

Article

Creating Choropleth Maps by Artificial Intelligence—Case Study on ChatGPT-4

Parinda Pannoos and Rostislav Netek * 

Department of Geoinformatics, Palacký University Olomouc, 779 00 Olomouc, Czech Republic;
parinda.pannoos01@upol.cz

* Correspondence: rostislav.netek@upol.cz

Abstract

This study explores the potential of ChatGPT-4, an AI-powered large language model, to generate thematic maps and compare its outputs to the traditional method in which maps are produced manually by humans using GIS software. Prompt engineering is a crucial methodology of large language models that can enhance output quality. The main objective of this study is to assess the capability of AI-generated maps and to compare the quality with a traditional method. The study evaluates two prompt patterns: basic (zero-shot prompts) and advanced (Cognitive Verifier and Question Refinement). The performance of AI-generated maps is assessed based on attempts, errors, incorrect results, and map completeness. The final stage involved evaluating AI-generated maps against cartographic rules to assess their suitability. ChatGPT-4 performs well in generating suitable choropleth maps but faced challenges in understanding the prompts and potential errors in the generated code. Advanced prompts reduced errors and improved the quality of outputs, particularly for complex map elements. This paper enhances the understanding of AI's role in cartography and further research in automated cartography. The study assesses cartographic aspects, offering insights into the strengths and limitations of AI in cartography, illustrating how large language models can process geospatial data and adhere to cartographic principles. The study also paves the way for future innovations in automated geovisualization.



Academic Editors: Wolfgang Kainz,
Andriani Skopeliti, Ioannis
Giannopoulos, Vassilios Krassanakis
and Guillaume Touya

Received: 29 September 2025

Revised: 21 November 2025

Accepted: 3 December 2025

Published: 9 December 2025

Citation: Pannoos, P.; Netek, R. Creating Choropleth Maps by Artificial Intelligence—Case Study on ChatGPT-4. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 486. <https://doi.org/10.3390/ijgi14120486>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

AI has played an increasingly important role in recent years. The role of AI is multi-faceted and continues to expand in various industries [1]. The journey of AI started in 1956, when John McCarthy introduced the concept “Artificial Intelligence” to describe an area of computer science focused on developing computers that behave like humans [2]. In recent years, AI has been widely leveraged because of its ability to analyze large amounts of data, recognize patterns, and make informed decisions. This shift is driven by the transition from rule-based systems to data-driven approaches, which are trained on vast amounts of pre-trained knowledge. In addition, powerful graphics processing units (GPUs) have reduced the time required for rendering images and performing machine learning tasks. In earlier periods, training an AI model on a central processing unit (CPU) took significantly longer, depending on the model’s complexity and the size of the dataset [3].

The launch of ChatGPT, available online in late 2020 [4], marked a crucial step in the evolution of AI technology. ChatGPT leverages the power of LLMs to understand and interact with human language in the form of an AI Chatbot [5]. Its natural language processing (NLP) capabilities enable it to generate meaningful and relevant responses to tasks such as software development [6], product designs [7], and text generation and translation [8]. By utilizing textual prompts, ChatGPT streamlines development workflows and minimizes manual effort.

In the domain of geoinformatics, AI technology has transformed the way geospatial data is processed. Traditionally, the first step in geospatial analysis involves collecting geospatial data and manually preprocessing it, which can be time-consuming. In later stages, experts perform analysis and visualization, utilizing their expertise to select the most suitable GIS tools for each sub-task. Nowadays, the integration of AI algorithms to process various types of big geospatial data—such as mobility data, temporal data, and high-resolution satellite imagery—has been increasing. This integration enhances AI's ability to perform spatial reasoning and location-based analytics, similar to human reasoning [9], enabling it to interpret complex geographic information, predict spatial trends, and provide highly accurate insights.

The application of AI, utilizing machine learning and deep learning algorithms, is being employed to address complex geographic problems across various domains, including remote sensing, Earth System Science, and cartography [10]. For example, AI can predict floods by analyzing remote sensing imagery [11], predict urban traffic volume using deep learning models such as Graph Neural Networks [12], and leverage the capabilities of LLMs like GPT-4o to recognize and interpret various types of thematic maps from image inputs [13]. The paper explores ChatGPT-4's capabilities in generating maps through prompt engineering. AI-generated maps are compared with human-generated maps to assess their limitations and quality, including the influence of prompts on map outputs, because large language models can produce different levels of detail in the map outputs. This paper mainly aims to utilize AI for creating maps by applying different prompt patterns based on Python scripts. The study sets specific objectives, which are to evaluate the functional capability and learning ability of the AI in producing maps, in both static and interactive maps; to analyze and evaluate different prompt patterns that influence map completeness; and lastly, to assess the map quality between maps generated by the AI and those produced through a traditional method, in order to identify strengths and limitations.

To evaluate how well AI-generated maps achieve cartographic aspects through different prompt patterns, the study evaluated the completeness of map compositions. In other words, the number of map compositions successfully created in each session. Quantitative metrics such as the number of prompts, error messages, and incorrect results on each pattern are used to measure how different prompt structures influence AI-generated maps. For map quality assessment, the output maps are benchmarked against predefined cartographic rules and specifications derived from authoritative cartography sources [14].

