

DataOps.Rocks by Saagie



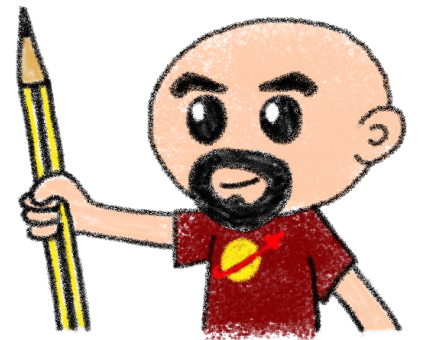
Industrialiser des projets IA avec OVHcloud

Elea Petton & Horacio González

2022-05-12



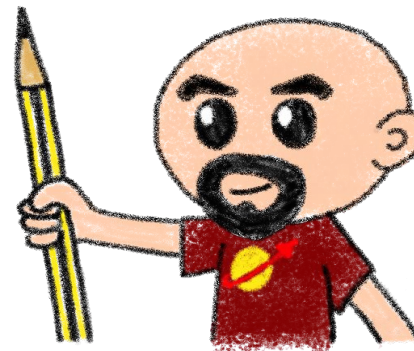
Eléa Petton
@EleaPetton



Horacio González
@LostInBrittany

Qui sommes nous ?

Et qui est OVHcloud ?



Eléa Petton

 **OVHcloud**
Machine Learning Engineer

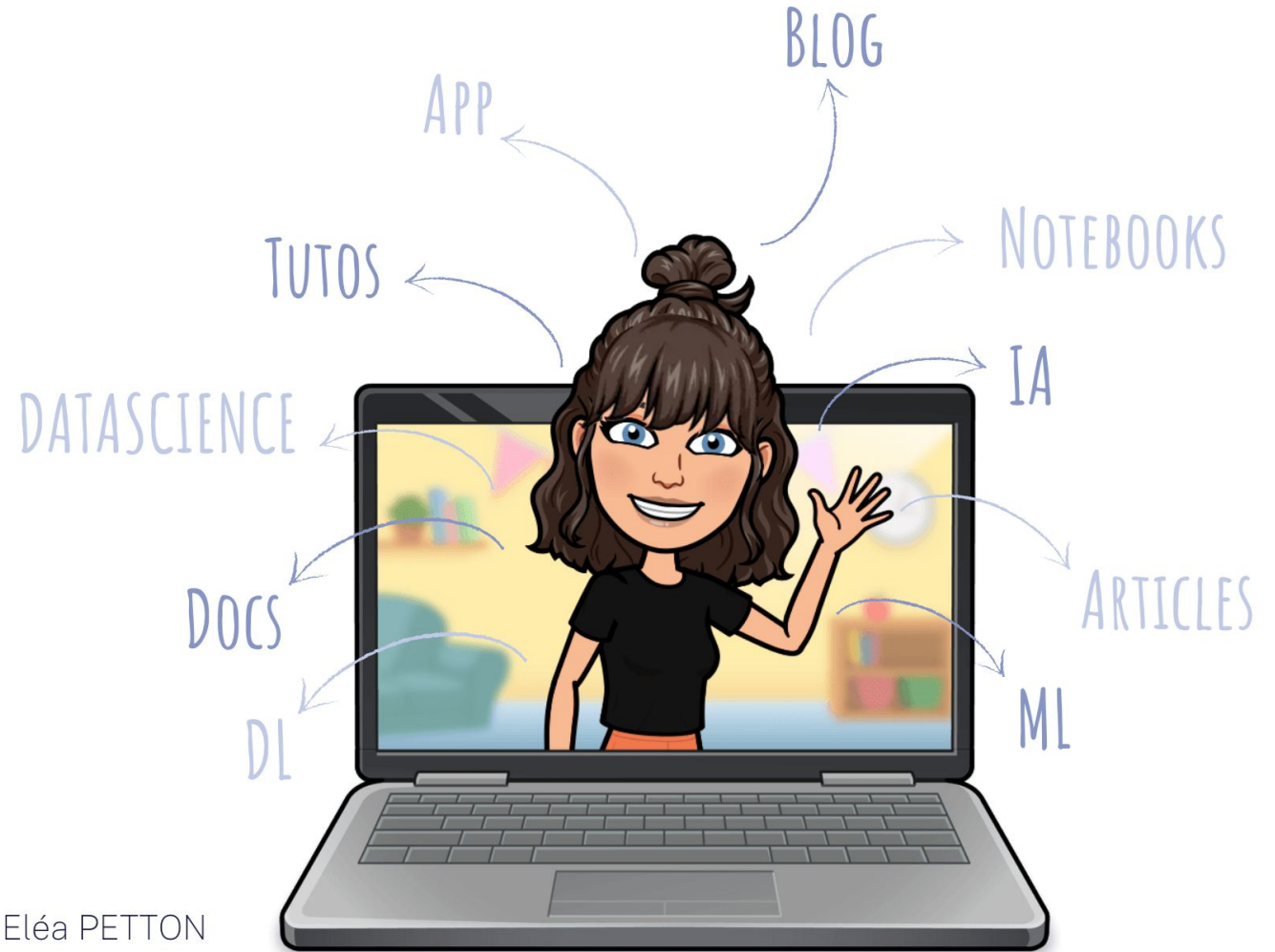
ISEN | 
ALL IS DIGITAL!
BREST
Etudiante ingénieure

 @EleaPetton

 eleapttn

 Eléa PETTON

 OVHcloud



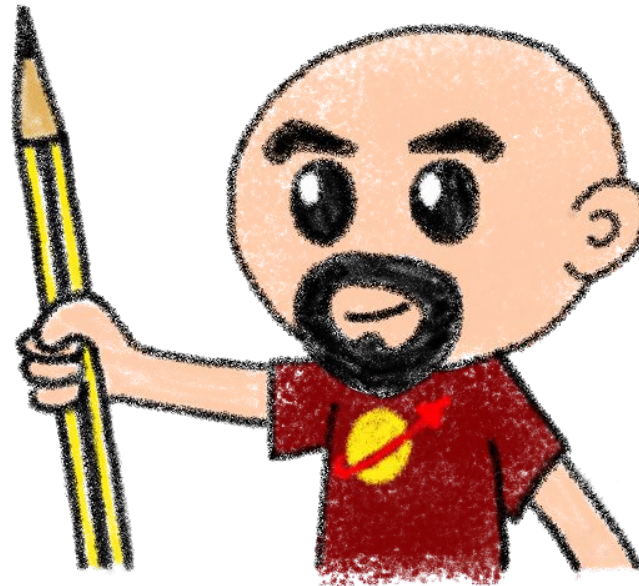
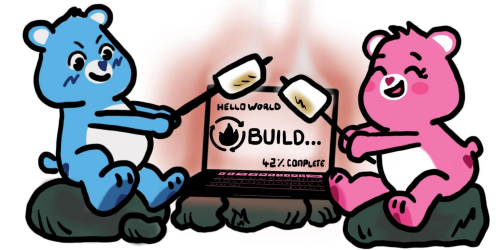
@EleaPetton 

@LostInBrittany 

Horacio Gonzalez

@LostInBrittany

Espagnol perdu en Bretagne...





