



RÉPUBLIQUE
FRANÇAISE **CNIG**

Liberté
Égalité
Fraternité

Conseil national
de l'information
géolocalisée



Intelligence artificielle & informations géolocalisées

Atelier – 8 janvier 2026

Atelier 3

Les perspectives ouvertes par les LLM pour faciliter l'accès aux données
(Exploitation des LLM pour requêter les jeux de données via langage naturel)

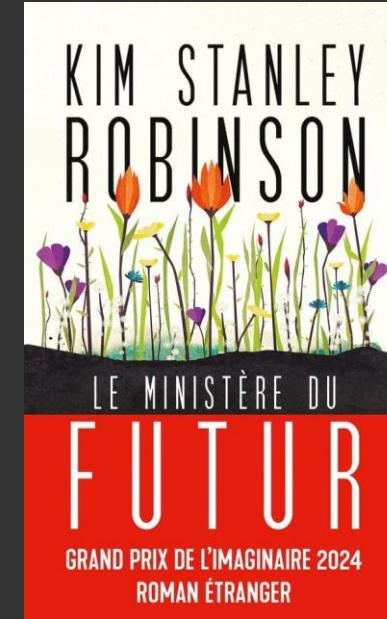
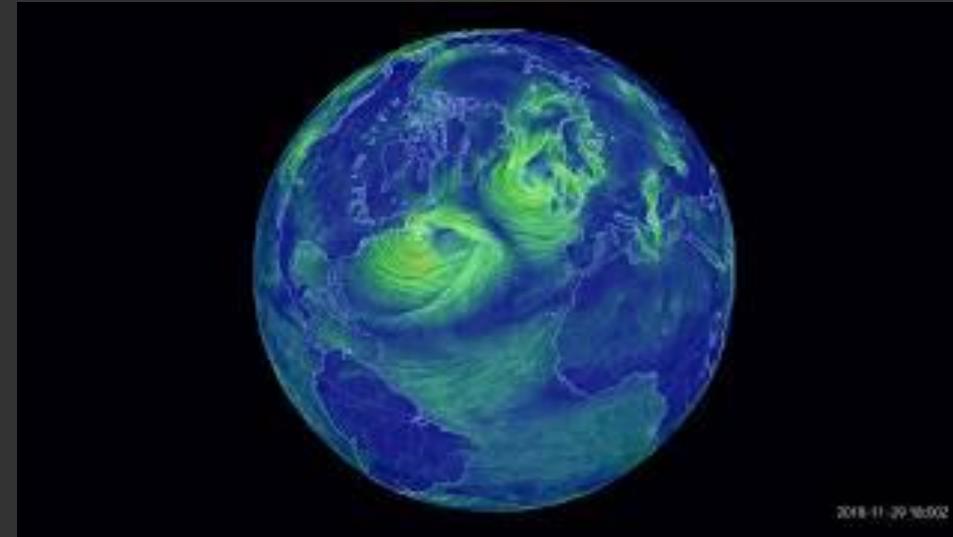
Florent Gravin, Camp to camp

Marc Poupée, Geodata Paris

Gaëtan Lavenu, ESRI

Modèles génératifs et IG





Objectifs de la séance

Les IAs génératives

Comprendre ce que sont ces nouvelles IAs génératives et pourquoi elles fascinent tant.



Les limitations

En déduire un ensemble de biais qui proviennent des modèles mais aussi des façons de les utiliser.



De nouveaux usages

Comment exploiter sainement ces nouveaux outils de façon générale ? Et spécifiquement pour l'information géographique.

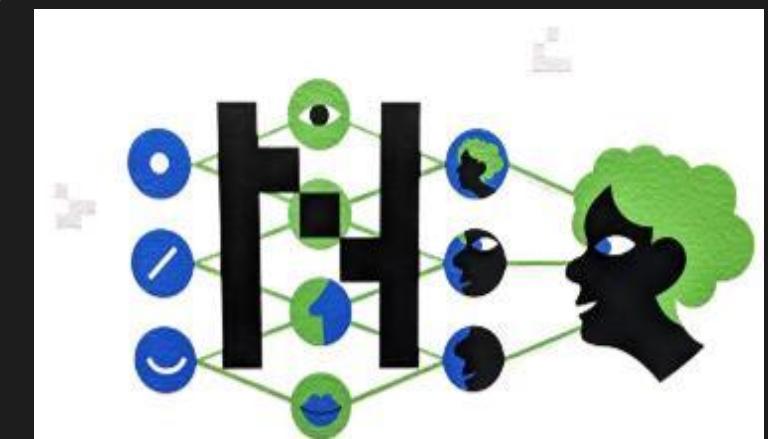
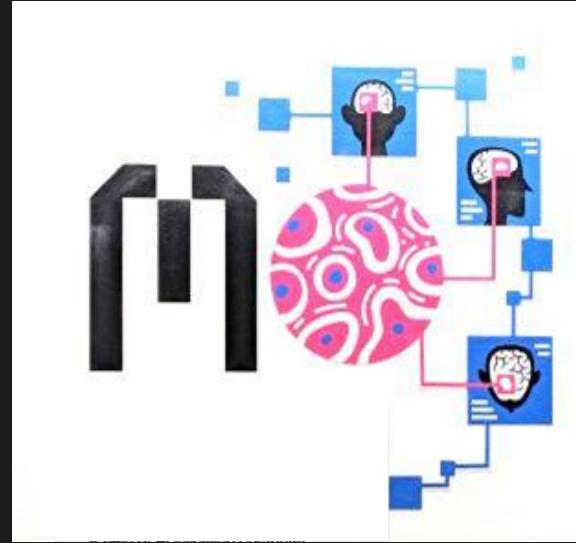


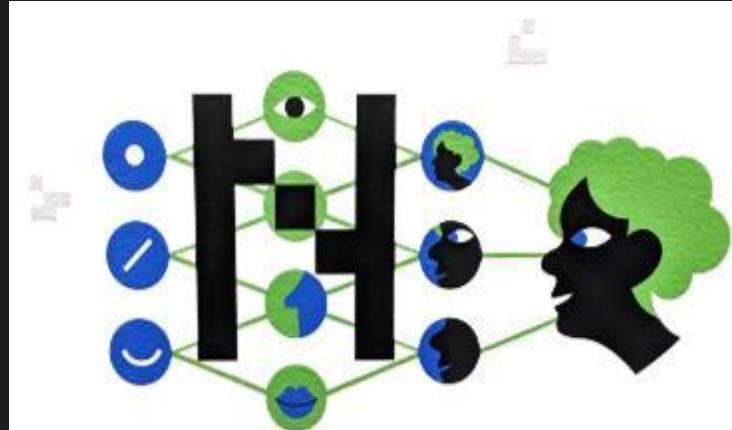
Les IAs génératives





Symbolique ou statistique





Un chat ...

noir
dort
est
blanc
mange





Dans le salon, un chat, un colis et une plante verte attendent patiemment...

Qui donc dort sur le canapé?



Dans le salon, un chat, un colis et une plante verte attendent patiemment...

Qui donc **dort** sur le canapé?

Query



Key : capacité de dormir

Dans le salon, un **chat**, un **colis** et une **plante** verte attendent patiemment...

Qui donc **dort** sur le canapé?

Query



Key : capacité de dormir

animal
objet
plante

x

dort
Query

=

fort
faible
moyen

La value “chat” sera la solution



Hugging Face

Humains
dédiés

Modèle de Fondation

Apprentissage
auto-supervisé

Apprentissage
supervisé

Paramétrage par
raisonnement

Corpus +/- internet

Paramétrage fin
par préférence

Raisonnements
connus

Utilisateurs



Deep Seek

Pour aller plus loin : David Louapre et sa chaîne Science étonnante



Les fameux biais

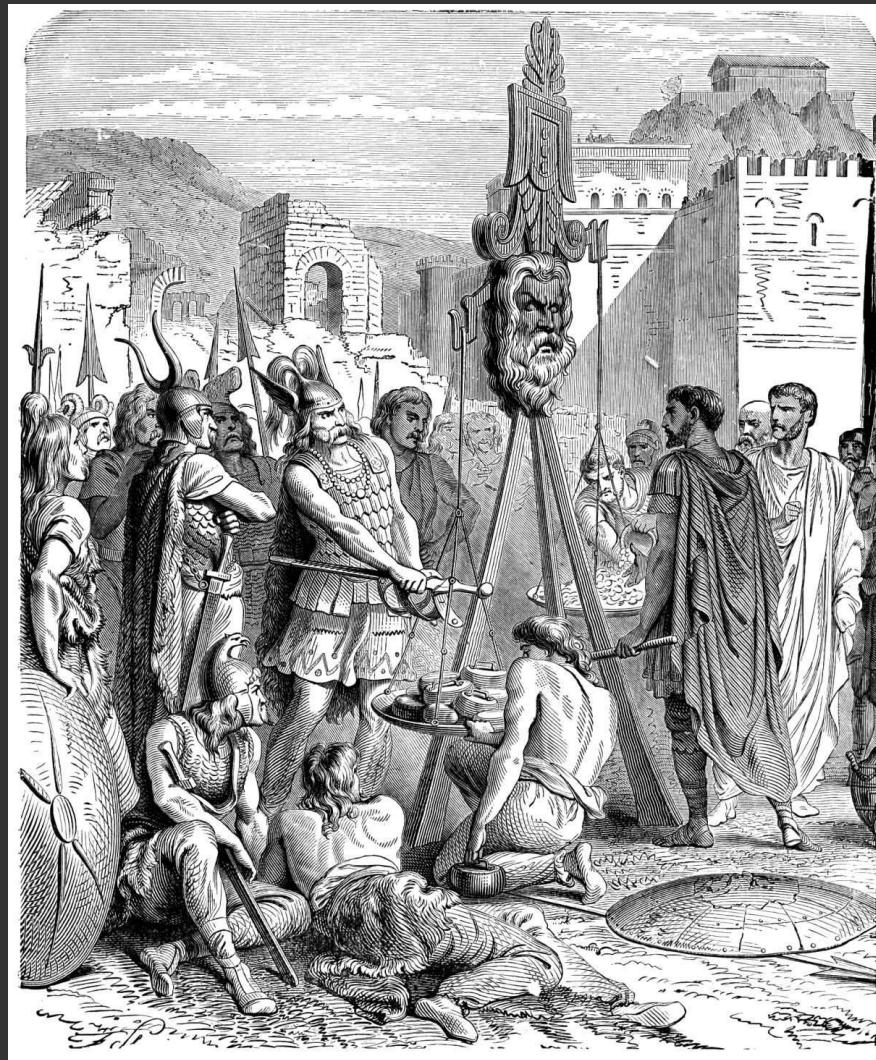




Des Biais... par l'effet Brenna !