2. Geo-Artificial Intelligence

The rapid growth of AI techniques and huge amounts of data has resulted in the combination of AI and geoinformatics technologies, which mainly aim to analyze, solve spatial problems, and derive insights from geospatial data. The interdisciplinary relationship between geospatial data and AI is not new. In the early stage of AI in the geospatial domain, Openshaw and Openshaw [15] introduced machine learning and deep learning techniques, which are fundamental AI technologies, integrated for geospatial analysis.

Richter and Scheider [16] stated that the term GeoAI is a combination of ‘geo’ as in ‘geographic’ or ‘geography,’ and ‘Artificial Intelligence’. The interdisciplinary relationship between geospatial data and AI is not new; the origins of GeoAI can be traced back to its earliest days of geographic information systems (GISs) and applied statistics. In 1965, Howard Fisher at Harvard Laboratory created computer map-making software for spatial analysis and visualization research [17]. GeoAI has its roots in the mid-1960s [18], but during that period, its applications were limited by computational power, data availability, and the development of machine learning algorithms.

In recent years, GeoAI has continued to evolve rapidly, driven by advancements in AI algorithms, particularly machine learning, sensor technology, and computational infrastructure. Additionally, spatial data are becoming more accessible, with an increase in the volume of real-time sensor observations, the variety of imagery data, and geotagged text data [19]. These data can be processed by GeoAI techniques such as the integration between GIS analysis and deep learning for extracting useful insights and automating processes. The study [16] discusses the evolution of GeoAI, highlighting the influx of new geographic information and advanced machine learning techniques that have expanded the scope of GIS research [20]. The application of NLP in deep learning for handling geographic information from unstructured textual data and interpreting narratives about landscapes can deal with question-answering tasks related to geographic information. The authors mentioned that GeoAI can be beneficial for solving complex tasks combined with diverse data. However, this field still faces challenges with model transparency issues and reliability, which come with the black box nature of machine learning methods.

Snow [21] discussed the transformative impact of GeoAI on Esri’s article, which states that the integration of AI, deep learning, and GeoAI can improve productivity and modernize mapping processes. The author emphasizes that GeoAI at the intersection of AI and geospatial technology offers significant opportunities for improvement in remote sensing and national mapping. It facilitates complex tasks across multiple domains such as aviation, topographic mapping, and disaster response. In recent years, AI map products have undergone significant evolution, reflecting broader trends in emerging GeoAI applications. These map products encompass a wide range of applications, including autonomous routing suggestion, land use–land cover extraction, thematic map creation, and more. They leverage technologies such as machine learning, large language models, and generative AI to provide a vast set of maps that would not be possible with traditional mapping technologies.

3. The Current State of Large Language Models in Cartography

Large language models (LLMs) are a type of artificial intelligence (AI) technology designed to understand, generate, and manipulate human language. Large language models are trained on billions of parameters and massive amounts of text data [22]. This training enables them to perform a wide range of natural language processing (NLP) tasks, such as text generation, translation, summarization, and question-answering. The interaction of LLMs allows a machine to understand or communicate with natural language as humans do. ChatGPT is one of the popular generative AI models in the form of a chatbot developed by OpenAI. According to [8], this generative AI has seen significant improvement, its advanced understanding and interaction capabilities enabling accurate, contextual text processing even with nuanced inputs.

AI has played an increasingly important role in the cartography domain, supporting tasks such as automated map creation [23], map color design [24], map labeling [25], and style transfer [26].

According to [27], the self-operating geographic information systems (GISs) integrated the GPT-4 API with a GIS, known as Autonomous GIS. It can perform spatial analyses by accepting inputs as natural language and generating results as code, map, and graph workflow. Through case studies, LLMs within GISs can generate maps successfully by automating intricate spatial analysis tasks. This integration could make GIS technologies more approachable for those without a GIS background, making a step towards the future of AI-driven autonomous GISs and significantly reducing manual operation time. However, there are several limitations, such as the inability to debug code during execution, which often makes it difficult to generate correct code in a single attempt. Tao and Xu [28] revealed that ChatGPT offers a potential alternative approach for mapping by utilizing prompts. It reduces the barrier to producing maps, leading to the enhanced efficiency of producing large volumes of maps and enabling an understanding of geographical spaces through spatial thinking capability. The authors highlighted that utilizing ChatGPT for mapping presents challenges, including unequal advantages and quality control issues across different user groups. Users should be aware when using ChatGPT for mapping tasks, particularly when dealing with unverified data sources. Also, the process of map improvement is not straightforward.

The advancement of AI integration in map production has evolved into a map agent that is not limited to a single task. MapGPT [23] utilizes an LLM to develop a framework that determines which tools to apply for generating specific map elements, such as map titles, map symbols, or map layers. The framework breaks down the complex map-making process into subtasks, with each tool responsible for a corresponding element. This demonstrates AI's ability to make decisions throughout the entire map-creation workflow and to support automated map modifications through interactive conversations with users.

In cartographic design, the potential of GPT-4o has also been examined through evaluations of text-to-text and image-to-text tasks, focusing on how well the model follows cartographic principles such as visual hierarchy, symbolization, and color theory. The evaluation demonstrated the tool's ability to support less-experienced users by helping them identify and address issues in their maps. Notably, the system effectively recommended color schemes aligned with color theory principles, contributing to the creation of clear and visually appealing maps [29].

