

# ESUP-DAYS **38**

---

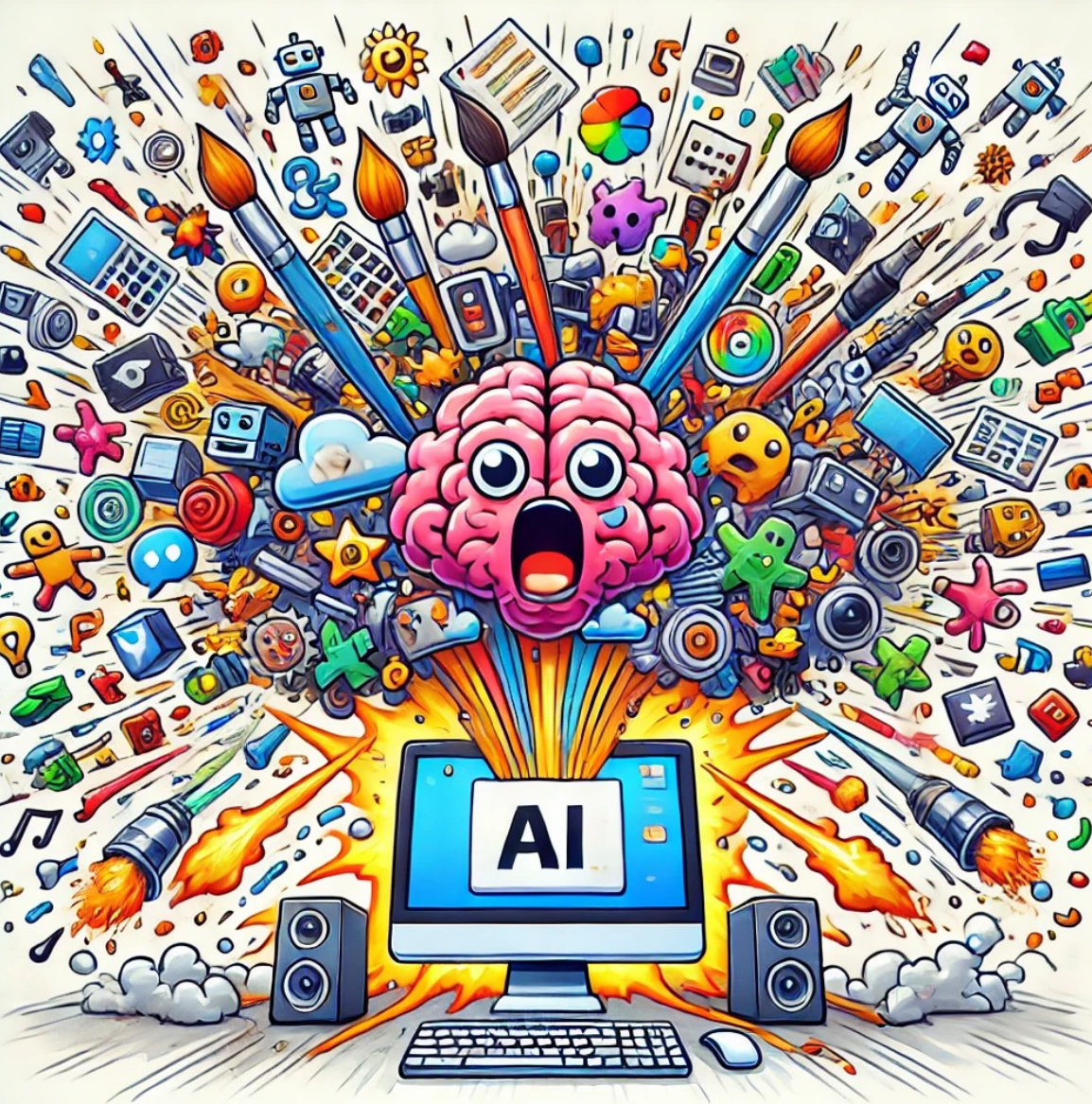
## ATELIER IA



Nicolas TRUCHAUD Université Jean Monnet – Saint Étienne

---

# CONTEXTE




# EXPLOSION DES IA GÉNÉRATIVES

Image : Dall-E

# DES ATTENTES FORTES DE SOLUTIONS SOUVERAINES ADAPTÉES AU CONTEXTE ACADÉMIQUE

Image : Dall-E



A person in a dark coat is walking away from the viewer on a long, narrow wooden pier that stretches into a vast, misty sea. The sky is a soft gradient of light blue and yellow, suggesting a sunrise or sunset. The overall mood is contemplative and serene.