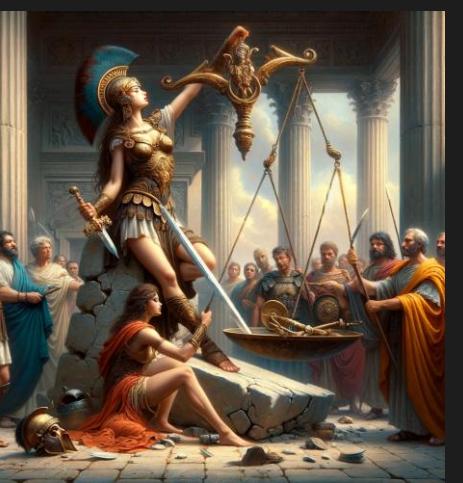
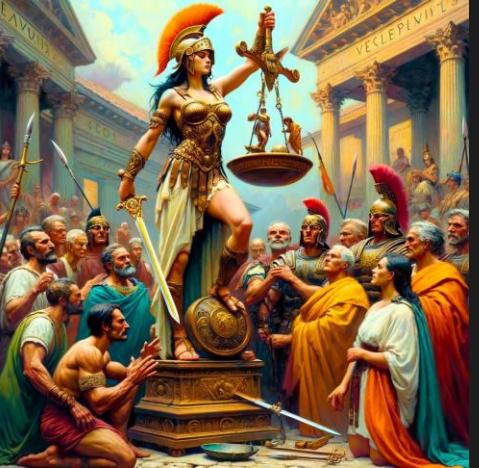
Jean Philippe Jaworsky



Par Paul Lehugeur (1854-1916) — "Histoire de France en cent tableaux" by Paul Lehugeur, Paris, 1886, Domaine public, <https://commons.wikimedia.org/w/index.php?curid=12381924>

Prompt :

A historically inspired fictional scene in the style of Eugène Delacroix, known for his dramatic and emotional compositions with bold, vibrant colors. Set in ancient Rome, it features a female Gaulish warrior, Brenna, placing her sword on a large traditional balance, symbolizing the 'Vae victis' moment. Roman senators in historical attire are presenting a tribute to her. Surrounding them are other Gaulish warriors, depicted in vivid colors and dramatic expressions, facing the senators. The background includes detailed Roman architecture, capturing the essence of Delacroix's historical and exotic themes.





Mistral vs ChatGpt
03 - 2025

Exemple : Est-il possible de segmenter par Mean-Shift une image avec les fonctions d'Orpheo Tool Box, si oui quelle est la ligne de commande ?

Oui, il est possible de segmenter une image en utilisant l'algorithme Mean-Shift avec l'Orfeo ToolBox (OTB). L'outil approprié dans OTB pour effectuer cette segmentation est `otbcli_MeanShiftSegmentation`.

Voici la ligne de commande pour exécuter une segmentation Mean-Shift avec OTB :

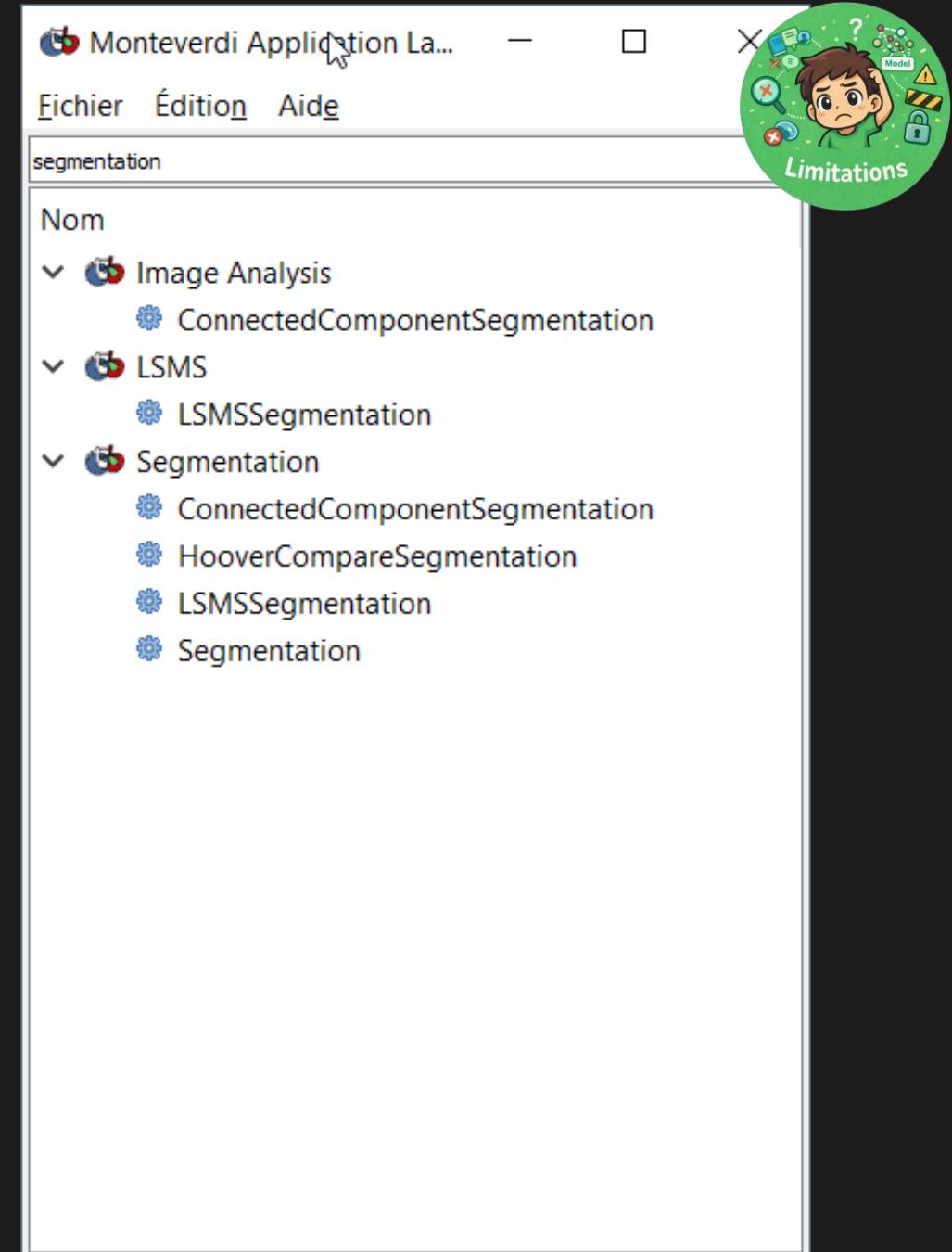
```
bash                                     Copier   Modifier
otbcli_MeanShiftSegmentation -in image.tif -spatialr 5 -ranger 15 -minsize 100 -mode raster -
```

Explication des paramètres :

- `-in image.tif` : Spécifie l'image d'entrée.
- `-spatialr 5` : Définit le rayon spatial du noyau Mean-Shift.
- `-ranger 15` : Définit le rayon dans l'espace des couleurs.
- `-minsize 100` : Fixe la taille minimale des segments, fusionnant les plus petits avec leurs voisins.
- `-mode raster` : Indique que le résultat sera un raster (image segmentée).
- `-mode.raster.out segmented.tif` : Définit le fichier de sortie contenant l'image segmentée.

Si vous souhaitez un **mode vecteur** (polygones), utilisez plutôt :

```
bash                                     Copier   Modifier
```





Mistral 03--2025



*Le système **maximise la vraisemblance** des items à l'intérieur de l'**enveloppe convexe d'apprentissage**.*

Nicolas Audebert

Les principales limites des modèles de fondation

- **biais du modèle statistique**
- **biais des données d'apprentissage**



Biais utilisateur



Mistral 03--2025

L'IA conversationnelle est-elle ?

- Une IAG / singularité ?
- Une sorte de Léonard de Vinci ?
- Une interface déterministe ?
- Un système stable ?
- Un système critique ?
- Un creuset d'intimité ?
- Un J.A.R.V.I.S. en gestation ?

De nouveaux usages



Usages :

Les LLMs sont de bons traducteurs

Parce ce qu'ils alignent des espaces sémantiques hétérogènes.

