

Etat de l'art

LLM et Géodonnées

Approches, outils et méthodologies, Version de Mai 2024



Contexte

L'État de l'art *LLM et Géodonnées* a été mandaté dans le cadre du plan d'action 2024 (action 4-24-02) de la Stratégie suisse pour la géoinformation (SGS) en collaboration avec l'Office fédéral de topographie (swisstopo). Ce rapport contient les observations, les avis, les résultats et les conclusions de chercheurs et d'experts dans le domaine de la technologie géospatiale et de l'intelligence artificielle.

Remerciements

Ageospatial souhaite remercier l'Office fédéral de topographie (swisstopo), COSIG (Coordination, Services et Informations Géographiques), l'Agence suisse pour l'encouragement de l'innovation (Innosuisse), EPFL Innovation Park (La Forge), l'Université de Genève (UNIGE), Geneva Responsable Entrepreneurship Center (GREC) et l'Institut des Sciences de l'Environnement (ISE).

TABLE DES MATIÈRES

| | |
|--|----|
| Préambule | 4 |
| Définitions..... | 5 |
| Introduction | 7 |
| Chapitre 1: Analyse de la Dimension Geographique des LLM..... | 9 |
| 1.1 Extraction de géodonnées intégrés dans les LLM..... | 10 |
| 1.2 Fine tuning et Geospatial Location Embedding..... | 14 |
| 1.3 Compréhension des principes SIG..... | 17 |
| Chapitre 2: Les LLM dans l'interprétation, la compréhension et l'exécution de tâches de géotraitement | 18 |
| 2.1 Traduction des commandes humaines en code..... | 19 |
| 2.2 Opérations séquentielles..... | 22 |
| 2.3 Automatisation de l'analyse géospatiale..... | 24 |
| Chapitre 3: Extraction et traitement des géodonnées à partir de Multimodal Large Language Models (MLLM) | 31 |
| 3.1 Dimension geographique des MLLM..... | 32 |
| 3.2 Les MLLM comme assistants intelligents..... | 35 |
| Conclusion..... | 36 |
| Liste des figures | 37 |
| Bibliographie | 39 |
| Lecture supplémentaire | 41 |

PRÉAMBULE

La “Stratégie suisse pour la géoinformation”, adoptée le 11 décembre 2020 par le Conseil fédéral et la Conférence suisse des directeurs cantonaux des travaux publics, de l'aménagement du territoire et de l'environnement (DTAP), a pour objectif de fournir des informations géolocalisées fiables, détaillées, actualisées et interopérables pour tous les utilisateurs.

Cette stratégie vise à intégrer la géoinformation dans tous les aspects de la société, de l'économie et de l'environnement, pour faciliter la prise de décision, promouvoir le développement durable et favoriser l'innovation. Elle souligne l'importance d'une gestion durable et concertée de l'espace et des ressources limitées en Suisse, par la création et le maintien de représentations numériques des espaces géographiques.

Le mandat du Conseil fédéral est clair : exploiter la numérisation pour rendre le cadre de vie plus attrayant et durable, et pour maintenir la compétitivité de la Suisse en tant que site économique et de recherche. Les sept champs d'action définissent les grandes orientations pour les prochaines années: promouvoir l'écosystème de la géoinformation, relier les géodonnées, faciliter les processus, développer la science des géodonnées, encourager l'innovation, acquérir et renforcer les compétences, et développer les plateformes numériques. Ces actions visent à maximiser l'utilité des géoformations et à encourager une approche agile et participative.

Une consolidation de l'infrastructure nationale de données géographiques (INDG) est actuellement en cours pour faciliter l'accès aux données géographiques via des interfaces standardisées (INDG 2.0) et pour préparer la mise en œuvre d'un réseau de plateformes collaboratives (INDG 3.0).

Les effets attendus de la stratégie comprennent un soutien accru à la prise de décision quotidienne, un accès à des géodonnées fiables et harmonisées, une meilleure connectivité entre les données non spatiales et les géodonnées, et une amélioration de la résilience et de la satisfaction des utilisateurs.

Pour garantir la mise en œuvre de cette nouvelle stratégie, des plans d'action annuels sont publiés afin d'établir une feuille de route pour les différentes parties prenantes. Ce rapport est réalisé dans le cadre de la Stratégie Suisse pour la Géoinformation (SGS) et de son plan d'action pour l'année 2024. Dans le champ d'action dédié à “Développer la science des géodonnées”, une initiative spécifique (4-24-02) porte sur l'utilisation de l'intelligence artificielle (IA) et des grands modèles de langage (LLM) dans les infrastructures de géodonnées. Cette initiative a pour objectif d'intégrer les capacités avancées des LLM au traitement et à l'analyse des données géospatiales.

DÉFINITIONS

Intelligence Artificielle: L'Intelligence Artificielle fait référence à la simulation de l'intelligence humaine par un programme qui a pour objectif d'effectuer des tâches qui incluent la prise de décision, la résolution de problèmes, la compréhension du langage et la reconnaissance d'objets.

Deep Learning: Le Deep Learning est une branche spécifique du Machine Learning qui utilise des réseaux de neurones artificiels avec de nombreuses couches. Ces réseaux de neurones sont capables d'apprendre à partir de grandes quantités de données non structurées ou non étiquetées. Le Deep Learning a permis des avancées importantes dans les domaines tels que la reconnaissance vocale, la traduction automatique, et la vision par ordinateur.

Foundation Model: Les modèles de fondation sont des modèles d'intelligence artificielle de grande taille, pré-entraînés sur une grande quantité de données non étiquetées. Ces modèles sont utilisés pour une large gamme d'applications, telles que le traitement du langage naturel (LLM), la vision par ordinateur, et d'autres domaines de l'IA (voir plus à la page 43).

LLM: Un Large Language Model (Grand Modèle de Langage, inusité en français) est un type avancé de modèle d'intelligence artificielle spécialisé dans la compréhension et la génération du langage naturel. Le LLM apprend à partir d'une quantité importante de données, lui permettant de saisir les nuances linguistiques ,les contextes variés, et

même les subtilités stylistiques du langage humain, facilitant ainsi des applications telles que la traduction automatique, la génération de texte, et les systèmes de questions-réponses.

Tokens : La tokenisation consiste à découper le texte en petites unités, appelées *tokens*. Ces *tokens* peuvent être des mots, des parties de mots, ou même des caractères, selon le niveau de détail que l'on souhaite obtenir. La tokenisation est l'une des premières étapes du traitement du texte : elle permet d'organiser et de découper le texte et est nécessaire pour que les modèles comprennent la structure du texte d'entrée.

Natural language Processing : Le *natural language processing* (NLP) désigne l'ensemble des techniques permettant aux ordinateurs de comprendre, d'interpréter et de manipuler le langage humain naturel.

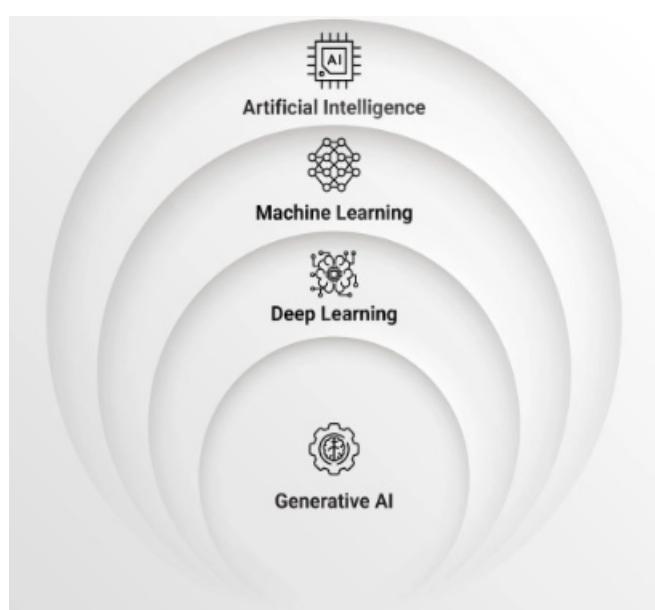


FIGURE 1 – Illustration de l'intégration de l'IA, du ML et du DL dans les systèmes modernes

MLLM: Un Multimodal Large Language Model (MLLM) est une évolution des LLM qui intègre et traite plusieurs types de données ou “modalités” pour comprendre et générer du contenu. Ces modalités peuvent inclure, mais ne se limitent pas à, du texte, des images, de l’audio, et des vidéos.

Agent Intelligent (LLM): Un agent intelligent est une entité autonome capable d’effectuer des actions en fonction de certains paramètres ou objectifs prédéfinis. L’agent permet d’utiliser un LLM pour résoudre des tâches complexes de manière séquentielle.

Fine-tuning: Le fine-tuning consiste à prendre un modèle pré-entraîné et à le réentraîner (ou affiner) sur un jeu de données plus petit et spécifique pour accomplir une tâche donnée. L’objectif est de transférer la connaissance générale acquise lors du préentraînement à un domaine spécifique ou à une tâche particulière sans repartir de zéro. Lors du fine-tuning, les poids et les plongements du modèle sont ajustés pour mieux s’adapter à la nouvelle tâche (voir plus à la page 42).

Information géospatiale: L’information géospatiale désigne des données associées à ou dérivées d’un emplacement spécifique sur la surface de la Terre, souvent représentées sous forme de coordonnées, d’adresses ou de caractéristiques spatiales.

Weights: Dans le contexte d’un Large Language Model (LLM), les “poids” sont des paramètres numériques qui déterminent la force des connexions entre les neurones au sein du modèle. Lors de l’entraînement d’un LLM, ces poids sont ajustés de manière itérative pour minimiser les erreurs dans les prédictions du modèle.

Embedding (Plongement): L’embedding est une méthode qui transforme des tokens en vecteurs dans un espace vectoriel continu de grande dimension. Chaque token est représenté par un vecteur, et ces vecteurs recueillent des informations sémantiques et syntaxiques sur les tokens. Les mots ayant des significations similaires se trouvent plus proches les uns des autres dans cet espace. L’embedding intervient après la tokenisation et est essentiel pour que les modèles comprennent le sens des mots.

Prompt engineering : Le prompt engineering, ou ingénierie de prompt, consiste à structurer un texte de façon à le rendre compréhensible pour un LLM. Cela peut consister en l’ajout de contexte, d’instructions, voire même un exemple de la réponse attendue.

RAG: Retrieval Augmented Generation est une technique qui permet au LLM de rechercher et exploiter des données provenant de sources externes (base de données, sites web...)

INTRODUCTION

Les Large Language Models (LLM) sont dérivés des foundation models (voir page 42), des modèles d'intelligence artificielle de grande échelle conçus pour l'exécution de nombreuses tâches. Leur développement a débuté avec des modèles comme BERT (Bidirectional Encoder Representations from Transformers) et GPT (Generative Pretrained Transformer), pré-entraînés sur de vastes ensembles de données non-libellés pour apprendre les structures linguistiques de manière non-supervisée. L'avantage des foundation models découle de leur capacité d'être affinés pour des tâches nouvelles et spécifiques sans avoir recours à un réentraînement complet. Les progrès réalisés dans le domaine des foundation models et l'accroissement des capacités de calcul ont permis de créer des modèles de plus en plus sophistiqués, menant à la conception de LLM tels que ChatGPT.

Le 30 novembre 2022, OpenAI a publié ChatGPT, représentant une avancée majeure dans le développement de l'intelligence artificielle générative et qui a atteint les 100 millions d'utilisateurs après seulement trois mois. Désormais, une intelligence artificielle était capable de comprendre et de générer du texte comme le ferait un être humain. Des chatbots avaient été développés avant ChatGPT, mais leurs réponses restaient très limitées et possédaient une compréhension très rudimentaire du langage naturel. Ce lancement a déclenché la mobilisation d'autres géants américains de la technologie, tels que Google et Meta, qui ne voulaient pas se laisser devancer dans le développement d'une intelligence générative artificielle.

Ainsi, depuis la sortie de GPT 3.5, puis de GPT 4.0, nous avons vu apparaître sur le marché toute une série de LLM, soit sous forme de logiciels open-source, soit sous forme de logiciels propriétaires. Gemini, Claude, Grok et Llama, plusieurs modèles sont lancés régulièrement pour égaler ou surpasser les modèles proposés par OpenAI, notamment GPT-4.

En parallèle, le domaine géospatial profite à la fois d'un regain d'intérêt stratégique pour l'orbite terrestre basse et d'une digitalisation de plus en plus avancée à l'échelle mondiale. Il en ressort un montant exponentiel de données à stocker, analyser et distribuer, mais aussi une complexité accrue de ces analyses et des outils liés. Des programmes satellites d'observation de la Terre comme Sentinel ou Landsat ont généré des pétaoctets de données, tandis que le nombre de données vectorielles accessibles dans des formats standardisés se démocratise sur le web. Ces opportunités et contraintes font du domaine géospatial un domaine propice à l'implémentation de technologies LLM, qui sont capables de comprendre aussi bien des données textuelles, même sous forme de géodonnées, que des images (par exemple des images satellites) et vidéos, et d'enchaîner ces différentes analyses pour aboutir au résultat voulu par l'utilisateur. Cet état de l'art vise justement à explorer les différents liens entre LLM et le domaine du géospatial, en évaluant différentes techniques et outils parmi les plus récents sur la connaissance géographique des LLM, leur utilité dans l'interprétation de données géographiques, et leur capacité à traiter ces informations.

L'entraînement d'un grand modèle de langage (LLM) est un processus complexe qui implique plusieurs étapes clés, s'appuyant sur des principes d'apprentissage automatique et d'intelligence artificielle. Le processus commence par la collecte et la préparation d'un vaste ensemble de données textuelles. Ces données peuvent provenir de diverses sources, telles que des livres, des sites web, des articles de journaux et des encyclopédies.

L'objectif est de couvrir une large gamme de sujets, de styles et de langues pour permettre au modèle de comprendre et de générer du texte dans différents contextes. Les données sont nettoyées et prétraitées pour enlever les erreurs, les doublons et les informations non pertinentes. Le texte est ensuite divisé en petites unités, appelées tokens, qui peuvent être des mots, des parties de mots ou même des caractères individuels. Cette étape permet de transformer le texte en une forme que le modèle peut assimiler. La tokenisation aide également à gérer la diversité des langues et des structures grammaticales. Un modèle de base est construit en utilisant une architecture de réseau de neurones et comprend plusieurs couches de neurones artificiels qui apprendront à reconnaître les tendances dans les données textuelles.

Durant cette phase, le modèle est entraîné sur l'ensemble de données préparées, en utilisant des techniques d'apprentissage non supervisé. L'objectif est d'amener le modèle à prédire un mot ou un token manquant dans une phrase donnée (modélisation de langue) ou à reconnaître si deux segments de texte sont logiquement reliés. Cet entraînement préliminaire permet au modèle d'acquérir une compréhension générale de la langue et de ses nuances.

Après l'entraînement préliminaire, le modèle peut être affiné pour des tâches spécifiques (*fine-tuning*, voir page 41) telles que la traduction, la réponse à des questions thématiques ou la recherche de données externes. Pour ce faire, le modèle est entraîné sur un ensemble de données propres au domaine ou à la tâche que l'on souhaite qu'il maîtrise. Une fois l'entraînement complété, le modèle est évalué pour mesurer sa performance. En fonction des résultats, des ajustements peuvent être effectués, et le modèle peut être entraîné à nouveau ou affiné davantage. Cette étape d'itération est indispensable pour optimiser et maximiser la performance du modèle et assurer que sa version finale puisse répondre aux exigences particulières des tâches pour lesquelles il a été affiné.

CHAPITRE 1 :

ANALYSE DE

LA DIMENSION

GEOGRAPHIQUE DES

LLM

1.1 EXTRACTION DE GÉODONNÉES INTÉGRÉS DANS LES LLM

Évaluation de l'étendue des connaissances géospatiales des LLM

Dans une première phase, des recherches ont été menées pour démontrer la capacité des premiers modèles de langage pré-entraînés (LLM) open-source à intégrer et manipuler des connaissances géospatiales.

Deux études notables, intitulées *Are Large Language Models Geospatially Knowledgeable* [1] et *Extracting Geographic Knowledge from Large Language Models: An Experiment* [2], ont exploré, d'une part, la précision des LLM à prédire les coordonnées géographiques des villes via des approches zero-shot et three-shot et, d'autre part, leur capacité à générer des géodonnées qualitatives sans entraînements spécifiques ni fine-tuning.