Prompt Engineering

Prompt engineering is the process of structuring input text for LLMs and is a technique integral to optimizing the efficacy of LLMs [30]. The structure and context of given prompts can affect the desired output, which we need to refine the prompt patterns to solve tasks. Chen et al. [30] discussed the following essential components for constructing a well-made prompt:

- (1) Giving instructions: When the model is prompted with basic commands, the model will have broad answers. Therefore, a comprehensive description is necessary to obtain more accurate and relevant results.
- (2) Be clear and precise: This approach entails crafting prompts to be more unambiguous and specific. When faced with vague prompts, the model tends to produce outputs that are broadly applicable and may not align in a particular situation. On the other hand, a prompt that is both detailed and precise allows the model to produce content that closely matches the specific demands.
- (3) Try several times: Due to the unpredictable behavior of LLMs, running the model several times with the same prompt to achieve the best output helps in exploring the variations in the model's responses, thereby boosting the probability of achieving a high-quality result.

- (4) Role-prompting: Role-prompting is when the model assigns a specific persona. This strategy helps the model's response to match the expected outcome. For example, by prompting the model to assume the role of a historian, it becomes more likely to offer responses that are both detailed and contextually precise regarding historical events.

Additionally, recent studies have integrated prompt engineering methods with map creation, as seen in the study conducted by Kang et al. [31], who generated high-quality image maps using DALL·E 2, relying on a specific prompt format. Although the AI model of DALL·E 2 can create realistic and diverse images from text prompts, some inaccuracies and misleading information occurred in AI-generated maps due to cartographic concepts and terminology indicated in the prompt. According to [32], the authors leveraged the generative AI capabilities of ChatGPT, combined with a multimodal pre-training method (BLIP-2), to enhance map-tagging accuracy in OpenStreetMap (OSM) by analyzing street-level images from Mapillary. Providing a detailed description of the source photograph and refining prompts with additional contextual information increased the accuracy of the model's outputs. Their method involved constructing a context message and instructing the model using a few-shot pattern, which provided a small set of examples. Based on these examples, the model was able to learn the intended structure and respond more appropriately.

Existing studies have integrated LLMs into new AI systems for performing general cartographic tasks and automated spatial analyses, but the study of prompt engineering approaches to create thematic maps remains limited in current AI integration. In particular, creating a choropleth map requires AI's understanding of concepts such as visual representation, color selection, and contextual data. Additionally, exploring different prompt patterns to assess ChatGPT's cartographic knowledge and to evaluate how it interprets natural-language instructions into a choropleth map visualization remains a gap to explore.

RQ1: Do basic and advanced prompts significantly influence the map completeness through prompt refinement?

RQ2: How effectively can AI generate suitable choropleth maps across various prompt patterns?

RQ3: How does the quality of AI-generated maps compare to human-made maps when evaluated against cartographic principles?

4. Methodology

4.1. Overview

This paper leverages the large language model of ChatGPT-4 to generate maps through Python code based on prompt inputs by dividing the procedures into two stages:

The first stage is conducted to evaluate the functional capability and learning ability of Artificial Intelligence (AI) in producing maps, as well as to assess how prompt patterns influence map completeness. Choropleth maps are created in both static and interactive versions using basic and advanced prompts, and the map generation is iterated five times with the same or similar prompts. The iterative process helps in exploring the variations in the model's responses and enables developers to assess limitations and try to push the boundaries of AI ability [33]. Each map component is generated through a series of prompts that are refined until the result is satisfactory. The remaining map components are then developed using the same approach until the entire map is complete. For example, the process begins by generating the map field and ensuring that it aligns with cartographic principles. Once this component is correct, the next prompts are used to generate additional elements such as the scale bar, title, or legend. To achieve a complete choropleth map according to cartographic principles, the fundamental map compositions were defined as

map field, legend, scale bar, title, and credits. The additional elements are data visualization, labels, subtitle, and basemap. Tooltips and control layers are also added as additional elements for interactive maps.

For the input, the data source is derived from the Global Wildfire Information System (GWIS). GWIS is a joint initiative of the Group in Earth Observations (GEO) and the Copernicus Work Programs [34]. GWIS provides data on wildfire trends, geographic distribution of fires, burned areas per country, and sub-national level for all countries globally. Burned area values are based on the product MODIS MCD64 A1. The data visualization on the maps 'Yearly Burned Area' is also derived from the GWIS Country Profile application in CSV format [35].

The burned area data was preprocessed and normalized by district area in ArcGIS Pro 3.0.0, exported as a shapefile, and then used as input for ChatGPT-4. Each map component (e.g., legend, map field, scale bar) in a choropleth map is generated by one prompt at a time for the basic prompt pattern. The outputs among the five maps were evaluated in the last step.

The second stage aims to evaluate the quality of AI-generated maps in comparison to human-generated maps. The maps from the traditional method are set as the reference for the suitability criteria; the reference consists of the most appropriate specifications according to cartographic rules [14,36]. By comparing maps with specifications, the strengths and weaknesses can be assessed in their quality; the more the map complies with the map's specifications or benchmark, the better the quality of the map [37]. Then, the outputs between AI and traditional methods can be evaluated according to three suitability levels. The workflow is shown in Figure 1.