Traduction entre monde de représentation

Les LLMs sont de bons synthétiseurs

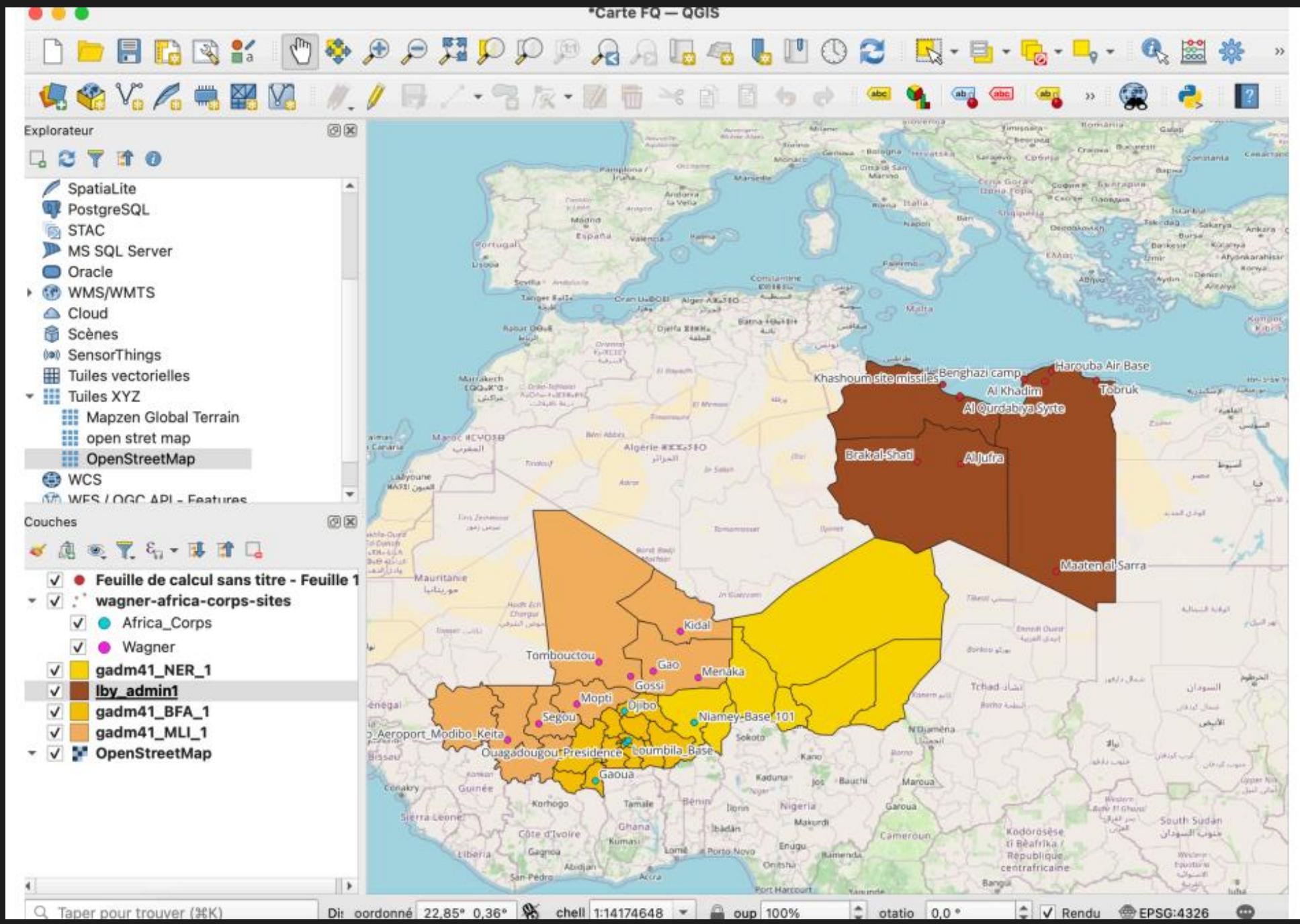
Parce qu'ils sont optimisés pour la pondération attentionnelle

Résumé spatial orienté intention

Les LLMs sont de bons défricheurs conceptuels

Parce qu'ils produisent des idées probables, mais navigables

Exploration assistée de problématiques territoriales



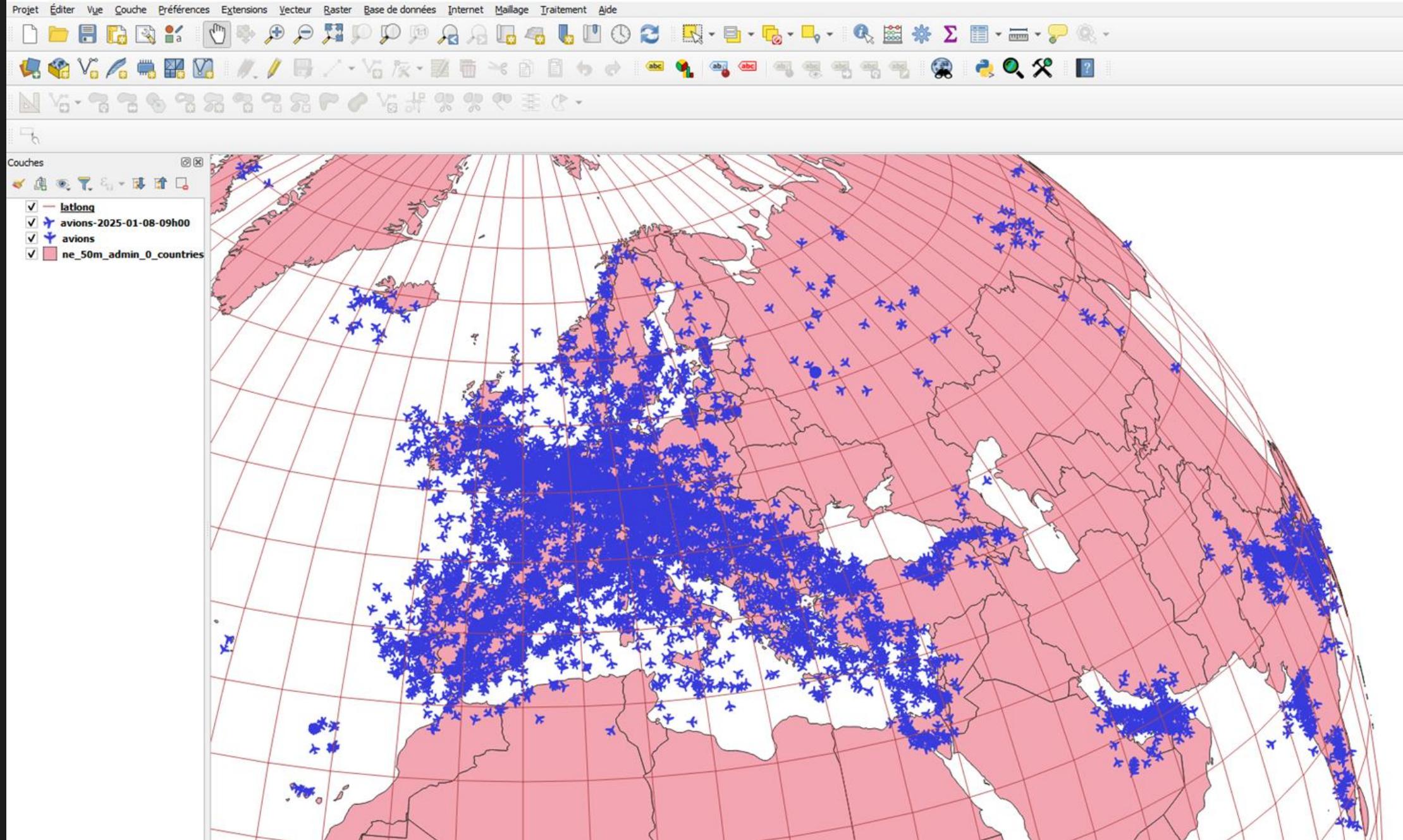
Usages :

Les LLMs sont de bons orchestrateurs d'outils

*Parce qu'ils savent décider quoi appeler et dans quel ordre
SIG piloté par intention*

Les LLMs sont de bons simulateurs narratifs

*Parce qu'ils peuvent maintenir une cohérence dans des mondes hypothétiques
Scénarisation territoriale qualitative*



Peut-on appliquer le même parallélisme à des modèles génératifs spatiaux ?

Ex : Un monde appris comme continuité, pas comme récit

Un LLM apprend un monde discret, symbolique, séquentiel.

Un modèle spatial apprend un monde continu, topologique, relationnel

Merci pour votre attention.