En appliquant des techniques de *prompt engineering* pour affiner la précision des réponses et en évaluant les performances sur les premiers LLM tels que BERT, RoBERTa ou les premiers modèles LLaMA, ces recherches illustrent le potentiel des LLM en tant qu'outils géospatiaux, capables de générer globalement des données précises.

| Model | Template | Prompt | Error (km) | P-Rate (%) |
|-------------|----------|--------|-------------|------------|
| Word2Vec | - | - | 2612 | - |
| BERT-L | - | - | 3077 | - |
| GPT-2 | - | - | 4498 | - |
| LLaMA (7B) | 2 | 0-shot | 521 | 10 |
| LLaMA (7B) | 2 | 3-shot | 1469 | 99 |
| LLaMA (13B) | 1 | 0-shot | 864 | 89 |
| LLaMA (13B) | 1 | 3-shot | 1069 | 99 |
| LLaMA (13B) | 2 | 0-shot | 386 | 31 |
| LLaMA (13B) | 2 | 3-shot | 1634 | 99 |
| Alpaca (7B) | 1 | - | 1799 | 76 |
| Alpaca (7B) | 2 | - | 2158 | 99 |

FIGURE 1.1 – Erreur moyenne de distance (en kilomètres) pour la prédiction de coordonnées géographiques et taux de Prédiction (P-Rate en %), indiquant la fréquence à laquelle les modèles génèrent une réponse avec des coordonnées pertinentes. Notamment, la méthode 3-shot peut présenter des résultats inférieurs car les exemples préalables peuvent influencer négativement sur la précision des prédictions [1]

Un des objectifs de ces études est notamment d'évaluer et comparer les différents modèles open-source pour établir des récurrences propres au LLM. Néanmoins les modèles utilisés dans ces études datent de 2022-2023 et sont limités dans les données qu'ils peuvent intégrer et utiliser, étant donné leur petite taille. Comme on peut le voir sur l'image ci-dessous, cela pose un problème au niveau de la qualité des résultats obtenus.

Pour obtenir des réponses plus précises, il semble pertinent de se tourner vers des LLM propriétaires, comme ceux d'OpenAI qui a développé des modèles plus complexes. En effet, d'autres essais ont été effectués sur le modèle ChatGPT-4 notamment dans la publication *GPT4GEO: How a Language Model Sees the World's Geography* [3]. Cette étude nous permet d'affiner nos hypothèses précédentes et d'obtenir une vision plus claire sur la dimension géographique des LLM.

L'article porte d'abord sur des tâches descriptives, allant de l'estimation d'indicateurs à la topographie mais également de la cartographie, avant de s'attaquer à des tâches basées sur des applications concrètes, telles que la navigation et la planification d'itinéraires.

Dans un premier test, les chercheurs ont évalué la compréhension par GPT-4 d'indicateurs socio-économiques à l'échelle nationale, notamment la population, l'espérance de vie et les émissions de carbone. Les résultats obtenus ont été relativement bons pour les deux premiers critères, avec une erreur moyenne relative à la vérité-terrain de 3,6% et 1,94% respectivement, mais nettement moins satisfaisants pour les émissions de carbone, avec une erreur moyenne de 21,91%. Il est à noter que le modèle commet davantage d'erreurs pour les pays peu peuplés, particulièrement en ce qui concerne les estimations de population.

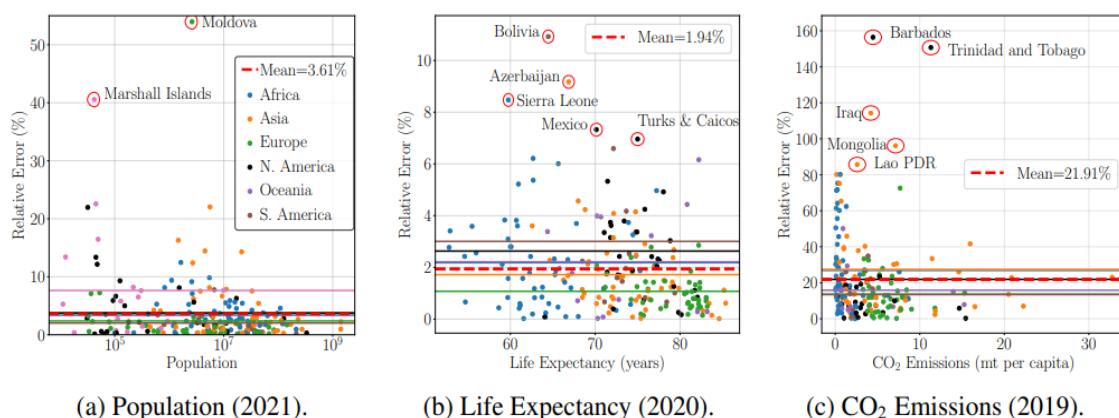


FIGURE 1.2 – Évaluation de la compréhension par GPT-4 d'indicateurs par pays tels que (a) le nombre d'habitants (b) les espérances de vie, et (c) les émissions de CO₂ par habitant. Les cercles rouges indiquent les données aberrantes [3]

Dans une autre approche, les chercheurs ont testé la capacité du LLM géospatial à fournir des données géospatiales brutes. Ils ont demandé à GPT-4 de prédire l'altitude le long de trois parcours différents situés dans les Alpes et le nord de l'Italie, en demandant des estimations d'élévation pour dix points uniformément espacés sur un tracé rectiligne (fig. 1.3).

Ces demandes ont été réitérées trois fois pour chacun des ensembles de points afin d'évaluer la cohérence des réponses. La vérification des prédictions de GPT-4 a été faite en utilisant le modèle numérique du terrain Copernicus comme référence pour obtenir les altitudes réelles de chaque point. Les résultats ont montré que, bien que GPT-4 ait une assez bonne compréhension des altitudes dans cette région spécifique, ses estimations manquaient de précision et ne pouvaient pas être utilisées dans une application concrète.

Dans l'objectif d'évaluer la capacité du modèle de langage à traiter et restituer des données brutes, les chercheurs ont adopté une démarche différente en lui demandant, via divers prompts, de définir les frontières administratives de certains continents et pays, ainsi que la localisation de rivières et de lacs, sur la

base de ses connaissances (fig. 1.4). Les résultats obtenus nous fournissent des indications sur la manière dont les modèles de langage comprennent l'espace géographique. Ils suggèrent que les réponses du modèle se fondent sur des repères majeurs connus, tels que des villes ou des éléments naturels (volcans, merveilles naturelles, montagnes...), pour établir des points de repère. Nous avons remarqué que le modèle identifie avec plus de facilité les frontières côtières, probablement parce que les villes côtières se trouvent invariablement sur ces frontières.

Cela est particulièrement vrai pour les États-Unis, dont les frontières est et ouest sont mieux délimitées que les frontières nord et sud. La densité des données disponibles joue également un rôle conséquent : plus une zone est pauvre en données détaillées, plus les résultats seront imprécis. Cela est illustré par les prédictions concernant les rivières, où l'on constate une différence notable entre la précision de la représentation de la Tamise, riche en détails, et celle de la Seine, simplifiée par une ligne droite. Cette observation souligne que les modèles sont encore fortement biaisés par les données d'entrée, favorisant le monde anglophone.

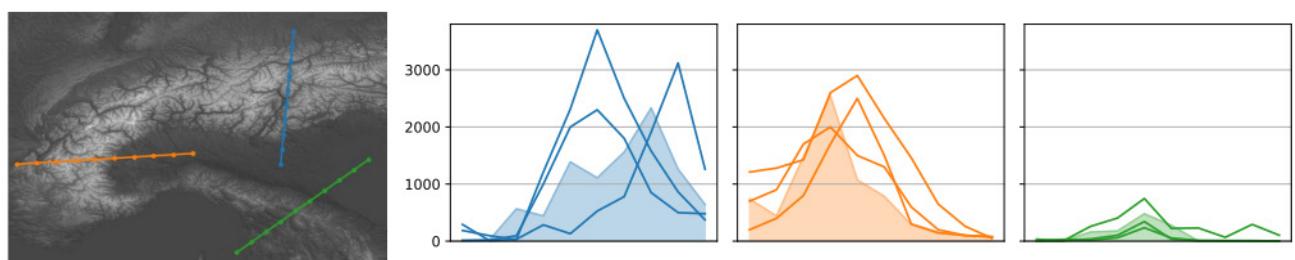


FIGURE 1.3 – Altitudes prévues (lignes) et réelles (zones ombrées) le long des trajectoires représentées à gauche (la région sous-jacente se trouve dans les Alpes, les zones claires représentent des altitudes plus élevées) [3]

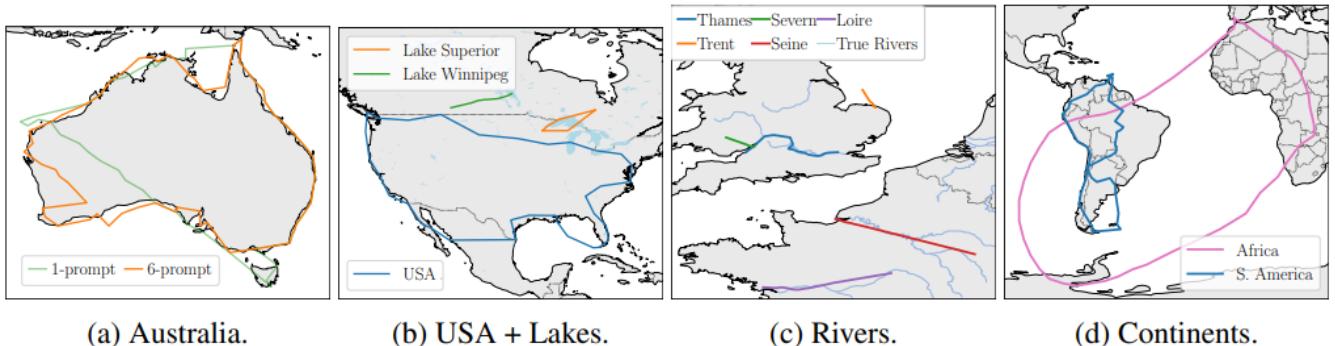


FIGURE 1.4 – Contours de divers éléments géographiques produits à l'aide des coordonnées fournies par GPT-4. L'affinement avec des prompts supplémentaires améliore les résultats [3]

Cette dernière expérience évalue la capacité de GPT-4 à comprendre et manipuler des données géospatiales dans des contextes connus et variables (fig. 1.5). En se concentrant sur la répartition des sous-espèces de tigres et les voies migratoires des oiseaux, les résultats montrent que le modèle identifie généralement correctement les zones géographiques. Cependant, il positionne parfois des habitats dans l'océan, révélant des difficultés pour reconnaître des incohérences logiques dans ses réponses. Les itinéraires migratoires, bien qu'approximatifs, indiquent que GPT-4 s'appuie sur des repères significatifs

et des connaissances partielles pour structurer les données géospatiales, adoptant une approche qui privilégie les informations vérifiables pour déduire les résultats dans les régions moins documentées. Cette observation suggère que, malgré un potentiel notable dans l'analyse des données géospatiales, les LLM nécessitent des instructions détaillées et pourraient gagner en précision avec l'intégration de bases de données géospatiales, afin d'améliorer la précision des résultats et fournir au modèle une vérité terrain d'actualité comme référence.

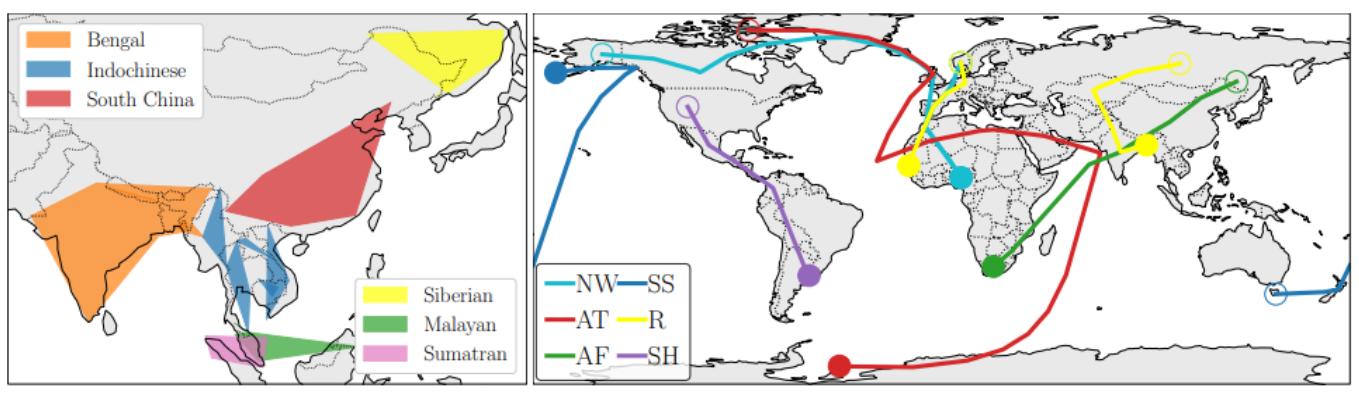


Figure 1.5- Aire de répartition des sous-espèces de tigres et routes migratoires des oiseaux estimées par ChatGPT-4.(Espèces d'oiseaux: NW=Traquet motteux, AT=Sterne arctique, AF=Faucon de l'Amour, SS=Puffin à bec grêle, R=Combattant varié, SH=Buse de Swainson) [3]

1.2 FINE TUNING ET GEOSPATIAL LOCATION EMBEDDING

Approches pour renforcer les connaissances géospatiales des LLM

Comme démontré précédemment, les LLM recueillent une quantité non négligeable d'informations géographiques. Cependant, l'exactitude de leurs réponses et leur pertinence restent problématiques pour des utilisations intolérantes à l'imprécision. En effet, les LLM que nous avons présentés ont tendance à inventer ou fournir un chiffre aléatoire pour répondre aux questions posées. Pour répondre à cette problématique, il est possible d'enrichir les connaissances géographiques d'un LLM de deux manières : en affinant (affinage) le modèle en l'entraînant davantage sur un jeu de données précis, ou en y connectant une base de données externe à laquelle le modèle pourra se référer.

Le fine-tuning d'un modèle pré-entraîné s'avère prometteur notamment pour les modèles les plus grands comme GPT-3.5. Dans l'exemple suivant, tiré de l'article *GeoLLM: extracting geospatial knowledge from large language models* [4], les chercheurs ont affiné plusieurs LLM (GPT-2, GPT-3.5, Llama 2, RoBERTa) et des modèles de machine learning (XGBoost, K-NN, MLP-BERT) avec plusieurs jeux de données provenant de sources différentes comme OpenStreetMap et WorldPop.

