

CONFIGURATION D'UN SERVEUR IA

PARCOURS

SISR ☒

SLAM ☐

Lieu de réalisation

Sikiwis UERP



Période de réalisation

Du : 13/01/2025

Au :

Modalité de réalisation

SEUL ☒

EN EQUIPE ☐

Intitulé de la mission

Configuration d'un serveur IA

Description du contexte de la mission

Installation, configuration et entraînement d'un modèle d'intelligence artificiel local et personnalisé de l'entreprise

Contraintes & Résultat

Ressources fournies / contraintes techniques / Résultats attendu

- Serveur
- Ollama
- LLama3.2
- Llama3.3
- OpenwebUI

Productions associées

Liste des documents produits et description

Modalités d'accès aux productions

Identifiants, mots de passe, URL d'un espace de stockage et présentation de l'organisation du stockage

Solution en local

Configuration d'un serveur IA

Introduction

Ce document présente la configuration d'un service d'IA local et dédié aux activités de Sikiwis.

Contexte

Sikiwis est une entreprise d'édition de logiciel, proposant également des activités de gestion de paie pour le compte d'autres entreprises.

Ces activités se basent en grande partie sur la législation en cours. Une connaissance de celle-ci est indispensable à la poursuite de ces activités.

Afin d'améliorer la productivité dans la recherche d'informations ayant rapport à la législation, il a été décidé de mettre en place un modèle d'IA, qui sera spécifiquement entraînée et formée à la législation française actuelle.

Objectifs

Les objectifs poursuivis sont :

- La mise en place d'un service IA sur une machine dédiée
- L'entraînement de celle-ci avec la législation en cours

Ressources

Les ressources utilisées sont :

- Un PC faisant office de serveur IA, aux spécifications suivantes :
 - RAM : 32 Gb
 - SSD : 512 Gb
 - GPU :
 - CPU :

Configuration d'un serveur IA

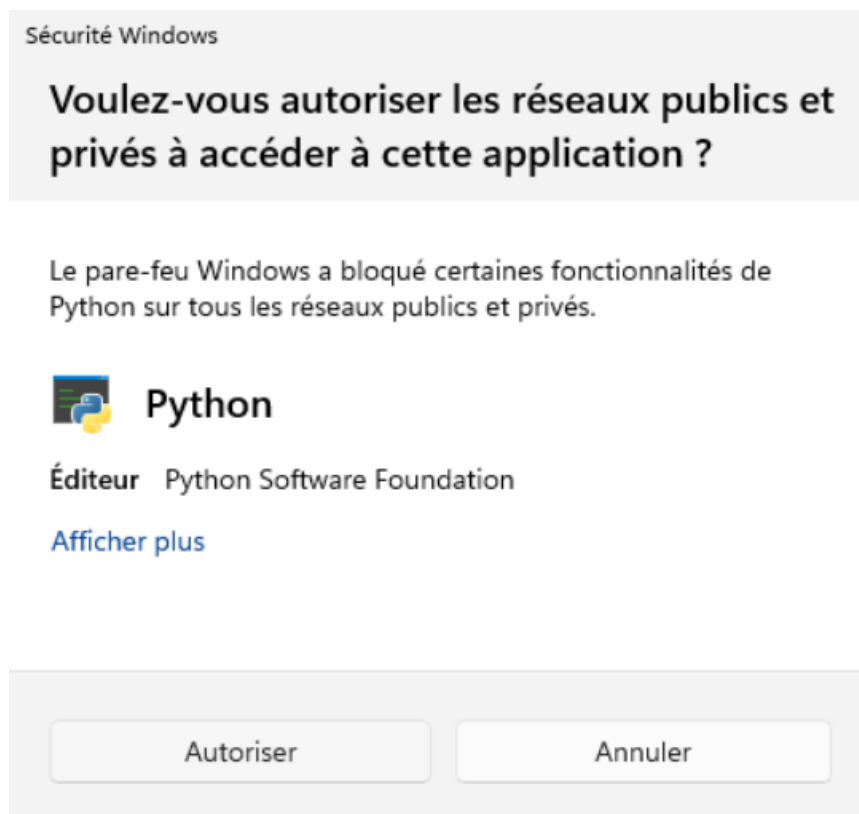
Résultats attendus

Les résultats attendus de la manipulation sont un modèle de langage capable de traiter des ensembles de données de la législation en cours afin d'aider l'entreprise dans la livraison de ses activités paie.