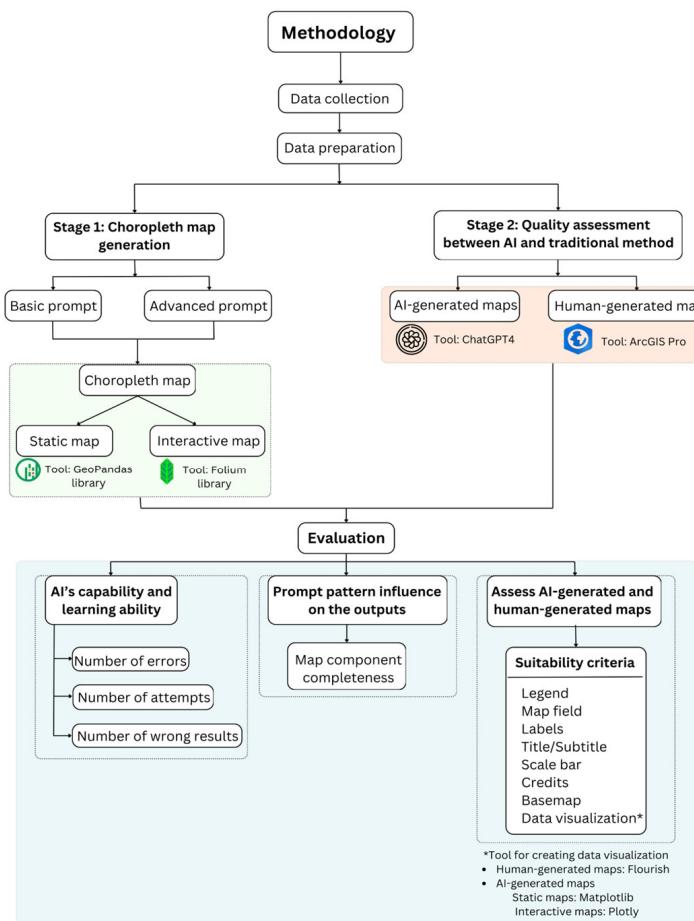


Figure 1. Workflow of the study, divided into two main stages and an evaluation.

4.2. Prompt Patterns

The AI-generated maps are performed by using two different patterns, which are basic and advanced prompt patterns. The basic prompt pattern refers to ‘Direct instruction’, also known as ‘Zero-shot’. It is the simplest type of prompt, without requiring any examples. The pattern consists only of instructions, directly as a question or request, stating what AI should do [33]. This is shown in Figure 2.

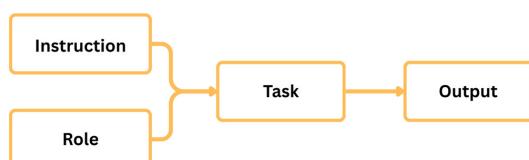


Figure 2. Basic prompt structure showing two main components—instruction and role.

Another prompt pattern is an advanced prompt that combines Cognitive Verifier and Question Refinement patterns. The Cognitive Verifier is used for generating map elements at the beginning of the process (Figure 3). The prompt can generate additional questions related to the original question, which potentially can return results exactly as specified in the answers. The Question Refinement will be used in the last step of the map development to adjust specific or nuanced details such as overlapping issues or positioning (Figure 4). The pattern refines the inputs or questions, reducing the gap between the LLM’s understanding and the user’s knowledge, thereby improving the quality of both input and output, meaning that the results can be more accurate and efficient [38].

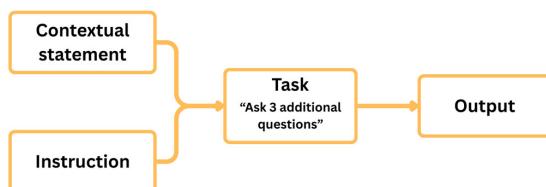


Figure 3. Cognitive Verifier prompt showing three components—contextual statement, instruction, and query guidelines.

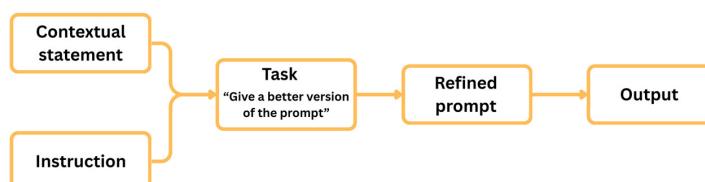


Figure 4. Question Refinement prompt showing three main components—contextual statement, instruction along with the query guidelines for the task. The prompt includes an extended request for an improved version of the prompt.

4.2.1. Basic Prompt Pattern

The basic prompt constructs the structure into three elements: instruction, role, and task, as shown in Figure 5. The instruction provides initial information about the given geospatial data, which enhances the LLM’s understanding of the context and ensures the data will be processed properly. Giving a role to LLMs enables them to dive into a specific domain, such as cartography or geoinformatics. Acting as a persona can provide outputs that the persona would create and help provide details to users who lack in-depth knowledge of the field [38]. In this prompt, the role of ‘*cartographer*’ is assigned to collect relevant cartographic knowledge for accomplishing a map-making task. The task is the primary command to straightforwardly inform what the expected output is that the LLM

should perform. In this pattern, the AI keeps training what it needs to create or adjust each element one at a time, until it gives the desired map outputs.