Nous pouvons voir, sur une série d'exercices, que GPT-3.5 présente les meilleurs résultats, car l'affinage s'adapte bien à la taille du modèle pré-entraîné.

| Task | Source | Samples | GPT-3.5 | Llama 2 | RoBERTa | GPT-2 | MLP-BERT | XGBoost-FT | XGBoost | k-NN | Nightlight |
|----------------|----------|---------|-------------|---------|---------|-------|----------|------------|---------|------|------------|
| Population | WorldPop | 10,000 | 1.02 | 1.12 | 1.44 | 2.32 | 1.49 | 1.56 | 1.65 | 1.81 | 1.28 |
| | | 1,000 | 1.13 | 1.40 | 1.75 | 2.69 | 1.81 | 1.88 | 1.92 | 2.26 | 1.38 |
| | | 100 | 1.38 | 2.22 | 2.47 | 3.86 | 2.48 | 2.26 | 2.18 | 2.48 | 1.55 |
| Asset Wealth | DHS | 10,000 | 0.94 | 0.97 | 1.06 | 1.64 | 1.22 | 1.19 | 1.20 | 1.21 | 1.34 |
| | | 1,000 | 1.02 | 1.24 | 1.33 | 1.94 | 1.44 | 1.48 | 1.45 | 1.51 | 1.42 |
| Women Edu | DHS | 10,000 | 0.89 | 0.92 | 0.98 | 1.36 | 1.09 | 1.00 | 1.01 | 1.00 | 1.27 |
| | | 1,000 | 0.95 | 1.08 | 1.13 | 3.62 | 1.14 | 1.23 | 1.19 | 1.17 | 1.30 |
| Sanitation | DHS | 10,000 | 1.44 | 1.49 | 1.46 | 2.44 | 1.60 | 1.57 | 1.64 | 1.57 | 1.96 |
| | | 1,000 | 1.46 | 1.86 | 1.81 | 2.98 | 1.93 | 1.96 | 1.93 | 1.90 | 2.08 |
| Women BMI | DHS | 10,000 | 0.37 | 0.37 | 0.39 | 0.50 | 0.41 | 0.39 | 0.39 | 0.39 | 0.51 |
| | | 1,000 | 0.40 | 0.41 | 0.45 | 0.60 | 0.47 | 0.46 | 0.45 | 0.43 | 0.54 |
| Population | USCB | 1,000 | 1.03 | 1.64 | 1.90 | 2.59 | 1.97 | 2.06 | 2.01 | 2.11 | 1.63 |
| | | 100 | 1.35 | 1.98 | 2.50 | 3.07 | 2.15 | 2.27 | 2.24 | 2.34 | 1.82 |
| Mean Income | USCB | 1,000 | 1.41 | 1.74 | 2.09 | 3.24 | 2.17 | 2.19 | 2.14 | 2.05 | 2.32 |
| | | 100 | 1.61 | 2.41 | 2.44 | 4.28 | 2.36 | 2.48 | 2.53 | 2.38 | 2.48 |
| Hispanic Ratio | USCB | 1,000 | 1.11 | 1.46 | 1.52 | 3.05 | 1.66 | 1.60 | 1.48 | 1.45 | 2.02 |
| | | 100 | 1.30 | 2.43 | 2.45 | 3.60 | 1.94 | 2.13 | 2.11 | 1.78 | 2.27 |
| Home Value | Zillow | 1,000 | 0.74 | 1.12 | 1.67 | 2.82 | 1.72 | 1.74 | 1.63 | 1.44 | 2.23 |
| | | 100 | 1.11 | 1.81 | 2.51 | 3.26 | 1.94 | 2.29 | 2.03 | 1.81 | 2.44 |

FIGURE 1.6 – Coefficient de détermination linéaire de Pearson pour tous les modèles, pour toutes les tâches et toutes les tailles d'échantillons d'entraînement [4]

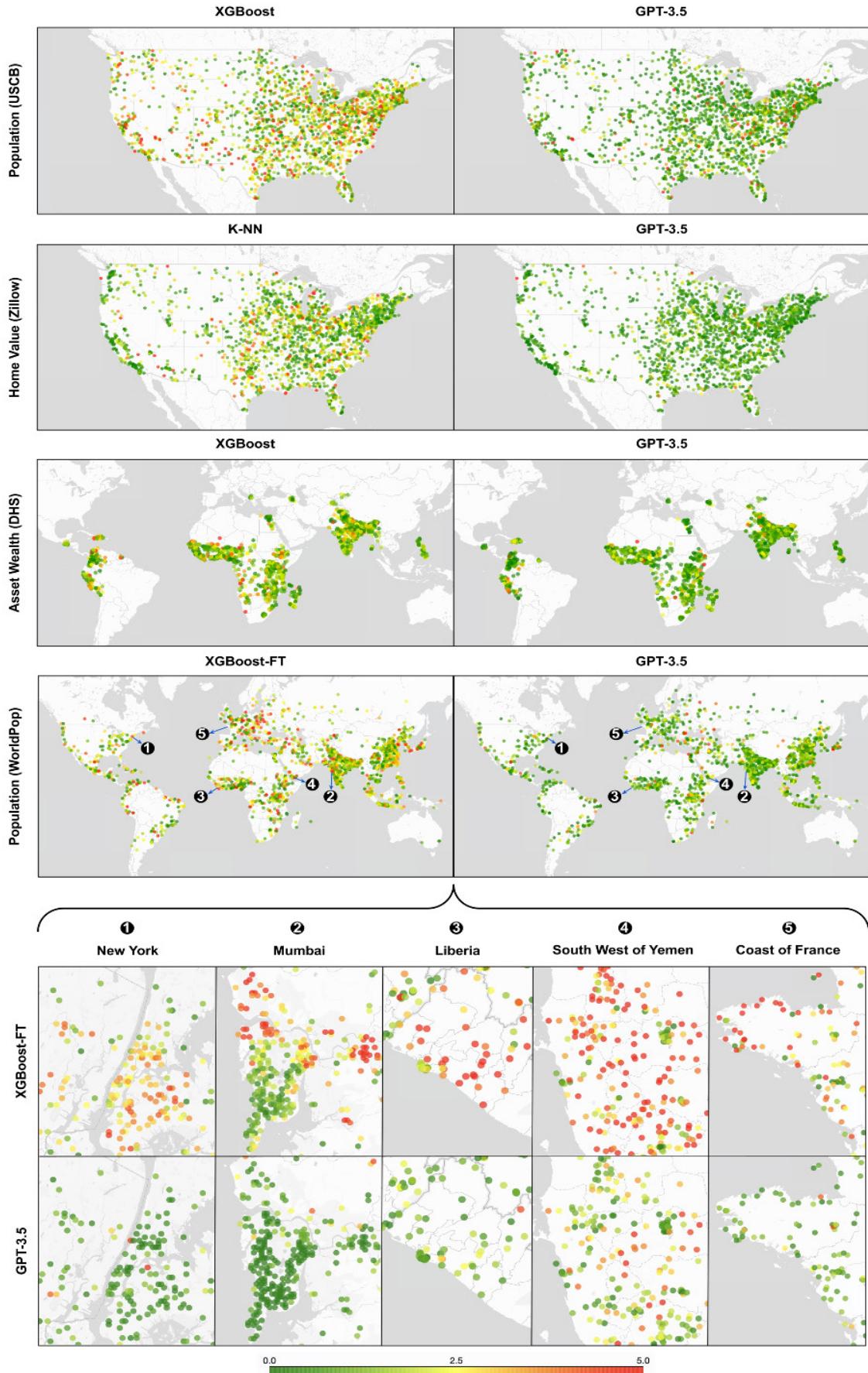


FIGURE 1.7 – Représentation graphique de l'erreur absolue comparant les machines learning models et GPT-3.5 sur l'exécution de tâches uniques à partir de plusieurs sources de données (1000 échantillons par source de données) [4]

Geospatial Location Embedding

Le *Geospatial Location Embedding* (GLE) est une technique qui permet d'aider les LLM à assimiler et analyser la donnée géospatiale. L'objectif serait de créer un *Spatial Foundational Model* qui est entraîné sur des données textuelles et spatiales en parallèle pour acquérir une compréhension approfondie et intégrée de l'information géographique dans le contexte des requêtes basées sur la localisation. L'incorporation du *Geospatial Location Embedding* dans un *Spatial Foundational Model* vise à enrichir les représentations vectorielles normalisées utilisées par les LLM, en leur permettant de comprendre les nuances et les complexités des données spatiales.

En se basant sur la méthodologie de MapGPT [5], l'objectif de GLE est d'encoder les attributs spatiaux, tels que les coordonnées géographiques, les distances, et les relations spatiales entre les lieux, dans une forme qui est compréhensible pour le LLM. Cela implique de créer des vecteurs qui intègrent à la fois des informations de localisation fixes et des attributs dynamiques, reflétant les événements ou les changements au sein d'un lieu spécifique. Ces vecteurs permettent au modèle de comprendre non seulement

le lieu mentionné dans une requête, mais aussi le contexte spatial et temporel associé, facilitant ainsi des réponses plus informées et précises. La combinaison des vecteurs de texte et de localisation géospatiale permettrait au *Spatial Foundational Model* de traiter de manière efficace les requêtes complexes qui nécessitent une compréhension à la fois du langage naturel et des données spatiales. Par exemple, le modèle pourrait non seulement identifier les endroits les plus touristiques dans une région donnée mais aussi comprendre et analyser les tendances spatiales, comme les mouvements de population ou les changements environnementaux, à partir d'une requête en langage naturel.

Néanmoins, l'intégration de GLE dans un *Spatial Foundational Model* présente des limites, notamment en termes de normalisation des représentations vectorielles spatiales pour qu'ils soient compatibles avec les architectures de LLM existantes. De plus, le fine-tuning d'un tel modèle nécessite un ensemble de données d'entraînement qui comprend une riche variété d'informations spatiales et textuelles, posant des défis relatifs à la collecte, à la gestion et à la protection de ces données.

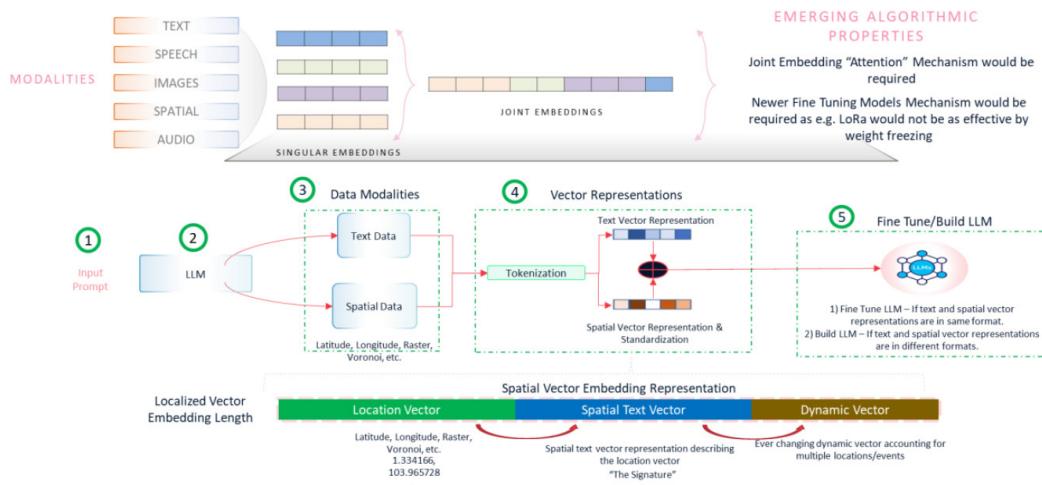


FIGURE 1.8 – Architecture technique proposée pour MapGPT avec intégration des vecteurs spatiaux avec des vecteurs textuels [5]

1.3 COMPRÉHENSION DES PRINCIPES SIG

Évaluation des connaissances théoriques sur les systèmes d'information géographiques

Le dernier point concernant l'analyse de la dimension géographique innée des LLM correspond à leur connaissance des principes et notions des systèmes d'informations géographiques (SIG), c'est-à-dire leur connaissance théorique sur les méthodes actuelles de traitement de la géodonnée. En effet, GPT-3.5 et GPT-4 sont déjà capables de passer l'examen du barreau aux États-Unis [6], obtenant des résultats similaires voire supérieurs à ceux de candidats humains.

Dans le même ordre d'idée, Peter Mooney et son équipe ont examiné GPT-3.5 et GPT-4 sur les principes fondamentaux des SIG [7] en leur posant des questions tirées d'un manuel scolaire intitulé *GIS Fundamentals: A First Text on Geographic Information Systems, 7th edition*. Les résultats montrent que GPT-3.5 est capable d'obtenir 63% de

réponses correctes, équivalent à un D sur le système de notation américain, tandis que GPT-4 obtient 88% de bonnes réponses, l'équivalent d'un B+. Cette différence s'explique notamment par le fait que GPT-4 est capable de traiter des données tabulaires et de résoudre des problèmes mathématiques à l'aide de son module code interpreter (voir page 40) qui lui permet de générer du code plutôt que d'halluciner une réponse aléatoire.

Dans tous les cas, ces résultats sont encourageants et démontrent les capacités de ces LLM dans la compréhension des théories et méthodes de base relatives aux SIG. Néanmoins, même si les résultats sont bons, ils ne sont pas exceptionnels, et certaines tâches plus complexes restent problématiques.

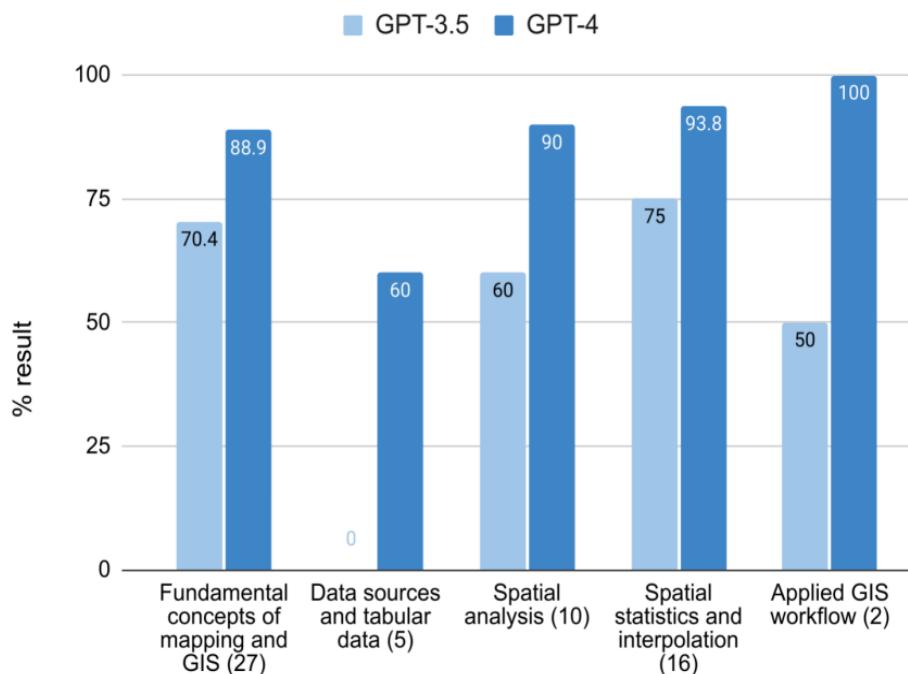


FIGURE 1.9 – Résultats des modèles en % sur les différentes catégories de l'examen [7]

CHAPITRE 2 :

LES LLM DANS

L'INTERPRÉTATION,

LA COMPRÉHENSION

ET L'EXÉCUTION

DE TÂCHES DE

GÉOTRAITEMENT

2.1 TRADUCTION DES COMMANDES HUMAINES EN CODE

Capacité des LLM à générer du langage informatique pour télécharger, traiter et visualiser la donnée géospatiale à partir d'une requête

Comme nous l'avons observé précédemment, les grands modèles de langage (LLM) offrent déjà un accès à certaines données géospatiales souvent intégrées durant la phase d'entraînement ou ajoutées ultérieurement lors d'un affinage. Cependant, la requête, le traitement et la visualisation de données spatiales ne peuvent pas être effectués directement en langage naturel.

En exploitant les capacités de modèles avancés comme GPT-4 dans la génération de code, des chercheurs ont réussi à convertir des requêtes formulées en langage naturel en instructions techniques précises, souvent sous forme de requêtes SQL ou SPARQL [8]. Autrement dit, ils ont développé une méthode pour convertir une requête formulée en anglais en un code utilisable par l'utilisateur ou le modèle de langage. Ce code permet ensuite d'appeler et de traiter des données géospatiales hébergées dans une base de données. L'exemple suivant met en avant un modèle affiné

de GPT-4 dont l'objectif est de servir d'intermédiaire entre un utilisateur et une base de données PostGIS. Cela permet à l'utilisateur de formuler une question géospatiale et de recevoir une réponse en langage naturel. Le LLM est ensuite responsable de la génération de requêtes SQL, de l'exécution de ces requêtes dans la base de données en back-end, et de l'analyse des résultats pour fournir une réponse pertinente.

Cette méthode démontre la capacité des LLM à créer du code SQL précis pour réaliser des analyses géospatiales de manière automatisée.

