



HAL
open science

Text-to-map generation: an exploration of current potentials

Azelle Courtial, Jérémy Kalsron, Bérénice Le Mao, Quentin Potié, Guillaume Touya, Laura Wenclik

► **To cite this version:**

Azelle Courtial, Jérémy Kalsron, Bérénice Le Mao, Quentin Potié, Guillaume Touya, et al.. Text-to-map generation: an exploration of current potentials. CartoAI: AI for cartography, Sep 2023, Leeds, United Kingdom. hal-04224973

HAL Id: hal-04224973

<https://hal.science/hal-04224973v1>

Submitted on 3 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Text-to-map generation: an exploration of current potentials

Azelle Courtial^{a,*}, Jérémy Kalsron^b, Bérénice Le Mao^a, Quentin Potié^a, Guillaume Touya^a, Laura Wenclik^a

^a LASTIG, Univ Gustave Eiffel, IGN-ENSG, F-77420 Champs-sur-Marne, France, firstname.lastname@ign.fr

^b Université Clermont Auvergne, CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, F-63000 Clermont-Ferrand

* Corresponding author, All the authors contributed equally to this research

Keywords: Deep learning, image generation, cartography

Abstract:

In recent years, we have seen the development of artificial machines capable of understanding requests formulated in natural language and generating a response in the form of images. These models are composed of a text encoder that understands the semantics of the request and a decoder that produces a graphic response (Ramesh, et al., 2022, .Yu, et al., 2022 and Saharia et al., 2022). They achieve impressive results for generating images of diverse domains from a request in natural language. But there are very few examples of successful generations of maps. These models seem to have a great potential to contribute to the significant objective of on-demand mapping, (Bailey & Regnaud, 2011), in which we expect to create a framework able to understand non-cartographer requirements, search the relevant information and represent it in the desired map automatically. Thus, we propose to review the potential of these models for generating cartographic knowledge. We would like to identify the relevance of text-to-image models in the process of generating a map. In the following paragraphs we discuss different angles of text-to-map generation and illustrate the strength and weakness of such model with examples produced using DALL-E (one of the most popular text -to-image models produced by OpenAI).

Describing a map with a text: The first step to generate a map from a text is to be able to precisely describe and specify the map with a text. This is not a straightforward task because of the complexity of a map. The text should ideally mention the extent, the scale, the represented features and their geometrical abstraction, and the style of the map. Our experiments show the influence of the text on the output image and research should explore how to effectively describe a map with a short text.

Generation of map looking image: In our experiments, we then aim to verify the ability of such model for producing realistic map looking images including content transformation, data symbolisation, map generalization, etc. Deep learning has already shown potential in generating map looking images in different styles (Christophe et al., 2022).

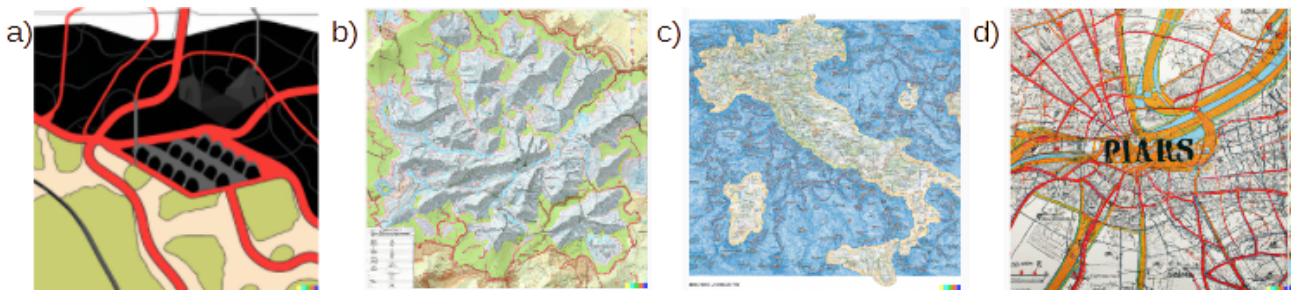


Figure 1. Some map looking images generated by Dall-E a) “A topographic map with black buildings aligned along a red road, with a forest nearby” b) “A topographic map of mountainous region with castles” c) “Topographic map of Italy with the main landmarks” d) “Topographic map of Paris, France with the main landmarks”.

In Figure 1, we observe the output images for some prompts in natural language asking for a map depicting diverse entities at local, medium and small scales. We observe that the model can generate a map in the relevant styles with various levels of abstraction; all predictions look like a map. For maps where the text describes only the content (a and b), the desired content is properly encoded and correctly depicted. However for prompts that specify a real place (c and d) we observe that the shape of generated entities sometimes does not reflect the ground truth and often appears distorted. This problem is more important for the maps at a large scale, probably because there are fewer examples of maps depicting these entities in the training set.

Spatial information rendering: In a second time we experiment with the generation of spatial informations in several manners: with in-painting, and out-painting requests and with spatial relation generation requests. Our goal was to test

the model ability to modify the render of a base map, to highlight/ hide or change the display of specific spatial informations.

