

Développer avec l'IA : et si c'était aussi simple qu'ajouter une librairie ?

Thierry Chantier / Stéphane Philippart



SNOWCAMP

Un Grand MERCI à nos sponsors 2025



CGI



ABYLSSEN



KAIZEN
SOLUTIONS



copen



MOODY'S



VISEO
— POSITIVE DIGITAL MAKERS —



CRITEO



elastic



clever cloud



AVISTO



KLS GROUP
La French Logistique



Bonitasoft



HAVANA
IT & APPS



zenika



salesforce



sopra
steria



bpifrance.iO



kelkoo group



MERCI !!!



 ESPACE-CONGRES 

 visiteur

 ULfrguBM





Faisons connaissance



slido.com



#snowcamp-2025



Thierry Chantier



Developer Advocate
@OVHcloud



 TitiMoby@mamot.fr

 TitiMoby

 <https://noti.st/titimoby>



Stéphane Philippart

  DeveloperAdvocate@OVHcloud 

 Co-créateur de [TADx](#) (meetup à Tours)

```
whoami
stef
```

 Padawan Intelligence Artificielle 

 [@wildagsx](#)

 <https://philippart-s.github.io/blog>

 <https://github.com/philippart-s/>

 <https://www.linkedin.com/in/philippartstephane/>

TAD 



Que va-t-on voir aujourd'hui ?

 Développer un chatbot en Java / LangChain4J / Quarkus

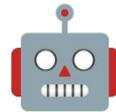
 Développer une analyse de texte (sentiments, émotions, ...) en JavaScript

 Développer une application de text to speech et de speech to text en Python





IA / ML / LLM



Les différentes catégories de l'IA

 Actuellement IA ~ Machine Learning

 Artificial Narrow Intelligence (ANI) / Weak AI

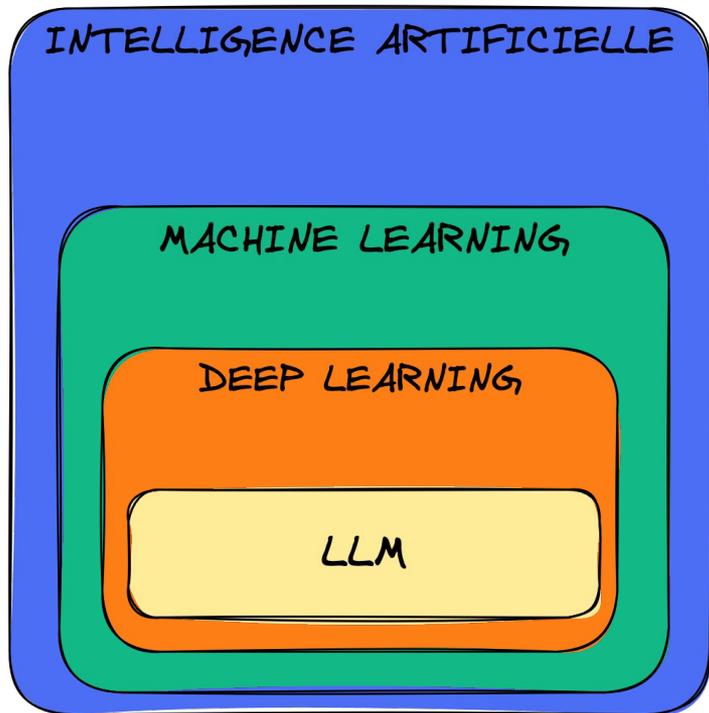
 Artificial General Intelligence (AGI) / Strong AI

 Artificial Super Intelligence (ASI)





Apprendre ...