Web Cloud & Telecom



Private Cloud



Public Cloud



Storage



Network & Security



30 Data Centers
in 12 locations



34 Points of Presence
on a 20 TBPS Bandwidth Network



2200 Employees
worldwide



115K Private Cloud
VMS running



300K Public Cloud
instances running



380K Physical Servers
running in our data centers



1 Million+ Servers
produced since 1999



1.5 Million Customers
across 132 countries



3.8 Million Websites
hosting



1.5 Billion Euros Invested
since 2016



P.U.E. 1.09
Energy efficiency indicator

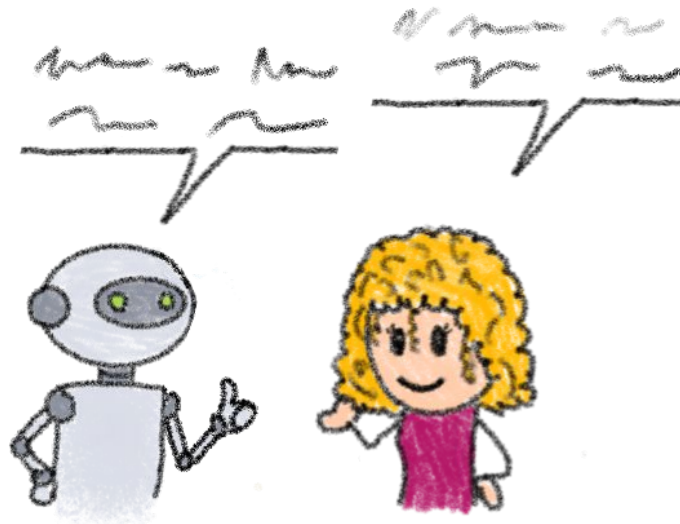


20+ Years in Business
Disrupting since 1999



Les différents visages de l'IA

Et des gens qui travaillent avec



On parle souvent de deux profils type

A



Doing AI

Startups, Tech companies
Public sector, governments
Services providers

I have data scientists in my company.
I need an **AI toolbox** to develop my projects

B



Using tools

« powered by AI »
Mid-Large companies
Public sector

I'm looking for **Out-of-the box data products powered by AI** ! A Chatbot, Fraud analyzer, a search engine, ...

Et, bien entendu

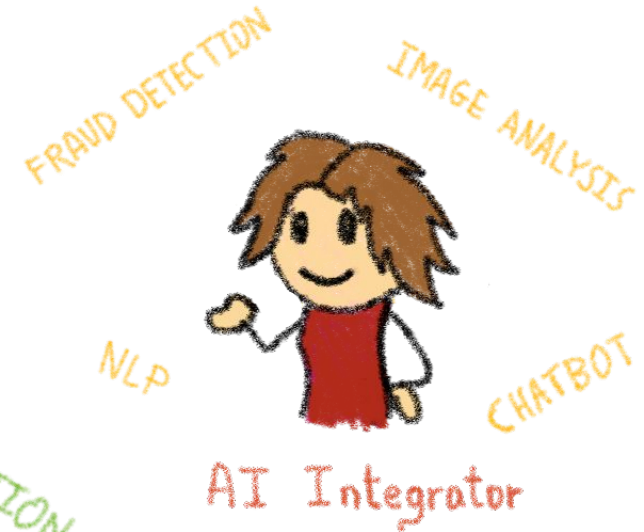
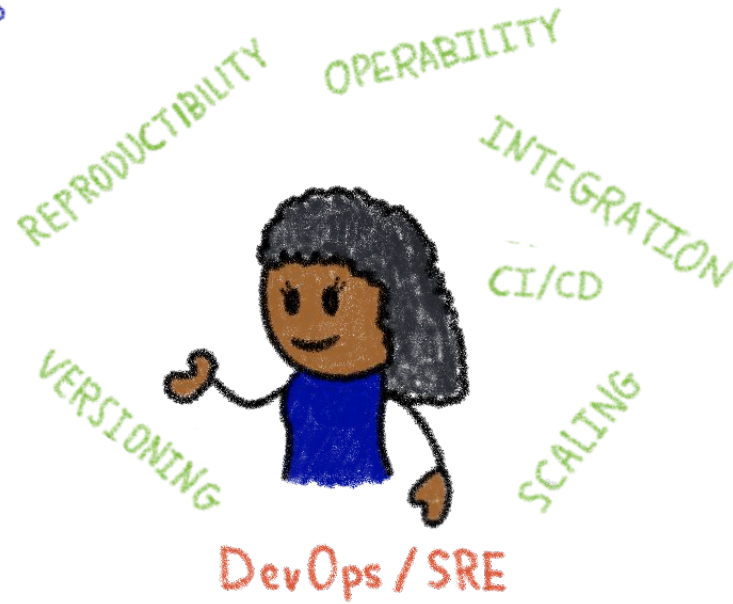
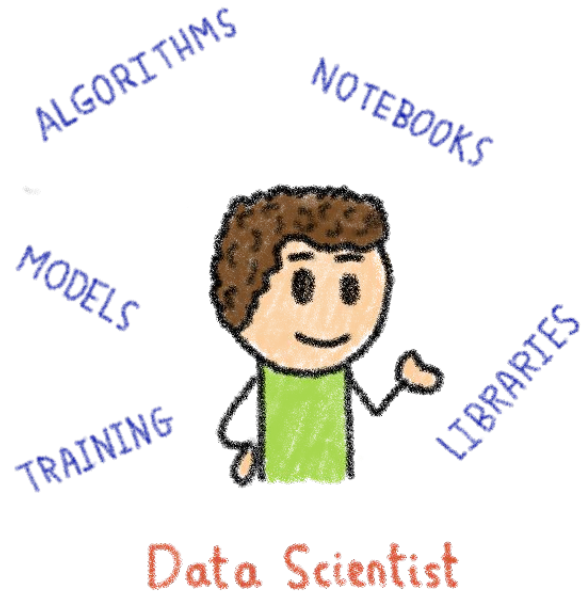
- Simple
- Sécurisé
- Respect du RGPD
- Coût maîtrisé



