



# AI ENDPOINTS

**OVHcloud COMME ALTERNATIVE SOUVERAINE AUX GAFAM**



# PRÉSENTATION

**Eléa Petton**

Machine Learning Engineer

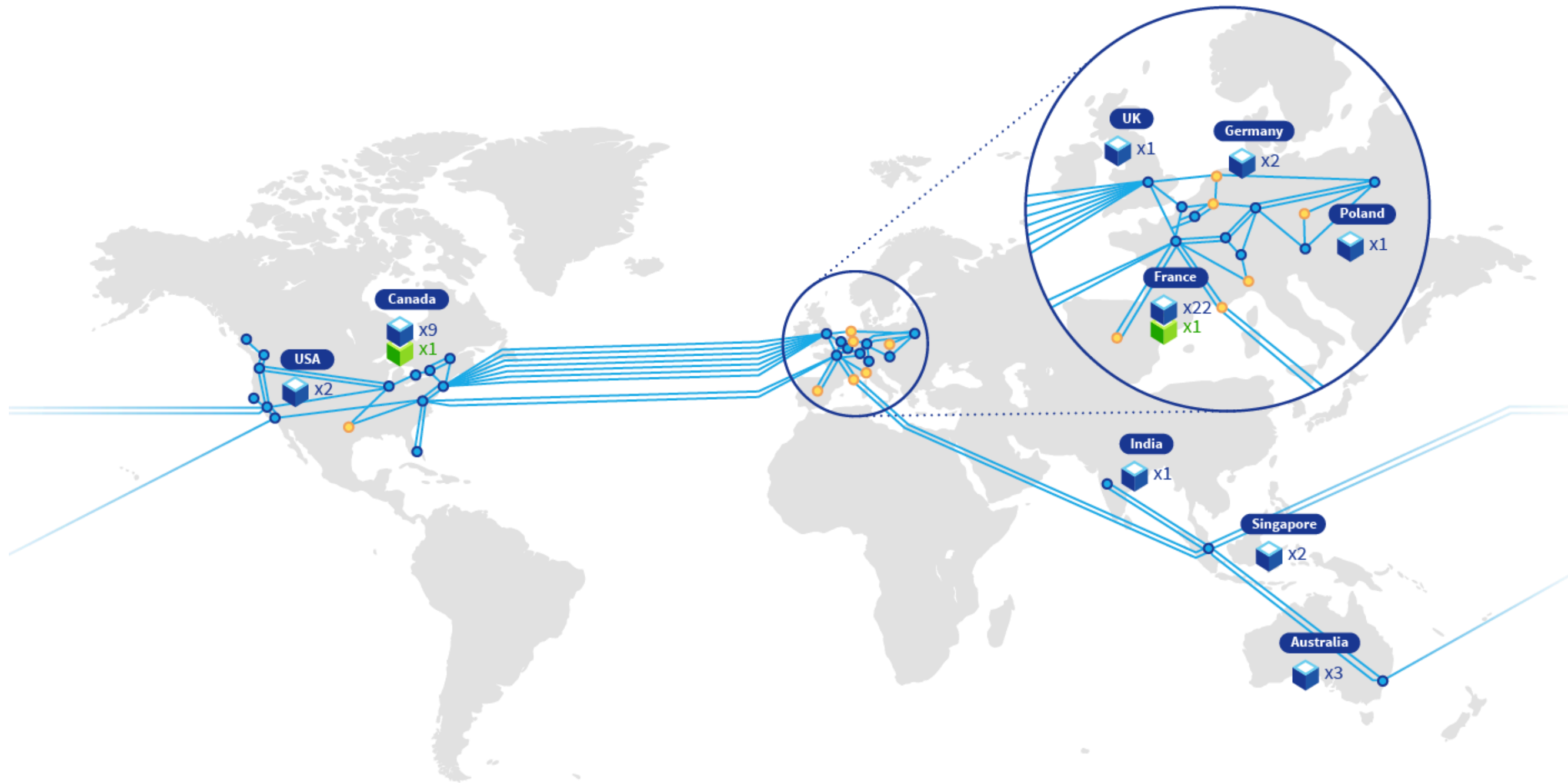
OVHcloud

AI Solutions Team



# OVHcloud en quelques mots

# 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.*



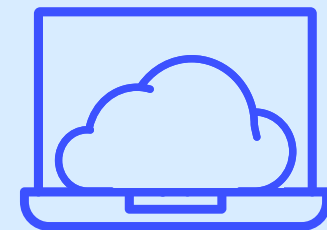
