

---

# Il s'agit d'une carte OpenStreetMap

## Description de cartes par des grands modèles de langage

Guillaume TOUYA<sup>1</sup>, Jérémy KALSRON<sup>1</sup>,  
Laura WENCLIK<sup>1</sup>, Quentin POTIÉ<sup>1</sup>,  
Bérénice LE MAO<sup>1</sup>

*LASTIG, Univ Gustave Eiffel, IGN, ENSG, Champs-sur-Marne, F-77420  
{prenom}.{nom}@ign.fr*

---

**ABSTRACT.** *Being able to extract the description of the map from the image of the map can be extremely useful to make the map accessible to users with visual impairment, or to users that need to keep their visual attention on a different task than reading the map. To extract this description from a map image, large language models could be used as they combine capabilities to analyse images and to generate structured textual responses. This paper presents a preliminary experiment to assess the current quality of the map descriptions generated by large language models. The results are very promising with detailed descriptions and few errors, which calls for a further investigation of the automated generation of accessible map descriptions.*

**MOTS-CLÉS :** *grand modèle de langage, carte, accessibilité, audio-description*

**KEYWORDS:** *LLM, map, audio-description, accessibility*

---

## 1. Introduction

Les cartes, notamment les cartes topographiques, sont des objets contenant de nombreuses informations à différents niveaux de lecture. Cependant, il n'est pas toujours possible de les explorer visuellement, soit du fait d'une déficience visuelle (Kalsron *et al.*, 2022), soit parce que l'attention visuelle est mobilisée sur d'autres tâches. C'est le cas par exemple lorsque nous naviguons en voiture ou en vélo. Rendre accessible le contenu d'une carte par une description écrite qui pourrait être lue par un dispositif audio (Robinson, Griffin, 2024) est donc un moyen prometteur pour rendre les connaissances incluses dans ces cartes plus utilisables.

Sonoriser des cartes n'est pas une idée nouvelle (Krygier, 1994 ; Edler *et al.*, 2019) et des propositions ont déjà été faites en utilisant les données vectorielles d'OpenStreetMap pour générer un contenu descriptif écrit/sonore (Boularouk *et al.*, 2017 ; Kalsron *et al.*, 2022). Mais quand il s'agit d'images sans accès direct aux données sous-jacentes, les grands modèles de langage (LLM) peuvent être modélisés pour générer une description adaptée (Robinson, Griffin, 2024). La problématique abordée par cet article est d'évaluer la qualité des descriptions générées par les LLM actuellement disponibles pour la population.

## 2. Protocole de test

Pour explorer cette problématique, nous avons adopté un protocole simple. Nous avons réalisé une capture d'écran d'OpenStreetMap au niveau de zoom 17 dans un format proche de ce qui est visible sur téléphone (Figure 1). Cette carte représente le Jardin Lecoq, dans le centre de Clermont-Ferrand. Puis, nous avons sélectionné 5 LLM capables d'analyser des images, dans leur version gratuite : ChatGPT, Gemini, Claude, DeepSeek et Le Chat. Pour chaque modèle, nous avons attaché l'image de la carte et posé la question suivante (en français) : "Peux-tu décrire le contenu de cette carte?". Nous avons enregistré la première réponse donnée, sans chercher à l'améliorer par de nouvelles questions.

## 3. Résultats

Les cinq LLM testés produisent des descriptions très intéressantes. A titre d'exemple, nous reproduisons ici celle de ChatGPT, en changeant légèrement la mise en page pour l'intégrer à l'article :

"cette carte représente le Jardin Lecoq, un parc situé à Clermont-Ferrand, en France. Voici les principaux éléments visibles sur la carte :

- Informations générales : Le parc est entouré par plusieurs routes (Boulevard Lafayette (au nord), Avenue Vercingétorix (à l'ouest), Cours Sablon (à l'est), Au sud, on trouve l'UCA - Campus Centre et des bâtiments associés).



FIGURE 1. Extrait cartographique issu d'OpenStreetMap, centré sur le Jardin Lecoq à Clermont-Ferrand, en France (©OpenStreetMap contributors).

- Points d'intérêt à l'intérieur du parc :
  - Bustes et statues : Buste de Pierre Alphonse Julien (au nord-ouest), Statue de H. Lecoq (au centre).
  - Lieux emblématiques : La mémoire des crapauds (un site particulier du parc), Tourincot, une structure près de l'entrée est, Tourelle d'angle du château de Bien Assis, une ruine historique près du sud-est.
  - Éléments naturels et de détente : un grand plan d'eau au centre-sud, plusieurs chemins en pointillés rouges qui traversent le parc, Théâtre de Verduze, un espace en plein air pour des représentations, De nombreux espaces verts accessibles pour la promenade et la détente.

Le parc est bordé par plusieurs institutions et bâtiments notables, dont le Musée Bargoin et la Bibliothèque du Patrimoine au nord."