# RÉPONSE PAR DES INITIATIVES LOCALES PLUS QUE PAR DES ÉQUIPES DÉDIÉES

Image : Office365

LE RISQUE : LAISSER  
LES UTILISATEURS  
ALLER VERS SOLUTIONS  
NON SOUVERAINES  
NON SÉCURISÉES

Image : Aria (Opera)





**IDÉE : FÉDÉRER LES  
INITIATIVES EN  
RASSEMBLANT LES  
COMPÉTENCES  
EXISTANTES**

Image : Office365

---

# OBJECTIFS



## DÉCLARATION D'INTENTION

- ◆ L'objectif de ce groupe est de rassembler les initiatives existantes autour des IA souveraines et open source, afin de favoriser la montée en compétences de tous et de concevoir des modèles d'infrastructures adaptables et personnalisables.
- ◆ Ces modèles permettront aux établissements d'enseignement supérieur de proposer à leurs utilisateurs des services d'IA exploitant leurs données de manière confidentielle et respectueuse de la vie privée.

## DÉCLINAISON OPÉRATIONNELLE

---

1. **Recenser les acteurs et les initiatives ;**
2. **Encourager l'acquisition et le renforcement des compétences techniques ;**
3. **Favoriser la collaboration entre établissements ;**
4. **Conduire une réflexion technique sur la souveraineté numérique, la protection des données, et les implications légales pour les universités ;**
5. **Esquisser les grandes lignes d'un framework open source déployable à grande échelle.**

---

# RESSOURCES

---

## RESSOURCES DE L'ATELIER

- ◆ Un canal dédié sur le rocketChat Esup : <https://rocket.esup-portail.org/channel/GT-IA>
- ◆ Un wiki : <https://www.esup-portail.org/wiki/x/BgDLVg>
- ◆ Une liste de discussion : [myia@groupe.renater.fr](mailto:myia@groupe.renater.fr)  
Pour s'abonner : <https://groupe.renater.fr/sympa/subscribe/myia>

Retrouvez tous ces liens sur la page de l'atelier sur le portail Esup :  
<https://www.esup-portail.org/atelier-ia>