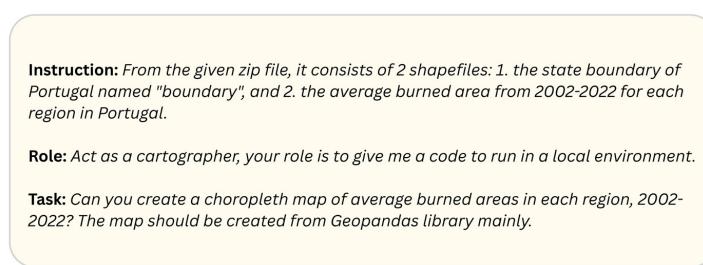


Figure 5. Basic prompt template consisting of instructions, role, and task.

4.2.2. Advanced Prompt Pattern

This study applies Cognitive Verifier and Question Refinement. The Cognitive Verifier can provide three sub-questions related to the user's command. Thereafter, the LLM is capable of combining users' answers and processing them into the final outputs. Another prompt is Question Refinement, which is used for refining map details such as color, placement, and text placement, including improving map compositions. The contextual refinement of the Question Refinement is described in a prompt (see Figure 6). Whenever it is asked to adjust a map, it should suggest a better version of the prompt. This study adapted the contextual statement from the study of White et al. [38].

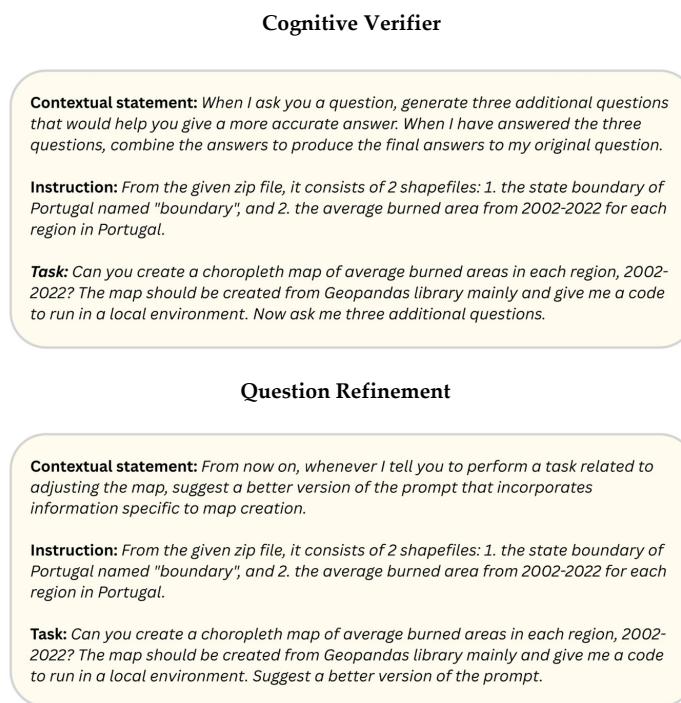


Figure 6. Advanced prompt templates consisting of contextual statement, instructions, and task.

These advanced capabilities enable them to provide refined prompts beyond simple text. The initial prompt assigns contextual statements to a prompt, which is a way to describe how a user and an LLM will communicate in a prompt. The advanced prompts consist of the following structures, as shown in Figure 6.

4.3. Software and Technical Tools

This study used essential tools for creating choropleth maps and visualizing additional data. For creating a choropleth map, we used the following tool:

1. ChatGPT-4: A generative AI trained to generate human-like text responses from given prompts. Since it can interpret natural language input, this allows users to interact with the prompts and context of the conversation. In this study, ChatGPT4 is used as an AI tool for generating code snippets of both static and interactive maps through textual prompts. In the context of geovisualization, libraries like Folium and GeoPandas play a crucial role in this research.
 - Folium: To maintain consistency in the outputs, Folium is used to generate interactive maps. Folium is a Python library used for creating interactive maps. It is built on top of the Leaflet JavaScript mapping library.
 - GeoPandas: GeoPandas is used to produce the static map versions. It is an open-source Python library that extends its capabilities to handle geospatial data and plot maps by leveraging Matplotlib and several libraries.
2. ArcGIS Pro: ArcGIS Pro is a GIS software developed by Esri for creating maps, managing geospatial data, and performing spatial analysis. For human-generated maps, ArcGIS Pro was used as the traditional GIS tool for comparison with the AI-generated outputs.

Besides creating the map extents, the map outputs also include data visualizations such as bar charts. The following tools were used in this study for both AI-generated and human-generated data visualizations:

1. AI-generated data visualization
 - Matplotlib: Matplotlib is a popular Python library used for creating static and interactive data visualization, such as charts or diagrams. In this study, data visualizations in the static maps were generated by Matplotlib.
 - Plotly: An open-source data visualization library for creating interactive data visualizations. It offers a range of visualization types, from basic charts to more complex plots. Data visualization in the interactive maps was generated by Plotly in this study.
2. Human-generated data visualization
 - Flourish: Flourish enables users to create interactive data visualizations. The platform supports various types of visualizations, including bar charts, pie charts, scatter plots, and more. Then, the chart for human-generated maps was created and exported by Flourish.

