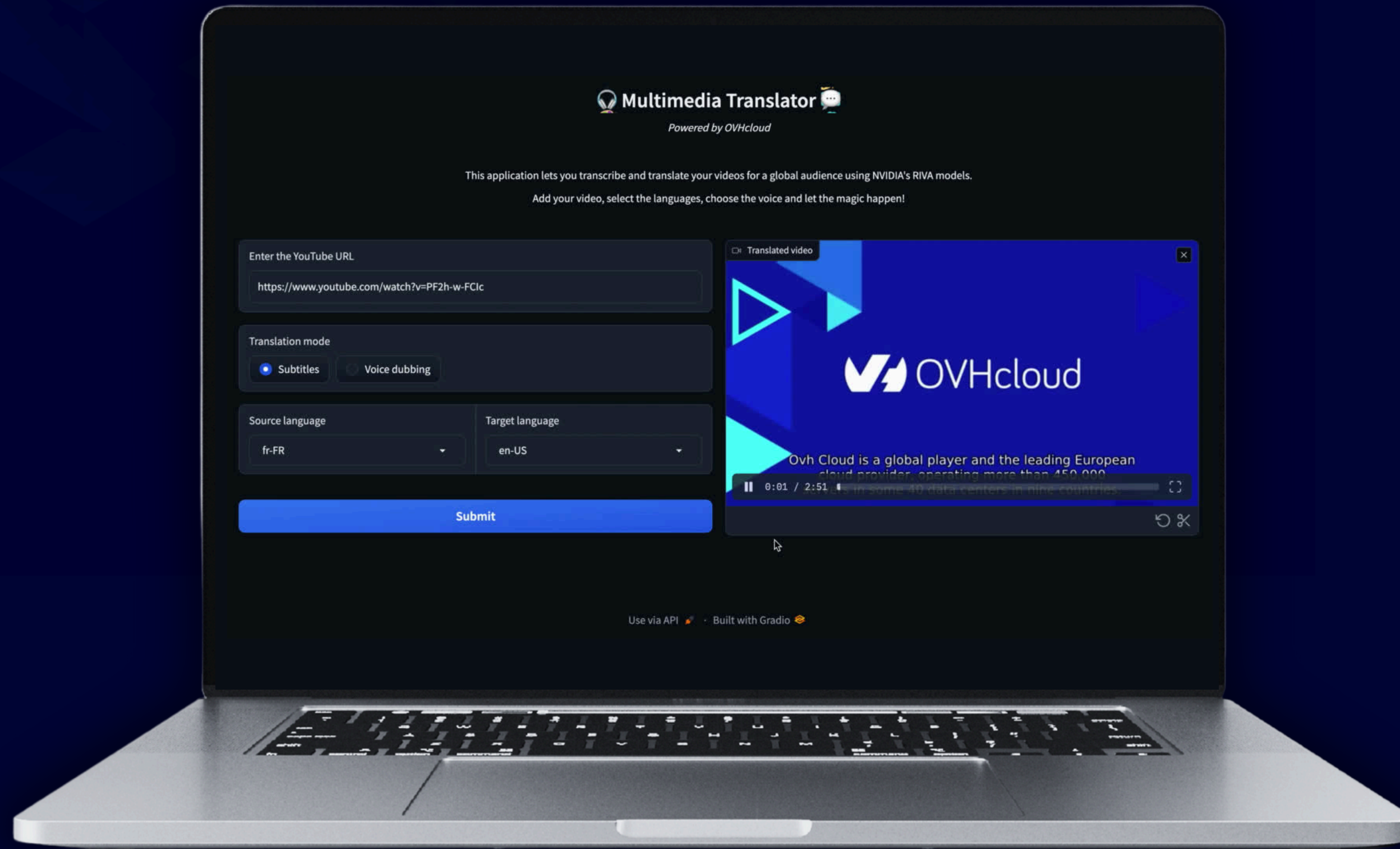
CHOOSE THE **AI ENPOINTS** ACCORDING TO YOUR NEEDS

DEPLOY **END-TO-END SOLUTION** THROUGH **GRADIO APP** IN THE CLOUD WITH HIGH AVAILABILITY



# DÉMO de l'app

# DÉMO DE L'APP



# DÉMO DE L'APP

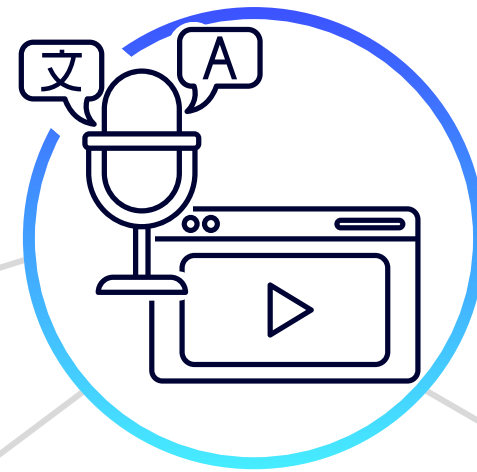


<https://bit.ly/ai-days-multimedia-translator>

# CONCLUSION

# CONCLUSION

Le “*multimode multimedia translator*” en résumé...



## SOLUTION DE TRANSCRIPTION ET TRADUCTION MULTIMÉDIA

- vidéos traduites et sous-titrées
- vidéos doublées dans une autre langue



## OVHcloud AI ENDPOINTS

- simple d'utilisation
- gain de temps dans le développement



## OVHcloud AI DEPLOY

- solution serverless
- top des GPUs (H100, A100, L4, L40S, V100S)



## ET CE N'EST PAS FINI...

- reconnaissance des émotions
- transcription en temps réel
- personnalisation de la voix
- enrichissement du vocab

# CONTACTEZ-NOUS !

## INFORMATIONS

 <https://www.ovhcloud.com/>

 [elea.petton@ovhcloud.com](mailto:elea.petton@ovhcloud.com)

 OVHcloud

 @OVHcloud



# À VOUS DE TESTER!



<https://bit.ly/ai-days-multimedia-translator>

MERCI !