INNOVATIVE SOLUTIONS
BY OPEN SOURCE EXPERTS

Les LLM au service du géospatial

CNIG

Atelier IA & Information Géolocalisée



Florent Gravin

Camptocamp

Directeur Technique



florent.gravin@camptocamp.com



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



<https://github.com/fgravin>



<https://twitter.com/fgravin>



<https://bsky.app/profile/fgravin.bsky.social>

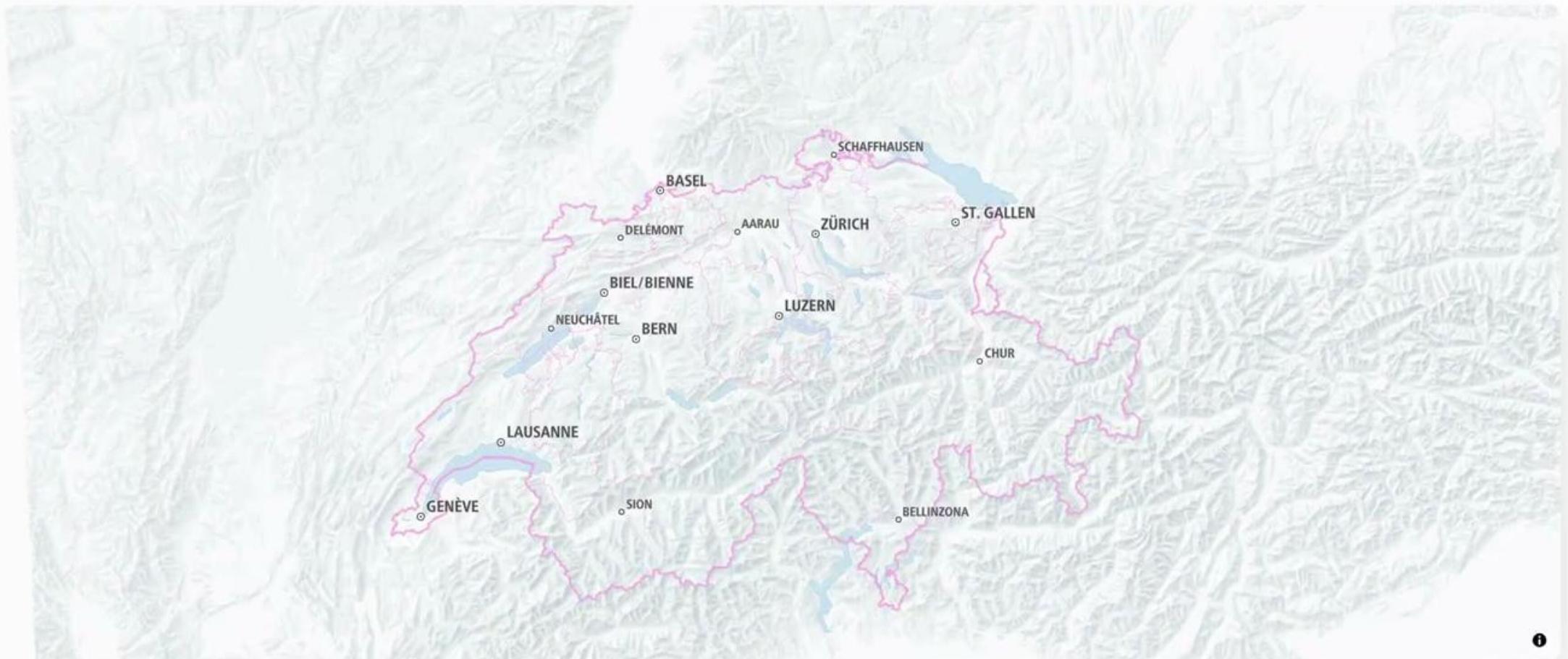


Qu'est-ce qu'un Large Language Model (LLM) ?

Un **système d'intelligence artificielle** capable de :

- Comprendre du texte
- Générer du texte





Que souhaitez-vous afficher ?



Ce que les LLM peuvent apporter au géospatial

- Améliorer l'expérience de recherche
- Chatbot, assistant (interaction avec une carte)
- Répondre aux problématiques d'intelligence géographique
- Enrichir les données
- Piloter vos outils SIG

Assistant GeoAI





La recherche

- Génération de synonymes
- Reformulation
- Recherche sémantique
- Ordonner les résultats



Enrichissement de données

VTT près de Luzerne

Toutes (253) Eté (206) Hiver (0) Lieux, lieux-dits et rues (47)

Suggestions pertinentes (7) La Suisse à VTT Localité: Luzern

Panorama Bike Rorschach - Montreux

820 Marbacher Panoramarunde Marbach - Marbach

Napf Bike Bern - Luzern

Tout afficher

La Suisse à VTT (199)

99 Dent de Vaulion Bike Vallorbe - Vallorbe

844 Les Prés-d'Orvin Bike Les Prés-d'Orvin - Les Prés-d'Orvin

592 Lac de Mayen Bike Leysin - Leysin

Randonnées hivernales Randonnées en raquettes Ski de fond Luge Eté



Route Description Extraction

Extract structured data from route descriptions using AI

Description Input

This scenic route takes you through the beautiful Alpine region, passing through dense forests and meadows. Along the way, you'll have stunning views of Lake Geneva and the Matterhorn in the distance. The path winds through several vineyards and crosses a picturesque river. There are multiple mountain huts and restaurants where you can stop for refreshments. In summer, you can take a refreshing swim in the lake or visit the thermal spa near the village.

Extract Data

Discovery Mode

Reset

Extract Data: Uses predefined categories
Discovery Mode: Discovers flexible categories

Translations

Translate the description to different languages

English

French

German

Italian

No translations yet. Click a language button to translate.

Extracted Data (Editable)

View On (distant views from the path)

Lake Geneva × Matterhorn ×

+ Add View

Pass Through (surrounding area and land usage)

Forest & Woods × Countryside (Meadows/Field: ×

Vineyards × River ×

+ Add Pass Through

Food & Drink (places for refreshments)

Restaurant / Cafe / Inn × Mountain Hut ×

+ Add Food/Drink

Swimming (places to swim or bathe)

Lake × Thermal Spa ×

+ Add Swimming

Save Extraction

Flattened Keywords

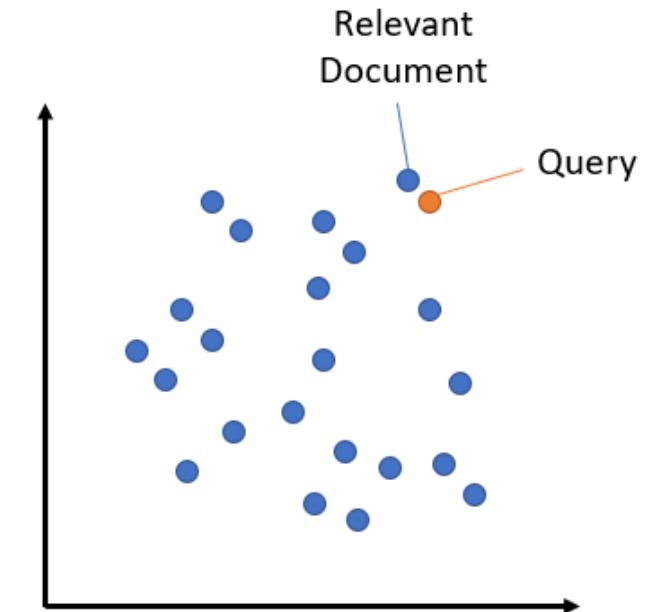
lake geneva matterhorn forest & woods countryside
(meadows/fields) vineyards river restaurant / cafe / inn
mountain hut lake thermal spa



La recherche sémantique

Une technique qui récupère des informations basées sur **le sens et le contexte** d'une requête, et non sur une correspondance exacte de mots-clés.

Cela s'appuie sur des **modèles d'embeddings** pour capturer les relations entre mots et concepts.



Consultez les données publiques des Hauts-de-France

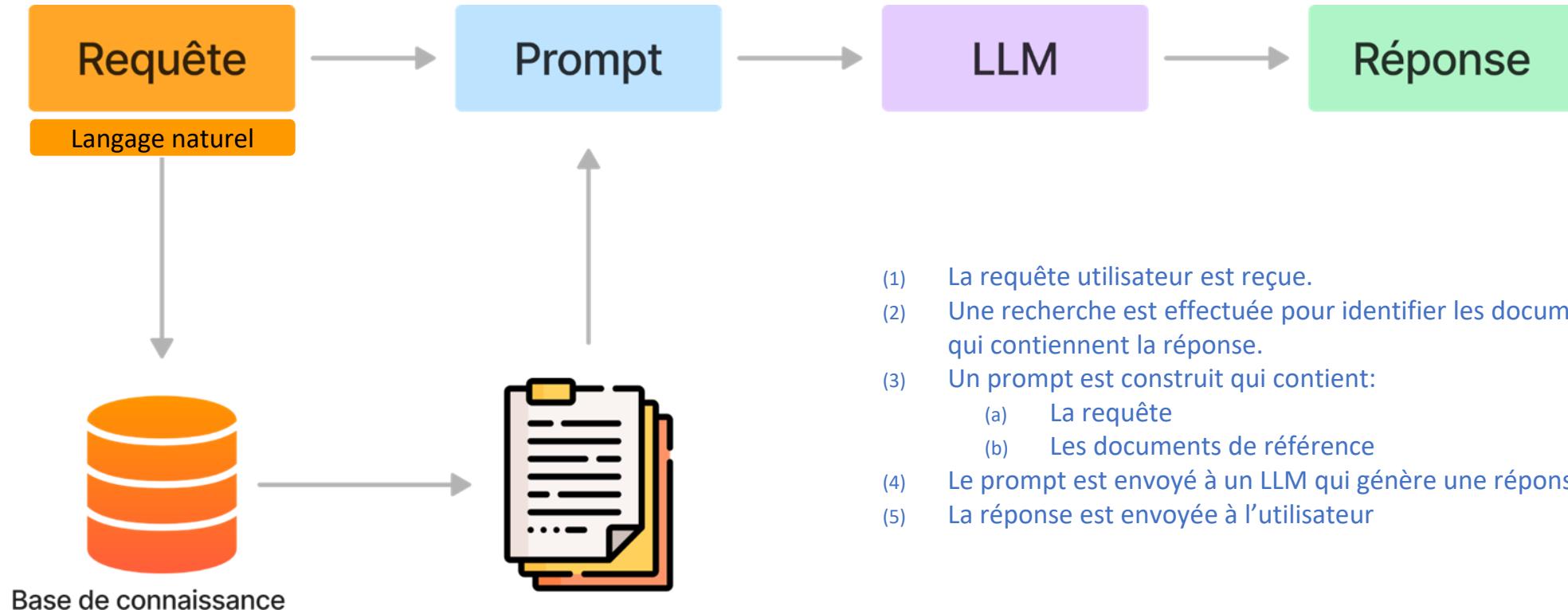
Rechercher une donnée ...