5. AI Map Generation

5.1. Basic Prompt on the Static Map

The concept of a basic prompt or direct instruction pattern is a method that directly instructs the model to follow instructions without providing any examples. A concise and uncomplicated statement that specifies what the AI should do can return accurate results. The advanced prompt aims to develop the quality of the map outputs and details, which aims to reduce the user's effort in creating a map. The effectiveness of the responses produced by a conversational LLM depends on the quality of the prompts [38]. The interactions between a user and an LLM can be developed to enhance its ability to solve a range of issues effectively.

5.1.1. Map Field

A choropleth map shows regional variations in burned areas, using a graduated color scheme according to the extent of land affected by fires. ChatGPT-4 usually gives the correct output for this thematic method since the choropleth map is a commonly used and basic

method in the GeoPandas library. The graduated-color scheme is assigned in the map field based on the burned area values.

The common error in the code given by AI is that the data is not normalized by the regional areas automatically. Therefore, the data either needs to be preprocessed before inputting into the ChatGPT or indicated in the prompt explicitly. Figure 7 illustrates the results from the AI that failed to generate discrete classes. Users need to specify the classification method to the AI (e.g., Equal Interval, Quantiles, etc.). In addition to the color scheme, the AI can also convey a cold color tone, even in the context of a wildfire incident (see Figure 8). In this case study, orange-red shades (OrRd) are the default color when the term ‘warm color scheme’ is specified in the prompt.

Prompt input

Anonymous

There is an error from previous code. Create a choropleth map of fire areas from the given files. The map will be created by Geopandas Python library

Output

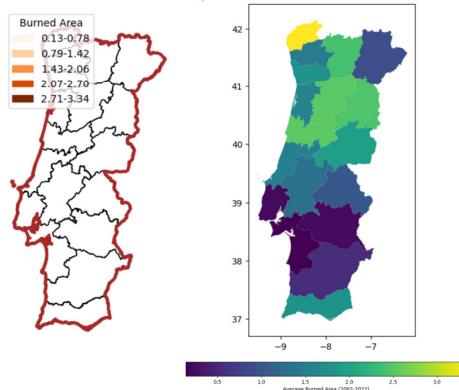


Figure 7. Inappropriate colors and data classification resulting from a misunderstanding of choropleth map principles.

Prompt input

Anonymous

```
↳ import geopandas as gpd
import matplotlib.pyplot as plt
# Assuming the paths have been correctly identified
boundary_path = os.path.join(subfolder_path, 'boundary.shp')
mainlandburn_path = os.path.join(subfolder_path, 'mainlandburn.shp') # Load the shapefiles...
```

From the previous code, can you change the continuous legend to a discrete legend ranging the value from high to low of 5 classes and change the map color to warm color scheme?

Output

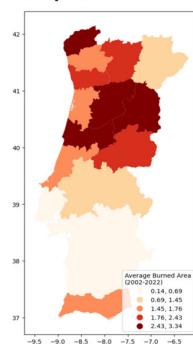


Figure 8. The basic prompt used to refine the map output by specifying changes to the color scheme and discrete legend (symbol # is used as user comment within the given code).

5.1.2. Legend

The default legend style given by ChatGPT-4 is a continuous bar. To modify the legend to five classes ranging from high to low values, the useful prompt is '*change the continuous legend to a discrete legend*', indicated in Figure 9. Then, the adjustment to the code will be made on the '*scheme*' parameter on the code, determining how the data is categorized into discrete bins.

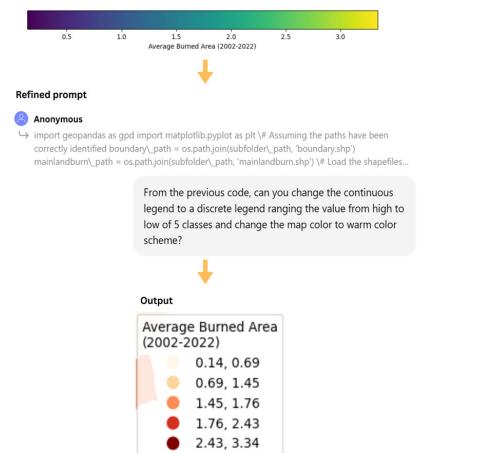


Figure 9. The refined prompt changes the inappropriate legend to one using intervals and warm colors (symbol # is used as user comment within the given code).

5.2. Advanced Prompt on the Static Map

5.2.1. Map Field

Using the Cognitive Verifier pattern can provide a comprehensive output when the AI is asked to create a choropleth map. The AI provides useful questions related to specific attributes. For example, the additional questions in Figure 10 allow it to create a map properly from the specification. In this case study, the specific burned area attribute, warm color scheme, and labels are specified at once. Compared to the basic prompt pattern, the AI can generate the map using a random column or color, so it may need to be refined repeatedly and require multiple attempts. Besides the map field, this pattern could suggest additional map elements such as labels and a legend.