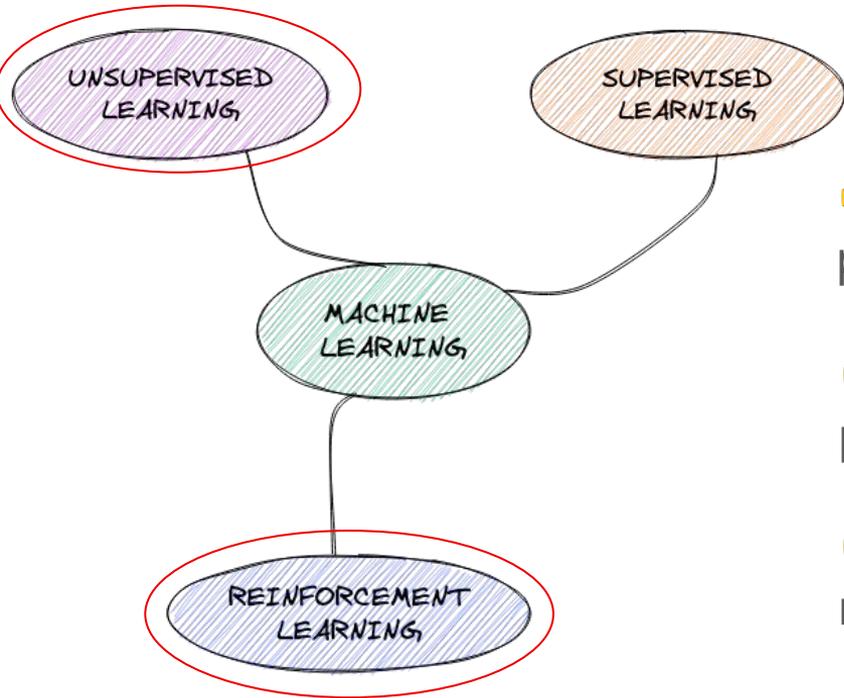
👉 IA : Techniques pour simuler les capacités “intelligentes” d’un être humain

👉 ML : Techniques pour qu’une machine apprenne sans algorithme supplémentaire par l’ingestion de données d’exemples

👉 DL : ML utilisant poussée des réseaux de neurones pour améliorer les performances dans les déductions



Le Machine Learning dans le détail



👉 Supervised learning : apprentissage proche de ce que l'on fait avec l'Homme

👉 Unsupervised learning : apprentissage par patterns et regroupements

👉 Reinforcement learning: apprentissage par renforcement avec système de "récompenses"



Sans donnée ... pas d'intelligence !



 La plupart du temps les données sont brutes : il faudra les nettoyer, les pré-traiter, les transformer, ...

 Il va falloir les transformer en Dataset (en gros une base de données pour IA) et les labelliser

 Un dataset contient les données d'apprentissage, de validations et de tests

 Enfin, ce sont des vecteurs et autres matrices qui seront manipulés par le modèle



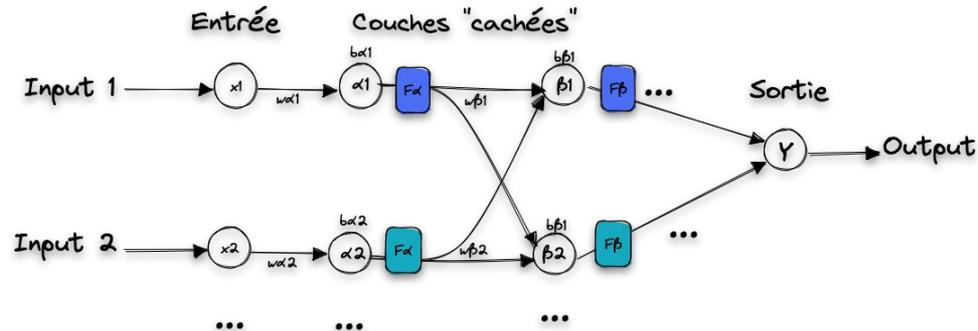


Deep learning & Réseau de neurones

 L'étape d'après ... Plus puissante, plus complexe

 Analogie avec les neurones humain plutôt fausse

↔ Essentiellement due au mécanisme d'inter-connexions



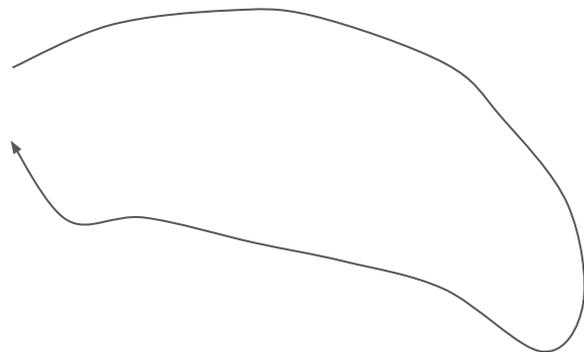
 w^* sont les poids et permettent d'ajuster le comportement du réseau

 b^* sont les biais pour rendre le modèle "plus réaliste" et influencer sur la fonction d'activation