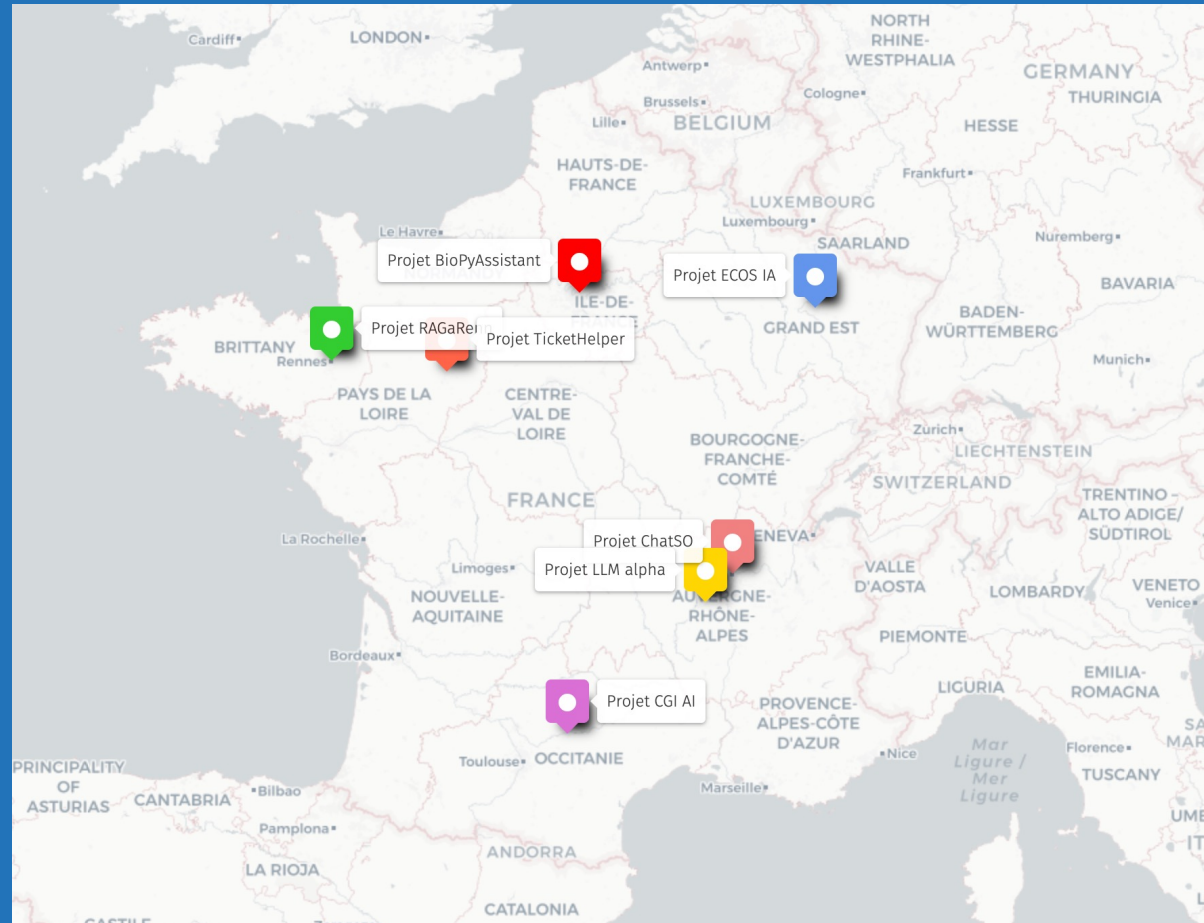


# EXEMPLES D'INITIATIVES



Carte : OpenStreetMap

[https://umap.openstreetmap.fr/fr/map/initiatives-esup-ia\\_1134449#6/47.130/2.219](https://umap.openstreetmap.fr/fr/map/initiatives-esup-ia_1134449#6/47.130/2.219)



---

## RÉPERTOIRE DES INITIATIVES

- ◆ **Projet BioPyAssistant - Université Paris Cité**
- ◆ **Projet CGI AI - École des Mines d'Albi**
- ◆ **Projet ChatSO - Réseaux des ateliers de la donnée**
- ◆ **Projet ECOS IA - Université de Lorraine**
- ◆ **Projet LLM alpha - Université Jean Monnet Saint Etienne**
- ◆ **Projet RAGaRenn - Université de Rennes**
- ◆ **Projet Ilaas - Université de Rennes**
- ◆ **Projet TicketHelper - Le Mans Université**

## PROJET BIOPYASSISTANT – UNIVERSITÉ PARIS CITÉ – PIERRE POULAIN

### Projet d'assistant conversationnel :

- ◆ Dédié à l'apprentissage de la programmation Python pour des étudiant·e·s biologistes.
- ◆ Disponible 24h24 et 7j/7 pour répondre aux questions les plus simples quand l'enseignant·e n'est pas disponible.
- ◆ Basé sur un cours existant en ligne, mais dense (385 pages dans la version PDF).
- ◆ Capable de renvoyer l'étudiant·e vers les sections du cours (en ligne) pertinentes par rapport à sa question.

### Technos :

- LangChain
- ChromaDB
- Gradio (interface web chatbot)

Statut : PROD

### Accès :

<https://biopyassistant.sdv.u-paris.fr/>

N'hésite-pas à le solliciter en lui posant tes questions.

Discuter avec le cours

Discuter avec le cours (battle)

Pose une question, compare les réponses fournies par deux modèles de langages tirés au hasard et vote pour la meilleure réponse.

Modèle A

Bonjour, comment puis-je t'aider ?

Comment afficher un float avec 2 chiffres avec la virgule ?

Pour afficher un float avec 2 chiffres après la virgule en Python, vous pouvez utiliser le formatage avec les *f-strings*. Voici comment vous pouvez le faire :

```
nombre = 0.4780405405405405  
print(f"{nombre:.2f}")
```

Cela affichera le float avec deux chiffres après la virgule, c'est-à-dire "0.48". Le format `{nombre:.2f}` indique que vous souhaitez afficher la variable `nombre` sous forme de float (`f`) avec une précision de deux décimales (`.2`).

Modèle B

Bonjour, comment puis-je t'aider ?

Comment afficher un float avec 2 chiffres avec la virgule ?

Pour afficher un float avec 2 chiffres après la virgule en Python, vous pouvez utiliser la méthode des *f-strings* et spécifier le format désiré. Voici comment procéder :

- Définissez votre variable float.
- Utilisez une *f-string* pour formater l'affichage de cette variable.

Dans le format spécifié, `.2f` indique que vous souhaitez deux chiffres après la virgule. Voici un exemple :

```
prop_GC = 0.4780405405405405
```

La réponse A est meilleure   Les deux réponses se valent   Les deux réponses sont mauvaises   La réponse B est meilleure

Pose-moi une question sur le cours.