Lexical Search
GeoNetwork - Elasticsearch

Semantic Search
LLM



Retrieval augmented generation





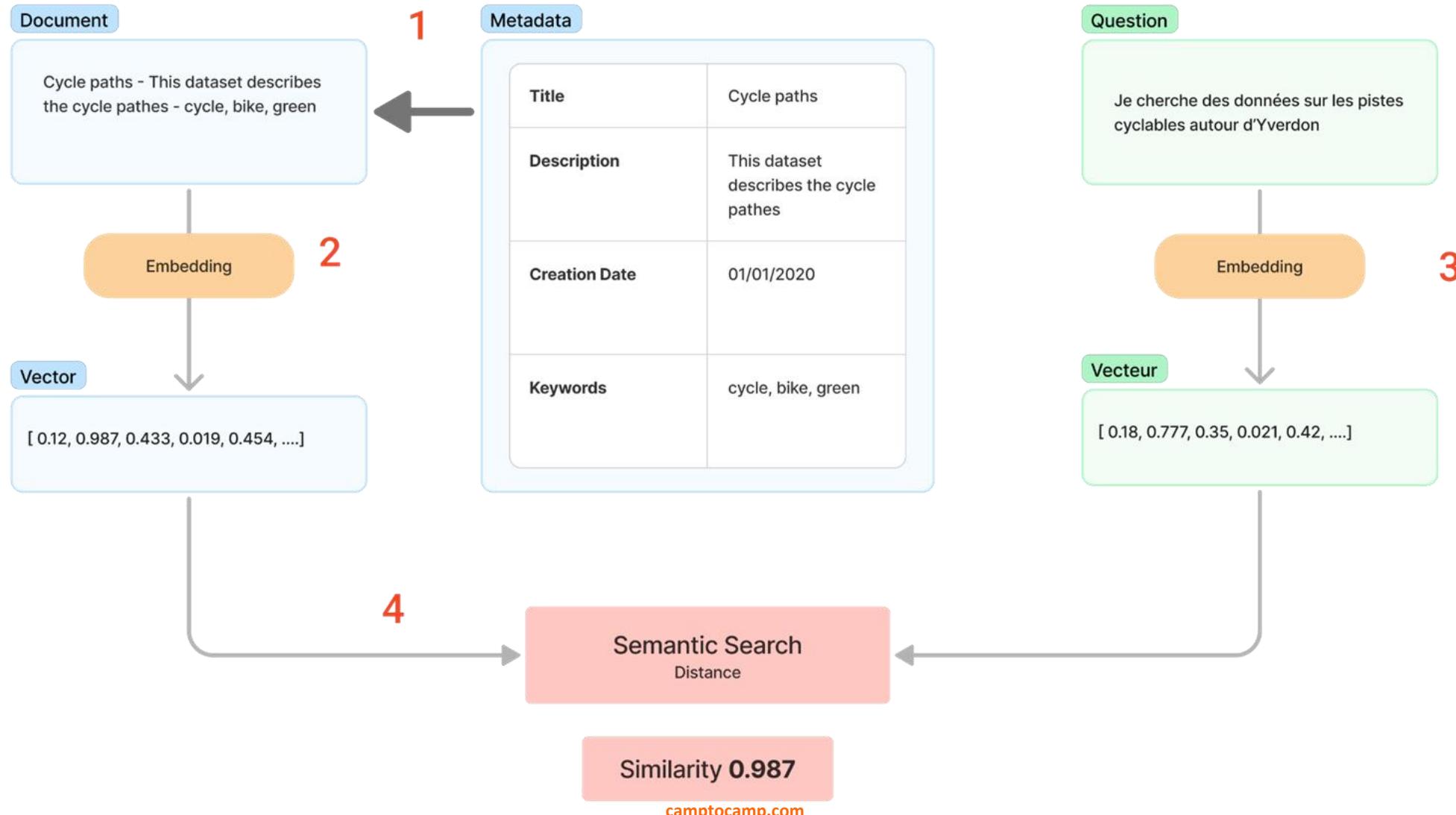
GG

Dans une application RAG, la qualité de la réponse dépend de la qualité de la recherche.

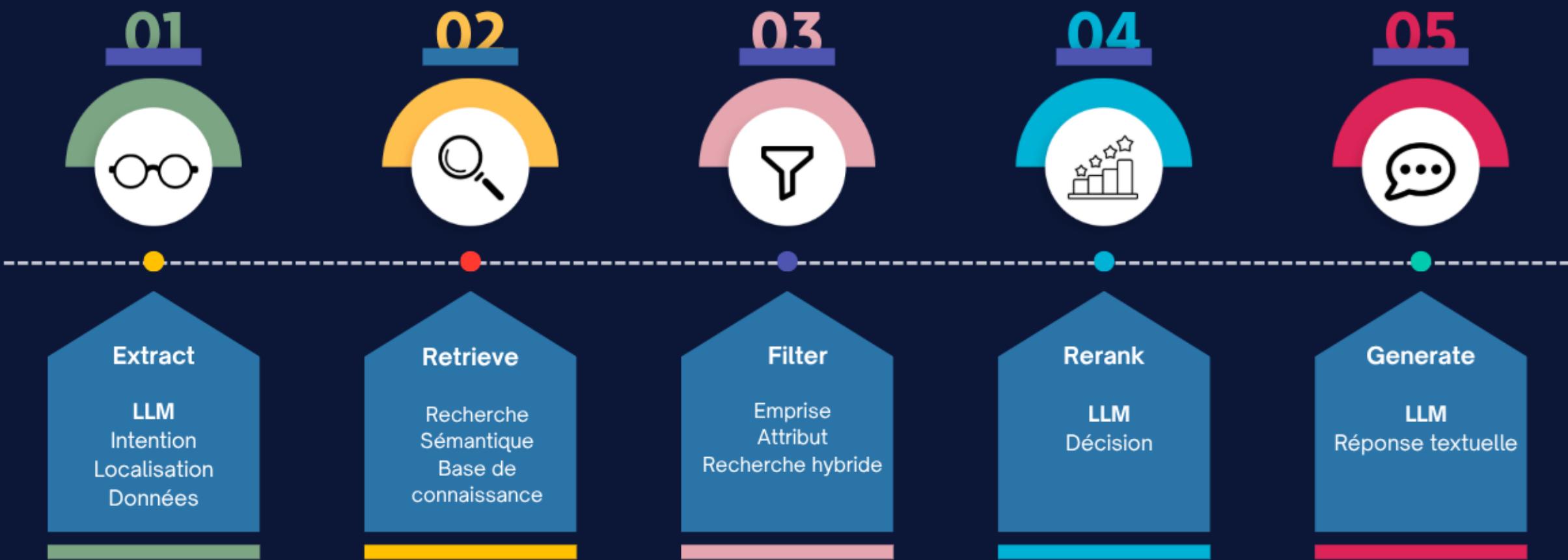
“ ”



RAG de recherche de données géo



CHATBOT - RAG





Swisstopo Meta Search



+ Nouveau chat

▫ Calcul de similarité

Invoke Chats

Où se trouvent les fontaines d'...

✓ Chat supprimé avec succès.

Bienvenue sur la Plateforme Nationale de Données Géographiques Comment puis-je vous aider ?

▢ Invoke

▢ Chat

Poser une question



Saisissez votre requête ou laissez-vous guider avec ces suggestions :

Montre-moi les possibilités de randonnée en hiver dans la région du Col des Mosses.

Essayer ↑



Où se trouvent les fontaines d'eau dans le canton de Fribourg ?

Essayer ↑

Quels sont les sentiers de VTT adaptés aux familles dans le Jura ?

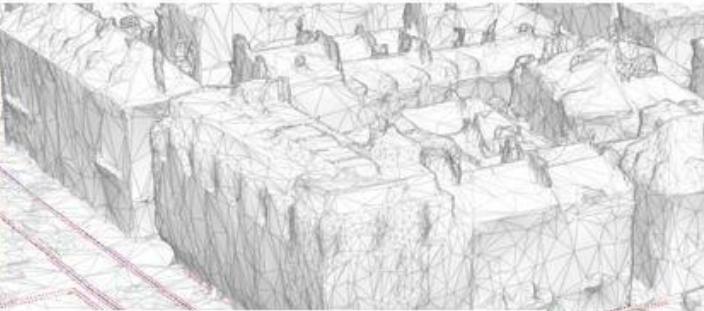
Essayer ↑

Quels sont les itinéraires de ski de fond dans la région du Valais ?