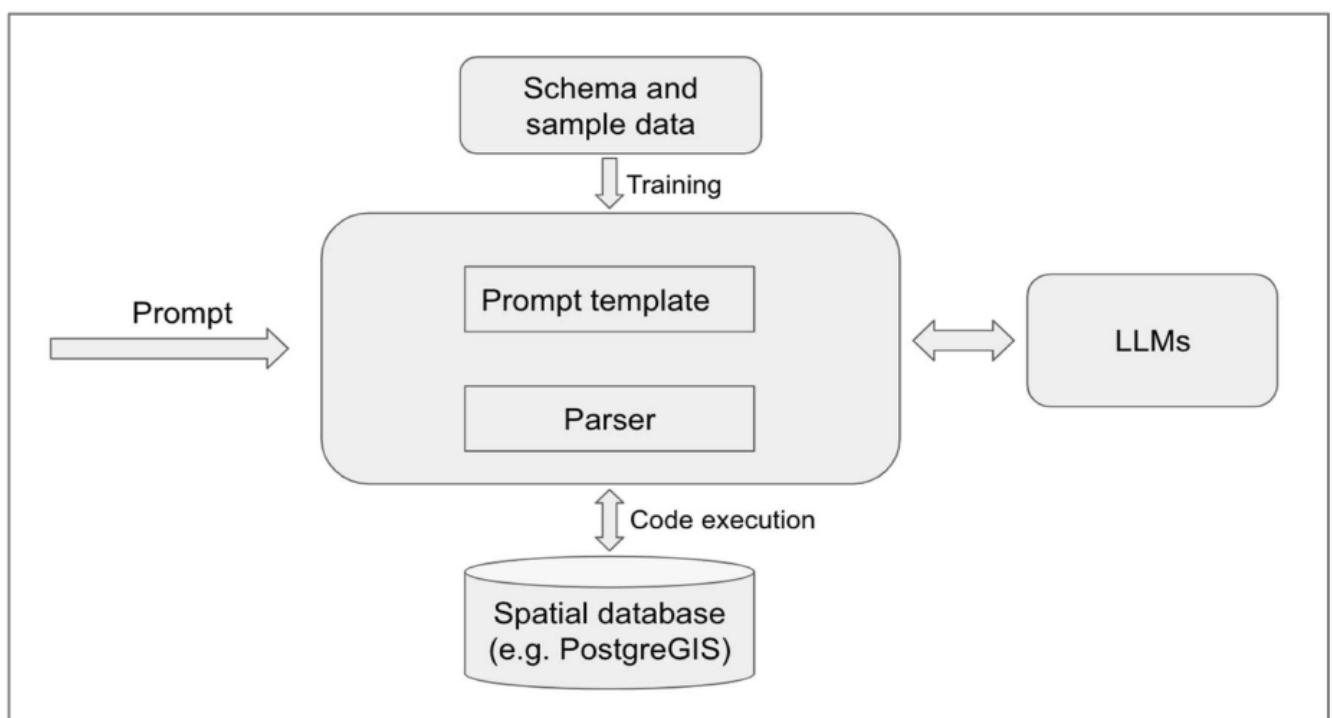


FIGURE 2.1 – Framework proposé pour la génération de code SQL avec un LLM pour la requête de données géospatiales [8]

SPARQL

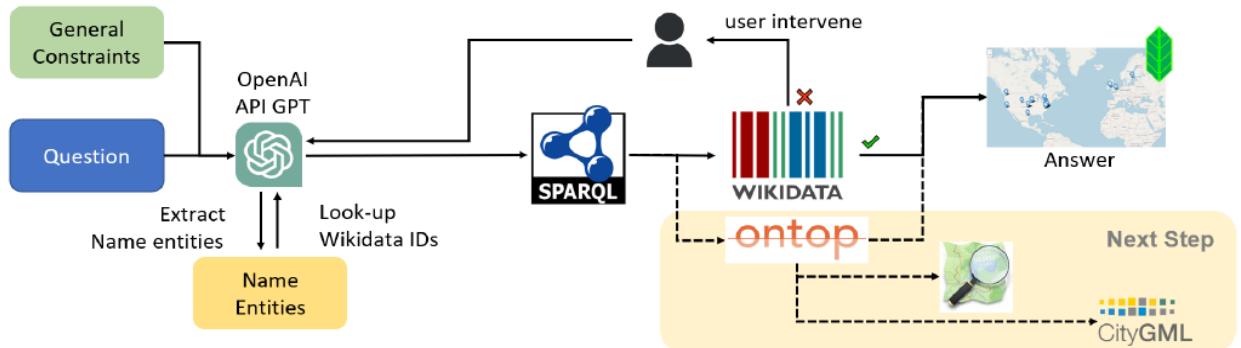


FIGURE 2.2 – Framework proposé pour la génération de code SPARQL avec un LLM pour la requête de données géospatiales [9]

ChatGPT peut aussi générer et comprendre d'autres langages de requête, dont SPARQL. Contrairement à SQL, utilisé pour exploiter les bases de données relationnelles présentant des tables structurées et des schémas prédéfinis, SPARQL permet d'exploiter des bases de données dont les informations sont stockées sous forme de graphes. Les auteurs de l'étude *GeoQAMap - Geographic Question Answering with Maps Leveraging LLM and Open Knowledge Base* [9] ont développé un système capable de

générer des réponses sous forme de carte interactive à partir d'une question en langage naturel. Ce système oblige le LLM à traduire les questions en requêtes SPARQL et à chercher les informations pertinentes dans des bases de données externes comme Wikidata. Cette expérience démontre la polyvalence des modèles de langage tels que ChatGPT, capables d'être affinés pour chercher des informations dans divers types de bases de données en utilisant une variété de langages de requête.

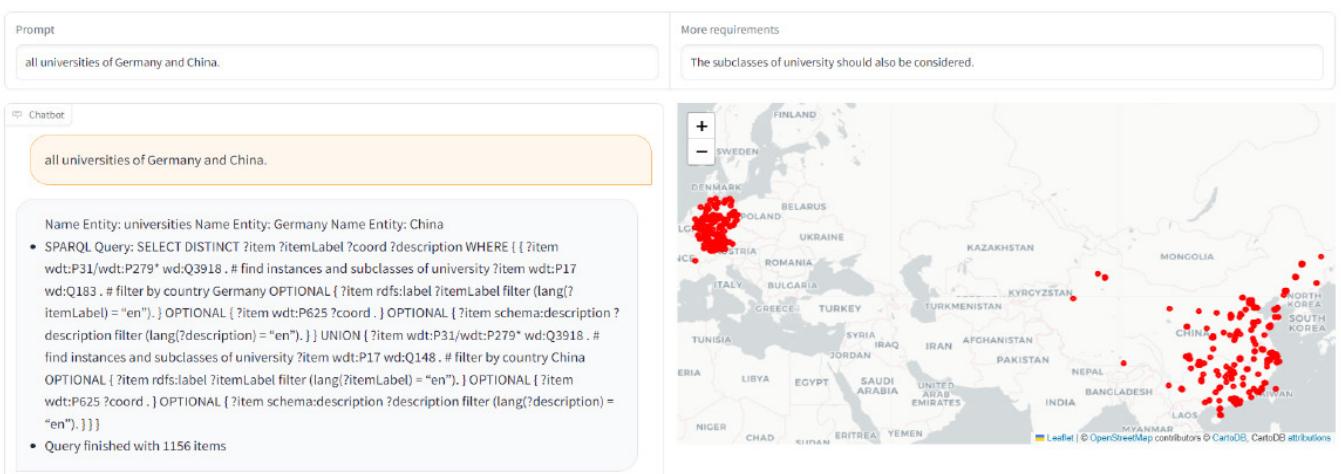


FIGURE 2.3 – Interface GeoQAMap [9]

Langage de Génération de Scénarios (SGL)

Pour rendre l'analyse géospatiale plus accessible et réduire la nécessité pour les utilisateurs d'avoir une expertise en systèmes d'information géographique (SIG) ou en langages comme SQL, les chercheurs Jonathan Frez et Nelson Baloian proposent l'utilisation d'un Langage de Génération de Scénarios (SGL) [10].

L'utilisation du SGL vise à simplifier l'analyse géospatiale et les processus décisionnels, le rendant particulièrement accessible et ouvrant ainsi l'analyse géospatiale à un public plus large, incluant les décisionnaires, les

planificateurs urbains et autres professionnels non-spécialistes en SIG. Le SGL, qui emploie un langage plus proche du langage naturel, facilite la formulation de requêtes et offre également la possibilité de générer des visualisations telles que des cartes interactives, améliorant ainsi l'interprétation des résultats. L'adoption du SGL permet de normaliser la manière dont les scénarios géospatiaux sont décrits au sein d'un projet ou d'une organisation, favorisant ainsi le partage, la compréhension et la reproduction des analyses.

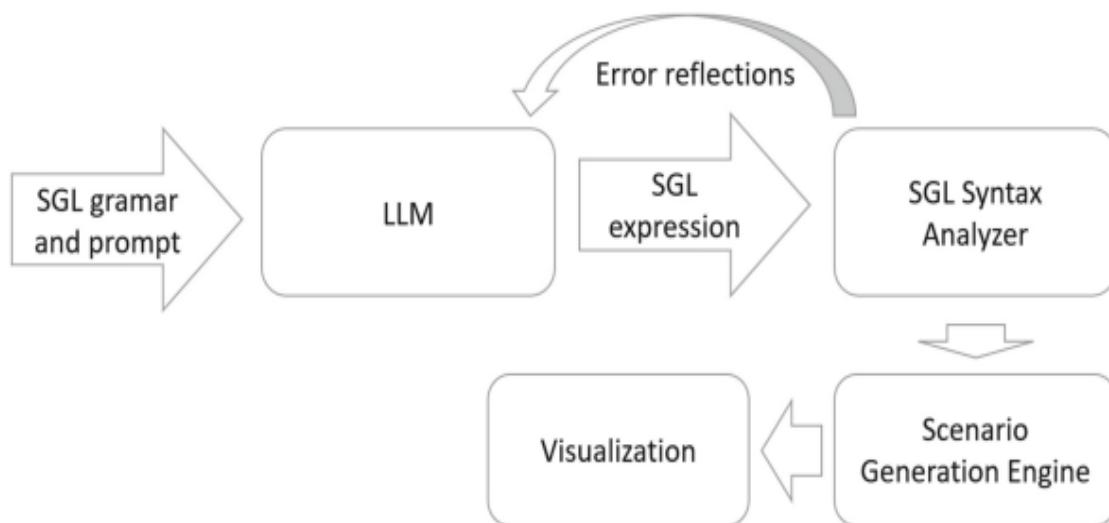


FIGURE 2.4 – Diagramme du processus de génération de scénarios [10]

2.2 OPÉRATIONS SÉQUENTIELLES

Solutions permettant au LLM de traiter les données géospatiales de manière séquentielle

Traditionnellement, l'analyse géospatiale a été une discipline presque exclusivement réservée aux experts en systèmes d'information géographique (SIG), utilisant des logiciels spécialisés pour importer, traiter et visualiser les données géospatiales. Cette tâche complexe implique de nombreuses sous-tâches, telles que le géotraitement de diverses géodonnées provenant de multiples sources. L'expert doit souvent planifier ses opérations et jeux de données bien à l'avance ou procéder par essais et erreurs jusqu'à obtenir les résultats souhaités. Toutefois, les progrès dans le domaine de la Geo IA ont commencé à automatiser certaines de ces opérations. Avec l'émergence de modèles de langage à grande échelle (LLM) comme GPT-4, qui montrent une capacité de raisonnement avancée, il est désormais possible de penser à automatiser l'analyse géospatiale pour la rendre accessible même aux

non-spécialistes. Cette automatisation s'appuie sur la faculté des LLM à interpréter et comprendre le langage naturel, à générer et exécuter du code (dans le cas de GPT-4) et à décomposer un problème complexe en plusieurs sous-tâches pour les traiter de manière séquentielle.

Dans l'exemple de GeoGPT, un modèle affiné de GPT-3.5 Turbo développé par Yifan Zhang et ses collègues, le modèle peut accéder à un ensemble d'outils pour réaliser son analyse géospatiale et répondre aux requêtes de l'utilisateur [11]. Afin de permettre à GeoGPT de saisir les capacités spécifiques de chaque outil, les chercheurs ont fourni pour chacun une description détaillée de leurs fonctionnalités. Ainsi, GeoGPT est capable de solliciter des données externes, de les traiter et de les visualiser de manière autonome en utilisant les outils adaptés pour chaque tâche.

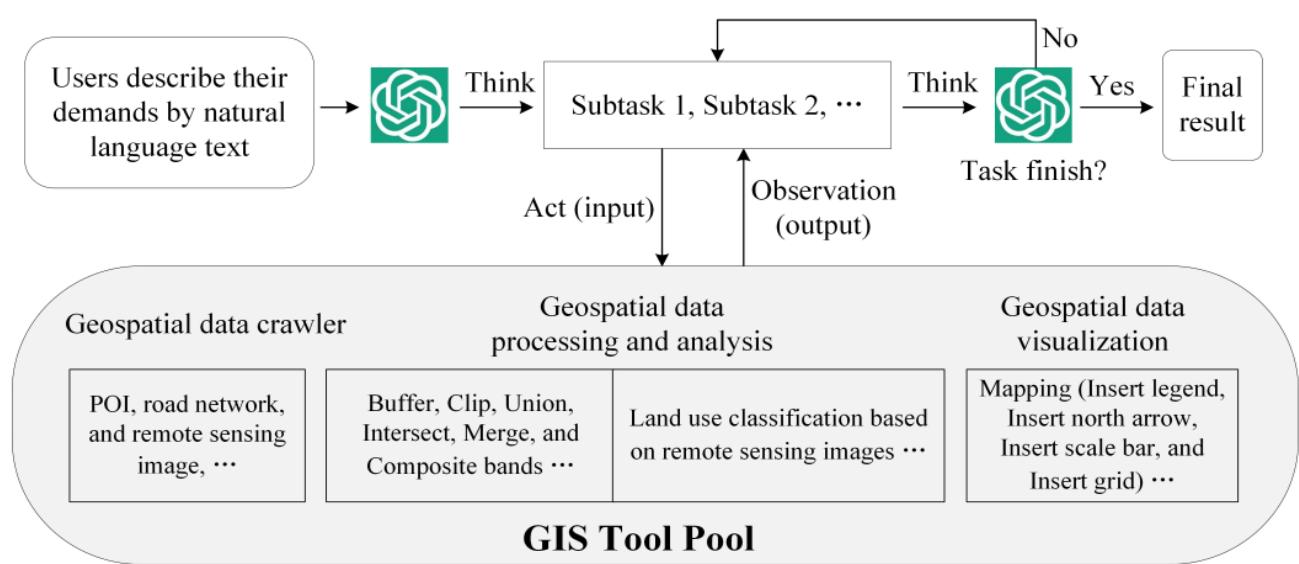


FIGURE 2.5 – Framework de GeoGPT. Un LLM est utilisé pour interpréter les demandes en langage naturel, et choisit et appelle les outils appropriés pour résoudre les tâches géospatiales [11]

Exécution en séquentiel à partir d'un graphe

Nous pouvons exploiter les capacités de programmation de GPT-4.0 pour générer, à l'aide du langage Python, des graphes où chaque nœud représente une donnée ou une opération de géotraitement. Ces graphes, en conjonction avec des agents intelligents, permettraient au LLM de suivre une séquence logique d'étapes pour la résolution de tâches géospatiales complexes. Cette méthode est présentée dans l'article scientifique *Autonomous GIS: the next-generation AI-powered GIS* [12], qui décrit le développement et l'essai d'un système

prototype nommé [LLM-Geo](#), conçu pour la génération autonome de graphes et leur utilisation comme guide pour coder une série d'opérations pour la requête, le traitement et la visualisation des données géospatiales. Ce prototype est décrit comme une première étape dans la création d'un système d'information géographique (SIG) autonome capable de générer et vérifier les données d'entrée et de sortie, d'exécuter des tâches géospatiales de manière séquentielle et d'apprendre et d'évoluer au fil du temps.

