

Running AI-LLM Application on Kata Containers

Patrick Beaucamp

Mail : patrick.beaucamp@bpm-conseil.com

Presentation Agenda

- Legal Context in Europe (and France)
- Migration to Kata Container
- Running AI Platform powered by LLM in a Kata cluster
 - Data4Citizen Platform
 - Data4Citizen Architecture
 - Data4Citizen & AI-LLM
 - Data4Citizen & DM-LLM

Part 1 : Legal Context in Europe

When too many rules and security kills the security !!!



Legal Context – NIS2



- As the cyber threat increases and information systems remain partly vulnerable, the NIS 2 (Network and Information Security) directive, published in the Official Journal of the European Union in December 2022, represents a unique opportunity.
- NIS2 implementation will enable thousands of entities that affect the daily lives of citizens to better protect themselves.

Legal Context – GDPR



- The EU general data protection regulation (GDPR) is the **strongest privacy and security law in the world**.
- This regulation updated and modernised the principles of the 1995 data protection directive. It was adopted in 2016 and entered into application on 25 May 2018.
- The GDPR defines:
 - individuals' fundamental rights in the digital age
 - the obligations of those processing data
 - methods for ensuring compliance
 - sanctions for those in breach of the rules

Some immediate Consequences

When recommendations turn as requests :

- Recommandation/Need to encrypt Virtual Machine
 - Impact on performance for standard déploiement
- Recommandation/Need to encrypt Database storage
 - Compliance : PostGreSql encryption with pgcrypto using AES256
- Recommandation/Need to Secure Docker Image
 - Compliance : running images in Kata Container

Part 2 : Migration and Running Kata

A short procedure to run your image on Kata



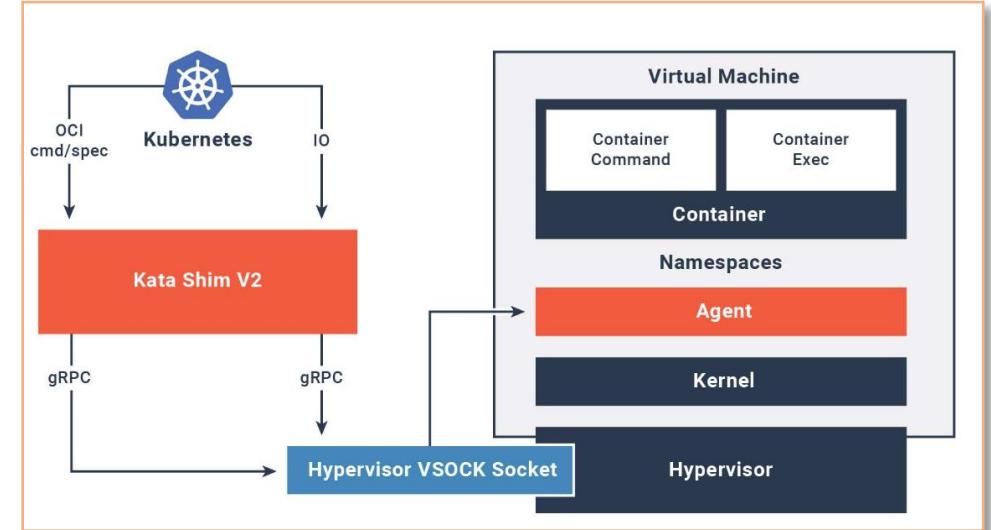
Install Kata Containers (kata-runtime, kata-proxy, kata-shim).
Configure Docker (daemon.json) or containerd.
Run a Kata container (docker run --runtime=kata-runtime).

Kata Containers - Overview

- **Kata Containers** combine the benefits of containers and virtual machines (VMs) to provide enhanced isolation. Here's how to build and run a Kata container on a Linux system.

Why Use Kata Containers?

- ✓ Enhanced security via VM isolation
- ✓ Compatible with Docker and containerd
- ✓ Ideal for sensitive workloads (cloud, edge computing, etc.)



Install Kata Containers

Prerequisites

- A Linux host (Ubuntu, Debian, CentOS, etc.)
- A compatible hypervisor (QEMU/KVM)
- A container runtime (Docker or containerd)

Exemple of Installation on Ubuntu/Debian

```
sudo apt-get update  
sudo apt-get install -y software-properties-common  
sudo add-apt-repository -y ppa:kata-containers/ppa  
sudo apt-get update  
sudo apt-get install -y kata-runtime kata-proxy kata-shim
```

```
kata-runtime --version
```

Configure Docker to Use Kata Containers

If you want to run Kata Containers with Docker, you need to add Kata as an **alternative runtime**.