**Private Cloud**



**Public Cloud**



**Bare Metal**



**Web Cloud**



**PaaS portfolio**

# À PROPOS D'OVHcloud

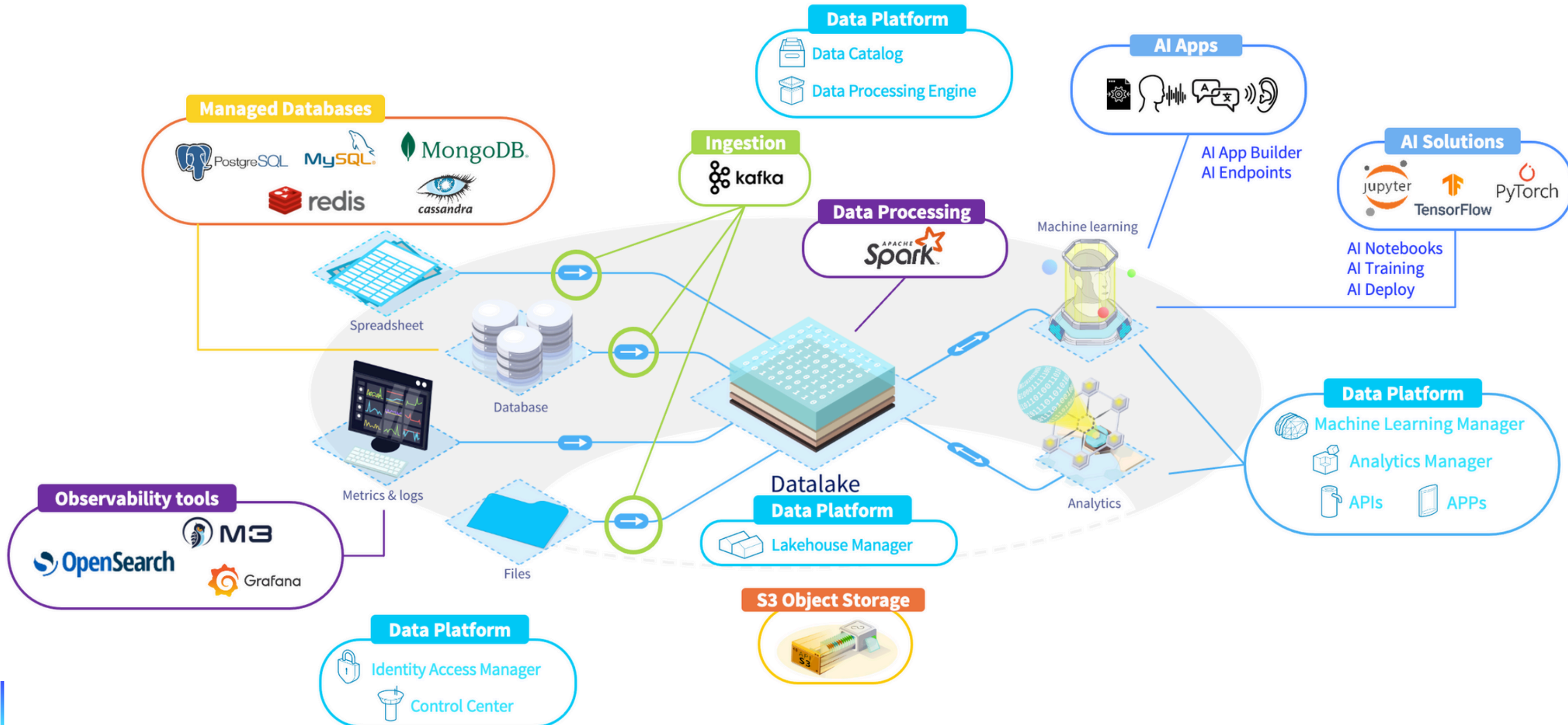
*OVHcloud est un fournisseur mondial de services de cloud computing spécialisé dans les solutions performantes et rentables pour mieux gérer, sécuriser et faire évoluer les données.*