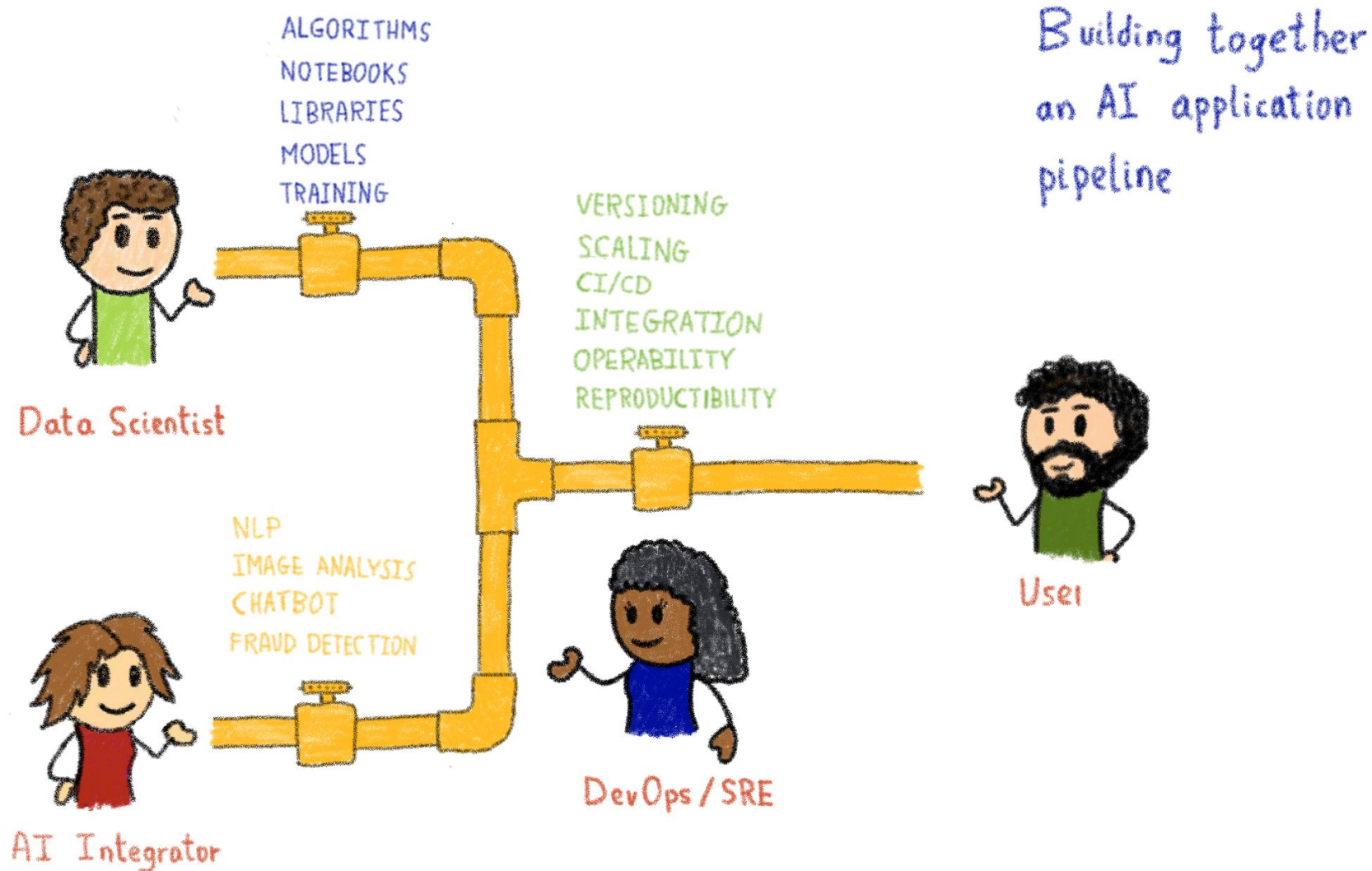
Mais il y a un troisième : DevOps/SRE/DataOps