Exemples de questions

Quelle est la différence entre une liste et un set ?   Comment écrire une boucle ?   Comment afficher un float avec 2 chiffres avec la virgule ?



## Fournir plusieurs services spécifiques :

- ◆ Outils de traduction et reformulation, (notamment traduction de PPTX)
- ◆ Discussion avec des PDFs (et résumé)
- ◆ Extraction de concepts depuis un cours PDF, construction de quizzes sur un cours donné en PDF
- ◆ Discussion avec des résultats de recherche dans HAL,

## Technos :

- LLM : Ollama
- RAG : LangChain/Langgraph
- BAML / Instruct
- PPTX translator : Clien Ollama
- HAL : Langchain RAG
- Frontend : NextJS/React

Statut : POC FONCTIONNEL

## Accès :

<https://cgi-ai.mines-albi.fr/> (auth)

## PROJET CHATSO - RÉSEAUX DES ATELIERS DE LA DONNÉE

**Fournir un chatbot simple, léger et adaptable, capable de répondre aux questions spécifiques d'un domaine tout en citant ses sources.**

Objectifs :

- ◆ Développer un chatbot spécialisé en Science Ouverte
- ◆ Offrir une solution simple à installer et peu gourmande en ressources
- ◆ Garantir la transparence et la fiabilité des informations via la citation des sources
- ◆ Créer un outil adaptable à d'autres domaines que la Science Ouverte

Technos :

- Serv inférence : Ollama
- Vectorisation : HuggingFaceEmbedding
- Création d'APIs : FastAPI
- DB : ChromaDB
- Front : VueJS 3 et Vuetify 3

Statut : version alpha

Accès :

<https://sbgvm-151-69.in2p3.fr/>

 Chat

 À propos

 Version

CH

quelles sont les données de la recherche dont tu disposes en sciences humaines ?

RDG

Bonjour ! En tant que chatbot expert en science ouverte, je suis ravi de discuter avec toi sur ce sujet passionnant. Les données de la recherche en sciences humaines sont extrêmement variées et peuvent concerner des domaines tels que l'histoire, les langues, les cultures, les sociétés, les psychologies, les éducations, etc. Voici quelques exemples de types de données que l'on peut trouver dans les sciences humaines : \* Les données textuelles : articles, livres, documents historiques, témoignages oraux, etc. \* Les données quantitatives : résultats d'enquêtes, statistiques, données démographiques, etc. \* Les données qualitatives : observations, entretiens, focus groups, etc. \* Les données multimédias : images, vidéos, sons, etc. Il est important de noter que les données en sciences humaines peuvent être sensibles et nécessiter une attention particulière pour protéger la confidentialité des participants ou des sources. Cependant, il est également essentiel de partager ces données pour favoriser l'apprentissage, la collaboration et le progrès scientifique. Pour en savoir plus sur les entrepôts de confiance pour le partage des données, je te recommande de consulter le lien suivant : <https://recherche.data.gouv.fr/fr/entrepots>.

### Évaluation de l'IA

Pourriez-vous attribuer une note, de 1 à 5, à la réponse fournie par notre IA ?

0

1

2

3

4

5

ENVOYER

Envoyer un message

quelles sont les données de la recherche dont tu disposes en sciences humaines ?

## PROJET ECOS IA - UNIVERSITÉ DE LORRAINE - OLIVIER ZILLER

### **Fournir aux étudiants un outil d'entraînement aux ECOS (Examens Cliniques Objectifs Structurés)**

#### Objectifs :

- ◆ IA simule un patient virtuel
- ◆ l'étudiant échange via un chat textuel
- ◆ puis analyse de cet échange en tenant compte de la grille d'évaluation de la station
- ◆ Suite : interaction orale de l'étudiant avec le patient virtuel

#### Technos :

- LLM : Chat 4-o
- RAG : Langchain
- DB : Supabase
- Front : Angular

Statut : MVP prévue fin 2024

## PROJET LLM ALPHA - UNIVERSITÉ JEAN MONNET SAINT ETIENNE – DAVID BOIT

### Explorer les possibilités liées aux LLM hébergés en local et leur intégration aux outils de messagerie instantanée existants (RocketChat)

#### Objectifs :

- ◆ Explorer les possibilités liées aux LLM hébergés en local
- ◆ Se positionner par rapport aux offres extérieures des GAFAM
- ◆ Etudier l'intégration à RocketChat