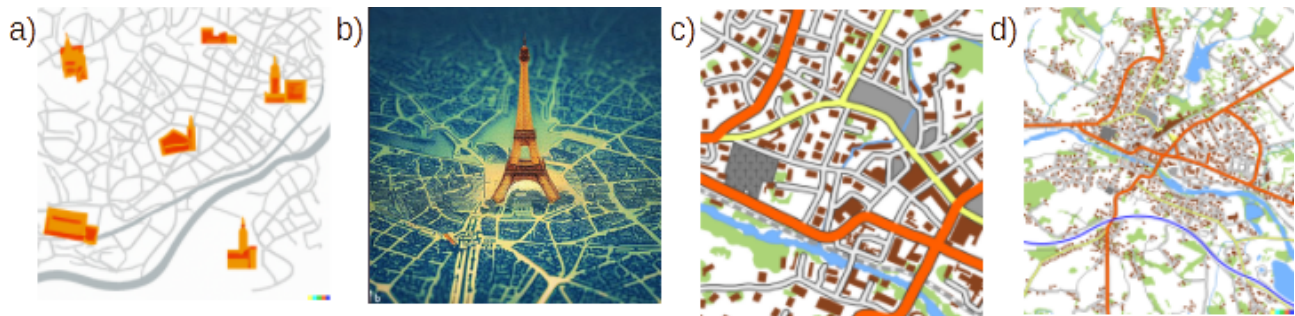


Figure 2 The prediction of Dall-E for some spatial rendering requests a) “A map of a city with several landmark buildings highlighted” b) “A IGN map of Paris at a scale of 1:2500 centred on the Eiffel tower” c) A base map of Orthez at the 1:50k scale designed by the authors with IGN data. d) The out-painting of a context around map c.

In Figure 2 we observe how such a model can be used to modify the rendering of spatial information on a map (e.g. to highlight some content; change the point of view or fill a part of the map). Indeed, more than allowing the generation of images from scratch (a and b) the model is able to deal with the transformation of a base image according to a request formulated in natural language (d), the in-painting and out-painting options can help to fill an incomplete map in a similar way to GANMapper (Wu & Biljecki, 2021). It seems also promising to change the style and content of various map images.

Conclusion: This preliminary experiment highlights the necessity of specific training for geo-information-related text-to-map models. Indeed, both the text encoding and the rendering part would benefit from fine-tuning with a specific dataset depicting maps, and map descriptions (for instance, constructed in the manner proposed by Juhász, Hochmair, & Guan (2023)). Moreover, these models are limited because they cannot find the appropriate dataset to include in the map. The integration of a data search and collection module, that can retrieve and integrate data, seem inevitable to produce reliable maps, in a process similar to what is proposed for geo-question answering (Scheider et al., 2021). Indeed, the information in the output is only issued from the image set used for training and such a model cannot be used to produce map representing specific phenomena or statistics. Finally, the question of ethic and reliability of such maps has to be examined (Zhao et al., 2021; Kang, Zhang, & Roth, 2023), and proper evaluation techniques or strategies to assess the quality of these generated maps are required. Due to these limitations, current text-to-map models can currently only be used to generate artistic maps.

Bibliography

- Balley, S., & Regnaud, N. (2011). Models and Standards for On-Demand Mapping. In 25th International Cartographic Conference., Paris, France.
- Christophe, S., Mermet, S., Laurent, M., & Touya, G. (2022). Neural map style transfer exploration with GANs. *International Journal of Cartography*, 8, 1–19. <https://doi.org/10.1080/23729333.2022.2031554>
- Juhász, L., Hochmair, H., & Guan, B. (2023). ChatGPT as a mapping assistant: A novel method to enrich maps with generative AI and content derived from street-level photographs.
- Kang, Y., Zhang, Q., & Roth, R. (2023). The Ethics of AI-Generated Maps: A Study of DALLE 2 and Implications for Cartography (arXiv:2304.10743). arXiv. <https://doi.org/10.48550/arXiv.2304.10743>
- Wu, A., & Biljecki, F. (2021). GANmapper: geographical content filling. ArXiv: 2108.04232 [Cs.CV].
- Zhao, B., Zhang, S., Xu, C., Sun, Y., & Deng, C. (2021). Deep fake geography? When geospatial data encounter Artificial Intelligence. *Cartography and Geographic Information Science*, 48(4), 338–352. <https://doi.org/10.1080/15230406.2021.1910075>
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M., (2022) Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv, <https://doi.org/10.48550/arXiv.2204.06125>
- Yu, J., Xu, Y., Koh, J., Luong, T., Baid, G., Wang, Z., Vasudevan, V., et al. (2022) Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. arXiv, <https://doi.org/10.48550/arXiv.2206.10789>
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Kamyar, S., Ghasemipour, S., et al. (2022) Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *Advances in Neural Information Processing Systems*, Oh, A., Agarwal, A., Danielle Belgrave, D., Cho, K., (eds).
- Scheider, Simon, Enkhbold Nyamsuren, Han Kruijer, Haiqi Xu (2021) Geo-analytical question-answering with GIS. *International Journal of Digital Earth* 14(1), 1-14. <https://doi.org/10.1080/17538947.2020.1738568>