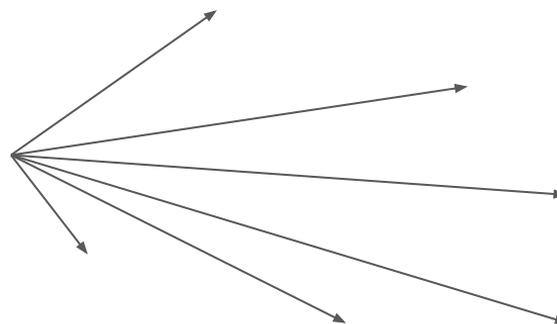
 F^* sont les fonctions d'activation



GPU vs CPU



CPU



GPU



AI Endpoints / Modèles



AI Endpoints en quelques mots

 LLM as a Service

 Beta ~ avril 2025 (gratuit)

 Dispo des LLM via des API

 Catégories des modèles

- Assistants : CodeLlama 13b, Llama 3 70b, Mixtral 8x22b, ...
- Embedding : BGE base, Multilingual E5, ...
- NLP: Bart, Bert, ...
- Translation : T5 large, ...
- Audio : Nvidia RIVA ASR
- Vision : Yolo V8, ...

 <https://endpoints.ai.cloud.ovh.net>



 Démo ! 



Cloud Development Environment



Cloud Development Environment

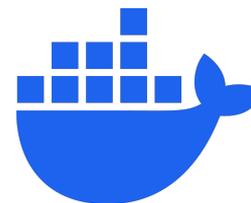
Le code du projet



Un éditeur de code



Un environnement d'exécution



A la demande

- Pour tester une idée
- Expérimenter différentes solutions
- Partager une session de travail entre collègues
- ... sky is the limit (et le coût de votre offre de CDE 😇)



Reproductible

- Environnement décrit précisément
- Configuration versionnée avec le code
- Cohérence entre l'environnement et le code lui même