Essayer ↑



FRANCE
2030

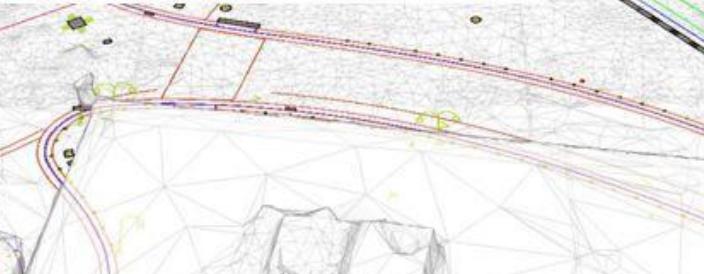


JUMEAU NUMÉRIQUE DE LA FRANCE ET DES TERRITOIRES

ANNEXE TECHNIQUE

PROGRAMME DE TRAVAIL

JANVIER 2025



IGN CHANGER
D'ÉCHELLE

Cerema
CLIMAT & TERRITOIRE DE DEMAIN

Inria

i-spatial
Your World Partner

camp to camp
PROFESSIONNELS
BY OPEN SOURCE EXPERTS

igo
geofit

SIRADEL
ENGIE

G Geometry
Factory

LuxCarta

cnes

cap-digital
Digital Future

Université
Gustave Eiffel

ifp Energies
nouvelles

© GEOFIT-IGO



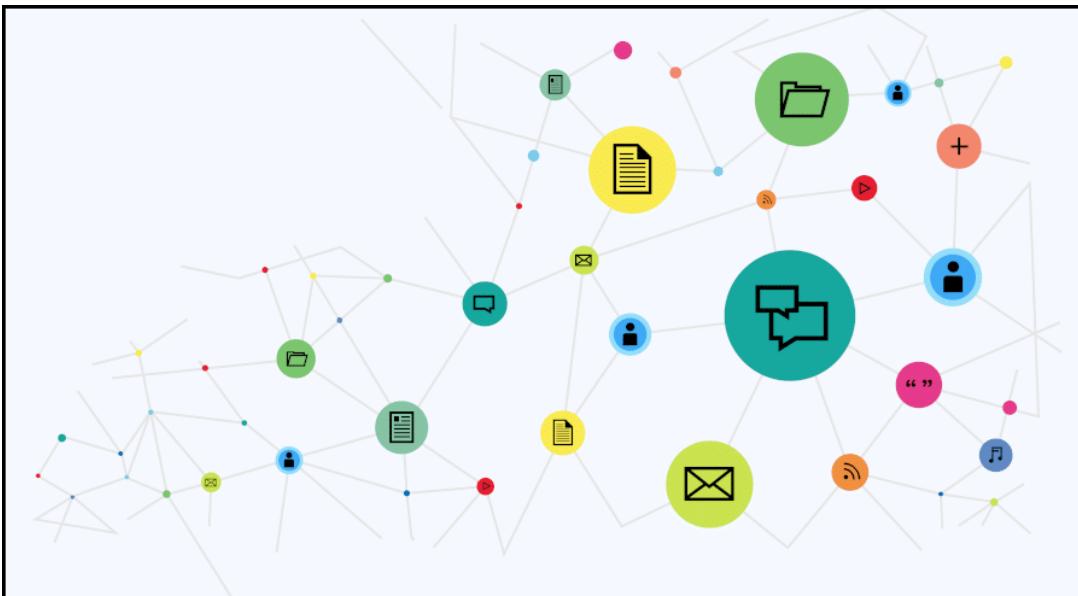
Aller plus loin

Anticiper les usages de demain

Ontologie : Lien entre données et connaissances

=> Contextualiser les informations

Graphe de connaissance



camptocamp.com

- Concepts
- Données
- Relations



Tools et function calling

Enter your query here

A quelle altitude se trouve Wabern?

Submit

- *Calling location_search with parameter(s): {'location_name': 'Wabern'}*
- *Calling get_height with parameter(s): {'x': 197615.8125, 'y': 601147.1875}*
- **Wabern se trouve à une altitude de 545,6 mètres.**

AGENT DE RAISONNEMENT



Dis-moi s'il y a des terrains
de sport de Lausanne à
moins de 500 mètres du lac

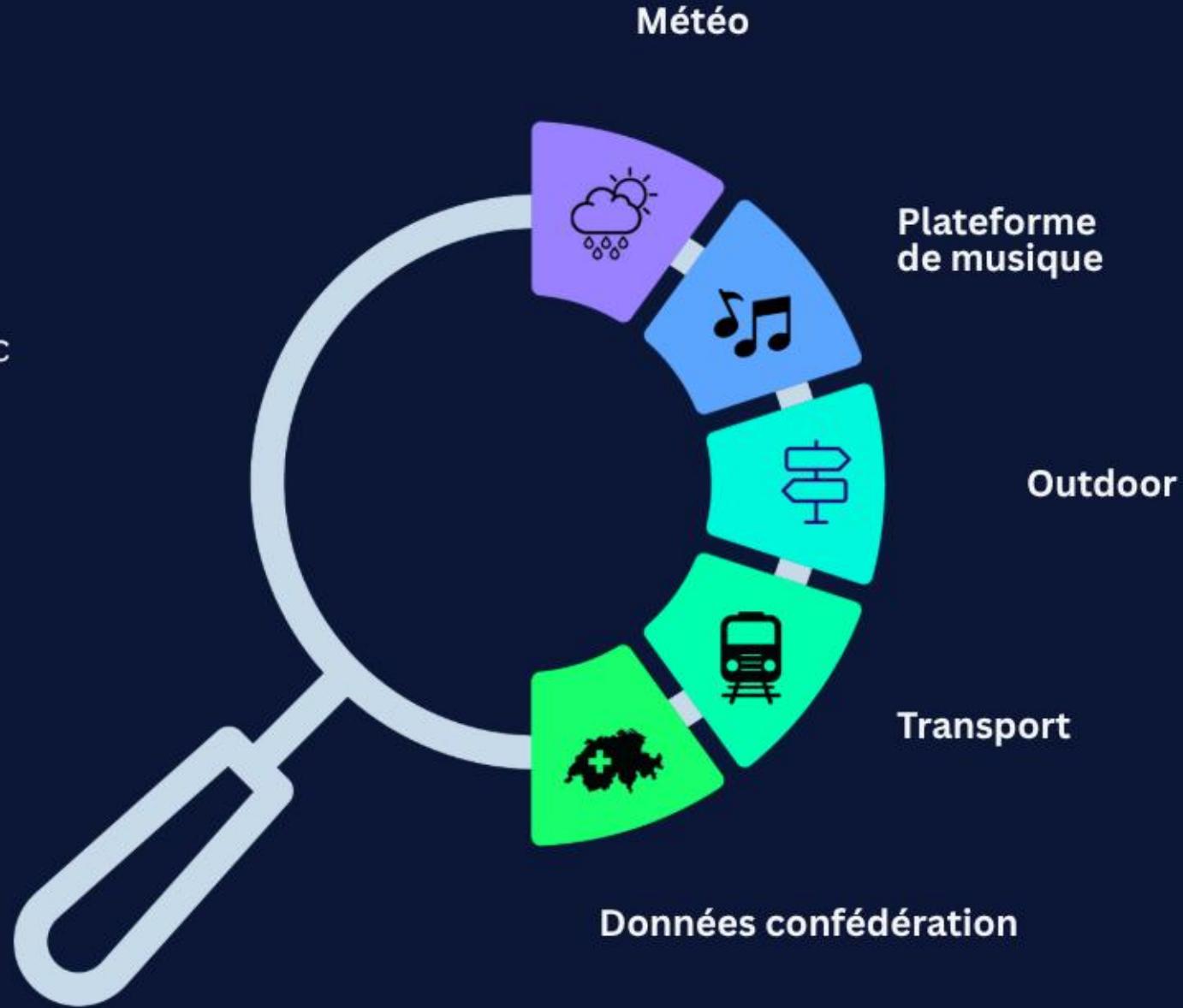
-  Téléchargement des équipements sportifs depuis OpenStreetMap
-  Téléchargement des contours du lac Léman depuis OpenStreetMap
-  Buffer
-  Intersection
-  Ajout des données à la carte

Voici les terrains de sports à moins de 500 m du lac, je
les ai affichés sur la carte. N'hésitez pas à zoomer
pour vérifier visuellement leur emplacement.



M o d e l C o n t e x t P r o t o c o l

Connecte un agent avec
n'importe quel service





GG

80% des projets échouent *
et seulement 30% dépassent le stade pilote **

“ ”



Clé du succès

- Ambitions mesurées
- Politique de la donnée adaptée
- Expertise technologique IA
- Expertise métier
- Tests et amélioration continue