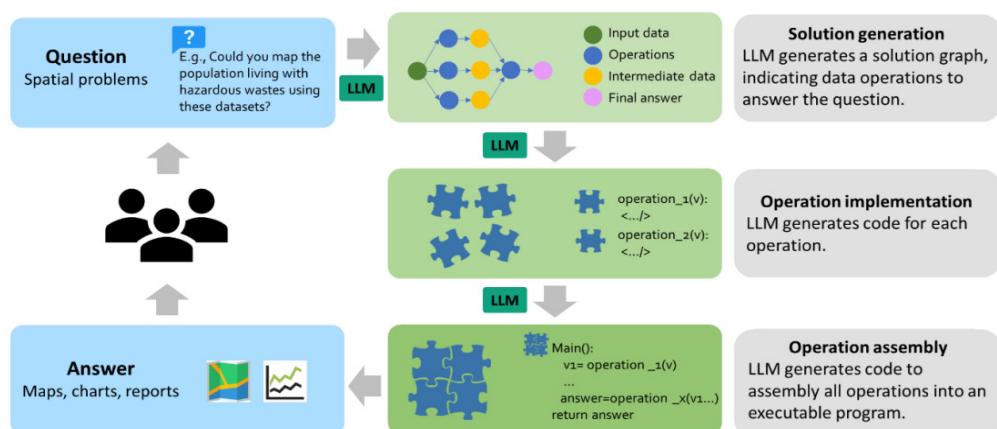


FIGURE 2.6 – Diagramme de flux de LLM-Geo [12]

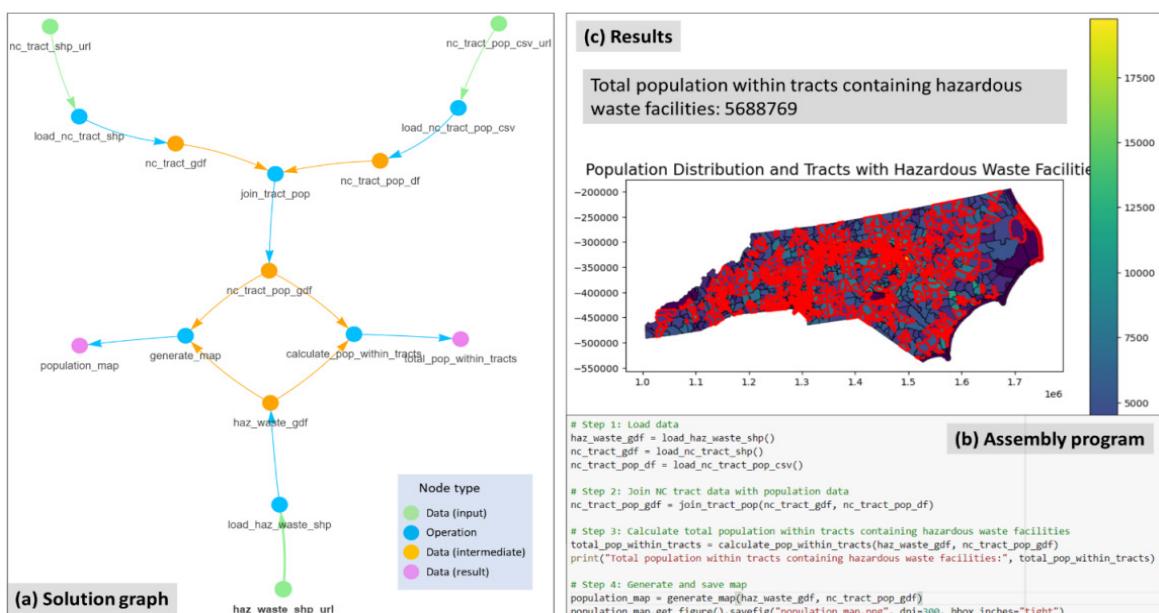


FIGURE 2.7 – Résultats générés automatiquement par LLM-Geo pour le dénombrement des populations vivant à proximité de déchets dangereux. (a) Graphe, (b) programme d'assemblage (codes Python), et (c) décompte de la population et carte générée [12]

2.3 AUTOMATISATION DE L'ANALYSE GÉOSPATIALE

Assistants intelligents et SIG autonomes

Jusqu'à présent, nous avons étudié différentes méthodes pour automatiser l'analyse géospatiale à partir d'une requête. Le LLM est en mesure de saisir la demande formulée en langage naturel, d'accéder aux données géospatiales, de réaliser une série de géotraitements et de présenter les résultats soit en langage naturel, soit sous forme de carte. Les mêmes chercheurs qui ont développé LLM-Geo, proposent deux options pour l'apparence de l'interface graphique des assistants SIG autonomes: une

plateforme indépendante ou une extension semblable à [Microsoft Copilot](#) intégrée dans un SIG existant (fig. 2.8). Nous pouvons aussi suggérer ChatGPT comme interface utilisateur, grâce à l'introduction des GPTs, qui permettent à tout utilisateur d'affiner GPT-4 sans code en lui fournissant des documents et des instructions précises. De plus, GPT-4 est capable de traiter des formats géospatiaux comme GeoJSON et à générer des cartes directement dans l'interface de chat.

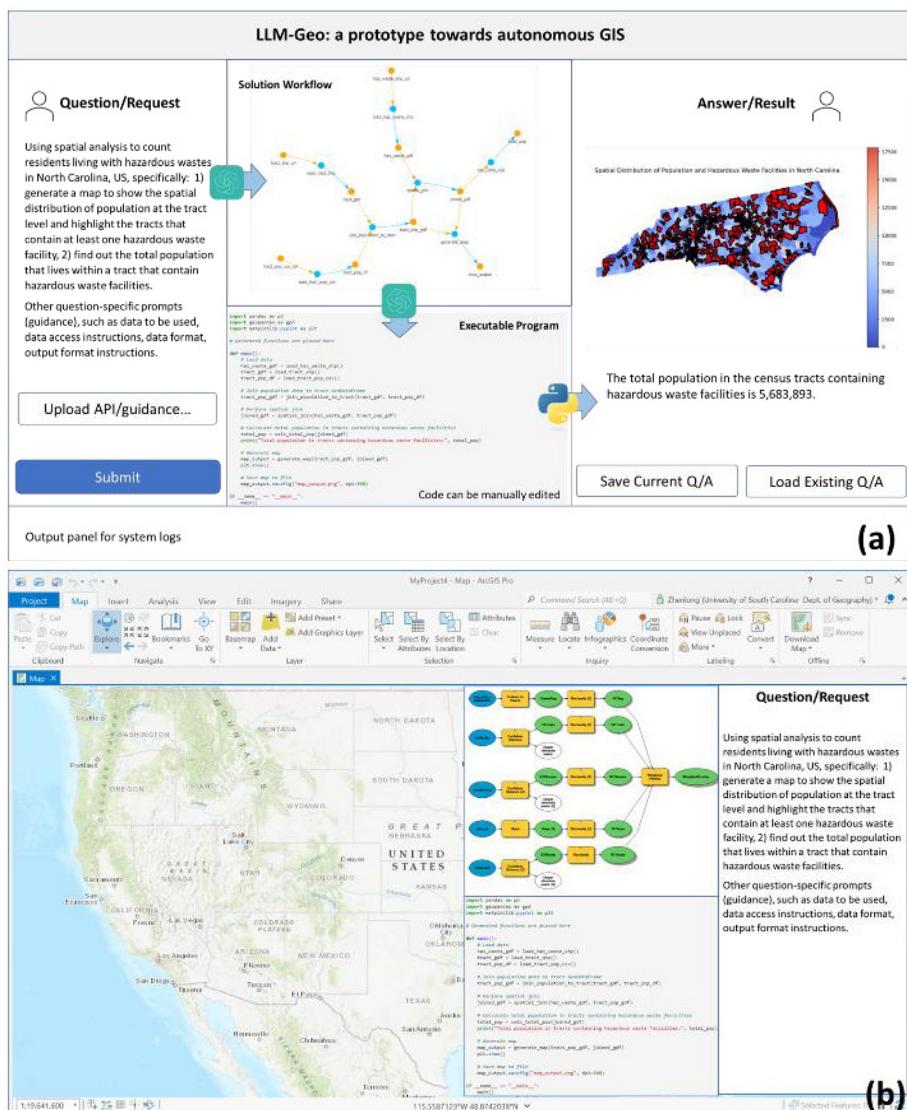


FIGURE 2.8 – Illustration de l'interface utilisateur pour (a) un SIG autonome, (b) un SIG autonome de type *Copilot* intégré dans un logiciel SIG existant (par exemple, ArcGIS Pro) [12]

ChatGPT et GPTs

Dans la vidéo ChatGPT make me a map de Franz's blog [13], le créateur nous montre comment ChatGPT peut lire un fichier .zip qui contient des shapefiles grâce à son module *code interpreter*. À partir de ces données, le créateur va pouvoir générer et éditer sa carte à l'aide de *prompts*.

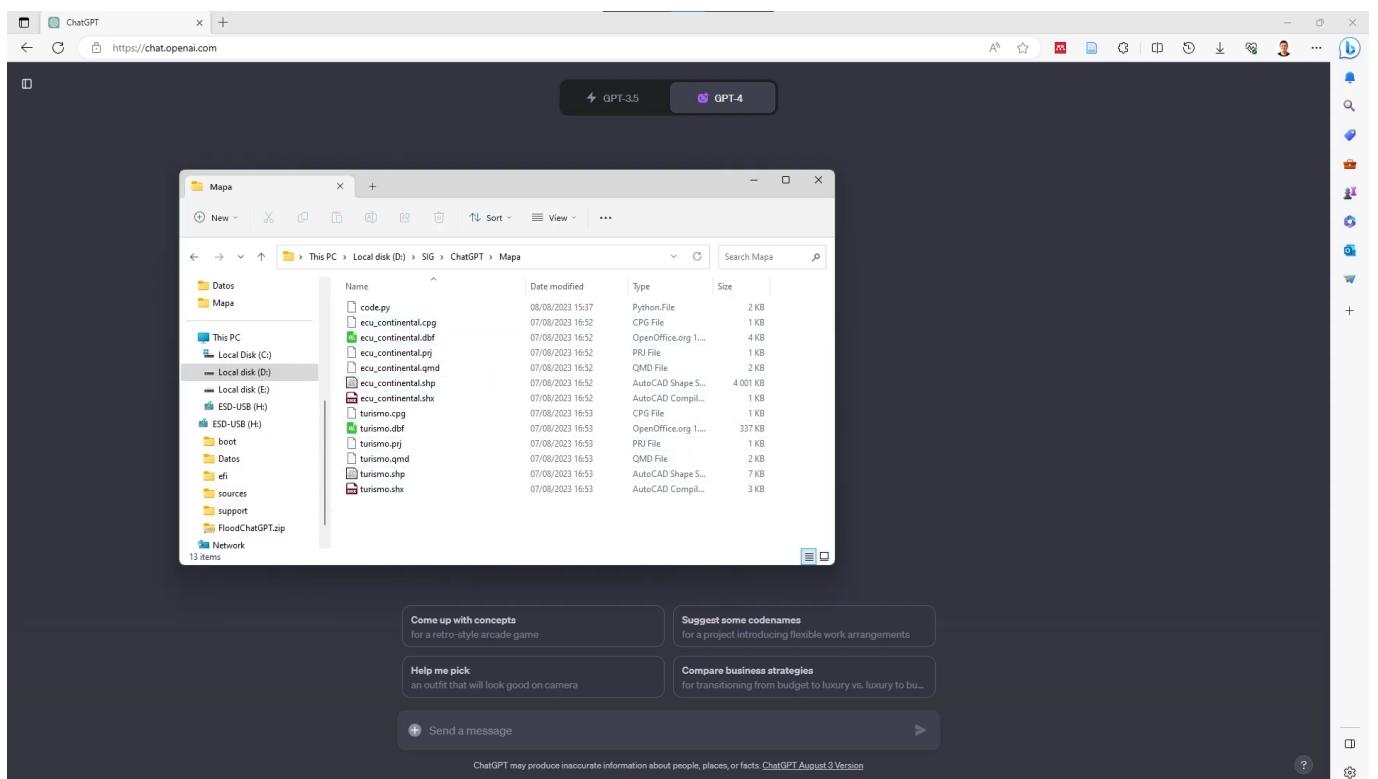


FIGURE 2.9 – Chargement des shapefiles dans ChatGPT [13]

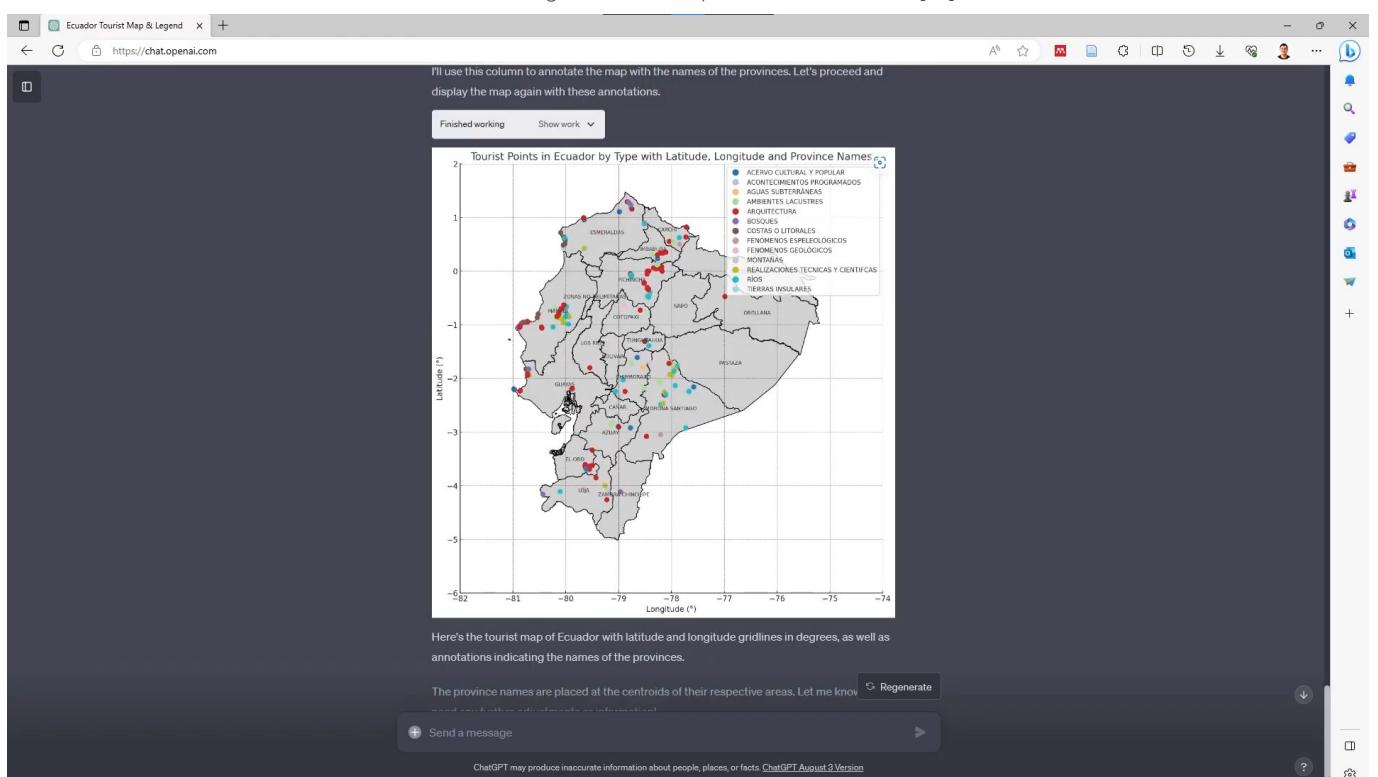


FIGURE 2.10 – Visualisation de la carte au sein de l'interface chat de ChatGPT [13]

Geo+

Conçu par Ageospatial, Geo+ est un GPT disponible sur le GPT Store d'OpenAI [14]. Geo+ permet d'appeler de la donnée géospatiale open-source, par exemple l'imagerie satellite Sentinel-2. Geo+ peut aussi effectuer des géotraitements simples, comme un clip à partir d'une limite administrative, et proposer des liens de téléchargement de la donnée, par exemple en format GeoParquet pour les empreintes des bâtiments. L'assistant est également capable de conseiller l'utilisateur pour des tâches géospatiales sur un SIG donné grâce à un grand nombre de données supplémentaires fournies au modèle.