Modify the Docker configuration file:

```
sudo mkdir -p /etc/docker  
cat <<EOF | sudo tee /etc/docker/daemon.json  
{  
    "runtimes": {  
        "kata-runtime": {  
            "path": "/usr/bin/kata-runtime"  
        }  
    }  
}  
EOF
```

Restart Docker:

```
sudo systemctl restart docker
```

Run a Kata Container with Docker

Once everything is set up, run a Kata container with:

```
docker run --rm -it --runtime=kata-runtime ubuntu bash
```

This will start an **Ubuntu container** but with **enhanced isolation** via a **Kata micro-VM**.

Use Kata with containerd (Optional)

If you are using containerd instead of Docker, edit its configuration file:

```
sudo nano /etc/containerd/config.toml
```

Add Kata as a runtime:

```
[plugins."io.containerd.grpc.v1.cri".containerd.runtimes.kata]
runtime_type = "io.containerd.kata.v2"
```

Restart containerd

```
sudo systemctl restart containerd
```

Then run a container:

```
ctr run --runtime io.containerd.kata.v2 -t
docker.io/library/alpine:latest kata-test sh
```

Part 3 : Running AI Platform Powered by LLM

Data4Citizen Platform

Data4Citizen Architecture

Demo 1 : Data4Citizen & AI-LLM - Dashboard

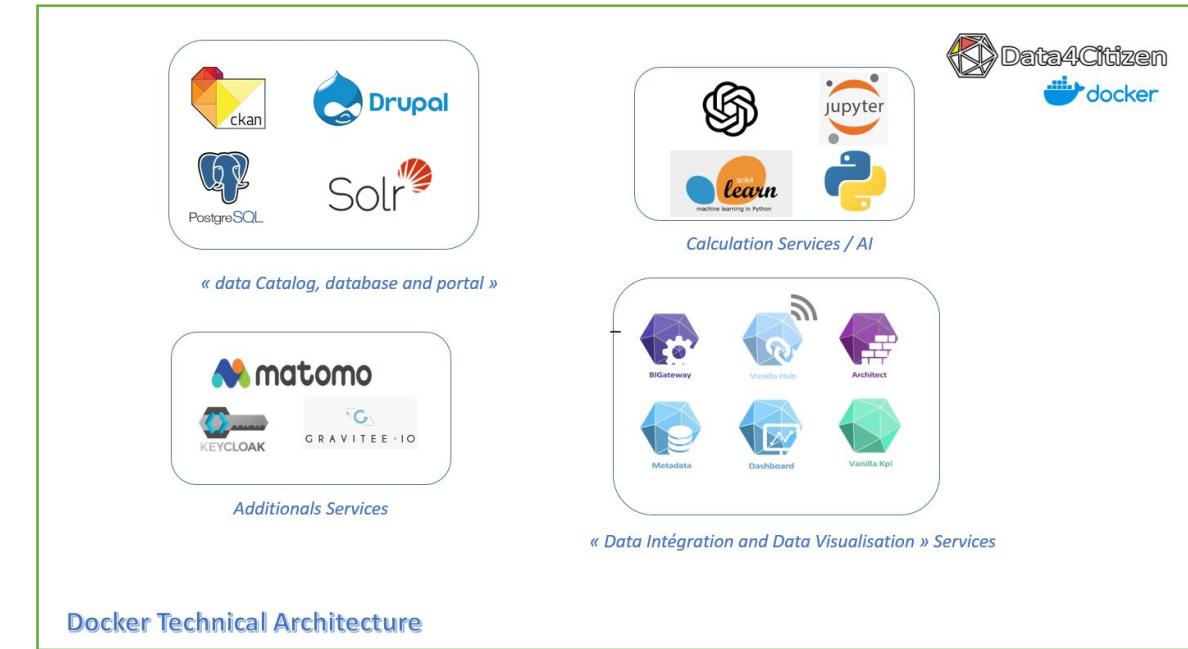
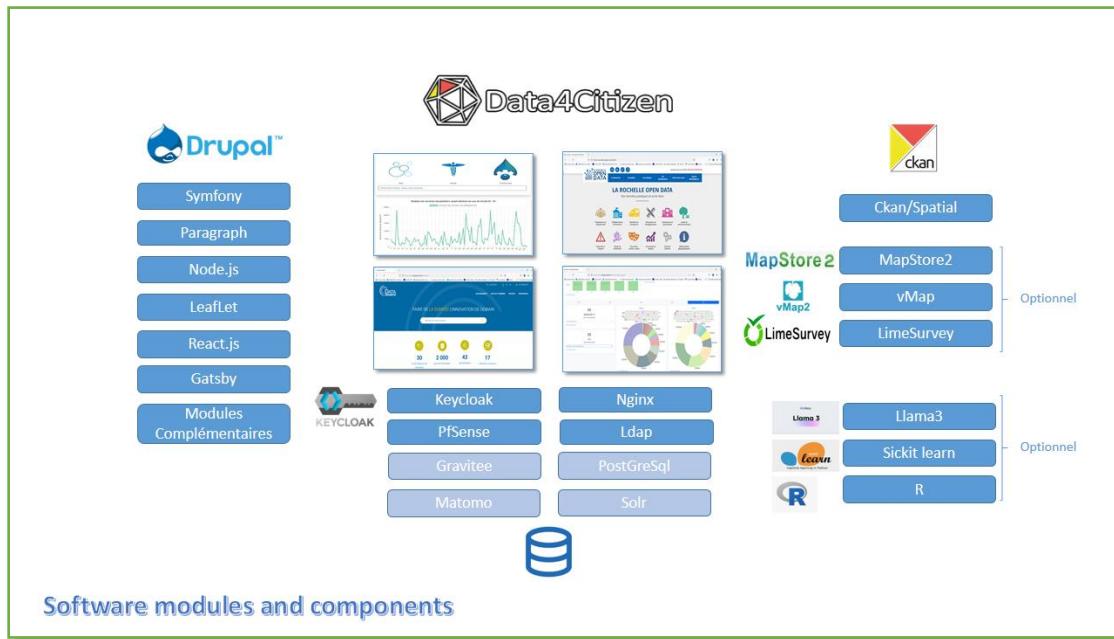
Demo 2 : Data4Citizen & DataMining