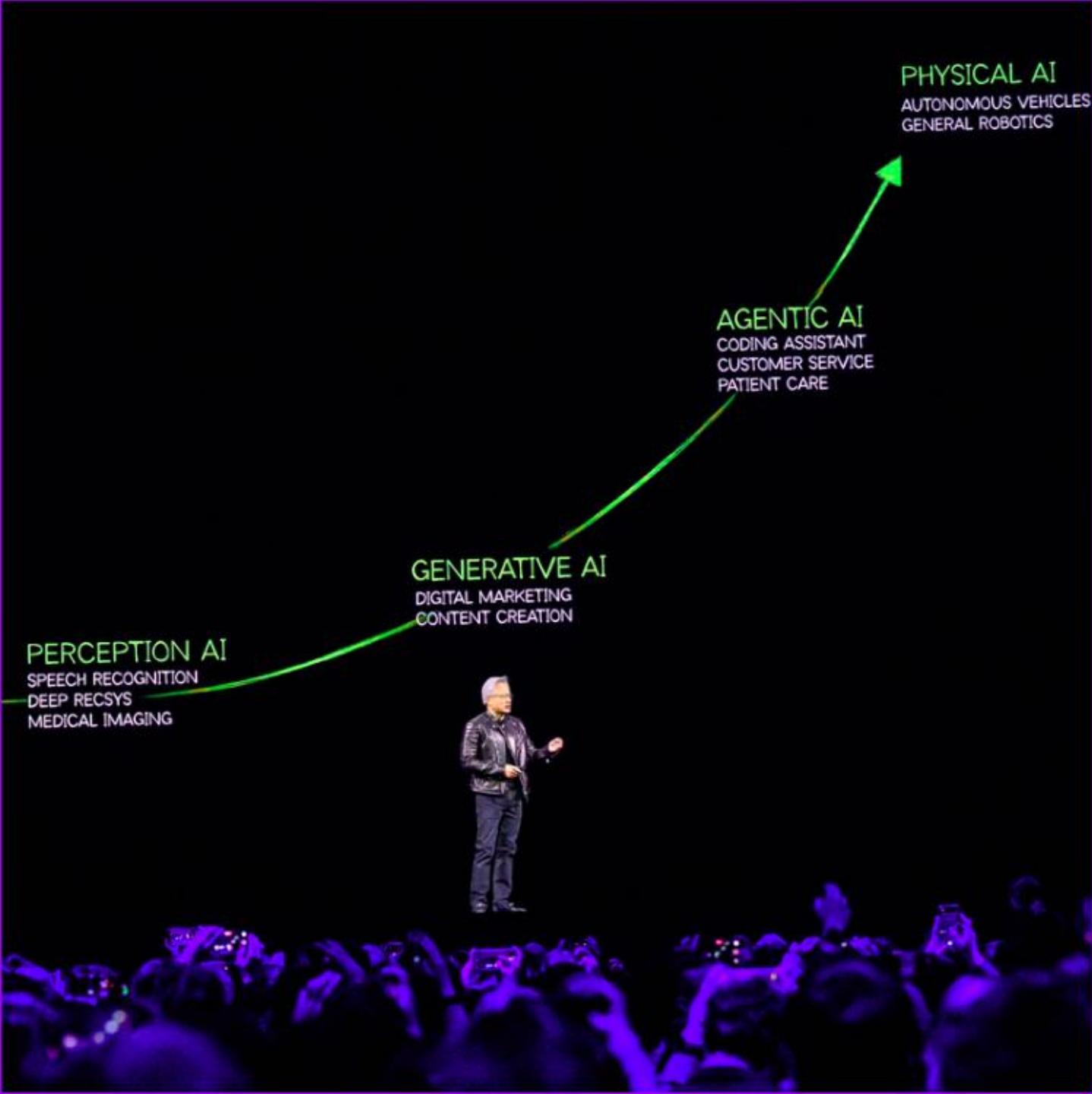
Nos réponses à certains enjeux

Souveraineté :

- Utiliser des modèles open source
- Contrôler leur hébergement

Eco-conception :

- Rationaliser et mutualiser les infrastructures.
- Optimiser la taille des modèles
- Minimiser les traitements LLM





INNOVATIVE SOLUTIONS
BY OPEN SOURCE EXPERTS

Intégrer les LLM aux SIG en toute confiance

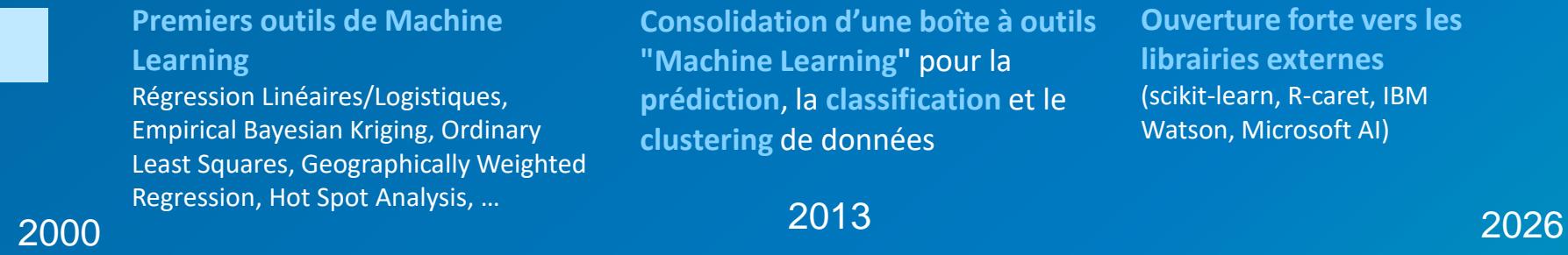
La vision d'un éditeur de logiciels

Gaëtan Lavenu

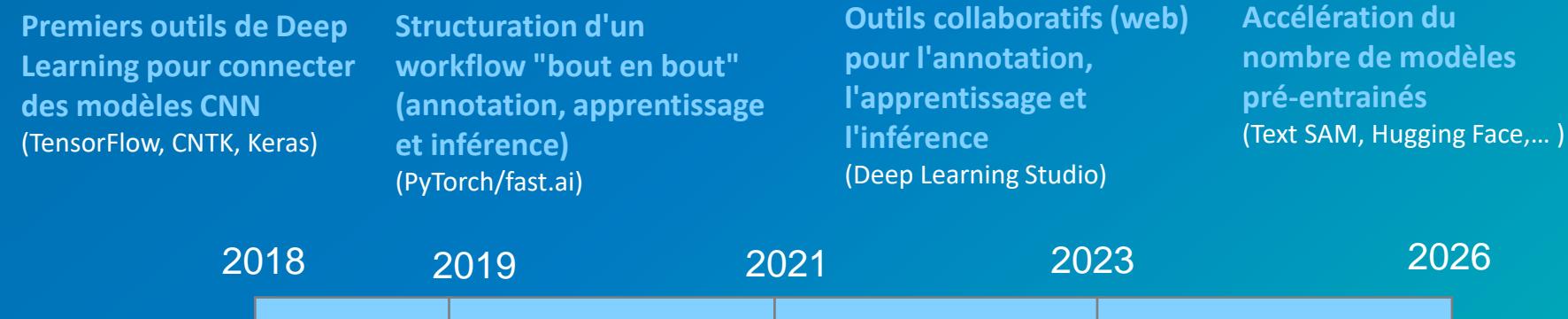
Atelier CNIG - IA et information géolocalisée - 08/01/2026

Un bref historique de l'intégration de l'IA dans nos technologies SIG

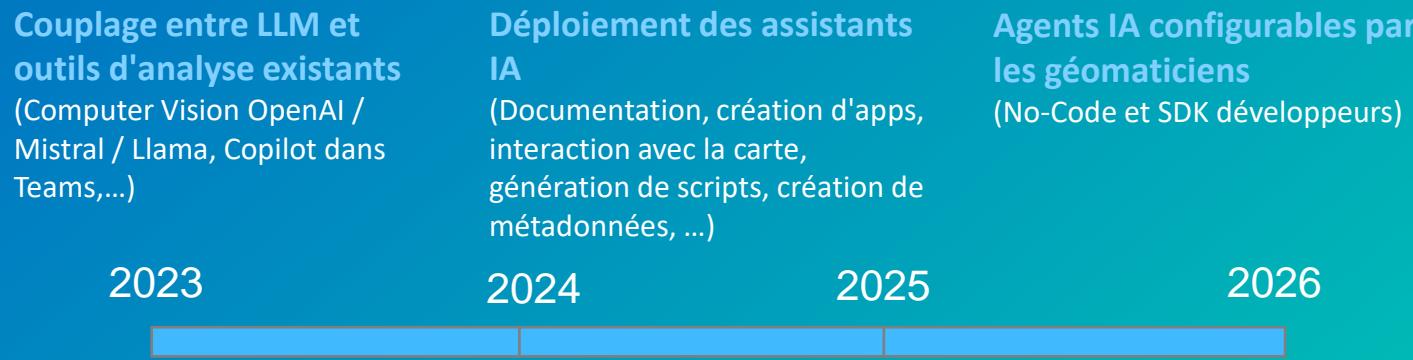
Phase 1 - Machine Learning



Phase 2 - Deep Learning et GeoAI



Phase 3 - IA Génération / Agents IA



Les bénéfices attendus de l'intégration des LLM dans les SIG

- Démocratiser l'accès aux SIG
 - Interrogation en langage naturel : “Montre-moi les communes où la population a baissé mais où l'artificialisation des sols progresse”
 - Abaissement de la barrière technique des outils (non-experts SIG, décideurs, ...)
- Accélération des flux de travail SIG
 - Génération de scripts (Python, Requêtes SQL, ...)
 - Assistance à la configuration d'application
 - Aide à l'exploration des données
- Enrichissement sémantique des données géographiques
 - Description automatique de couches et génération de métadonnées
 - Résumés intelligents pour le catalogage des jeux de données
 - Mise en relation de sources similaires / hétérogènes
- Nouveaux usages décisionnels
 - Interfaces conversationnelles géospatiales (Dialoguer avec la carte)
 - Aide à la formulation de scénarios
 - Support à la planification territoriale, urbaine, environnementale

Les Challenges qui persistent encore

- Absence de compréhension spatiale native
 - Pas de représentation intrinsèque des géométries, des relations topologiques
 - Raisonnement spatial simulé, indirect
- Problèmes de précision et de reproductibilité
 - Réponses probabilistes
 - Variabilité des résultats
 - Difficulté à garantir des résultats déterministes
- Dépendance aux données d'entrée
 - Qualité des prompts
 - Qualité des données exposées au LLM
 - Risque de sur-interprétation ou de simplification abusive
- Performances et passage à l'échelle
 - Volumétrie des données SIG incompatible avec un chargement direct
 - Nécessité de mécanismes intermédiaires : RAG (Retrieval Augmented Generation), APIs géospatiales et index spatiaux