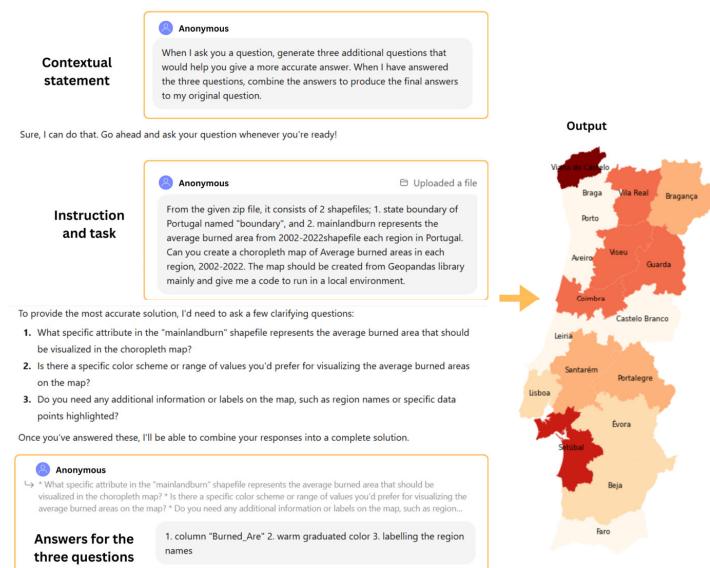


Figure 10. Additional questions and given answers in the Cognitive Verifier pattern for generating an initial map (symbol * is used as a new sub-question within the prompt).

5.2.2. Legend

The default legend labels often duplicate the lower value with the upper value of the next class. To avoid duplicating intervals of adjacent classes, the answers given to the AI can be ‘edit the legend of the lower values to be unique from the upper values of the next class’. Then, the AI could subtract a value (0.01) from the upper boundary of each class, except for the last one. Additionally, Cognitive Verifier can provide relevant questions about decimal numbers, legend position, legend colors, and legend title within a single prompt (see Figure 11).

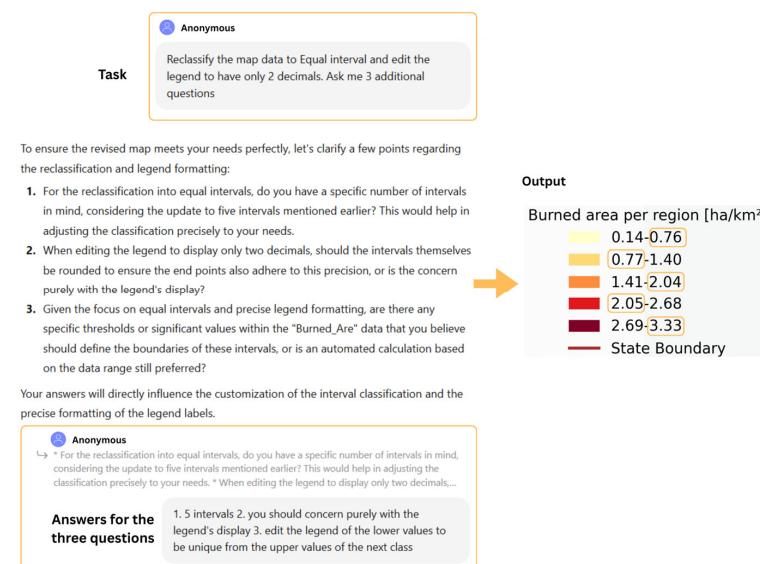


Figure 11. Additional questions and answers in the Cognitive Verifier pattern to create proper interval values in the legend (symbol * is used as a new sub-question within the prompt).

5.3. Basic Prompt on the Interactive Map

An interactive map created using Folium offers a user-friendly way to display geospatial data. The interactive ability allows maps to embed additional information through pop-ups and tooltips that allow users to interact with data dynamically, such as a zoom button, dynamic scale bar, and toggle layers to customize the view. The interactive choropleth map can be generated using the same prompts as the static version, only changing the library’s name and specifying the ‘interactive’ term in the prompt. The primary map composition consists of a map field, legend, scale bar, title, and credits. The additional elements for the interactive version are data visualization, labels, subtitle, basemap, tooltips, and layer control.

5.3.1. Map Field

The interactive map is barely successfully created on the first try; ChatGPT-4 often raises an error while trying to load the shapefiles. This can happen due to the complex geometries. The language used in the prompt needs to be clear and concise by specifying the thematic method, number of classes, and the data classification methods (see Figure 12). Then, the AI will adjust the library’s parameters correctly from the given prompt. It calculates the minimum and maximum values of the ‘burned_area’ attribute to determine intervals, then creates a list of breaks that defines the five classes. The breaks are passed to the bins argument of the ‘folium.Choropleth’ variable to specify the desired interval classes. However, the AI does not consider the data normalization since Folium requires only two data frames, which are geographical coordinates and burned area attributes. Therefore, the normalized data needs to be processed before creating the map in ChatGPT-4.