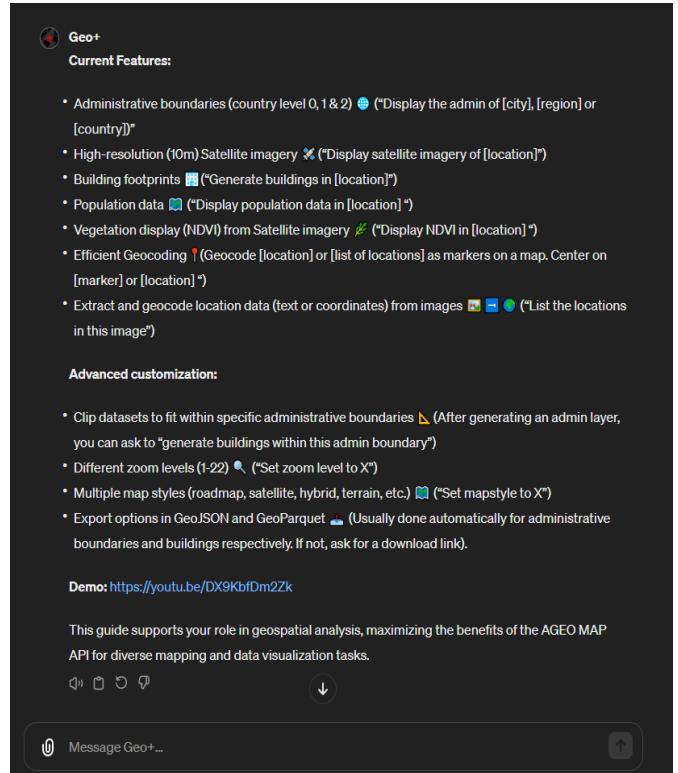


FIGURE 2.11 – Liste des fonctionnalités de Geo+ [14]

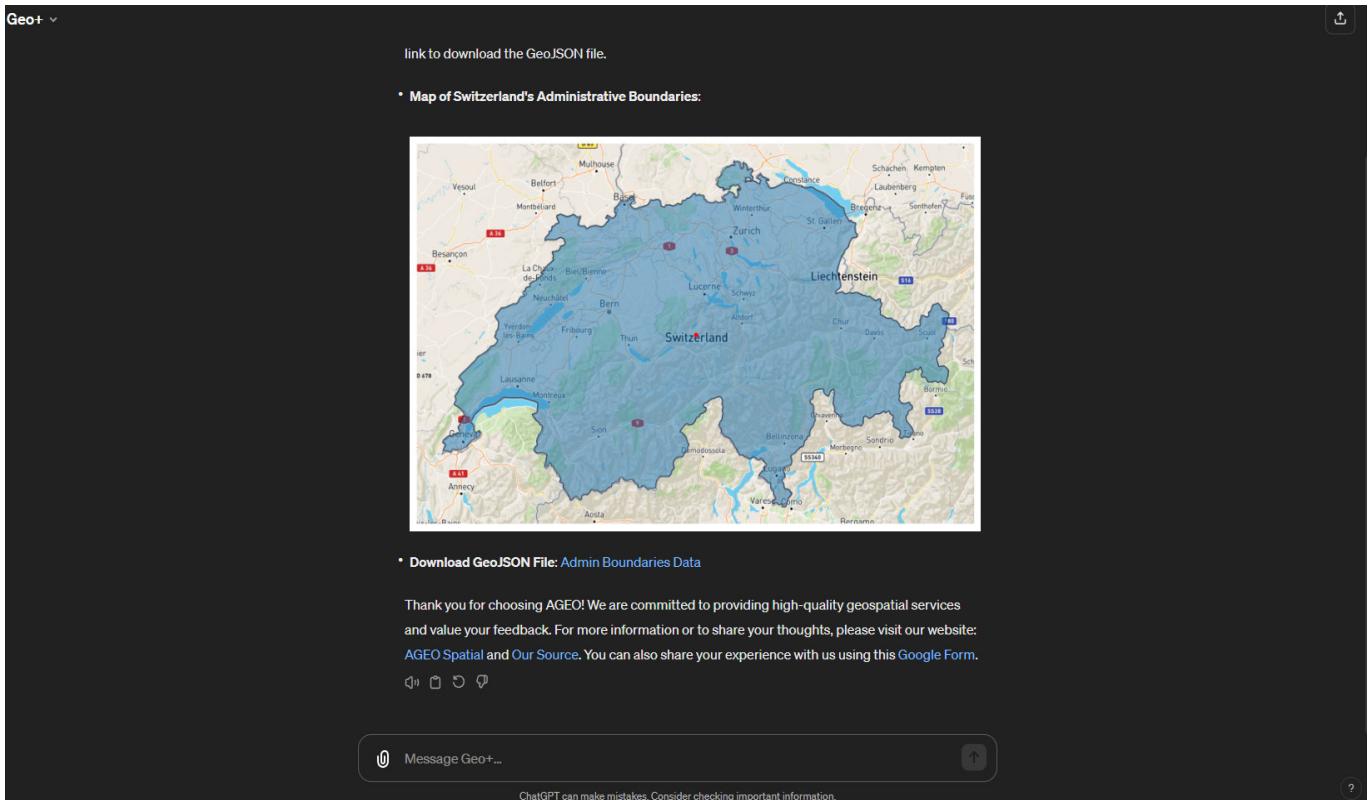


FIGURE 2.12 – Génération de frontières administratives avec Geo+ [14]

Assistants IA de type Microsoft Copilot

Le 14 février 2024, lors de la conférence ESRI Federal GIS, ESRI a annoncé le développement d'un assistant IA destiné à être intégré dans leur gamme de produits, tels que ArcGIS Hub et Survey123 [15]. À ce stade, ESRI n'a pas révélé quel LLM est utilisé pour cet assistant IA, ni si c'est un modèle qu'ils ont développé en interne ou s'ils ont affiné un modèle pré-entraîné.

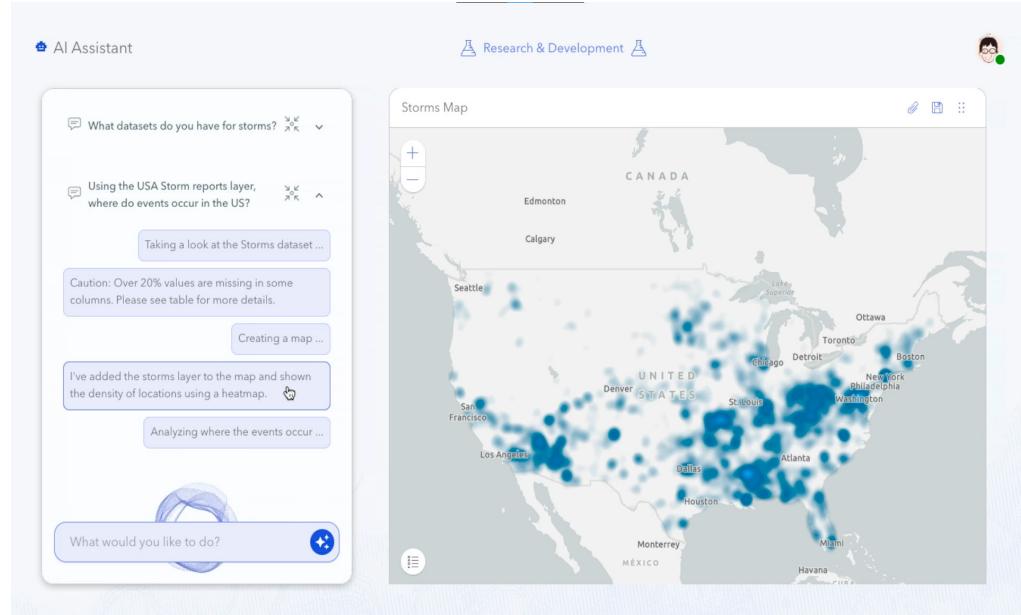


FIGURE 2.13 – ArcGIS experience builder, ArcGIS and Generative AI Assistants [15]

Dans la communauté open-source, des créateurs de plugins pour QGIS offrent déjà des assistants IA, intégrés directement dans l'interface utilisateur de QGIS [16]. Par exemple, un assistant appelé QGPT Agent est capable d'importer automatiquement des données d'OpenStreetMap (OSM) directement dans QGIS.

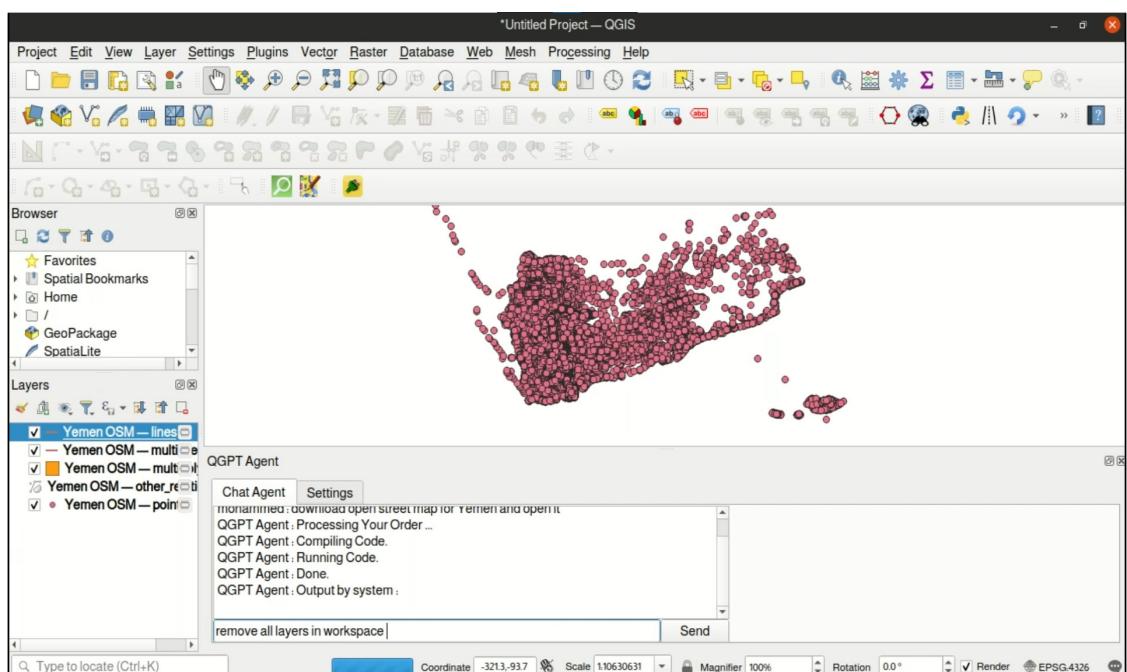


FIGURE 2.14 – Interface utilisateur de QGPT Agent [16]

Plateformes indépendantes

Afin de conserver une meilleure maîtrise sur la technologie et le LLM, des startups et chercheurs développent des plateformes indépendantes visant à automatiser l'analyse géospatiale. Geoforge est une plateforme web développée par Ageospatial capable d'appeler et traiter un grand nombre de données géospatiales open-source à partir du prompt de l'utilisateur [17]. Pour traiter la donnée vectorielle ou matricielle, le LLM génère automatiquement un graphe semblable à celui de LLM-Geo (voir page 22) et réalise des géotraitements séquentiellement et de manière autonome.

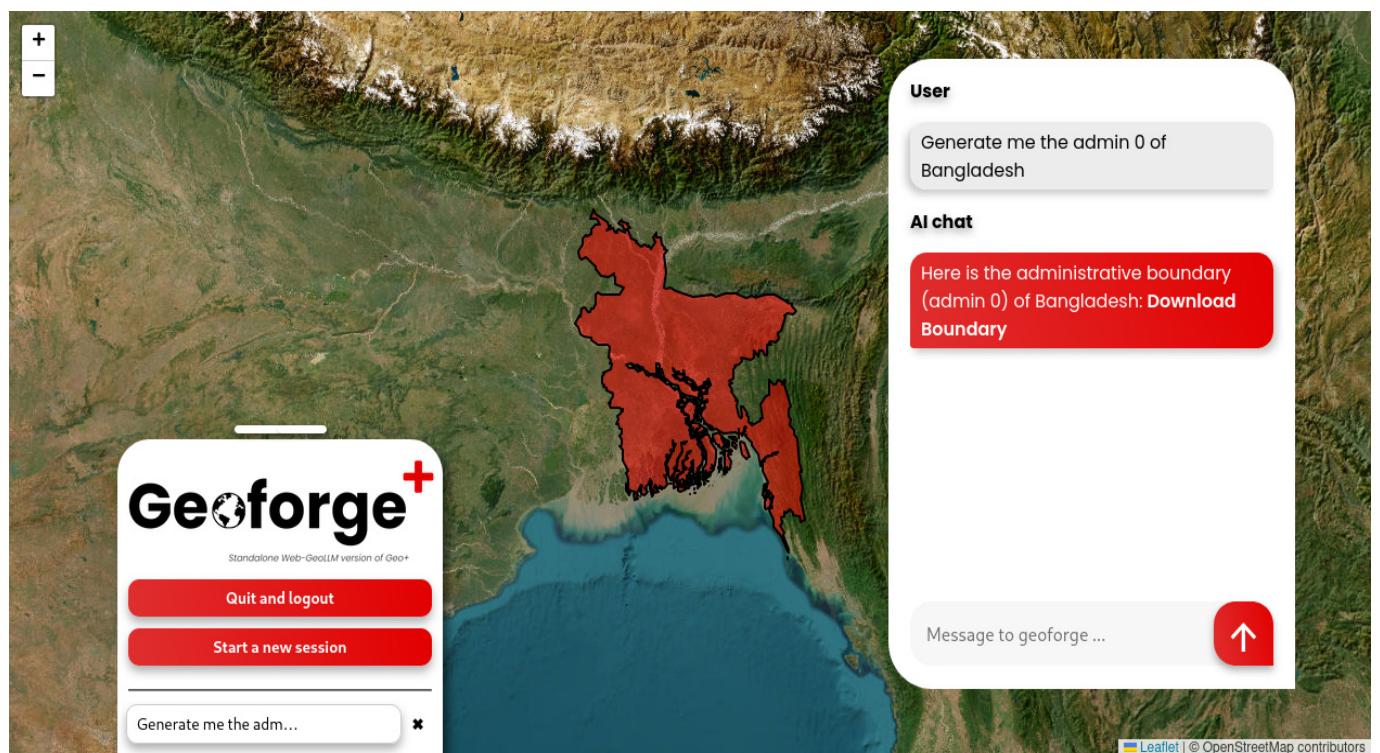


FIGURE 2.15 – Interface utilisateur GeoForge [17]

Aino est une plateforme web collaborative qui permet d'appeler et de visualiser de la donnée géospatiale pour répondre à des problématiques d'urbanisme et d'aménagement du territoire [18].

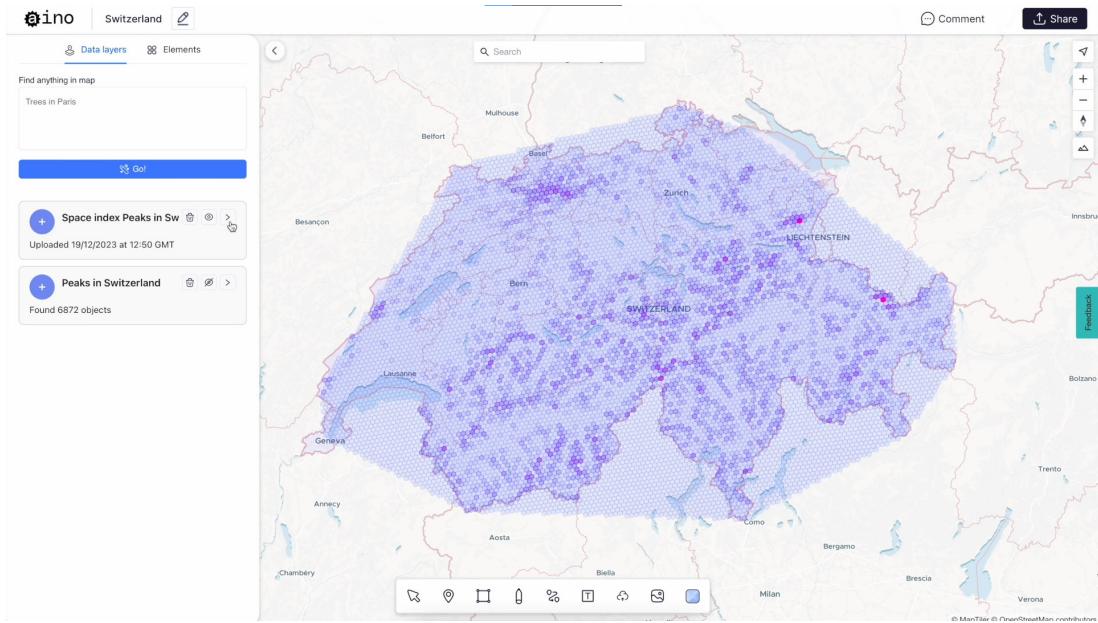


FIGURE 2.16 – Interface utilisateur Aino [18]

GeoDa.AI est une plateforme web open-source développée par l'Université de Chicago qui permet de visualiser et traiter de la donnée géospatiale tabulaire à l'aide d'un chatbot [19].