AI Solutions



# OVHcloud DATA & IA



# AI SOLUTIONS

et sa gamme complète de GPUs



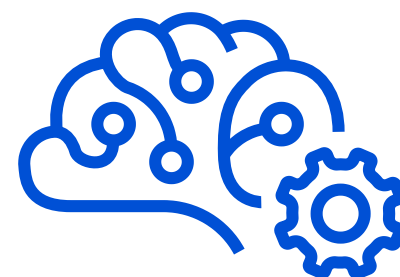
# OVHcloud AI SOLUTIONS

## AI NOTEBOOKS



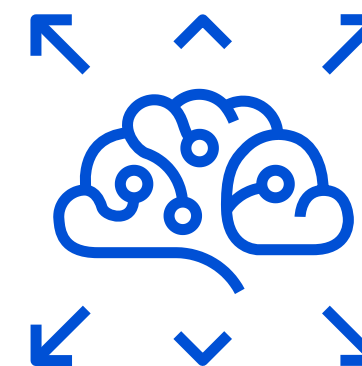
**Notebook As A Service**  
Environnement prépackagé  
Persistence du workspace

## AI TRAINING



**Container As A Service**  
Entraînement de modèles  
Facturation à la minute

## AI DEPLOY



**Container As A Service**  
Inférence de modèles  
Haute disponibilité - autoscaling

# OVHcloud ET SA GAMME DE GPUS

## H100

extreme fine-tuning and training, demanding model inference, LLM projects



**ALREADY AVAILABLE**

## A100 L40S

deep-learning training or inference, generative AI, LLM projects



**A100 - ALREADY AVAILABLE**  
**L40S - ALREADY AVAILABLE**

## V100S

advanced training and inference of AI models



**ALREADY AVAILABLE**

## L4

enhanced AI capabilities workloads



**ALREADY AVAILABLE**

# OVHcloud AI SOLUTIONS

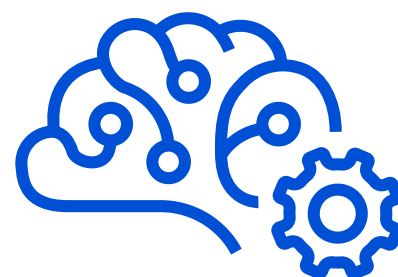
## AI NOTEBOOKS



### **Notebook As A Service**

Environnement prépackagé  
Persistence du workspace

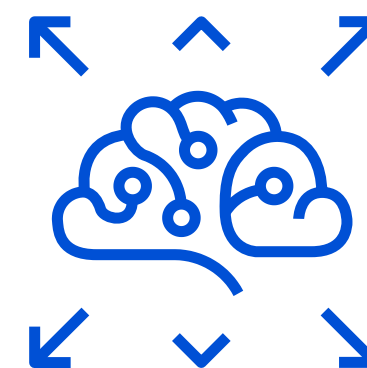
## AI TRAINING



### **Container As A Service**

Entraînement de modèles  
Facturation à la minute