Démarche

1. Installation de Python

La première étape concernant la mise en place d'un serveur IA est l'installation du langage Python. Python est un langage informatique, mis au point dans les années 90, permettant une manipulation plus facile des données, et est très utilisé dans l'édition de logiciels d'IA.



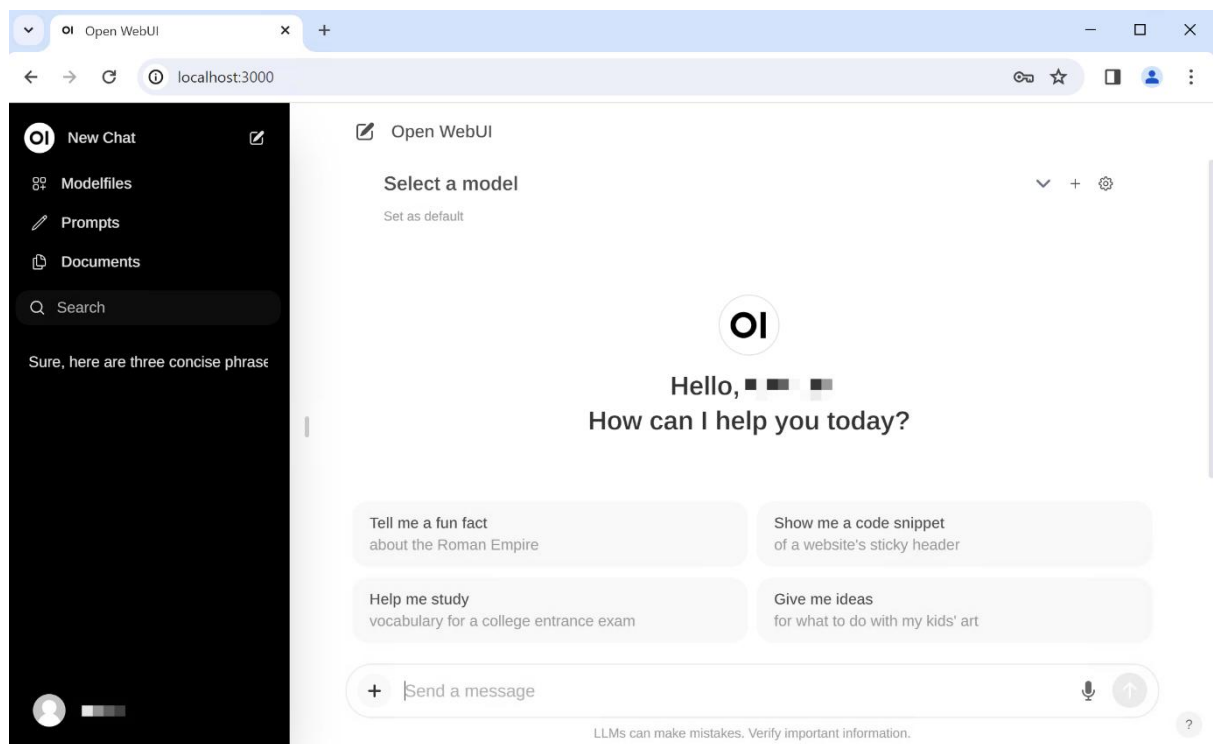
Configuration d'un serveur IA

2. Installation d'un IDE et d'un logiciel de partage de code

La seconde étape consiste en l'installation d'un IDE, c'est-à-dire un environnement de travail dans lequel on peut manipuler des fichiers et coder, ainsi qu'un logiciel de partage de code, avec lequel on peut télécharger des paquets. Nous allons installer des solutions Microsoft, Visual Studio Code et Github Desktop.

3. Installation d'OpenWebUI

OpenWebUI est un logiciel offrant une interface graphique pour des modèles d'IA conversationnels.



4. Installation d'Ollama

Ollama est un logiciel permettant d'exécuter des modèles de langage de grande taille (LLM) localement sur sa machine.