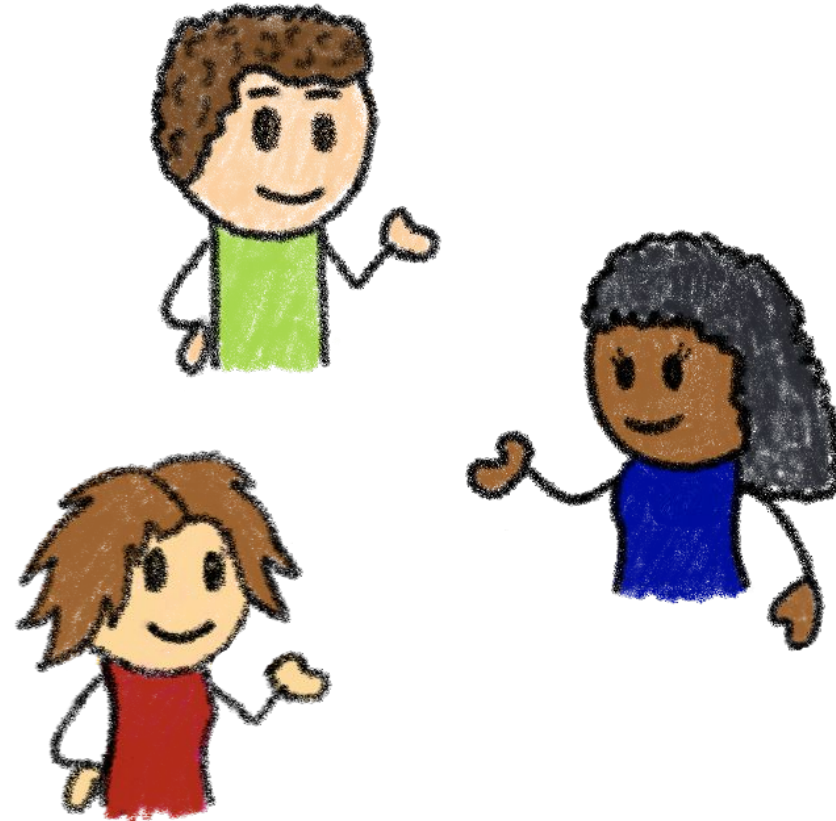
Ils parlent des langages différents



Mais ils ont besoin de travailler ensemble



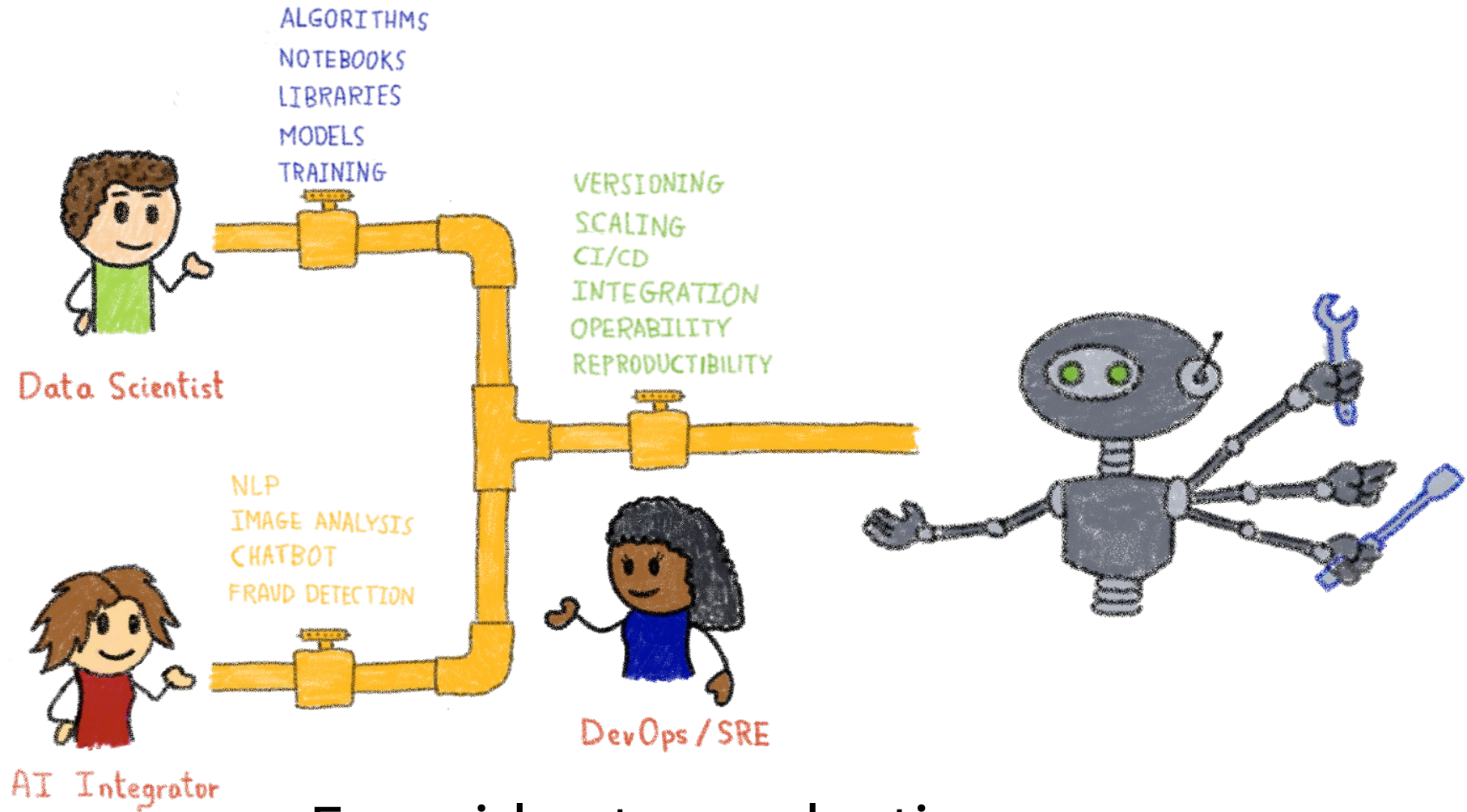
Le challenge de l'intégration



Intégrer les processus et outils
des équipes IA/ML, Dev & DataOps

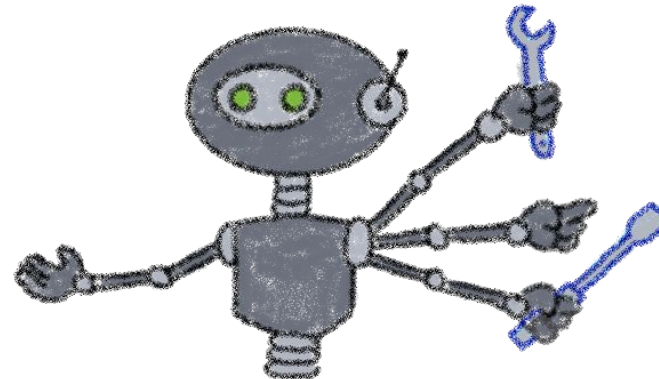


Automatiser le pipeline de bout en bout



OVHcloud & IA

Notre réponse à l'automatisation du pipeline de l'IA



Partie de notre gamme de solutions data



New Database as a Service

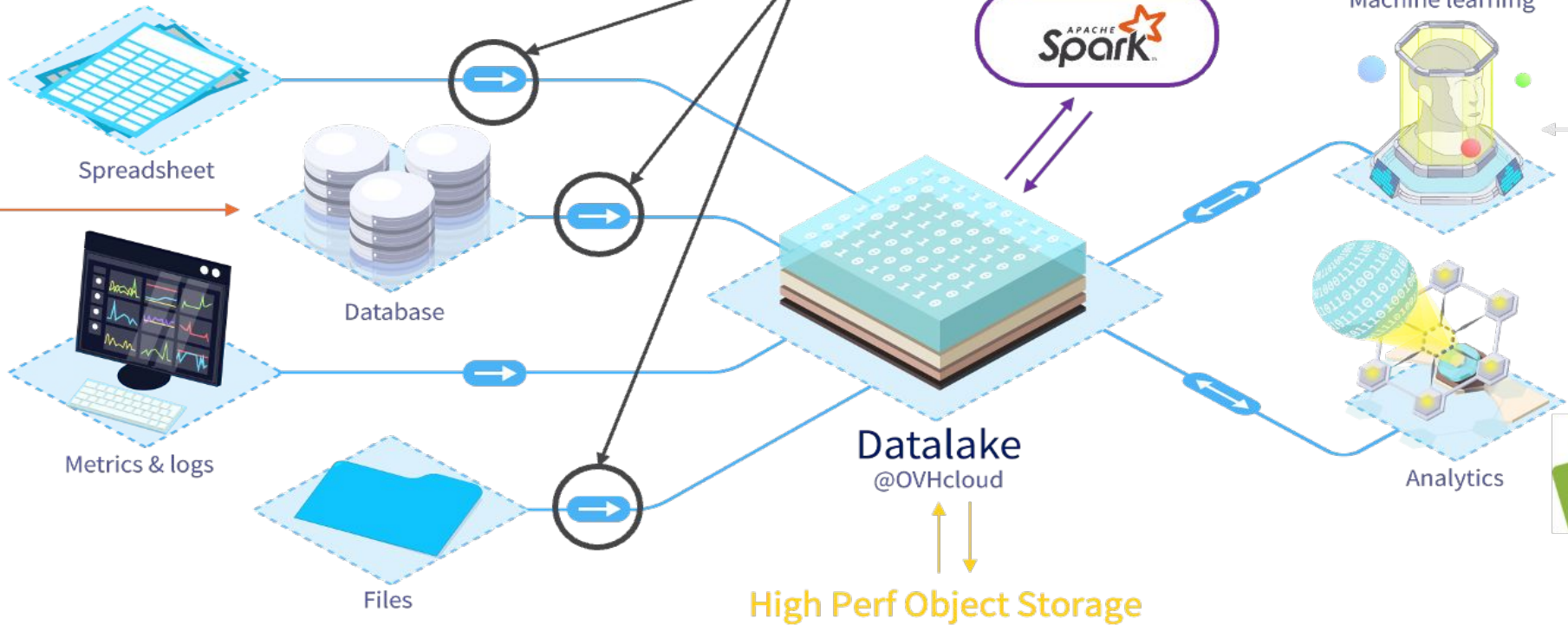
PostgreSQL, OpenSearch, MongoDB, MySQL, cassandra, M3, Grafana