## AI DEPLOY



### **Container As A Service**

Inférence de modèles  
Haute disponibilité - autoscaling

## AI ENDPOINTS



### **AI As A Service**

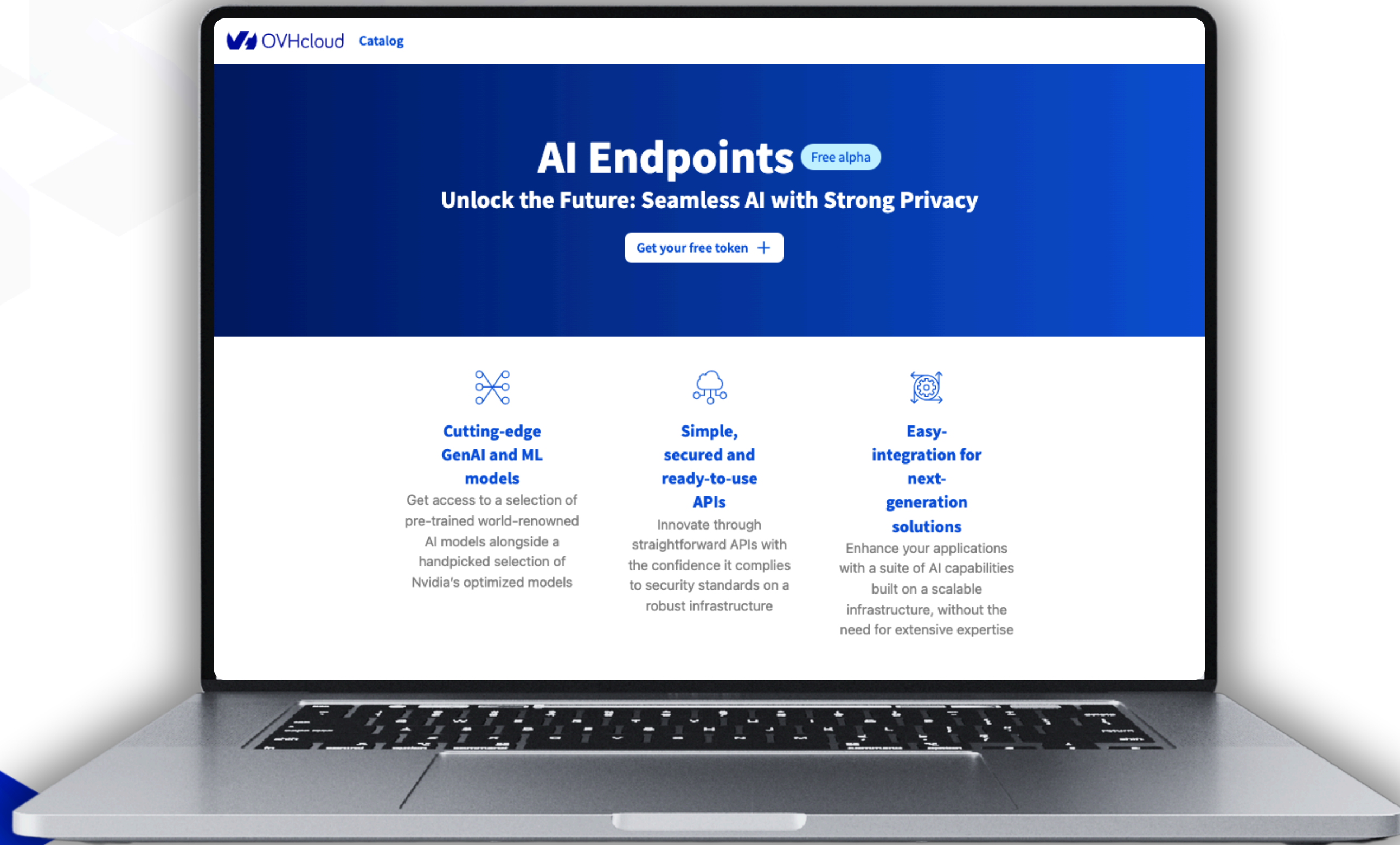
APIs d'IA prêtes à l'emploi  
Solution serverless, simple, sécurisée

**NEW - alpha**

# AI ENDPOINTS

“seamless AI with strong privacy”

# OVHcloud AI ENDPOINTS



# DÉMOCRATISER L'IA

“ Améliorez vos applications avec **AI Endpoints** ”



## CONÇU POUR LES DÉVELOPPEURS

avec une documentation complète, des API simples et des exemples de code



## ENGAGÉ ENVERS LA CONFIDENTIALITÉ

nous ne stockons pas et ne partageons pas vos données pendant ou après l'utilisation du modèle



## LISTE ORGANISÉE DE MODÈLES D'IA

mise à disposition des derniers modèles, optimisés pour des performances et une précision maximales



## TECHNOLOGIE SANS VERROUILLAGE

grâce à notre transparence sur les modèles d'IA utilisés, les clients peuvent mettre en œuvre ces modèles sur leur propre infrastructure ou sur d'autres services cloud

# AI ENDPOINTS

## et son catalogue de modèles

# CATALOGUE DE MODÈLES

“ Une **plate-forme serverless** donnant accès à des modèles d'IA avancés, comme les **LLM**, le NLP, la traduction, la reconnaissance vocale ou encore la reconnaissance d'images. ”