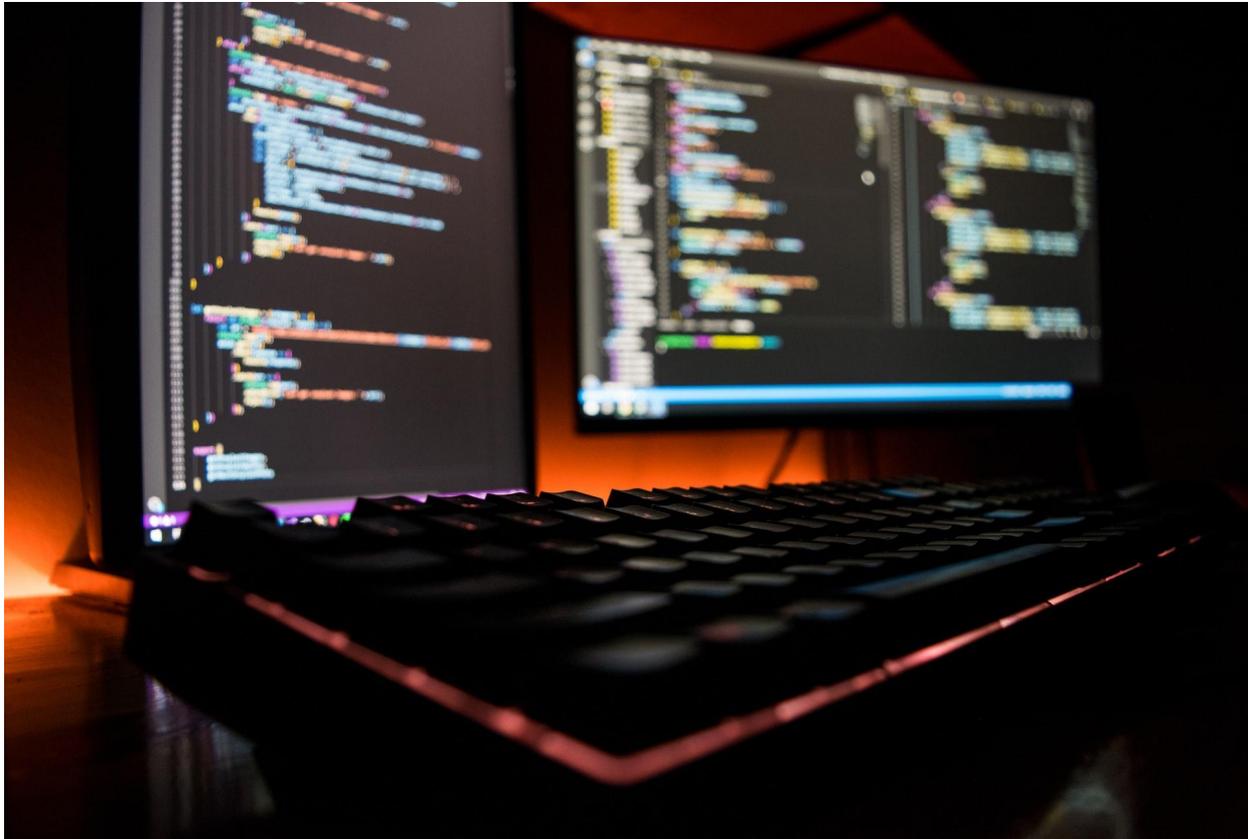
Le CDE pour cet atelier : Coder

- Hébergé chez OVHcloud pour les besoins de cet atelier
- Accessible par : <https://workshop.labdevrel.ovh/>
- Deux fichiers  de configuration :
Dockerfile
main.tf
- 3 modes d'accès :
 - code-server
 - VS Code Desktop
 - Terminal





Workshop time !



☁ Démarrage de Coder 

 : <https://workshop.labdevrel.ovh/>

 : emailX@email.com

 : PasswordX



Fork du repository GitHub

 Projet à forker : <https://github.com/devrel-workshop/ai-as-lib-workshop>



<https://ovh.to/DGo3c1a>



Récupération du token

 Facultatif

 Permet d'augmenter les rates limits

 1 requêtes / ip /seconde ou 12 par minutes pour les utilisateurs anonymes (ou pour les tokens expirés)

 250 requêtes / seconde et 1000 requêtes / minute pour les utilisateurs authentifiés

 Token valide x jours

 Nécessite un projet Public Cloud

<https://ovh.to/8stq8bE>



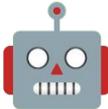
<https://endpoints.ai.cloud.ovh.net/>





Et si on vous laissait choisir ?



 Chatbot 



Java : 101 pour cet atelier



the java logo, stable diffusion on AI Endpoints



Java

 Langage objet ... datant de 1996

 Basé sur une machine virtuelle (JVM) et un development kit (JDK)

 Largement représenté en entreprise

 Depuis quelques années : une release tous les 6 mois (une LTS tous les 3 ans)



LangChain4J & Quarkus

[LanChain4j](#)

 Dérivé du projet originel LangChain

 Accès à des providers pré-définis : OpenAI, Mistral, Ollama, ...

 RAG / Templating de prompt / Streaming / ...