New Ingestion
kafka

Data Processing
APACHE Spark

New AI Tools
learn, jupyter, smiley

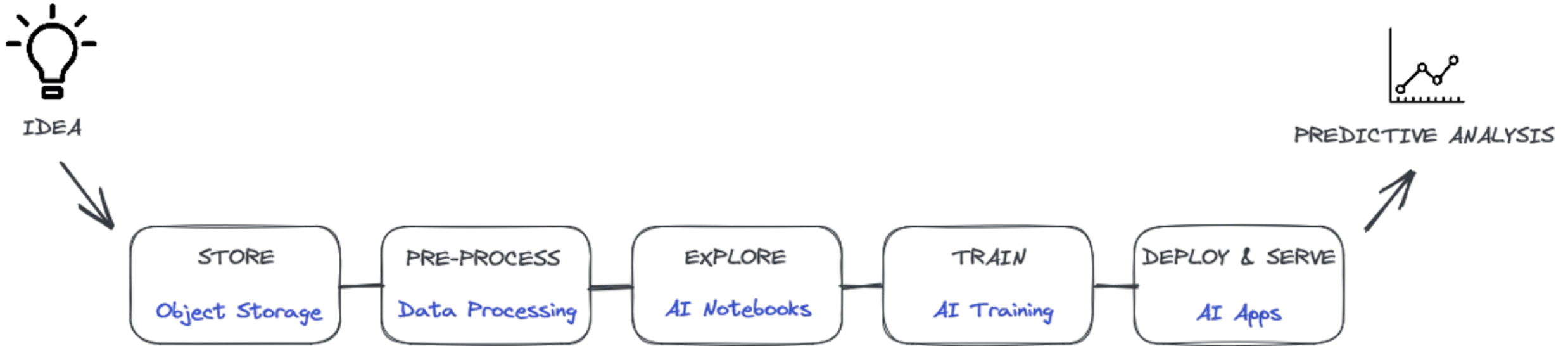
AI Notebooks
AI Training
AI Apps



New High Perf Object Storage
S3



OVHcloud AI Solutions



Un **ensemble complet de services managés** qui offrent un moyen rapide et facile aux data scientists et développeurs de travailler avec **des modèles ML de l'idée à la production**

Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite



Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite
- **Nettoyez et traitez vos données** avec **Data Processing**, solution construite sur Apache Spark

Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite
- **Nettoyez et traitez vos données** avec **Data Processing**, solution construite sur Apache Spark
- **Développez vos modèles d'IA** avec **AI Notebooks**, en démarrant instantanément un notebook Jupyter ou VS Code avec des ressources GPU/CPU



Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite
- **Nettoyez et traitez vos données** avec **Data Processing**, solution construite sur Apache Spark
- **Développez vos modèles d'IA** avec **AI Notebooks**, en démarrant instantanément un notebook Jupyter ou VS Code avec des ressources GPU/CPU
- **Entraînez vos modèles d'IA** avec **AI Training**, sans configuration d'installation complexe, permettant la parallélisation CPU/GPU

Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite
- **Nettoyez et traitez vos données** avec **Data Processing**, solution construite sur Apache Spark
- **Développez vos modèles d'IA** avec **AI Notebooks**, en démarrant instantanément un notebook Jupyter ou VS Code avec des ressources GPU/CPU
- **Entraînez vos modèles d'IA** avec **AI Training**, sans configuration d'installation complexe, permettant la parallélisation CPU/GPU
- **Déployez vos modèles d'IA** dans des modèles de production ou pré-formés ou même des applications très facilement avec **AI Apps**

Un pipeline complet pour faire de l'IA



- **Centralisez vos données** au sein d'un datalake tel qu'**Object Storage**, pour bénéficier d'un haut débit et une latence réduite
- **Nettoyez et traitez vos données** avec **Data Processing**, solution construite sur Apache Spark
- **Développez vos modèles d'IA** avec **AI Notebooks**, en démarrant instantanément un notebook Jupyter ou VS Code avec des ressources GPU/CPU
- **Entraînez vos modèles d'IA** avec **AI Training**, sans configuration d'installation complexe, permettant la parallélisation CPU/GPU
- **Déployez vos modèles d'IA** dans des modèles de production ou pré-formés ou même des applications très facilement avec **AI Apps**

Ces services sont construits sur notre **Public Cloud**, fiable et sécurisé, respectueux des valeurs et du droit européens, et basé sur le socle open source **Openstack**



Pay as you go, simple, et avec un prix ultra-compétitif



GRA +
BHS

Object storage

Prix par GB, a partir de:

0,01 € HT /mois /GB stockage
+ 0,01€ HT /GB trafic sortant

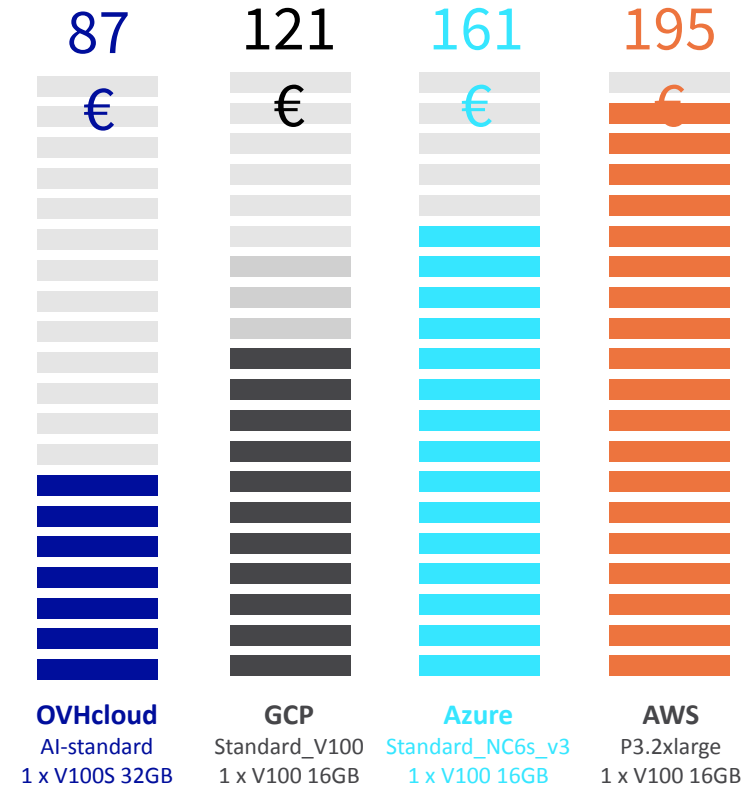
(exemple: 10TB = 100€ HT/mois)

AI Notebooks / AI Training

Prix par GPU par minute, à partir de :
1,75€ /heure /gpu (NVIDIA V100s 32GB)

Pris par CPU par minute, à partir de :
0,03€ /heure /cpu (Intel Xeon 1vCPU + 4GB)

Combien ça coûte 50 hours de
AI Notebook avec 1 x NVIDIA V100 GPU ?

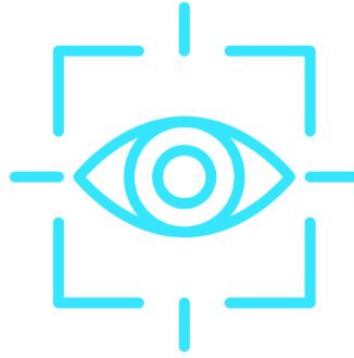


Prices in EU datacenters, without storage attached,
no period commitment.

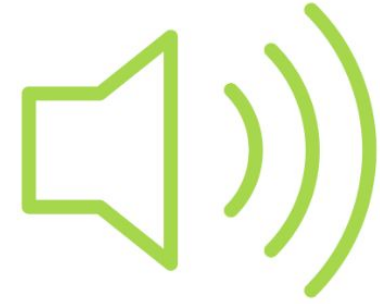
Différentes familles en IA



NLP



Computer Vision



Audio



Pourquoi le son en IA ?