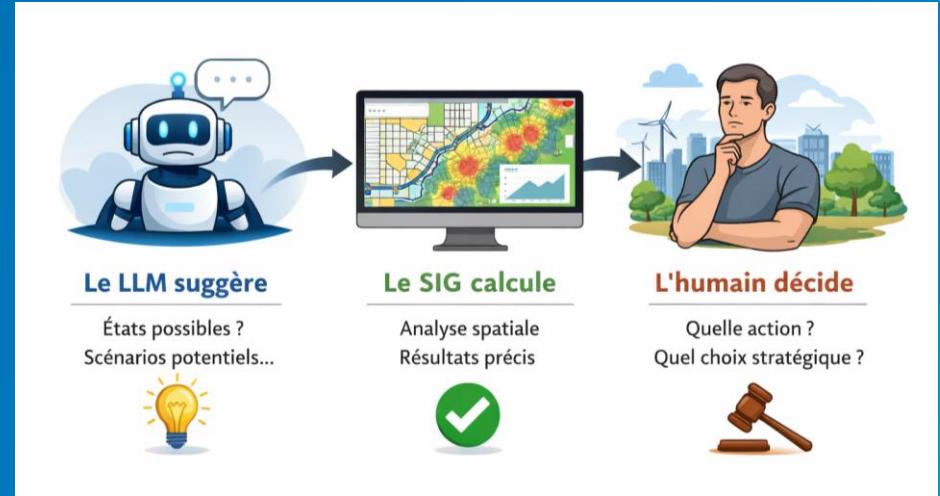
Une question centrale : la confiance

- Hallucinations et fabrication de réponses
 - Création de résultats plausibles mais faux
 - Cas critiques en : Aménagement, Gestion des risques, Environnement
- Traçabilité et explicabilité
 - Difficulté à répondre à ces questions : “D'où vient cette information ?”, “Sur quelles données s'appuie cette réponse ?”
 - Incompatibilité avec les exigences SIG classiques : traçabilité, auditabilité et reproductibilité
- Biais et représentations erronées du territoire
 - Biais culturels et géographiques des corpus d'entraînement
 - Vision non neutre des territoires
 - Risque de décisions biaisées
- Responsabilité et gouvernance
 - Qui est responsable d'une erreur ? Le modèle ? Le fournisseur ? L'analyste ?
 - Problèmes juridiques et réglementaires
 - Acceptabilité institutionnelle



Stratégie pour instaurer la confiance

- LLM comme assistant, pas comme moteur décisionnel
 - Principe “Human-in-the-loop”
 - Validation humaine systématique
 - Positionnement clair dans la chaîne de valeur
- Architecture maîtrisée
 - Couplage LLM + moteurs SIG déterministes
 - Séparer : interprétation (LLM), calcul spatial (SIG), RAG géospatial contrôlé
- Transparence et pédagogie
 - Description des sources (moteur et données d'apprentissage)
 - Justification des résultats
 - Indication du niveau de confiance et biais potentielles
 - Règles de confidentialités appliquées
- Évaluation continue
 - Tests de robustesse
 - Jeux de données de référence
 - Monitoring des dérives



Exemple dans le système ArcGIS

Documentation de nos modèles de GeoAI pré-entraînés

The screenshot shows the ArcGIS Model Catalog interface. At the top, there are various navigation icons: Tous, Tendances, Fonds de carte, Imagerie, Frontières, Personnes, Infrastructure, and Environnement. Below these are search filters: Tous les types de contenu, Tout le temps, Toutes les régions, Contenu exclusif à Esri, Contenu réservé aux sources officielles, and Trier par: Pertinence. A search bar labeled "Exemples de recherche" is present. The main area displays a grid of 12 model cards:

- Classification des mangroves (Landsat 8) - Module d'apprentissage profond par esri_analytics
- Classification des infrastructures piétonnes - Module d'apprentissage profond par esri_analytics
- Classification de la couverture terrestre (Sentinel-2) - Module d'apprentissage profond par esri_analytics
- Classification de la hauteur de la canopée à haute résolution - Module d'apprentissage profond par esri_analytics
- Classification de la couverture terrestre (Landsat 8) - Module d'apprentissage profond par esri_analytics
- Classification des maladies des feuilles des plantes - Module d'apprentissage profond par esri_analytics
- Classification HF Zero-Shot - Module d'apprentissage profond par esri_analytics
- Classification des points d'arbre - Module d'apprentissage profond par esri_analytics

Each card includes a small thumbnail, a title, a detailed description, and three interaction buttons: Abonné, Autoritaire, and ...

Classification de la couverture terrestre (Sentinel-2)



Modèle d'apprentissage profond pour la classification de la couverture terrestre sur les images Sentinel-2.

package d'apprentissage profond | Officiel Atlas vivant Abonné

Mise à jour de l'élément : 4 nov. 2025 ▾

Description

L'occupation du sol décrit la surface terrestre. Les cartes d'occupation du sol sont utiles en urbanisme, en gestion des ressources, en détection des changements, en agriculture et dans de nombreuses autres applications nécessitant des informations sur la surface terrestre. La classification de l'occupation du sol est un exercice complexe, difficile à réaliser par les méthodes traditionnelles. Les modèles d'apprentissage profond sont particulièrement performants pour appréhender ces subtilités et obtenir des résultats supérieurs.

En utilisant le modèle

Suivez le [guide](#) pour utiliser le modèle. Avant de l'utiliser, assurez-vous que les bibliothèques d'apprentissage profond compatibles sont installées. Pour plus d'informations, consultez [le programme d'installation des bibliothèques d'apprentissage profond pour ArcGIS](#).

Mise au point du modèle

Ce modèle peut être affiné à l'aide de l'outil « Entrainer un modèle d'apprentissage profond ». Suivez le [guide](#) pour affiner ce modèle.

Saisir

Données raster, mosaïques ou service d'imagerie. (Taille de cellule préférée : 10 mètres.)

Remarque : Ce modèle est entraîné à fonctionner sur des ensembles de données d'imagerie Sentinel-2 qui sont dans le système de coordonnées WGS 1984 Web Mercator (sphère auxiliaire) (WKID 3857).

Exemple dans le système ArcGIS

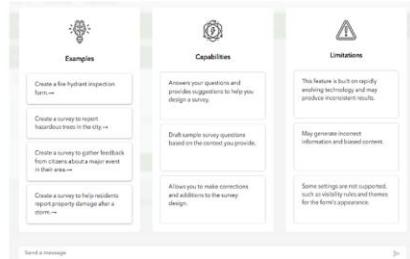
Fiches de transparence pour nos assistants IA

AI Transparency Card: ArcGIS Survey123 – Web designer assistant



Section	Description	Response
Product - Name	ArcGIS product name	ArcGIS Survey123
Product - Certification	Certification status of the ArcGIS Product or its subprocessors	Certified: FedRAMP Moderate
Product - Deployment	Deployment model of the product	SaaS
Name	AI feature name in the product	ArcGIS Survey123 – Web designer assistant (Preview)
Purpose	Actions AI feature is expected to perform within the product	Helps you jump-start new Survey123 forms through a conversational experience. It guides you in shaping your survey's structure, allowing you to transform ideas into functional forms, make real-time adjustments to questions, and transition your draft into the Survey123 web designer for further customization.
Release Status	Release status of AI feature	Preview
Certification	Certification status of AI feature or its subprocessors	None
Deployment	AI feature provided via what deployment model	SaaS
Management	How AI feature can be enabled or disabled?	Opt-in by AGO Administrator
Management – Feedback	Can/how user AI feedback be enabled or disabled?	Opt-in by User
Management - Telemetry	How user AI telemetry data can be enabled or disabled?	EUEI: Admin Config, Feature Specific: None, Subprocessors: Required
Prompt Stored	Are prompts submitted to the AI stored?	Not by default: Only when feedback provided, Retention: 30 days, Storage Purpose: Specific Improvement
Response Stored	Are AI-generated responses stored?	Not by default: Only when feedback provided, Retention: 30 days, Storage Purpose: Specific Improvement
Personal Data	Is personal data in training, testing, or validation datasets?	No
Processing Location	Where data is processed across the product, feature, and LLM levels, including details on any subprocessors	Product: US AGO Infrastructure, Feature: US AGO Infrastructure, LLM: US Microsoft Azure OpenAI LLM subprocessor.
Intended Users	Primary intended users of the AI feature	GIS Analysts, Survey Creator
Out-of-Scope Uses	Scenarios AI feature may not perform accurately or reliably	Create surveys with highly specific domain knowledge not covered in its training.
Key Function	Key capabilities and how the AI feature enhances workflows	Augment: Assists in generating concepts into surveys from natural language prompts.
Model Type & Technique	AI model type and technique	Generative AI: Transformer
Model Used	Specific model(s) used, such as GPT-4, T5, etc.	Microsoft Azure OpenAI – GPT-4o
Esri AI Framework Use	Indicate if the AI feature utilizes the Esri AI Framework	No

USER INPUT SCREEN



USER OUTPUT SCREEN



Model License	License of AI model powering the AI feature	Licensed
Training Data Sources	Data sources used for development of the AI feature	Mix: Microsoft Azure OpenAI service, Esri Content (documentation)
Human-in-the-Loop	Indicate if users can review or modify AI-generated outputs	Yes. Users can review and modify AI-generated outputs from the AI assistant before storing or modifying customer content.
Input/Output Formats	Input and output formats that the model can handle	Input: Text / Output: Rich Text and Imagery
Bias/Ethical Mitigations	Detail how biases are managed, especially in the data	Bias managed via Microsoft's Content Filter and avoiding customer data in training.
Security Mechanisms	Security mechanisms beyond guardrails and encryption	Deployed within Esri's ArcGIS Online infrastructure. Outside services are used for the LLM, and user inputs are processed temporarily without being stored.
Guardrails	Help provide secure AI (Four types)	LLM Integrated: Integrated LLM only Adversarial Attack: Integrated LLM only Objectionable Content: Content filter enabled Intellectual Property: Checks not enabled
Encryption at Rest	Is AES-256 or better used for at rest encryption?	Yes
Encryption in Transit	Is TLS 1.2 or better utilized for encryption in transit?	Partially: Full backend transport encryption planned for GA release.
Adversarial Testing	Adversarial robustness tests / Protection against manipulation	General holistic testing is conducted ad-hoc.
Key Metrics	Performance metrics relevant to ArcGIS	In-Progress: To be completed for General Availability.
Evaluation/Testing	Testing methodologies and how Esri evaluates performance	Functional Automation Testing is conducted.
Provenance	Mapping conducted that tracks data from source to output?	In-Progress: To be completed for General Availability.
Best Practices	Links to AI feature tutorials, and best practices	Limited: See initial help documentation here .
Continuous Improvement	Improvement plans (e.g. audits, bias checks, and transparency)	Product is currently in Preview and General Availability improvements are planned.