#### Technos :

- Serv inférence : Ollama
- LLM : codellama, llama3, llava, mistral, mixtral
- Front : Open WebUI

Statut : POC fonctionnel

**OI** Nouveau chat

**Fichiers de modèle**

Prompts

Documents

Recherche

LLM Base Services Offered

"Pole Exploitation Services"

"Compile Programming Insights Ne"

"User Missing from POD"

"Roman Empire Secrets"

## Sélectionner un modèle

Définir par défaut

**OI**

# Bonjour, Nicolas Truchaud

## Comment puis-je vous aider aujourd'hui ?

Tell me a fun fact about the Roman Empire

Show me a code snippet of a website's sticky header

Help me study vocabulary for a college entrance exam

Give me ideas for what to do with my

+ Envoyer un message

Les LLMs peuvent faire des erreurs. Vérifiez les informations impor

**NT** Nicolas Truchaud

## Sélectionner un modèle

Search a model

- codellama:7b-code-q4\_K\_M 7B ⓘ
- codellama:latest 7B ⓘ
- llama3:latest 8B ⓘ
- llava:latest 7B ⓘ
- mistral:latest 7B ⓘ
- mixtral:latest 47B ⓘ
- nomic-embed-text:latest 137M ⓘ

## PROJET RAGARENN - UNIVERSITÉ DE RENNES – OLIVIER WONG HEE KAM

**Plateforme permettant à un utilisateur d'exploiter en autonomie ses propres documents ou du contenu public (via URL), en choisissant parmi plusieurs LLM tout en maîtrisant leur paramétrage**

Objectifs :

- ◆ Tester une solution d'IA générative la plus sécurisée (voire souveraine) possible
- ◆ Expérimenter le déploiement d'une infrastructure locale pour en savoir plus sur les aspects techniques, budgétaires et d'impact environnemental
- ◆ Mener une réflexion sur les questions d'opportunités et de en s'appuyant sur des mesures statistiques sur des usages réels et des recueils auprès d'utilisateurs (enquêtes, ateliers participatifs)

Technos :

- Serveurs d'inférence : Ollama et vLLM
- RAG : Langchain
- Front : Open WebUI adapté
- LLM : Huggingface et Ollama

Statut : PROD

Accès :

<https://ragarenn.eskemm-numerique.fr>

## FÉDÉRATION ESR DE SERVEURS MUTUALISÉS POUR L'INFÉRENCE LLM « AS A SERVICE » (ILAAS)

### ◆ **Contexte : mutualisation au « juste nécessaire »**

- ◆ Enjeux : confidentialité, souveraineté, soutenabilité, impact environnemental
- ◆ Infrastructures ESR existantes

### ◆ **Objectifs**

- ◆ Fédérer des serveurs d'inférence avec Aristote-dispatcher : résilience, optimisation (pics, temps de réponse)
- ◆ Utiliser la fédération via API : services « consommateurs » (Crisalid + qui veut bien tester)

### ◆ **Phasage**

- ◆ POC : fédération initiale (Universités de Rennes, Lille, Paris 1 Panthéon-Sorbonne, Reims Champagne-Ardenne et CentraleSupélec Paris-Saclay)
- ◆ Extension : industrialisation (distribution logicielle), tableaux de bord (MCO, impacts environnementaux), modèle économique

Pour plus d'informations : [olivier.wong-hee-kam@univ-rennes.fr](mailto:olivier.wong-hee-kam@univ-rennes.fr)



## PROJET TICKETHELPER - LE MANS UNIVERSITÉ – PIERRE CARTIER

### Proposer des suggestions de réponses aux techniciens en charge des tickets DNUM/DSI

#### Objectifs :

- ◆ Affiner les prédictions du LLM en utilisant la base de connaissance technique (FAQ) interne
- ◆ Proposer automatiquement une réponse à l'utilisateur (en mentionnant clairement « Réponse générée par IA »)
- ◆ Projet de se brancher sur tout l'historique de ticket après offuscation des données sensibles

#### Technos :

- Inférence : Ollama
- LLM : Gemma2 et Ministral
- RAG : Langchain
- Source : CSV extraits de GLPI

Présentation aux JRES 2024 à Rennes en décembre

---

# À SUIVRE

## QUELQUES IDÉES

- ◆ Des webinaires « on ouvre le capot » ?
- ◆ Des webinaires thématiques sur une technique : le RAG, la fédération de serveurs d'inférence, ...
- ◆ Des formations

... vos idées et suggestions