Classification des données acoustiques



Classification des sons environnementaux



Classification du langage naturel



Classification de la musique

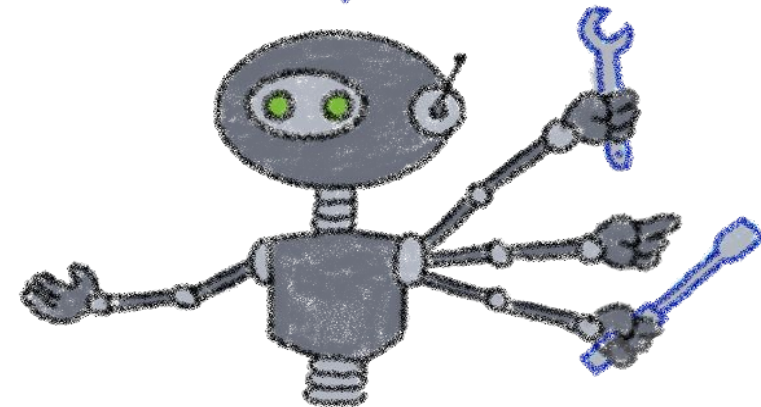


Sons de mammifères marins



Watkins Marine Mammal Sound Database

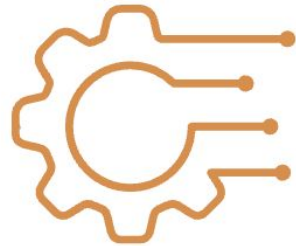
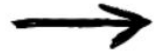
On va essayer de différencier
45 mammifères marins différents