Demo 3 : Data4Citizen & DM-LLM

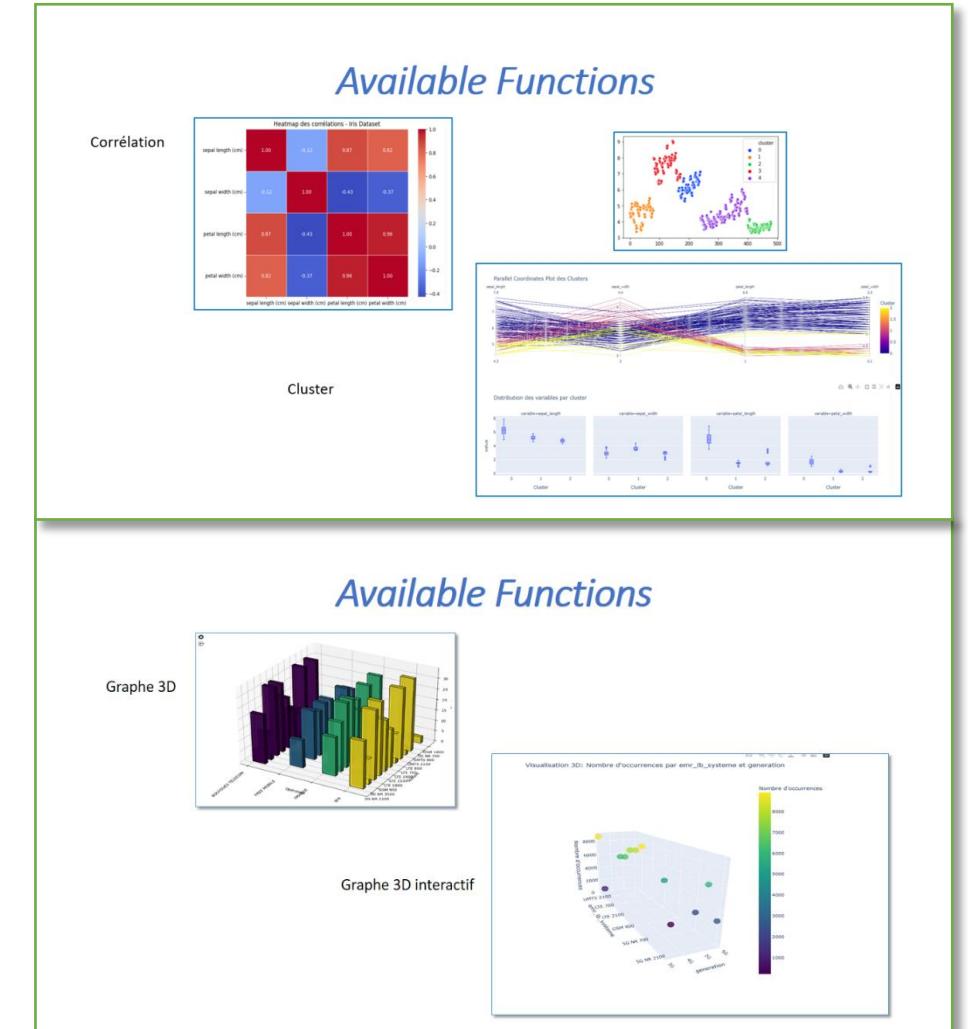
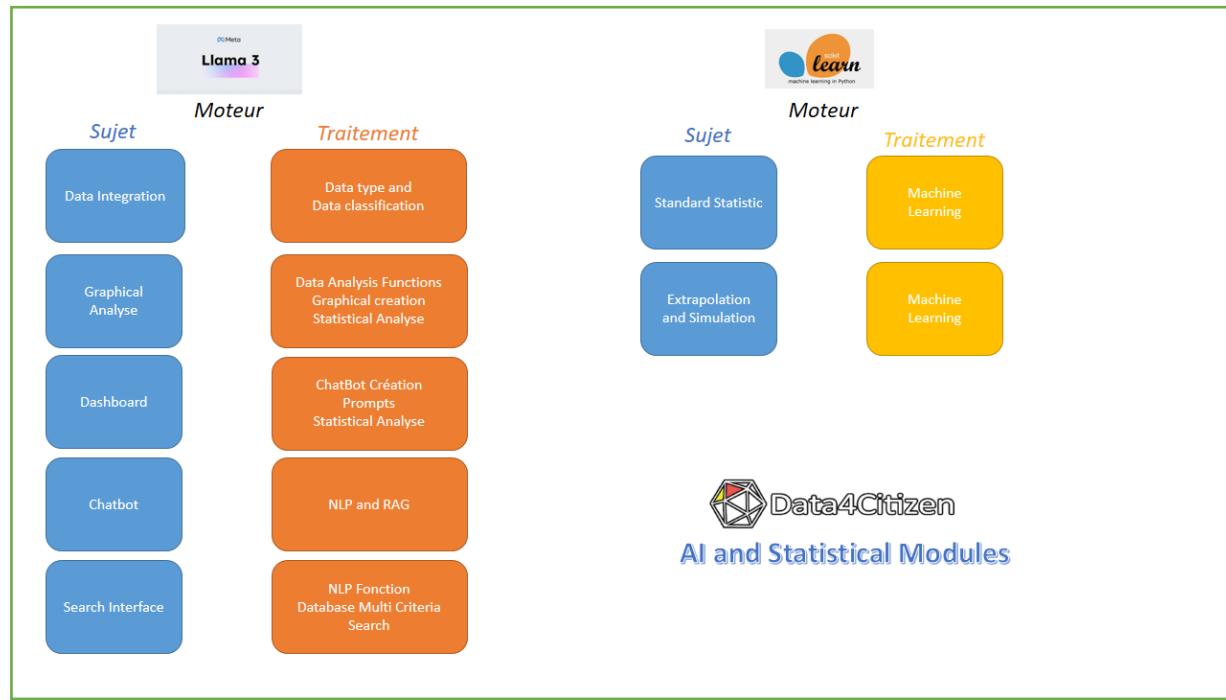
Demo 4 : Data4Citizen & Jupyter