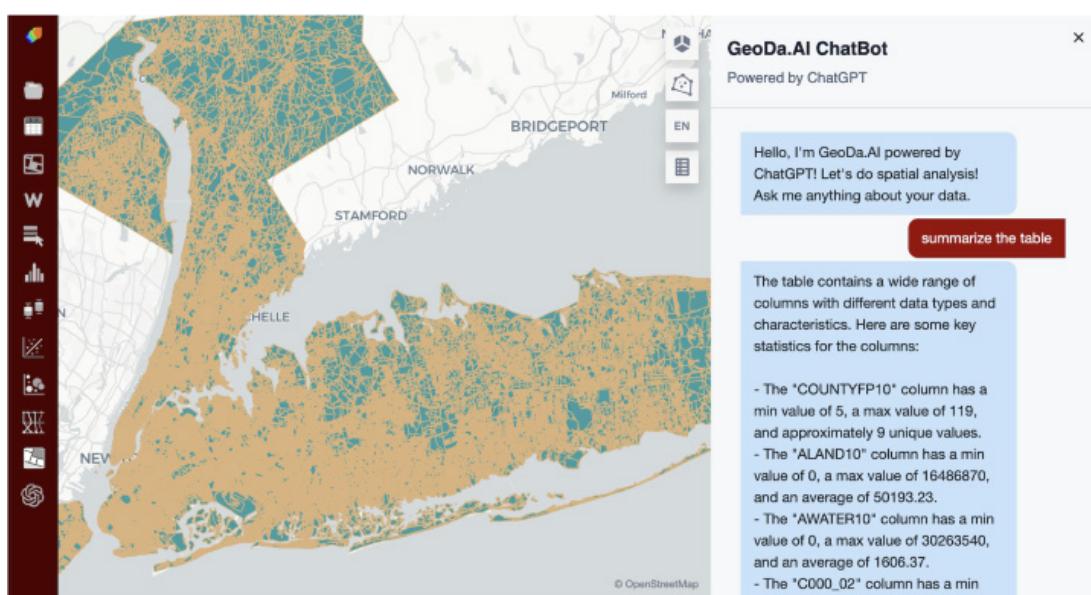


FIGURE 2.17 – Interface utilisateur Geoda.AI [19]

CHAPITRE 3 :

EXTRACTION ET TRAITEMENT DES GÉODONNÉES

À PARTIR DE MULTIMODAL LARGE LANGUAGE MODELS (MLLM)

3.1 DIMENSION GEOGRAPHIQUE DES MLLM

Analyse des compétences géospatiales des MLLM

Dans les exemples suivants, extraits de l'article *The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision)* [20], les chercheurs posent des questions à GPT-4V concernant des images. Les exemples sélectionnés possèdent tous une dimension géographique, et offrant ainsi un nouvel aperçu des capacités de GPT-4 en matière de connaissances géospatiales.

Dans la section *Science and Knowledge*, deux cartes différentes sont fournies au modèle : une carte du monde avec les méridiens et parallèles et une carte des États-Unis où sont affichés quatre noms d'États. Pour la carte du monde, la question posée est "Lequel de ces océans est traversé par le méridien d'origine ?".

GPT-4V répond correctement en disant que, d'après la carte, l'océan Atlantique, l'océan Arctique et l'océan Austral sont traversés par le méridien d'origine. Pour la carte des États-Unis, la question posée est "Quel est l'état le plus au nord ?". GPT-4V réussit à lire la rose des vents en bas à droite (sans cet élément, GPT-4V répond plutôt "North Carolina") de la carte et déduit correctement que l'État le plus au nord est le Delaware.

Cet exercice est intéressant car il permet d'imaginer l'application de cette capacité à lire des cartes pour l'auto-vérification et l'auto-correction des cartes produites automatiquement par le biais du LLM, ainsi que pour faciliter l'édition des couleurs ou des symboles présents sur la carte.

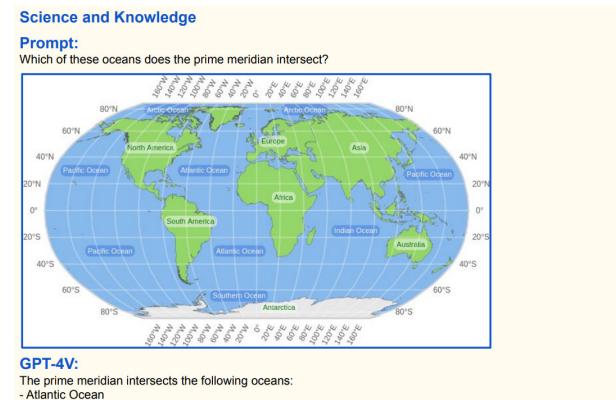


FIGURE 3.1 – Première tâche : "lequel de ces océans est traversé par le méridien d'origine ?" [20]



FIGURE 3.2 – Seconde tâche : "Quel est l'état le plus au Nord ?" [20]

Le prochain exercice consiste à demander à GPT-4V de décrire le monument présent dans l'image. Le modèle réussit à reconnaître les monuments dans l'image et décrit même les autres éléments figurant dans l'image comme les espaces verts et la verdure.

Landmark Recognition and Description

Prompt:

Describe the landmark in the image.



GPT-4V:

The landmark in the image is the Space Needle, an observation tower located in Seattle, Washington. It stands 605 feet tall and offers panoramic views of the city and surrounding areas. The tower was built for the 1962 World's Fair and has since become a symbol of Seattle and a popular tourist attraction. The tower has a futuristic design with a large circular observation deck at the top and a restaurant that rotates 360 degrees. The base of the tower is surrounded by a park with trees and green spaces.

Prompt:

Describe the landmark in the image.



GPT-4V:

This image shows the Kinkaku-ji Temple, also known as the Golden Pavilion, located in Kyoto, Japan. The temple is a three-story Buddhist temple that is covered in gold leaf and sits on a pond, creating a stunning reflection on the water. The surrounding landscape is filled with trees and rocks, creating a peaceful and serene atmosphere. It is a popular tourist attraction and a designated National Special Historic Site.

FIGURE 3.3 – Reconnaissance des monuments par GPT4-V [20]

Dans la dernière image, les chercheurs ont fourni au modèle des photographies de jeunes mariés en vêtements traditionnels et lui ont demandé de décrire l'image en anglais et dans la langue correspondant à la géolocalisation représentée sur l'image. Grâce à cet exercice, les chercheurs ont démontré la capacité du modèle à reconnaître un lieu géographique même si aucun marqueur géographique explicite n'est présent dans l'image, ainsi que sa capacité à générer du texte en plusieurs langues.

Sec. 4.5 Multilingual Multimodal Understanding

Multilingual Multicultural Understanding

Prompt:

Describe the image in both English and the local language corresponding to the geolocation depicted in the image.



GPT-4V:

English: This image shows a couple dressed in traditional Chinese wedding attire, sitting in front of a wall of red roses and a gold Chinese character for "double happiness."

Chinese (Simplified): 这张图片展示了一对穿着中国传统婚礼服装的夫妇，坐在红色玫瑰和金色“双喜”汉字前面。(Zhè zhāng tǐpiàn zhǎnshíle yī duì chuānzhe zhōngguó chuántǒng hūnlǐ fúzhuāng de fūfù, zuò zài hóngsè méiguī hé jīnsè "shuāng xǐ" hànzì qiánmiàn.)

Prompt:

Describe the image in both English and the local language corresponding to the geolocation depicted in the image.



GPT-4V:

English: This image shows a Japanese couple in traditional wedding attire. The man is wearing a black kimono with a white hakama and the woman is wearing a white kimono with a large headpiece.

Japanese: この画像は伝統的な結婚式の衣装を着た日本のカップルを示しています。男性は白い袴と黒い着物を着ており、女性は大きな頭飾りと白い着物を着ています。(Kono gazō wa dentō-teki na kekkon-shiki no ishō o kita Nihon no kappuru o shimeshiteimasu. Dansei wa shiroi hakama to kuroi kimono o kite ori, josei wa ōkina kazari to shiroi kimono o kiteimasu.)

Prompt:

Describe the image in both English and the local language corresponding to the geolocation depicted in the image.



GPT-4V:

English: The image shows a couple wearing traditional Ethiopian clothing, standing in a garden with a tree in the background. The woman is wearing a white dress with colorful patterns and the man is wearing a white suit with similar patterns.

Amharic (Ethiopia's official language): የሸጻ ማስፈልግ በኢትዮጵያ በሀበሻ አገልግሎት ገዢዎችና ለሰራተኞች በግብር ደንብ ይታወቁ:: ስትኩቸው በተለይነት ያለ አለባለሁበር ልብስና ለሰላለቷቸው እና ወንደሞተም በመሳሳይ አለባለሁበር ስት ለሰላለቷቸው::

Prompt:

Describe the image in both English and the local language corresponding to the geolocation depicted in the image.



GPT-4V:

English: This image shows a newlywed Indian couple sharing a kiss. The bride is wearing a traditional red sari and the groom is wearing a red turban and white suit. They are sitting on a decorated stage with flowers.

Hindi (local language in India): यह छवि एक नवविवाहित भारतीय जोड़े को एक चुबन शेयर करते हुए दिखाती है। दुल्हन एक परंपरागत लाल साड़ी पहन रही है और दुल्हा एक लाल पगड़ी और सफेद सूट पहन रहा है। वे फूलों से सजे हुए एक मंच पर बैठे हैं।

FIGURE 3.4 – Reconnaissance d'une localisation à partir d'une image de jeunes mariés en vêtements traditionnels [20]

GPT-4V est aussi capable de déterminer la localisation d'une image sans qu'un monument ou repère soit présent dans l'image [21]. Toutefois, le modèle présente des difficultés à reconnaître la localisation d'images provenant de certains pays, comme la Chine, probablement en raison d'un manque

de données d'entraînement provenant de cette région. Le modèle est également performant pour les tâches de télédétection telles que l'analyse et la description d'images satellites. Dans la figure 3.7, le modèle prédit correctement la saison de l'année pour chaque image satellite.

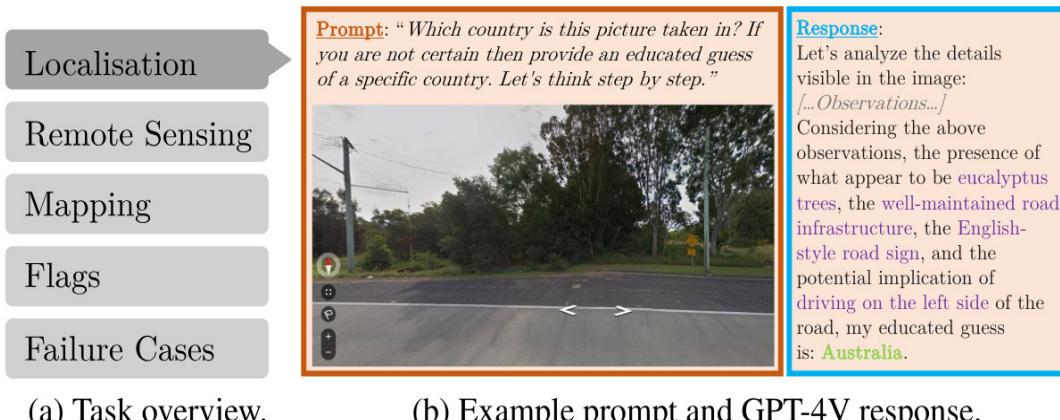


FIGURE 3.5 – Exemple de prompt pour demander la localisation d'une image [21]

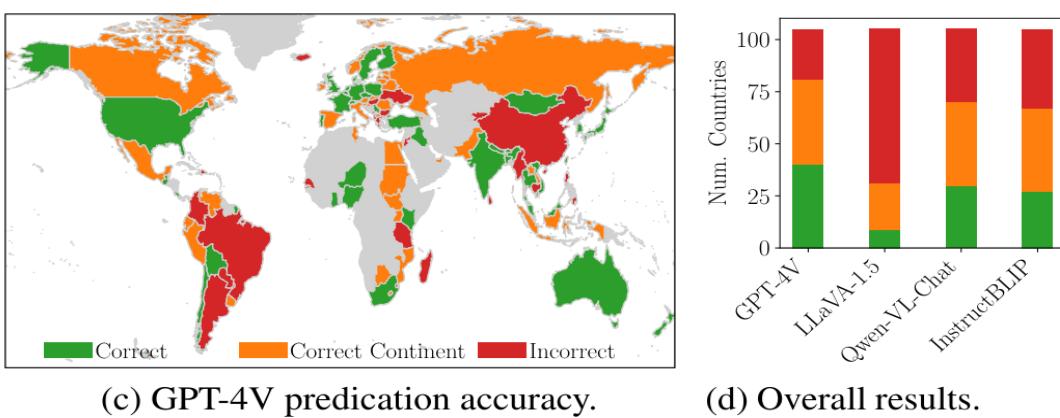


FIGURE 3.6 – Précision de prédition de GPT-4V par pays et comparé à d'autres LLM [21]



FIGURE 3.7 – Prédiction de la saison à partir d'une image satellite [21]

3.2 LES MLLM COMME ASSISTANTS INTELLIGENTS

Mobiliser le raisonnement multimodal des MLLM pour l'automatisation des tâches

Les Multimodal Large Language Models ont la capacité, notamment grâce à la vision par ordinateur, d'apporter une aide précieuse à certaines tâches géospatiales, voire de les automatiser. Il pourrait s'agir d'automatiser la création de balises OpenStreetMap sur la base d'une photographie au niveau de la rue [22], comme nous pouvons le voir ci-contre, où le MLLM peut déduire de l'image le nombre de voies, la surface de la route, l'éclairage public et d'autres éléments.

Il peut également s'agir de la segmentation automatique d'objets présents dans une image à partir d'un prompt. Cela est possible grâce au modèle SAM (Segment Anything Model) développé par Meta [23], qui est un *Foundational Model for Image Segmentation*. La segmentation d'image est un processus utilisé en computer vision et qui consiste à diviser une image en parties ou en segments pour en faciliter l'analyse. En divisant une image en segments, il est plus facile de se concentrer sur des zones ou des objets spécifiques de l'image et de les analyser individuellement.

ESRI a déjà intégré SAM dans ArcGIS Pro [24], ce qui permet aux spécialistes en SIG de segmenter automatiquement des objets tels que des bâtiments ou des arbres à partir d'une image satellite et d'un simple prompt.

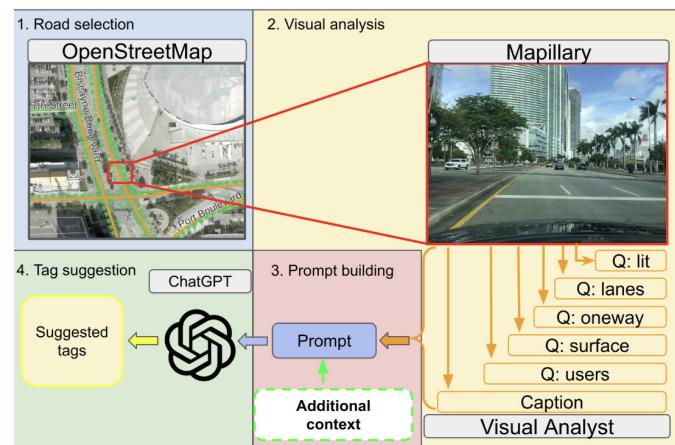
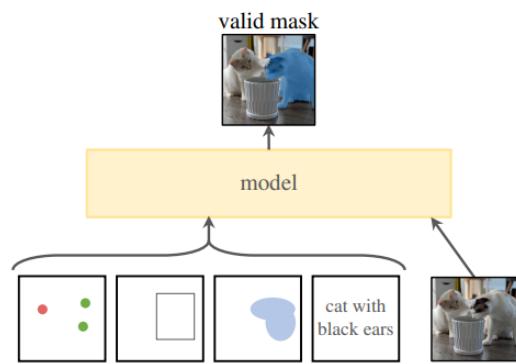


FIGURE 3.8 – Diagramme de flux pour la suggestion de balises OSM [22]



(a) Task: promptable segmentation

FIGURE 3.9 – Segmentation d'image à partir de prompt avec le modèle SAM [23]

CONCLUSION

Dans cet état de l'art, nous avons examiné la dimension géographique dans les grands modèles de langages et leur capacité à effectuer des analyses géospatiales de manière autonome. Le développement en parallèle des systèmes d'information géographique (SIG) et de l'intelligence artificielle générative révèle des synergies importantes qui menacent de perturber le statu quo dans les méthodes conventionnelles d'exécution de l'analyse géospatiale.

Les LLM sont capables d'intégrer des connaissances géospatiales lors de leur entraînement ou suite à une procédure d'affinage, et d'utiliser cette connaissance pour répondre à des questions ou effectuer des tâches géospatiales simples comme générer les coordonnées d'une ville. L'étude des différents modèles, y compris GPT-4, a révélé des variations significatives dans leur capacité à fournir des réponses géospatiales, suggérant un besoin d'affinage et d'intégration de données externes pour améliorer la précision et la reproductibilité des réponses.

L'utilisation des LLM pour la compréhension et l'exécution autonome d'opérations de géotraitement s'est révélée prometteuse. Nous avons étudié comment les LLM peuvent servir d'intermédiaires entre les utilisateurs non spécialisés et des bases de données géospatiales, rendant l'analyse géospatiale plus accessible. Les études montrent les bases du développement des SIG automatisés, qui ont commencé à être lancés en 2024 sous forme de plateformes indépendantes ou

d'assistants IA intégrés dans des SIG comme QGIS. Ces assistants ont pour objectif de simuler et reproduire le raisonnement d'un expert en système d'information géographique, et d'encadrer le LLM pour assurer la fiabilité et la reproductibilité de l'analyse géospatiale et des données.