Pré-traitement des données audios



audios .wav



paramètres audio
Librosa



26 paramètres audio Librosa

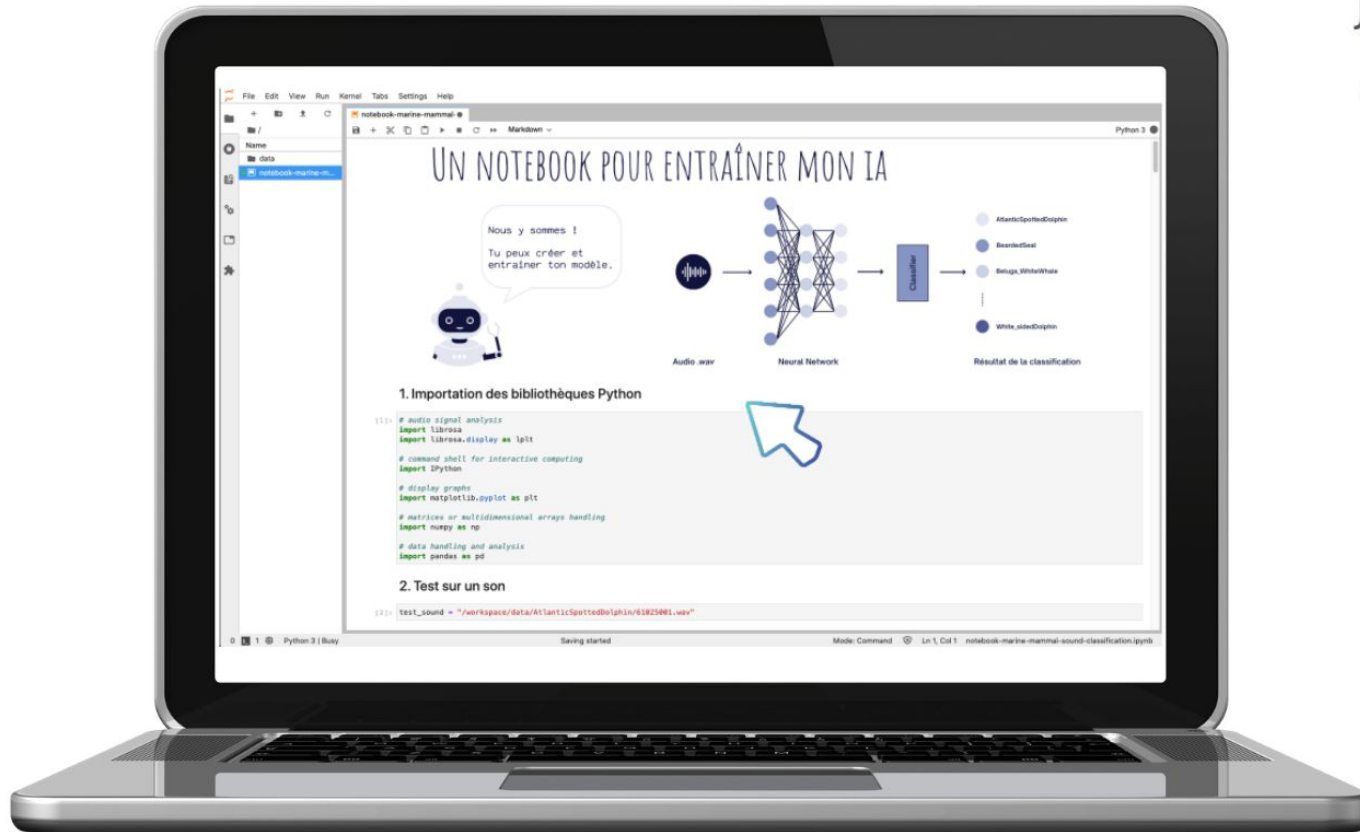


name	length	label

fichier .csv

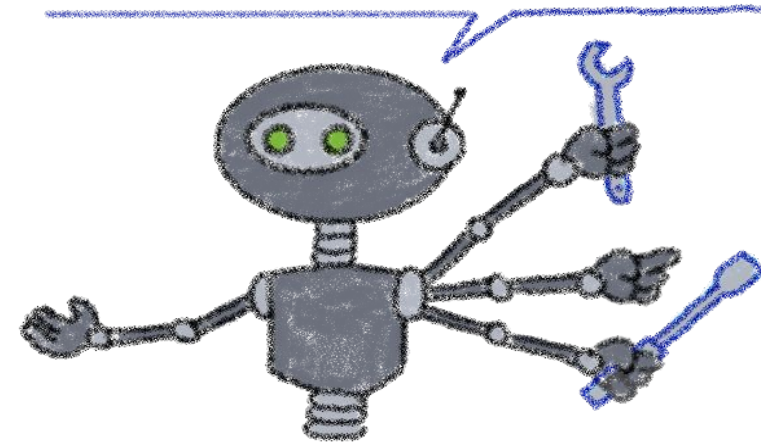


Un notebook pour entraîner mon IA



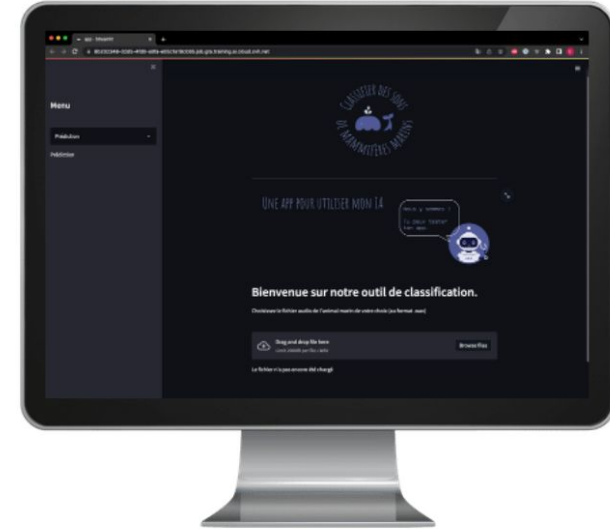
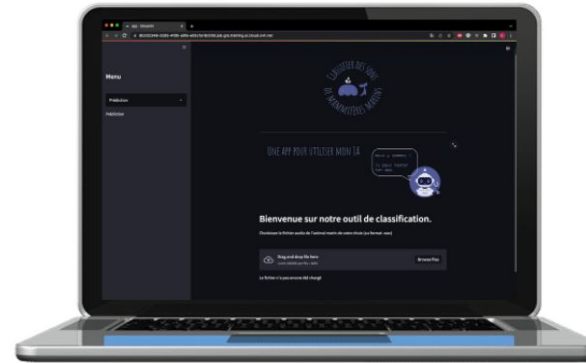
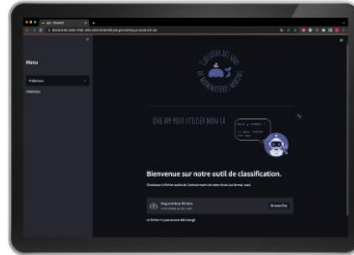
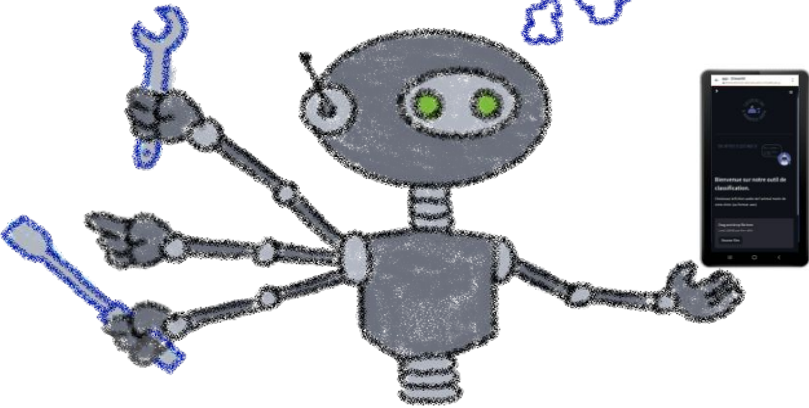
TensorFlow

Tu veux ouvrir ton notebook?
Click sur l'écran



Une app pour utiliser mon IA

Imaginez que votre
IA soit accessible
et utilisable par
tout le monde



Comment créer une App ?



Modèle d'IA



Code Python



Dockerfile



Job AI
Training

- Entraînement
- Export vers Object Storage

Comment créer une App ?



Modèle d'IA



Code Python



Dockerfile



Job AI
Training

- Framework Streamlit 
- Transformation du son 
- Classification 

Comment créer une App ?



Modèle d'IA



Code Python



Dockerfile



Job AI
Training

- Image Docker de base
- Répertoire courant
- Dépendances Python
- Commande à exécuter

Comment créer une App ?



Modèle d'IA



Code Python



Dockerfile



Job AI
Training

- Image Docker personnalisée
- Port d'accès du job
- CPU / GPU
- Conteneur objet du modèle

Comment créer une App ?



Modèle d'IA



Code Python



Dockerfile



Job AI
Training

Test de mon App

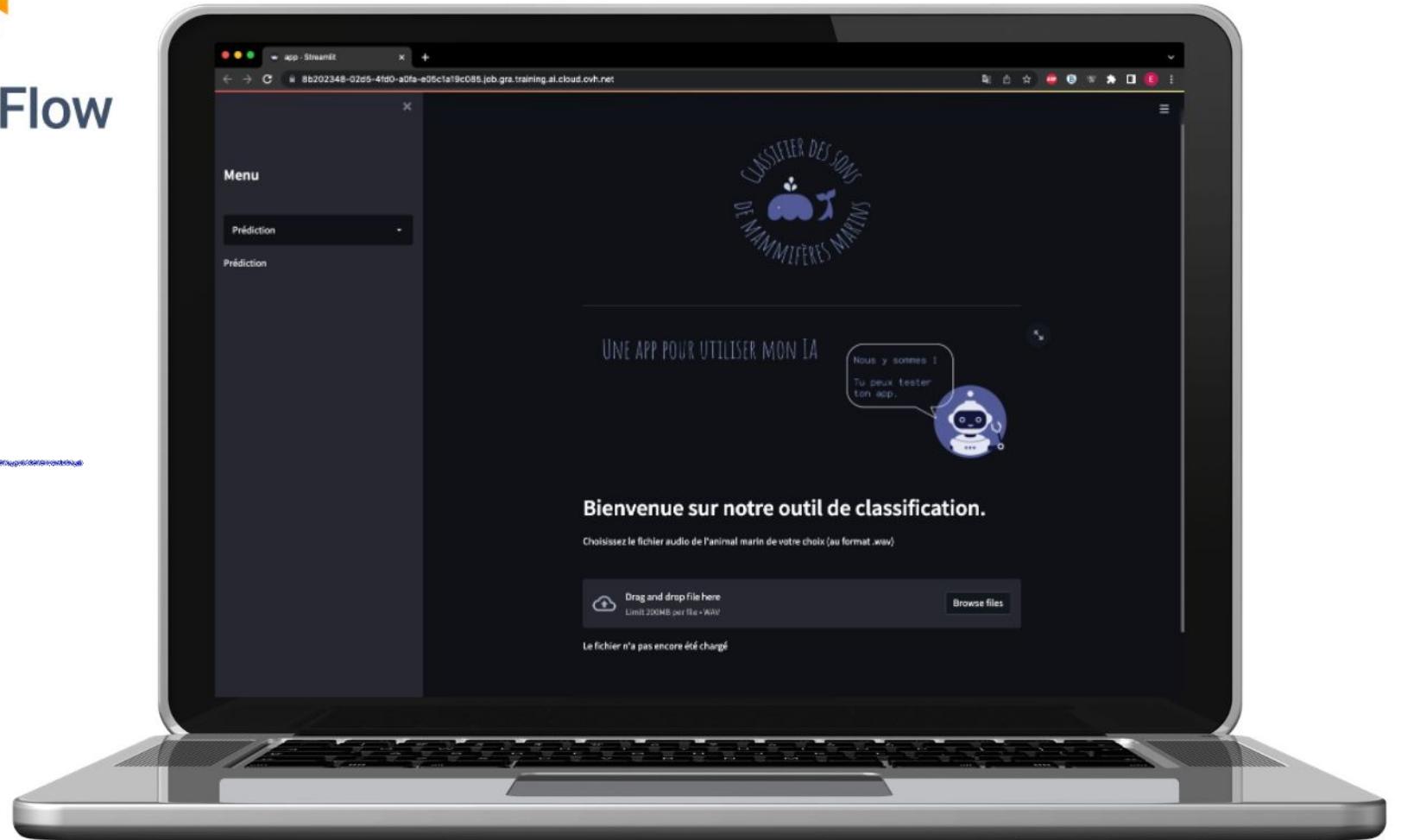
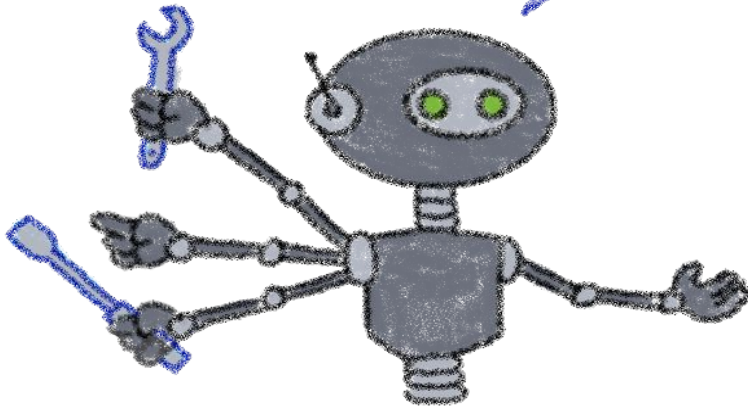