Data4Citizen Platform

Majority of projects are governmental projects : high recommendation to be GDPR and NIS2 compliant



Data4Citizen IA-LLM Architecture



Data4Citizen IA-LLM in Action – Dashboard - 1

Data Integration – Data Analysis and Dashboard Creation

The diagram illustrates the Data4Citizen platform's capabilities for data integration, analysis, and visualization. On the left, a 'Dashboard generation from file' window shows a dataset named '25050303-ds9' and its corresponding CSV file path. Below it is the 'scikit-learn machine learning in Python' logo. The main workspace on the right features a map of the Pyrenees region with various data points marked by colored circles (orange, yellow, red) and numerical values. Two bar charts are also displayed, showing data across different categories like mobile operators (Bouygues Telecom, Free Mobile, Orange, SFR) and network technologies (5G NR 2100, 5G NR 3500, 5G NR 7000, GSM 200, LTE 1200, LTE 2000, LTE 700, LTE 2100, UMTS 2100, UMTS 900). The sidebar on the right lists various content types: Contenu, Image, Graphique, Carte, Indicateur, Données, Document, HTML, Slider, AI Analysis, and Python.

Data4Citizen IA-LLM in Action – Dashboard - 2

Data Analysis inside Dashboard using LLM

The screenshot illustrates the Data4Citizen dashboard's AI integration. On the left, the 'Assistant AI' panel features a toolbar with icons for adding filters, content, graphs, KPIs, and an AI analysis, with the 'Ajouter une analyse IA' button highlighted by a green box. Below it is a text input field labeled 'Type your prompt here'. A blue arrow points from this panel to the central chart area. The central area contains two bar charts comparing mobile network technologies across operators. The top chart shows the number of accounts for 2G, 3G, 4G, and 5G technologies. The bottom chart shows the number of accounts for various LTE and 5G NR bands. To the right, a detailed analysis section provides a summary of the data and specific insights for each operator. A green arrow points from the top chart to this analysis section. At the bottom, a map displays the geographical distribution of network coverage or account density across a region, with data points overlaid on the terrain.