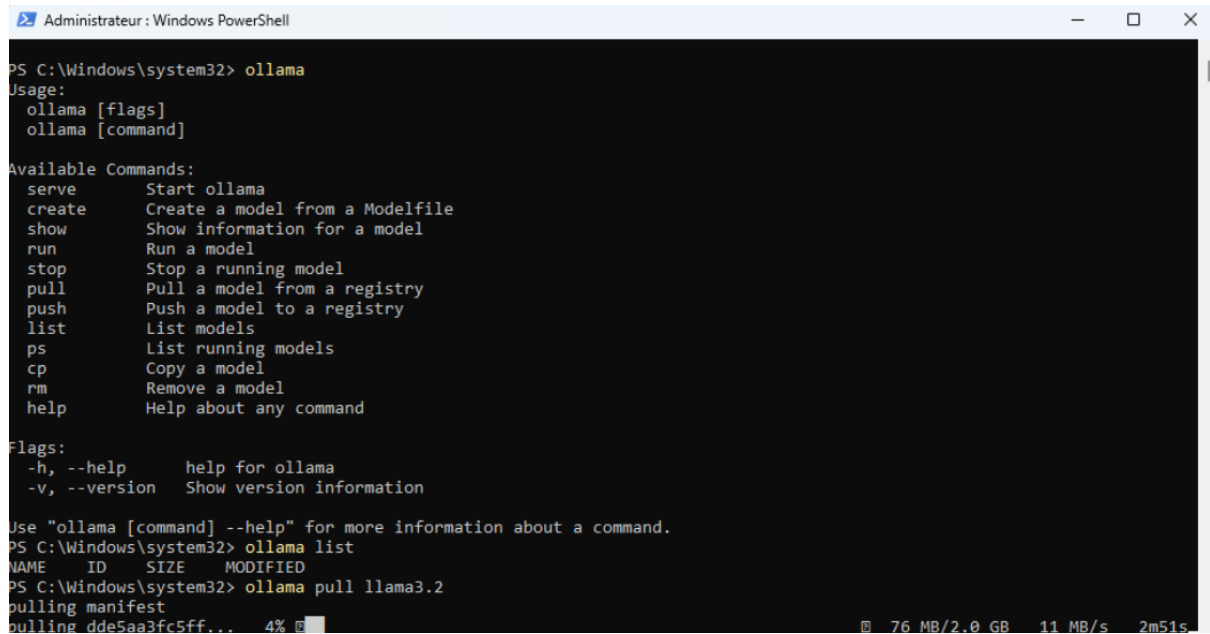
5. Installation des modèles de langage

La prochaine étape consiste en l'installation de modèles de langage. Un modèle de langage est un ensemble d'algorithmes possédant un grand nombre de paramétrage, entraîné sur de grandes quantités de documents. Les grands modèles de langage sont abrégés en LLM en anglais (large language model).

Les plus connus d'entre eux sont GPT, propriété d'OpenAI, Gemini, propriété de Google, ou encore Llama, propriété de Meta.

Configuration d'un serveur IA

On décide d'installer des modèles de Meta : Llama3.2 et 3.3.



```
Administrateur : Windows PowerShell
PS C:\Windows\system32> ollama
Usage:
  ollama [flags]
  ollama [command]

Available Commands:
  serve      Start ollama
  create      Create a model from a Modelfile
  show        Show information for a model
  run         Run a model
  stop        Stop a running model
  pull        Pull a model from a registry
  push        Push a model to a registry
  list        List models
  ps          List running models
  cp          Copy a model
  rm          Remove a model
  help        Help about any command

Flags:
  -h, --help      help for ollama
  -v, --version    Show version information

Use "ollama [command] --help" for more information about a command.
PS C:\Windows\system32> ollama list
NAME      ID          SIZE  MODIFIED
PS C:\Windows\system32> ollama pull llama3.2
pulling manifest
pulling dde5aa3fc5ff... 4% 76 MB/2.0 GB 11 MB/s 2m51s
```

6. Entraînement de l'IA : Fine tuning

L'entraînement d'une IA se nomme le « fine tuning ». Le fine-tuning (ou ajustement fin) est une méthode en apprentissage automatique qui consiste à adapter un modèle pré-entraîné sur un ensemble de données spécifique afin de l'optimiser pour une tâche particulière.

En l'occurrence, nous allons utiliser les logiciels Hugging Face pour procéder au fine tuning de notre LLM.

7. Constitution de modèles d'entraînement : les Datasets