Le développement des MLLM et leurs applications pour des cas d'usage tels que la cartographie automatisée et l'extraction de données géospatiales à partir de contenus multimodaux nécessitent des recherches plus approfondies. Toutefois, la vision par ordinateur des MLLM comme ChatGPT-4V est déjà très performante dans l'analyse du contenu d'une image et l'intégration de la reconnaissance d'images dans l'analyse géospatiale ouvre des possibilités intéressantes pour l'auto-vérification et l'auto-correction des contenus générés par le MLLM.

Pour conclure, même si l'association des technologies LLM et système d'information géographique produit des résultats très prometteurs pour un grand nombre de cas d'usage, il est important de rappeler que l'intelligence artificielle générative est encore dans ses premières phases de développement et qu'il reste du travail à accomplir pour renforcer sa fiabilité dans l'exécution de tâches géospatiales complexes qui demandent des réponses précises et reproductibles. Il est donc essentiel d'examiner soigneusement les limites des modèles et méthodologies existants afin de déterminer les axes d'innovation et résoudre les limites actuelles de la technologie.

LISTE DES FIGURES

FIGURE 1 – Illustration de l'intégration de l'IA, du ML et du DL dans les systèmes modernes

FIGURE 1.1 – Erreur moyenne de distance (en kilomètres) pour la prédiction de coordonnées géographiques et taux de Prédiction (P-Rate en %), indiquant la fréquence à laquelle les modèles génèrent une réponse avec des coordonnées pertinentes. Notamment, la méthode 3-shot peut présenter des résultats inférieurs car les exemples préalables peuvent influencer négativement sur la précision des prédictions

FIGURE 1.2 – Évaluation de la compréhension par GPT-4 d'indicateurs par pays tels que (a) le nombre d'habitants (b) les espérances de vie, et (c) les émissions de CO₂ par habitant. Les cercles rouges indiquent les données aberrantes

FIGURE 1.3 – Altitudes prévues (lignes) et réelles (zones ombrées) le long des trajectoires représentées à gauche (la région sous-jacente se trouve dans les Alpes, les zones plus claires représentent des altitudes plus élevées)

FIGURE 1.4 – Contours de divers éléments géographiques produits à l'aide des coordonnées fournies par GPT-4. L'affinement avec des prompts supplémentaires améliore les résultats

Figure 1.5 – Aire de répartition des sous-espèces de tigres et routes migratoires des oiseaux estimées par ChatGPT-4.(Espèces d'oiseaux: NW=Traquet motteux, AT=Sterne arctique, AF=Faucon de l'Amour, SS=Puffin à bec grêle, R=Combattant varié, SH=Buse de Swainson)

FIGURE 1.6 – Coefficient de détermination linéaire de Pearson pour tous les modèles, pour toutes les tâches et toutes les tailles d'échantillons d'entraînement

FIGURE 1.7 – Représentation graphique de l'erreur absolue comparant les machines learning models et GPT-3.5 sur l'exécution de tâches uniques à partir de plusieurs sources de données (1000 échantillons par source de données)

FIGURE 1.8 – Architecture technique proposée pour MapGPT avec intégration des vecteurs spatiaux avec des vecteurs textuels

FIGURE 1.9 – Résultats des modèles en % sur les différentes catégories de l'examen

FIGURE 2.1 – Framework proposé pour la génération de code SQL avec un LLM pour la requête de données géospatiales

FIGURE 2.2 – Framework proposé pour la génération de code SPARQL avec un LLM pour la requête de données géospatiales

FIGURE 2.3 – Interface GeoQAMap

FIGURE 2.4 – Diagramme du processus de génération de scénarios

FIGURE 2.5 – Framework de GeoGPT. Un LLM est utilisé pour interpréter les demandes en langage naturel, et choisit et appelle les outils appropriés pour résoudre les tâches géospatiales

FIGURE 2.6 – Diagramme de flux de LLM-Geo

FIGURE 2.7 – Résultats générés automatiquement par LLM-Geo pour le dénombrement des populations vivant à proximité de déchets dangereux. (a) Graphe, (b) programme d'assemblage (codes Python), et (c) décompte de la population et carte générée

FIGURE 2.8 – Illustration de l'interface utilisateur pour (a) un SIG autonome, (b) un SIG autonome de type Copilot intégré dans un logiciel SIG existant (par exemple, ArcGIS Pro), Autonomous GIS: the next-generation AI-powered GIS

FIGURE 2.9 – Chargement des shapefiles dans ChatGPT

FIGURE 2.10 – Visualisation de la carte au sein de l'interface chat de ChatGPT

FIGURE 2.11 – Liste des fonctionnalités de Geo+

FIGURE 2.12 – Génération de frontières administratives avec Geo+

FIGURE 2.13 – ArcGIS experience builder, ArcGIS and Generative AI Assistants

FIGURE 2.14 – Interface utilisateur de QGPT Agent

FIGURE 2.15 – Interface utilisateur GeoForge

FIGURE 2.16 – Interface utilisateur Aino

FIGURE 2.17 – Interface utilisateur Geoda.AI

FIGURE 3.1 – Première tâche : “lequel de ces océans est traversé par le méridien d'origine ?”

FIGURE 3.2 – Seconde tâche : “Quel est l'état le plus au Nord ?”

FIGURE 3.3 – Reconnaissance des monuments par GPT4-V

FIGURE 3.4 – Reconnaissance d'une localisation à partir d'une image de jeunes mariés en vêtements traditionnels

FIGURE 3.5 – Exemple de prompt pour demander la localisation d'une image

FIGURE 3.6 – Précision de prédiction de GPT-4V par pays et comparé à d'autres LLM

FIGURE 3.7 – Prédiction de la saison à partir d'une image satellite

FIGURE 3.8 – Diagramme de flux pour la suggestion de balises OSM

FIGURE 3.9 – Segmentation d'image à partir de prompt avec le modèle SAM

BIBLIOGRAPHIE

- [1] Bhandari, P., Anastasopoulos, A., & Pfoser, D. (2023). Are large language models geospatially knowledgeable? In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23). Association for Computing Machinery. <https://doi.org/10.1145/3589132.3625625>
- [2] Salmas, K., Pantazi, D.-A., & Koubarakis, M. (2023). Extracting geographic knowledge from large language models: An experiment. In KBC-LM/LM-KBC@ISWC, 2023.
- [3] Roberts, J., Lüddecke, T., Das, S., Han, K., & Albanie, S. (2023). GPT4Geo: How a language model sees the world's geography. arXiv preprint arXiv:2310.14478.
- [4] Manvi, R., Khanna, S., Mai, G., Burke, M., Lobell, D., & Ermon, S. (2023). GeoLLM: Extracting geospatial knowledge from large language models. arXiv preprint arXiv:2310.14852.
- [5] Fernandez, A., & Dube, S. (2023). Core Building Blocks: Next Gen Geo Spatial GPT Application. arXiv preprint arXiv:2310.11029.
- [6] Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). GPT-4 passes the bar exam. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 382(2270), 20230254. <https://doi.org/10.1098/rsta.2023.0254>
- [7] Mooney, P., Cui, W., Guan, B., & Juhász, L. (2023). Towards understanding the geospatial skills of ChatGPT: Taking a geographic information systems (GIS) exam. Association for Computing Machinery.
- [8] Jiang, Y., Yang, C., & Xiao, G. Is ChatGPT a Good Geospatial Data Analyst? Exploring the Integration of Natural Language into Structured Query Language within a Spatial Database. Geospatial Information Science.
- [9] Feng, Y., Ding, L., & Xiao, G. GeoQAMap - Geographic Question Answering with Maps Leveraging LLM and Open Knowledge Base.
- [10] Frez, J., & Baloian, N. (2023). Bridging the gap: Enhancing geospatial analysis with natural language and scenario generation language. In J. Bravo & G. Urzáziz (Eds.), Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023) (pp. 252-263). Springer Nature Switzerland.
- [11] Zhang, Y., Wei, C., Wu, S., He, Z., & Yu, W. (2023). GeoGPT: Understanding and processing geospatial tasks through an autonomous GPT. arXiv preprint arXiv:2310.14992.

- [12] Li, Z., & Ning, H. (2023). Autonomous GIS: The next-generation AI-powered GIS. arXiv preprint arXiv:2305.06453. <https://doi.org/10.48550/arXiv.2305.06453>
- [13] Franz's blog. (2023). ChatGPT make me a map. <https://www.youtube.com/watch?v=65k-pmMf2Vg&>
- [14] Ageospatial. (2024). Geo+. <https://chatgpt.com/g/g-ST9I22xLI-geo>
- [15] ESRI. (2024). ArcGIS and Generative AI Assistants. In ESRI Federal GIS Conference 2024.
- [16] Nasser, M. (2023). QPT Agent. Retrieved from <https://github.com/momaabna/QGPTAgent>
- [17] Ageospatial. (2024). GeoForge. Retrieved from <https://www.ageospatial.com/products>
- [18] Aino. Aino for QGIS. <https://aino.world/qgis/>
- [19] University of Chicago. GeoDa.AI. <https://geoda.ai/>.
- [20] Yang, Z., Li, L., Lin, K., Wang, J., Lin, C.-C., Liu, Z., & Wang, L. (2023). The dawn of LLMs: Preliminary explorations with GPT-4V(ision). arXiv preprint arXiv:2310.14478.
- [21] Roberts, J., Lüddecke, T., Sheikh, R., Han, K., & Albanie, S. (2024). Charting new territories: Exploring the geographic and geospatial capabilities of multimodal LLM. arXiv preprint arXiv:2401.12345.
- [22] Juhász, L., Mooney, P., & Hochmair, H. H. (2023). ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs. In Spatial Data Science Symposium 2023.
- [23] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. arXiv preprint arXiv:2301.10923.
- [24] ESRI. ArcGIS pretrained models. <https://doc.arcgis.com/en/pretrained-models/latest/imagery/introduction-to-segment-anything-model-sam-.htm>.

LECTURE SUPPLÉMENTAIRE

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998-6008).

Tao, R., & Xu, J. (2023). Mapping with ChatGPT. ISPRS International Journal of Geo-Information, 12(7), 284. <https://doi.org/10.3390/ijgi12070284>.

Tucker, S. (2024). A systematic review of geospatial location embedding approaches in large language models: A path to spatial AI systems.

Yan, Y., Wen, H., Zhong, S., Chen, W., Chen, H., Wen, Q., Zimmermann, R., & Liang, Y. (2024). UrbanCLIP: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web.

European Space Agency. (2024). Building ChatGPT-style tools with Earth observation.

Code Interpreter (ChatGPT): L'outil *code interpreter*, directement intégré dans ChatGPT, permet au LLM d'interpréter, de visualiser, d'analyser et de générer du code. Grâce à cet outil, les utilisateurs de ChatGPT peuvent, au sein de l'interface chat, analyser et visualiser des données, convertir des fichiers, développer du code et résoudre des équations mathématiques.

<https://platform.openai.com/docs/assistants/tools/code-interpreter>

<https://research.aimultiple.com/chatgpt-code-interpreter/>

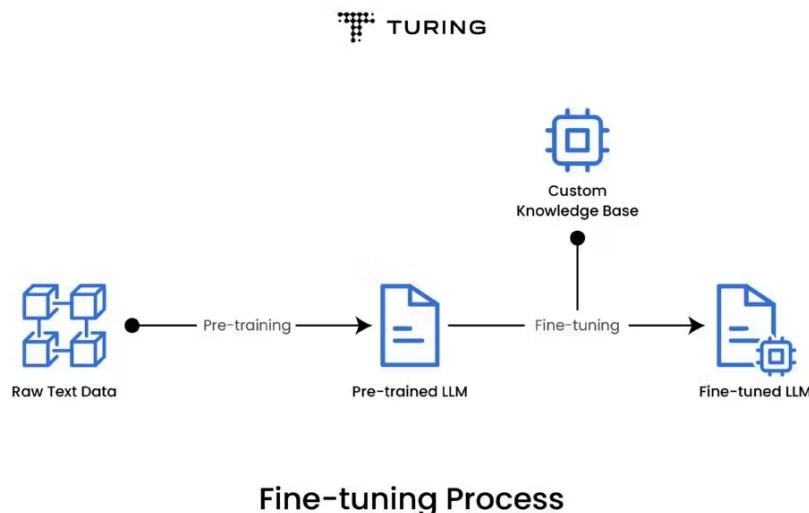
https://www.linkedin.com/posts/miles-lee-81994175_ai-geoai-geogpt-activity-7165835012167200768-t3XT

https://www.linkedin.com/posts/giswqs_geemap-jupyter-geospatial-activity-7173840581188595714-IQga

Fine-tuning

L'illustration ci-dessous présente le processus de fine-tuning des LLM, qui comprend les étapes suivantes :

1. Pré-entraînement: Le LLM est d'abord pré-entraîné sur une vaste quantité de données textuelles brutes, ce qui lui permet d'acquérir une compréhension générale du langage.
2. Base de connaissances personnalisée: Une base de connaissances spécialisée est ensuite utilisée pour affiner le modèle. Cela inclut des données spécifiques au domaine ou à la tâche que le modèle doit maîtriser.
3. Fine-tuning: Le modèle pré-entraîné est ajusté en utilisant ces nouvelles données, ce qui permet de créer un modèle affiné (*fine-tuned LLM*) capable de meilleures performances sur une ou plusieurs tâches prédéfinies.



Turing, Fine-Tuning LLM : Overview, Methods, and Best Practices, <https://www.turing.com/resources/finetuning-large-language-models>.

LangChain est un *framework* souvent recommandé pour le fine-tuning des LLM car il permet de gérer facilement les instructions données aux modèles, d'optimiser leurs réponses et propose des outils pour surveiller et améliorer continuellement les performances des modèles affinés.

Lang-chain:

- <https://docs.smith.langchain.com/old/cookbook/fine-tuning-examples/export-to-openai>
- <https://docs.smith.langchain.com/old/cookbook/fine-tuning-examples/lilac>

Foundation models

Le terme de “*foundation models*” a été utilisé pour la première fois par des chercheurs de Stanford en 2021 pour décrire un nouveau sous-ensemble de modèles d'intelligence artificielle générative. Le nom d'intelligence artificielle générative fait directement référence à la capacité des modèles à générer de nouvelles données pour accomplir les tâches qui leur sont confiées. Ces modèles sont une évolution directe des modèles d'intelligence artificielle précédents, qui n'étaient capables que d'exécuter des tâches uniques et ne pouvaient pas être affinés et apprendre de nouvelles tâches sans un ré entraînement complet du modèle.

- IBM, What are foundation models? <https://research.ibm.com/blog/what-are-foundation-models>.
- AWS, What are foundation models? <https://aws.amazon.com/what-is/foundation-models>.
- NVIDIA, What Are Foundation Models?, 2023. <https://blogs.nvidia.com/blog/what-are-foundation-models>.

L'utilisation des foundation models pour les opérations géospatiales et l'observation terrestre représente un important potentiel sous-exploité. L'intelligence artificielle géospatiale (*Geospatial AI*) actuelle se base sur des modèles deep learning formés utilisant l'apprentissage supervisé sur des données libellées et pour accomplir des tâches spécifiques. Des efforts concertés ont toutefois été déployés pour développer des geospatial foundation models, tels que Prithvi 100M, développé dans le cadre d'une collaboration entre IBM et la NASA. Ce modèle, entraîné sur un jeu de données harmonisé d'images satellites Landsat et Sentinel, peut être affiné pour la détection automatique de l'étendue des inondations par exemple.

- Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., & Lao, N. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence.
- Mendieta, M., Han, B., Shi, X., Zhu, Y., & Chen, C. (2023). Towards geospatial foundation models via continual pretraining.
- Jakubik, J., et al. (2023). Foundation models for generalist geospatial artificial intelligence.

Le développement en parallèle des LLM et des *geospatial foundation models* présente des opportunités intéressantes. En effet, nous pouvons envisager d'intégrer les deux modèles dans le même système et permettre au LLM d'utiliser le foundation model comme un outil, comme présenté dans l'article *GeoGPT: Understanding and processing geospatial tasks through an autonomous GPT* à la page 21. Par exemple, Ageospatial développe actuellement ce système avec un LLM open-source et le modèle Prithvi 100M.