Assistant AI

- Ajouter un filtre
- Ajouter un contenu
- Ajouter un graphique
- Ajouter une carte
- Ajouter un KPI
- Ajouter une analyse IA

Type your prompt here

AI ANALYSIS

Le graphique représente les données de comptes pour quatre technologies de réseau mobile : 2G, 3G, 4G et 5G, fournies par les opérateurs BOUYGUES TELECOM, ORANGE, SFR, et FREE MOBILE.

En analysant les données, on observe que ORANGE se positionne comme le leader sur les trois premières générations de réseau (2G, 3G, 4G), avec un nombre de comptes de 88 pour la 2G, 167 pour la 3G et 491 pour la 4G. Cependant, la transition vers la 5G montre une baisse notable du nombre de comptes (78), ce qui pourrait indiquer une hésitation des utilisateurs à adopter cette nouvelle technologie.

BOUYGUES TELECOM suit avec des valeurs respectives de 59, 157 et 395 pour les 2G, 3G, et 4G, et une diminution similaire pour la 5G qui atteint 63. Bien que moins dominant qu'ORANGE, BOUYGUES montre une performance relativement solide dans les réseaux établis.

SFR affiche des performances légèrement inférieures avec 64, 161 et 420 pour les générations précédentes, et une baisse à 60 pour la 5G, ce qui révèle une tendance similaire à celle observée chez BOUYGUES et ORANGE.

MAP

20 km
10 m

Parc national des Pyrénées (Arie d'assistan)

Parc naturel régional des Pyrénées catalanes

Parc naturel régional des Corbières

Parc naturel régional de la Narbonnaise en Méditerranée

Parc naturel marin du Golfe du Lion

Lesotho

Tarbes

Saint-Gaudens

Vielha

Carcassonne

Lézignan-Corbières

Narbonne

Limoux

Perpignan

Saint-Cyprien

Parc naturel régional des Pyrénées ariégeoises

Parc naturel régional Corbières-Tenouïdes

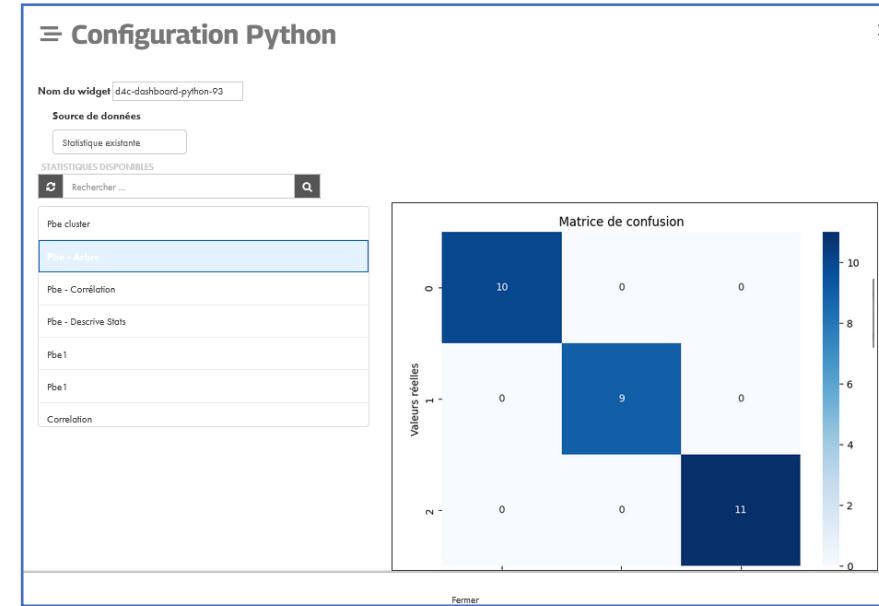
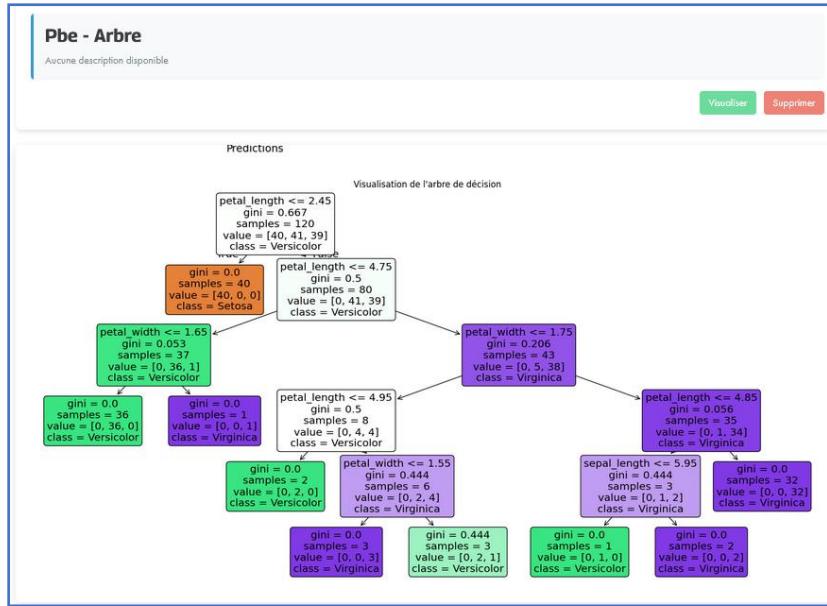
Parc naturel régional de la Narbonnaise en Méditerranée