## ASSISTANT



## AUDIO ANALYSIS



## COMPUTER VISION



## EMBEDDINGS



## NLP



## TRANSLATION





# COMMENT UTILISER CES ENDPOINTS ?



**PARCOURIR**  
le catalogue  
d'AI Endpoints



**TESTER**  
le playground depuis  
l'interface utilisateur



**INTÉGRER**  
les call API dans  
son code



**PERFECTIONNER**  
son application avec  
des API d'IA optimisées



**PROFITER**  
de sa solution à la  
pointe de la technologie

# DEMO TIME

“AI Endpoints in action”

# DEMO 1 - USAGE DE BASE

“ Un unique Endpoint en tant que **réponse** à votre besoin ! ”



# DEMO 1 - USAGE DE BASE

“ Un unique Endpoint en tant que **réponse** à votre besoin ! ”

INTÉGREZ UN ENDPOINT **LLM** À VOTRE SOLUTION

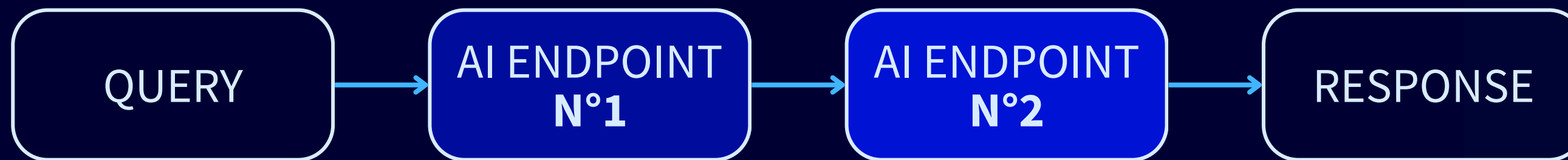


“INTÉGREZ SIMPLEMENT  
UN AI ENDPOINT À VOTRE  
SOLUTION EN QUELQUES  
LIGNES DE CODE”

# DEMO 2 - COMBINER LES ENDPOINTS À L'INFINI

“ Speech-To-Text, LLM, Translation, Text-To-Speech...

Comment assembler ces Endpoints pour obtenir une **solution** complète et adaptée à mon besoin ? ”



# DEMO 2 - COMBINER LES ENDPOINTS À L'INFINI

“ Speech-To-Text, LLM, Translation, Text-To-Speech...

Comment assembler ces Endpoints pour obtenir une **solution** complète et adaptée à mon besoin ? ”



# DEMO 2 - COMBINER LES ENDPOINTS À L'INFINI

“ Speech-To-Text, LLM, Translation, Text-To-Speech...”

Comment assembler ces Endpoints pour obtenir une **solution** complète et adaptée à mon besoin ? ”

