International Cartographic Association
Association Cartographique Internationale

Exploring the use of XAI for Automated Generation of Map Descriptions with Language Models

Güren Tan Dinga [a]*, Jochen Schiewe [a]

[a] *HafenCity University Hamburg, Lab for Geoinformatics and Geovisualization (g2lab) - gueren.dinga@hcu-hamburg.de, jochen.schiewe@hcu-hamburg.de*

* Corresponding author

Keywords: CartoAI, Explainable AI (XAI), Accessibility, Large Language Models (LLM), Map Descriptions

Abstract:

Despite significant advancements in mapping technologies in recent decades, accessibility for blind and visually impaired individuals remains an unmet need (Robinson and Griffin (2024)). This means that blind and visually impaired individuals continue to face barriers in accessing essential information. The European Parliament, Council of the European Union (2019), which establishes accessibility requirements for products and services, aims to eliminate and prevent these barriers. It states that products placed on the market in the EU after 28 June 2025 must be designed and produced in a way that maximizes their use by persons who have long-term physical, mental, intellectual or sensory impairments which includes blind and visually impaired persons. However, according to Chapter I, Article 2, Item 4 of the directive, *"online maps and mapping services, if essential information is provided in an accessible digital manner for maps intended for navigational use"* are excluded from these requirements. This exclusion indicates the complexity of the issue of generating meaningful map descriptions and highlights the research gap in creating efficient and effective automated solutions.

Vision Language Models (VLMs) offer a promising approach to address the challenge of automated map description generation. VLMs are multi-modal Large Language Models (LLMs). This means that VLMs are designed to accept both image and text data as inputs. While VLMs are capable of generating image descriptions, the quality of their output is heavily dependent on the model architecture, the data used during training and carefully engineered prompts. Xu and Tao (2024) emphasize that predicting the behaviour of such large models is challenging due to their complexity. This leads to difficulties in understanding and interpreting the output of such networks. Hence, we approach the challenge of generating reliable map descriptions from another perspective: using Explainable AI (XAI) to better understand the relationship between the inputs and generated map descriptions when using VLMs.

In natural language processing, text is typically represented as a sequence of subwords or word pieces. The process of splitting text into subwords and word pieces, which then are used as input to the model, is called "tokenization". In this context, subwords or word pieces, are referred to as "tokens". During inference, such as when generating map descriptions, LLMs tokenize the input text and later de-tokenize it to produce coherent and meaningful text outputs. By utilizing the concept of tokens, we aim to explore the potential of XAI to provide deeper insights into the generation process of VLMs, with the goal of enhancing both the interpretability and reliability of their outputs.

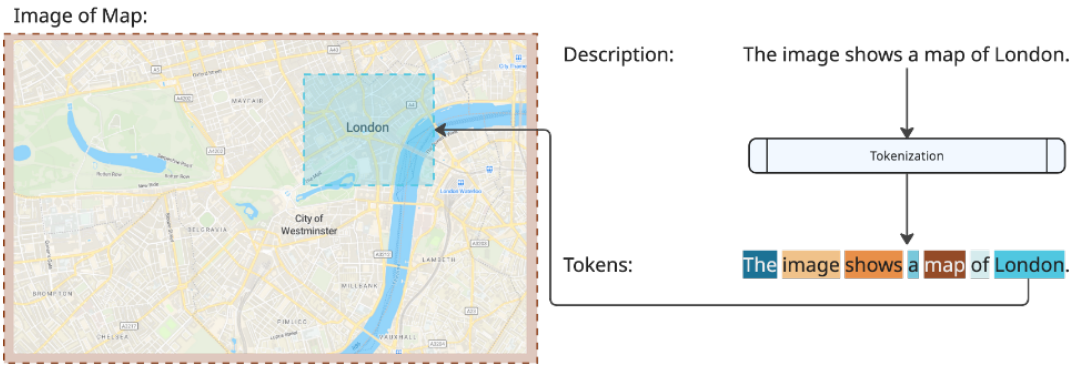

Figure 1. The simplified workflow to connect tokens to specific image regions. The exemplary workflow consists of generating a map description for an excerpt of a digital map showing London (©MapTiler ©OpenStreetMap contributors). The tokens "London" has a high marginal contribution to the image region containing the map label "London".

Figure 1 illustrates the workflow using an excerpt from a digital map of London. This cartographic map, in combination with a prompt, serves as input to a VLM. The output of the VLM, segmented into tokens, is then associated with specific image regions which are identified as particularly influential in generating the corresponding token.

To achieve this, two XAI methods are suitable in particular: GradCAM and the calculation of Shapley values. GradCAM operates by computing the gradient of the loss function with respect to each feature of the input image, highlighting areas of importance for the model's predictions. In contrast, Shapley Values assign a marginal contribution score to each feature, representing its individual impact on the model's output relative to all other features. Based on our prior experience with comparable tasks, we have chosen to employ Shapley Values for this study.

Early results indicate that the VLM we have tested is particularly effective in performing Optical Character Recognition (OCR) tasks. For instance, the token "London" shows the highest marginal contribution for an image region containing a map label reading "London". Tokens representing more general terms such as "map" show equally distributed marginal contribution values across the image.

Future experiments will focus on extending the evaluation to a broader range of VLMs and multiple datasets to assess the generalizability of our findings. Further, we aim to investigate marginal contributions of individual map components such as map labels, street networks and building polygons. The initial outcomes of our use of XAI in the context of generating map descriptions with language models will be shared at the upcoming workshop.

References

European Parliament, Council of the European Union, 2019. Directive (eu) 2019/882 of the european parliament and of the council of 17 april 2019 on the accessibility requirements for products and services. https://eur-lex.europa.eu/eli/dir/2019/882/oj/eng.

Robinson, A. C. and Griffin, A. L., 2024. Using AI to Generate Accessibility Descriptions for Maps. *Abstracts of the ICA* 7, pp. 1–2.

Xu, J. and Tao, R., 2024. Map Reading and Analysis with GPT-4V(ision). *ISPRS International Journal of Geo-Information* 13(4), pp. 127.