Golf du Lion

Live Demo

Data4Citizen Just DM in Action – Iris Dataset

DataSet – Sickit-learn Analysis



Live Demo

Data4Citizen DM-LLM in Action – Chat AI - 1

DataSet – Chat AI using Python, Sickit-learn and LLM

retour aux résultats de recherche f in e

20250302-ds09

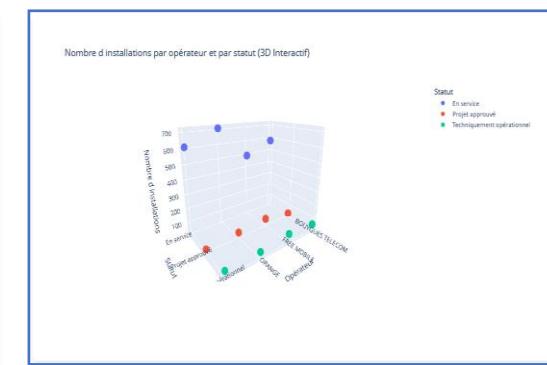
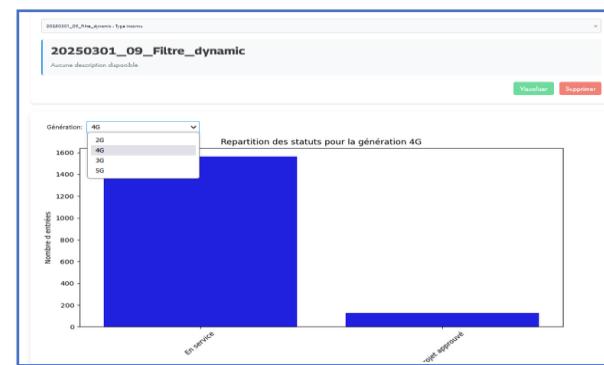
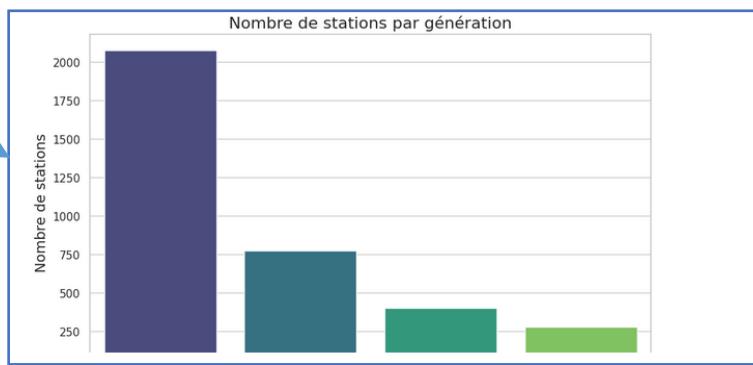
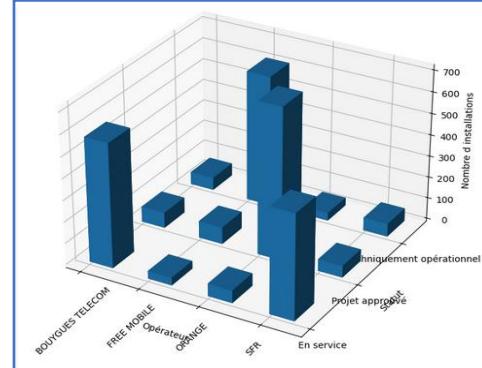
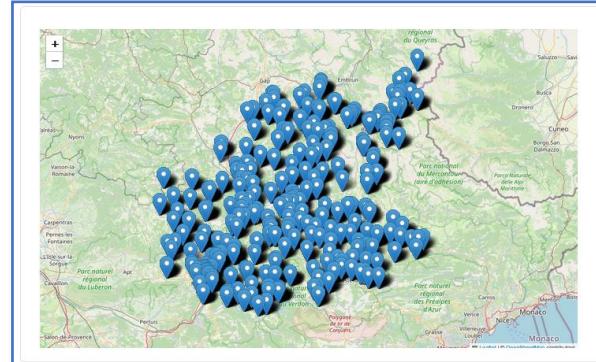
Informations Tableau Carte Analyse Exploitations Export Réalisations Administration