Nous avons compilé des éléments comparatifs pour les cinq descriptions (cf. tableau 1). Par exemple, trois LLM sur cinq trouvent la localisation complète de la carte (le Jardin Lecoq à Clermont-Ferrand, en France), et Claude ne donne pas la ville mais trouve le nom du jardin et le pays.

TABLEAU 1.

| LLM            | ChatGPT | Gemini | Claude   | DeepSeek | Le Chat |
|----------------|---------|--------|----------|----------|---------|
| nb mots        | 177     | 196    | 143      | 200      | 275     |
| localisation   | oui     | oui    | à moitié | non      | oui     |
| nb erreurs OCR | 1       | 0      | 1        | 6        | 0       |

Un élément particulièrement saillant dans ces descriptions est la reproduction de la quasi totalité des toponymes de la carte qui montre la capacité des modèles à bien reconnaître le texte dans des images. Toutefois, ces capacités (OCR pour "Optical Character Recognition") ne sont pas encore parfaites, comme le montre l'exemple du manège Tournicoti qui est reconnu comme "Tourincot" par ChatGPT et "Totumicon" par DeepSeek. Le contraste au niveau de ce toponyme rend en effet sa détection particulièrement complexe et seul Le Chat l'a reconnu correctement.

Ces résultats montrent aussi l'importance du mécanisme dit de "Retrieval-Augmented Generation" (RAG) qui permet aux LLM d'effectuer des requêtes sur le web pour récupérer des informations complémentaires. Par exemple, la Tourelle d'angle du château de Bien Assis est reconnue comme une ruine par ChatGPT, information trouvable sur Wikipedia, mais pas représentée sur la carte (le pictogramme correspond à une tour fortifiée).

Les résultats montrent également la capacité des LLM à utiliser des modèles de vision capables de segmenter l'image en objets. Par exemple, la plupart des modèles reconnaissent la zone bleue dans le parc comme un étang, et les lignes pointillées rouges comme des chemins.

Plus surprenant encore, les descriptions contiennent des relations spatiales comme des relations de cardinalité (Tournicoti est "situé vers le nord-est du jardin" pour Le Chat) ou des relations de proximité (le buste de Pierre-Alphonse Julien est situé "près" de la sortie nord-ouest du parc pour Le Chat). Certains modèles décrivent même des relations topologiques implicites dans la carte : Claude identifie les points d'intérêt situés "dans" le parc, et les routes qui "bordent" le parc. Claude est enfin le seul LLM à identifier que cette carte est issue d'OpenStreetMap car il retrouve "le codage couleur et l'iconographie typiques".

Enfin, contrairement à ce que l'on pourrait attendre, nous ne constatons que très peu d'hallucinations dans les descriptions produites. La seule erreur que nous pouvons qualifier d'hallucination est produite par Le Chat qui décrit

la "rue Blatin au sud du parc". La rue au sud du parc n'est pas nommée dans l'image donnée, et il s'agit en fait du Boulevard François Mitterrand.

#### 4. Conclusion

Cette première expérience montre que la génération de descriptions accessibles de cartes topographiques par des modèles LLM est tout à fait possible à court terme car ces premiers résultats sont globalement bons. Nous prévoyons de poursuivre ces recherches en suivant une expérience similaire de plus grande ampleur, avec plus de modèles, plus de cartes, et un protocole d'évaluation et de comparaison plus poussé. En particulier, nous prévoyons de demander à plusieurs personnes de décrire les mêmes cartes pour identifier les différences avec les descriptions issues des LLMs.

#### 5. Remerciements

Ce projet a reçu un financement du European Research Council (ERC) dans le cadre du programme de recherche et d'innovation Horizon 2020 de l'Union Européenne (grant agreement No. 101003012).

#### Bibliographie

- Boularouk S., Josselin D., Altman E. (2017). Ontology for a Voice Transcription of OpenStreetMap Data: The Case of Space Apprehension by Visually Impaired Persons. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 11, n° 5, p. 585–589.
- Edler D., Kühne O., Keil J., Dickmann F. (2019, mai). Audiovisual Cartography: Established and New Multimedia Approaches to Represent Soundscapes. *KN - Journal of Cartography and Geographic Information*, vol. 69, n° 1, p. 5–17.
- Kalsron J., Favreau J.-M., Touya G. (2022, juin). CrossroadsDescriber – Automatic Textual Description of OpenStreetMap Intersections. In *AGILE GIScience Series*, vol. 3, p. 40. Vilnius, Lithuania, Copernicus Publications.
- Krygier J. B. (1994, janvier). Sound and Geographic Visualization. In A. M. MacEachren, D. R. F. Taylor (Eds.), *Modern Cartography Series*, vol. 2, p. 149–166. Academic Press.
- Robinson A. C., Griffin A. L. (2024, septembre). Using AI to Generate Accessibility Descriptions for Maps. *Abstracts of the ICA*, vol. 7, p. 1–2.