INTERROGEZ UN **LLM** VERBALEMENT GRÂCE AU **SPEECH AI**

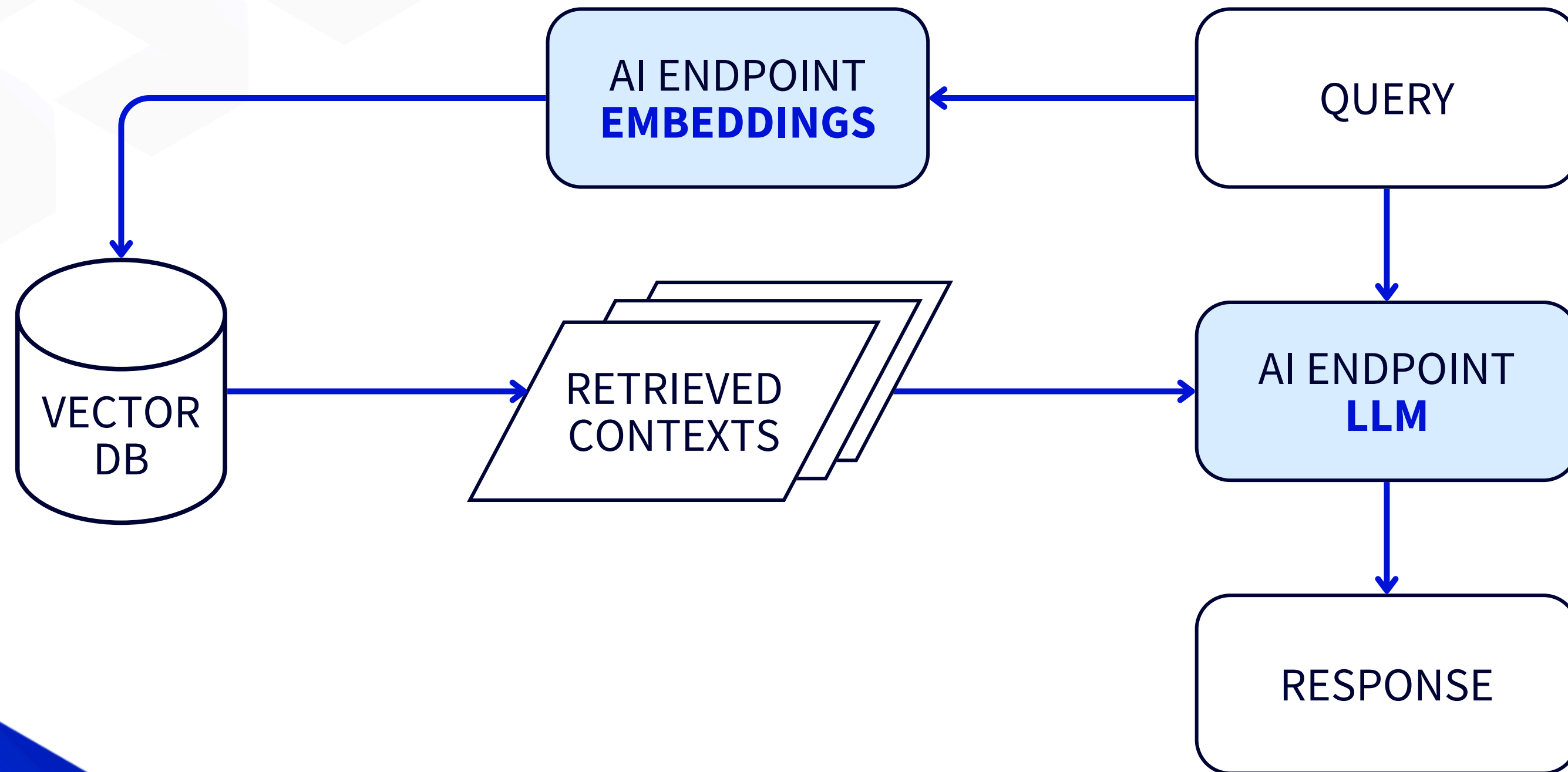




“ASSEMBLEZ CES  
AI ENDPOINTS  
ENTRE EUX POUR UNE  
SOLUTION INNOVANTE  
ET SUR-MESURE”

# DEMO 3 - UN CHATBOT INCOLLABLE SUR VOS DONNÉES

“ Utiliser **AI Endpoints** et **LangChain** pour créer son propre chatbot, c'est possible ! ”



“CRÉEZ VOTRE PROPRE  
RAG AVEC AI ENDPOINTS  
ET LANGCHAIN POUR UN  
CHATBOT PERSONNALISÉ”

# CONCLUSION

que retenir de AI Endpoints ?

# CONCLUSION

Que retenir de la solution *AI Endpoints* ?



## CONFIDENTIALITÉ

Avec AI Endpoints nous **ne stockons pas** et **ne partageons pas** vos données pendant ou après l'utilisation du modèle.



## TRANSPARENCE

Grâce à la **transparence** d'AI Endpoints sur les **modèles d'IA** utilisés, les clients peuvent mettre en œuvre ces modèles sur leur propre infrastructure ou sur d'autres services cloud.



## ACCESSIBILITÉ

AI Endpoints est conçu dans le but de **simplifier** et de **démocratiser** l'IA et permet aux développeurs de tous niveaux d'utiliser des APIs de pointe, **sans aucune expertise** nécessaire.



## IA OPTIMISÉES

AI Endpoints met à disposition les **derniers modèles les plus puissants**, optimisés par les ingénieurs en IA d'OVHcloud pour des **performances** et une **précision maximales**.

# CONCLUSION

*Que retenir de la solution **AI Endpoints** ?*



**TESTEZ PAR VOUS MÊME !**

MERCI !

# CONTACTEZ-NOUS !

## Informations

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

 OVHcloud

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

 @OVHcloud