Administration

- Éditer les métadonnées et les ressources
- Configuration de la ressource
- Transformation du jeu de données
- Modifier directement les données [ajout de lignes, modifier valeurs]

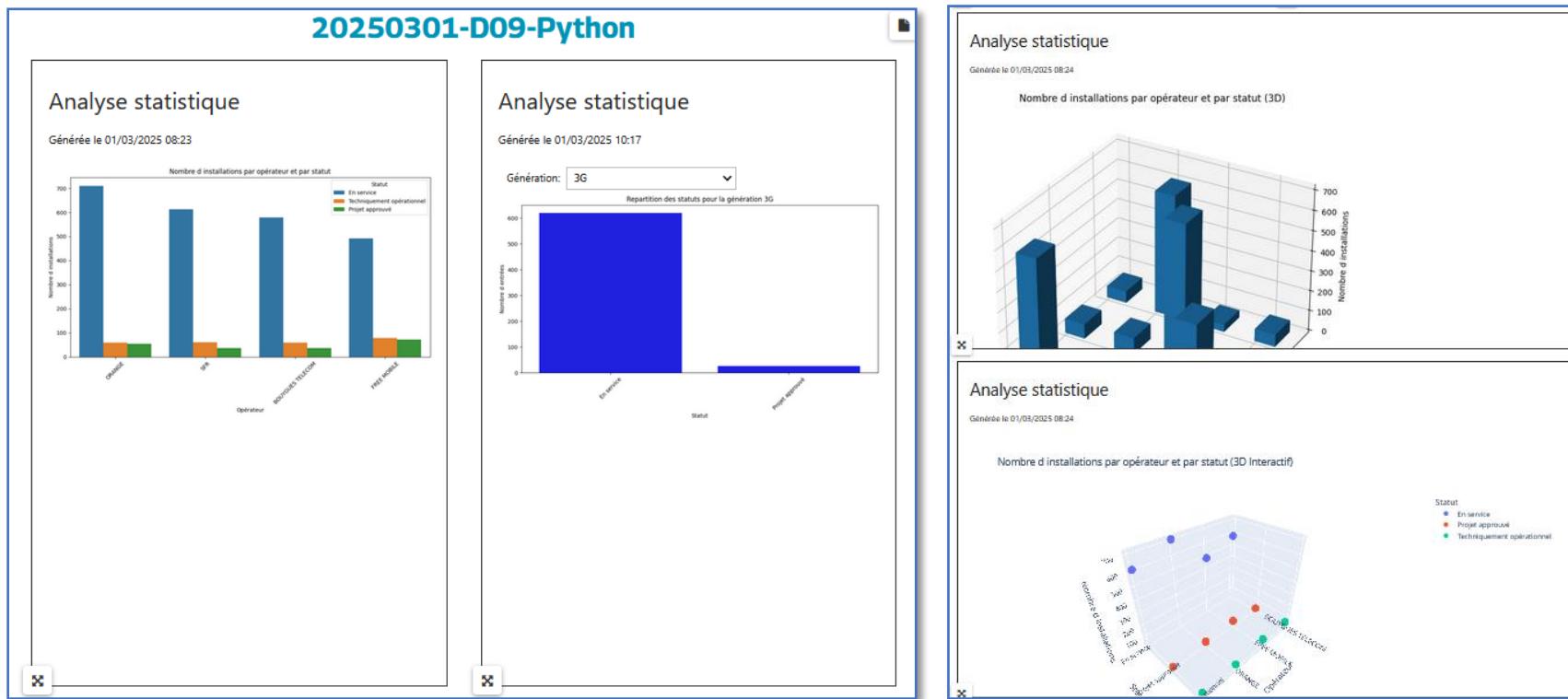
Gestion du jeu de données

- Extrapolation du jeu de données
- Simulation sur le jeu de données
- Gestion des analyses IA