TensorFlow

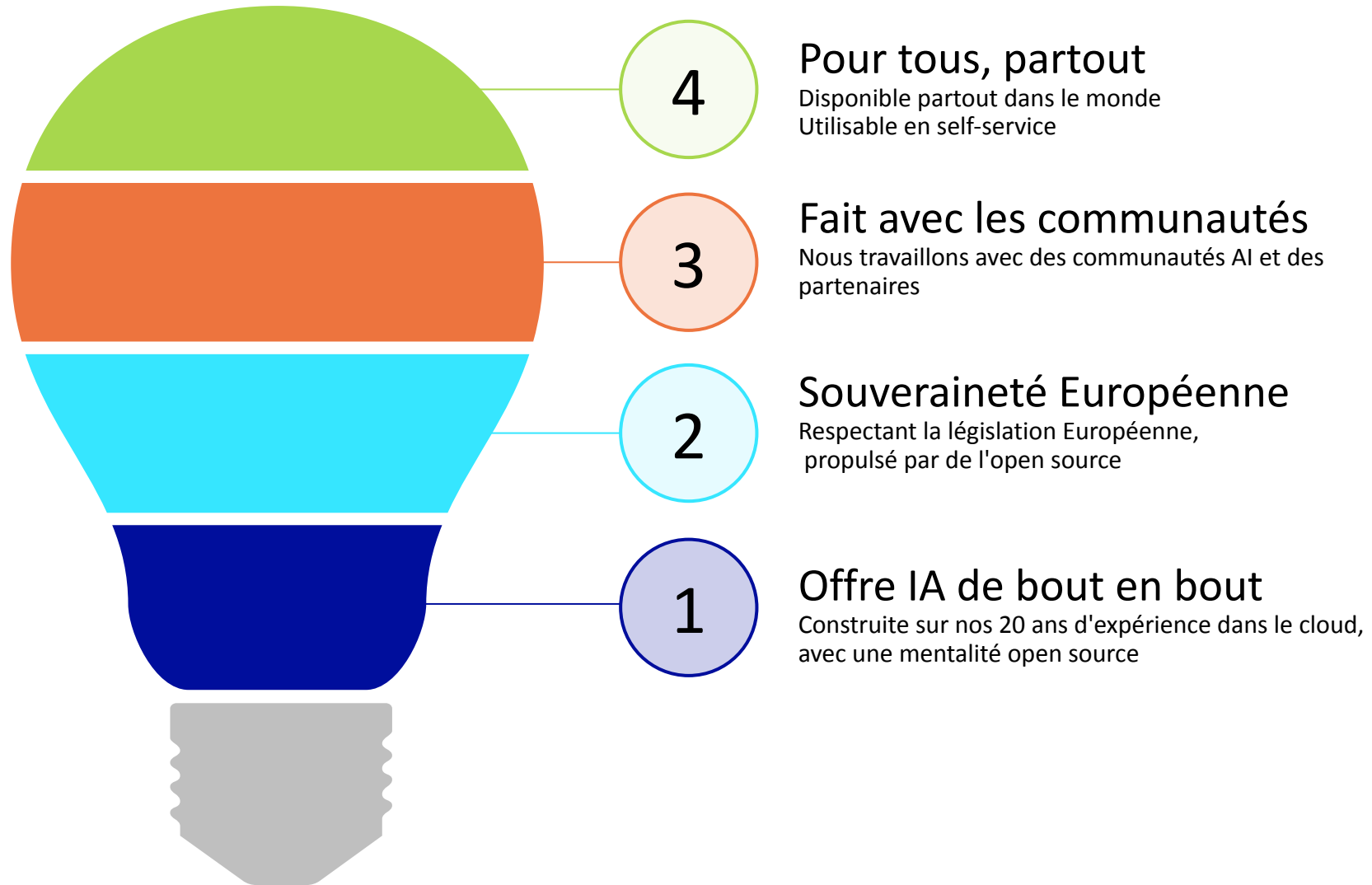


Streamlit

Tu veux tester ton app ?
Click sur l'écran



OVHcloud & IA: conclusion



Pour tester

- Le notebook :

<https://github.com/ovh/ai-training-examples/blob/main/notebooks/tensorflow/tuto/notebook-marine-sound-classification.ipynb>

- La doc du notebook :

<https://docs.ovh.com/gb/en/publiccloud/ai/notebooks/tuto-marine-mammal-sounds-classification/>

- L'app :

https://github.com/ovh/ai-training-examples/tree/main/jobs/streamlit/marine_sounds_classification_app

- La doc de l'app :

<https://docs.ovh.com/gb/en/publiccloud/ai/training/tuto-streamlit-sounds-classification/>

- L'article de blog :

<https://blog.ovhcloud.com/ai-notebooks-analyze-and-classify-sounds-with-ai/>



- TensorFlow : <https://www.tensorflow.org/guide?hl=fr>
- JupyterLab : <https://jupyter.org/>
- Pandas : <https://pandas.pydata.org/>
- Sklearn : <https://scikit-learn.org/stable/>
- Numpy : <https://numpy.org/>
- Matplotlib : <https://matplotlib.org/>
- Dataset des sons de mammifères marins :
<https://www.kaggle.com/datasets/shreyj1729/best-of-watkins-marine-mammal-sound-database>
- Informations sur la dataset : <https://cis.who.edu/science/B/whalesounds/index.cfm>
- Classification de musique avec un CNN :
<https://blog.clairvoyantsoft.com/music-genre-classification-using-cnn-ef9461553726>
- Classification des genre musicaux :
<https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8>
- Pré-traitement des données en Machine Learning :
<https://towardsdatascience.com/introduction-to-data-preprocessing-in-machine-learning-a9fa83a5dc9d>