 Quarkus pour la facilité les développements Java

 Cloud native

 Utilisation de l'extension quarkus-lanchain4j





Les concepts utilisés aujourd'hui



Maven pour la gestion des dépendances



Une classe avec un main



Quarkus

- Notion d'interface
- Notion de services HTTP



Pré-requis

 Java 21 LTS

 Maven 3.9.x

 Quarkus CLI



Modèles utilisés

 Mistral 7B Instruct



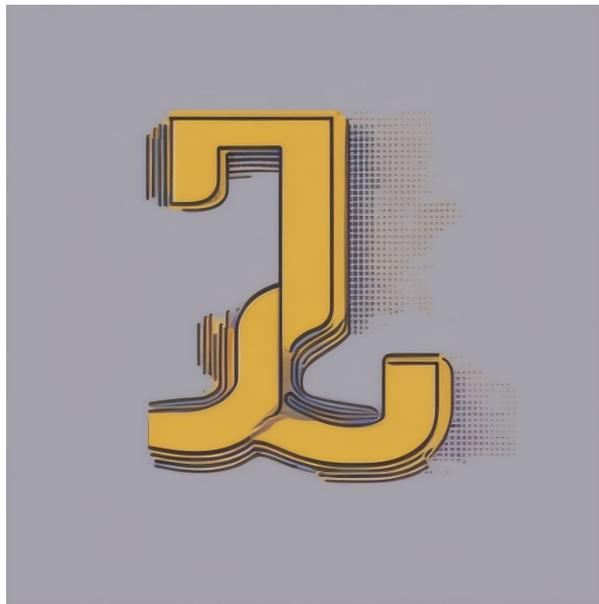
 1, 2, 3 ... Développez ! 



Analyse de texte



JavaScript : 101 pour cet atelier



the JavaScript logo, stable diffusion on AI Endpoints



JavaScript

 Non ce n'est pas basé sur Java

 Présent depuis très longtemps sur nos machines et nos navigateurs

 S'émancipe du navigateur pour devenir “full stack”

 Largement représenté en entreprise

 Simple pour faire une approche script (ce qui sera le cas ici)





Les concepts utilisés aujourd'hui



NPM pour la gestion des dépendances



Un script par exercice



Pré-requis

 NodeJS 20 LTS



Modèles utilisés

 SamLowe/roberta-base-go_emotions

 nlptown/bert-base-multilingual-uncased-sentiment

 dslim/bert-base-NER



 1, 2, 3 ... Développez ! 

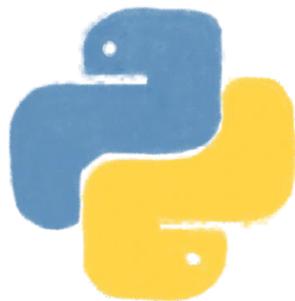




Audio



Python : 101 pour cet atelier



Python : pourquoi ce choix ?

-  Langage simple d'approche mais qui reste complet
-  “All batteries included”
-  Choix des communautés data science et data analysis



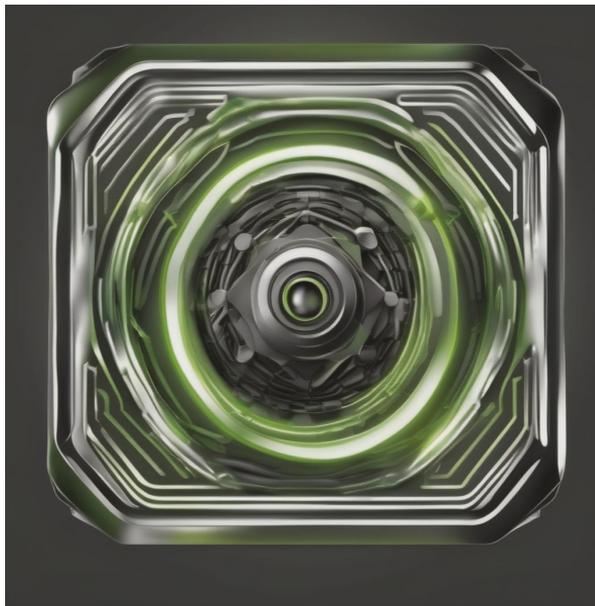
Python : concepts pour aujourd'hui

-  Les fichiers *requirements.txt*
-  L'instruction *import*
-  Venv : isolation de votre Python



🧠 Modèles utilisés

💬 📝 NVIDIA Riva



the NVIDIA logo, stable diffusion on AI Endpoints



 1, 2, 3 ... Développez ! 



Merci !!!!



Slides

<https://ovh.to/z84N7LF>



Feedbacks

<https://ovh.to/fRZWY5y>



Code source

<https://ovh.to/DGo3c1a>





Liens

-  [Documentations](#) OVHcloud univers IA
-  [Enhance your applications with AI Endpoints](#)
-  [How to use AI Endpoints and LangChain4j](#)
-  [LLMs streaming with AI Endpoints and LangChain4j](#)
-  [AI Endpoints](#)
-  [Repository](#) GitHub du workshop