Data4Citizen DM-LLM in Action – Chat AI - 1

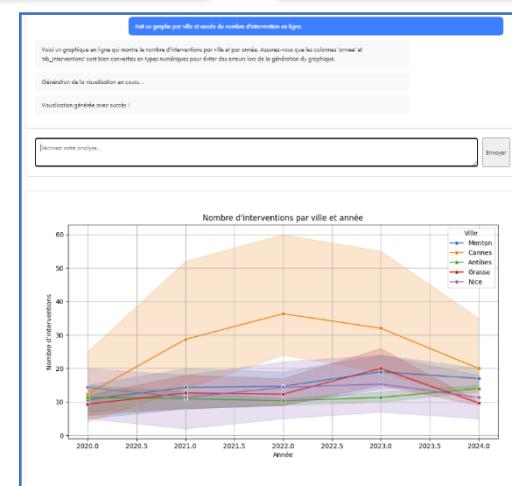
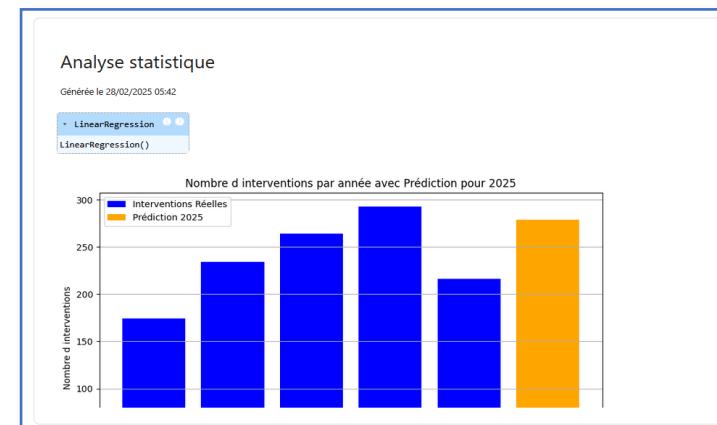
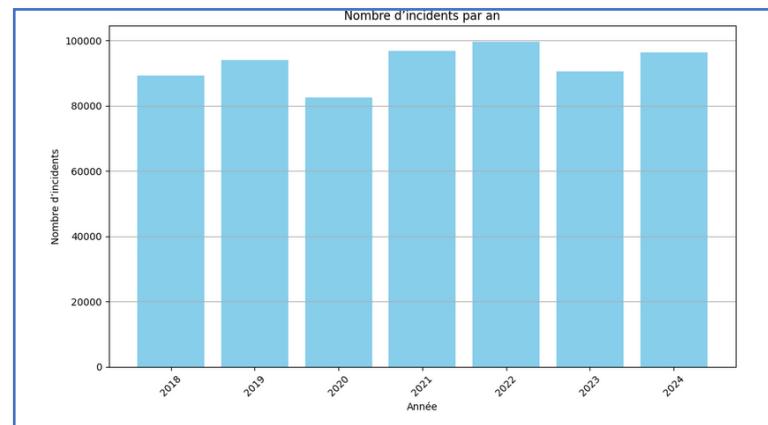
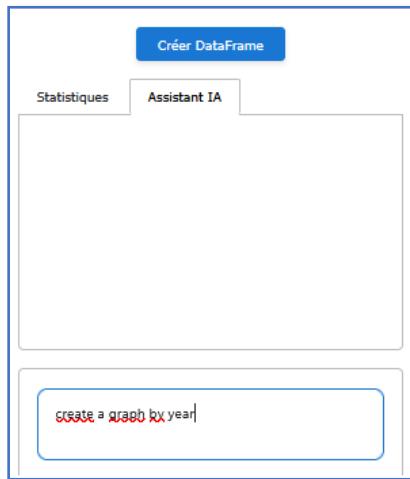
When AI-Analysis turns back inside a Dashboard



Live Demo

Data4Citizen DM-LLM in Action – Jupyter

Jupyter – Chat AI using Python, Sickit-learn and LLM



Data4Citizen DM-LLM in Action – Jupyter

When Notebook turns back inside a Dashboard

The image shows a screenshot of the Data4Citizen DM-LLM dashboard interface. On the left, a modal window titled "Configuration Python" is open, displaying a list of notebooks and a preview of a 3D bar chart titled "Analyse statistique". The main dashboard on the right features a sidebar with categories: Contenu, Image, Graphique, Carte, Indicateur, Données, Document, HTML, Slider, AI Analysis, and Python. The "Python" section contains two items: a 3D bar chart titled "Python configuration" and a map titled "Nombre d installations par opérateur". The map includes a legend for "Statut" with three categories: "En service" (blue dot), "Projet approuvé" (red dot), and "Techniquement opérationnel" (green dot). A bottom navigation bar includes links for "Leaflet", "Map data © OpenStreetMap contributors", and "Live Demo".

Live Demo