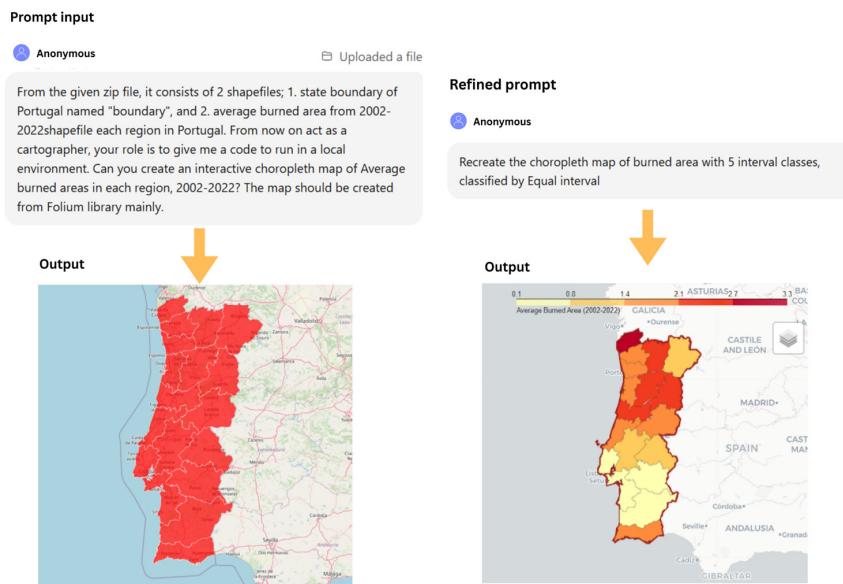


Figure 12. Basic prompts for refining an incorrect visualization into a correct choropleth map.

5.3.2. Legend

The default legend is generated by Folium's Choropleth method by automatically adding a legend based on the '`fill_color`' parameter and the '`bins`' used for classification, without requiring manual specification. Additionally, the interval values and colors always correspond to the map field. Conversely, the legend in the static map needs to be adjusted either in the color scheme or the legend labels several times. Therefore, the map legend for the interactive version requires fewer prompts than the static map.

5.4. Advanced Prompt on the Interactive

5.4.1. Map Field

When creating a choropleth map, AI not only generates a map field that represents the average burned area but also offers additional map features when using Cognitive Verifier prompts. The AI overlays the state boundaries from the '`Boundary`' shapefile and includes tooltips displaying region names and the fire attribute when hovered over each region. ChatGPT-4 utilizes a warm color scheme required in the prompt, as shown in Figure 13. Then, the map represents the burned areas, which is visually intuitive and shows intensity.

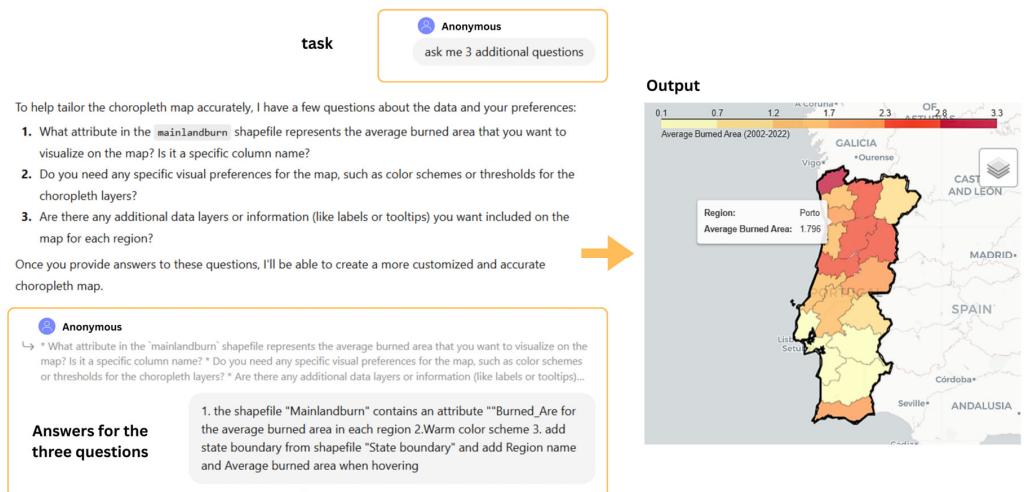


Figure 13. Cognitive Verifier prompt for creating a map field, along with interactive tooltips (symbol * is used as a new sub-question within the prompt).

5.4.2. Legend

Folium's built-in function provides legends along with maps. The legend will automatically adjust to match the map colors and classes. From the three answers given to the prompt in Figure 13, ChatGPT-4 can correctly return the five intervals based on the map classification with a warm color scheme.

5.5. Map Compositions

Besides the legend and map field, other map components are also generated by both the basic and advanced prompt, structured in the same way.

5.5.1. Title and Subtitle

The default titles provided by the AI do not accurately indicate the thematic content of where and when a phenomenon occurred. The AI usually parses the text from a given prompt directly and generates a title based on the input file's name. To solve this issue, a prompt can explicitly assign the text titles within the prompt. However, the subtitle always overlaps with the main title in a static version, and this can be resolved by describing a specific location within a map layout, such as '*moving the subtitle below the main title*'. Question Refinement prompts are useful for addressing this issue, as the AI replaces the original prompt with clearer and more precise instructions. Alternatively, an exact spacing parameter can also be specified between a title and a subtitle to prevent overlapping (see Figure 14). For an interactive map, there are no overlapping issues between the title and subtitle, because Folium typically adds them through a custom HTML script with style attributes for proper positioning.

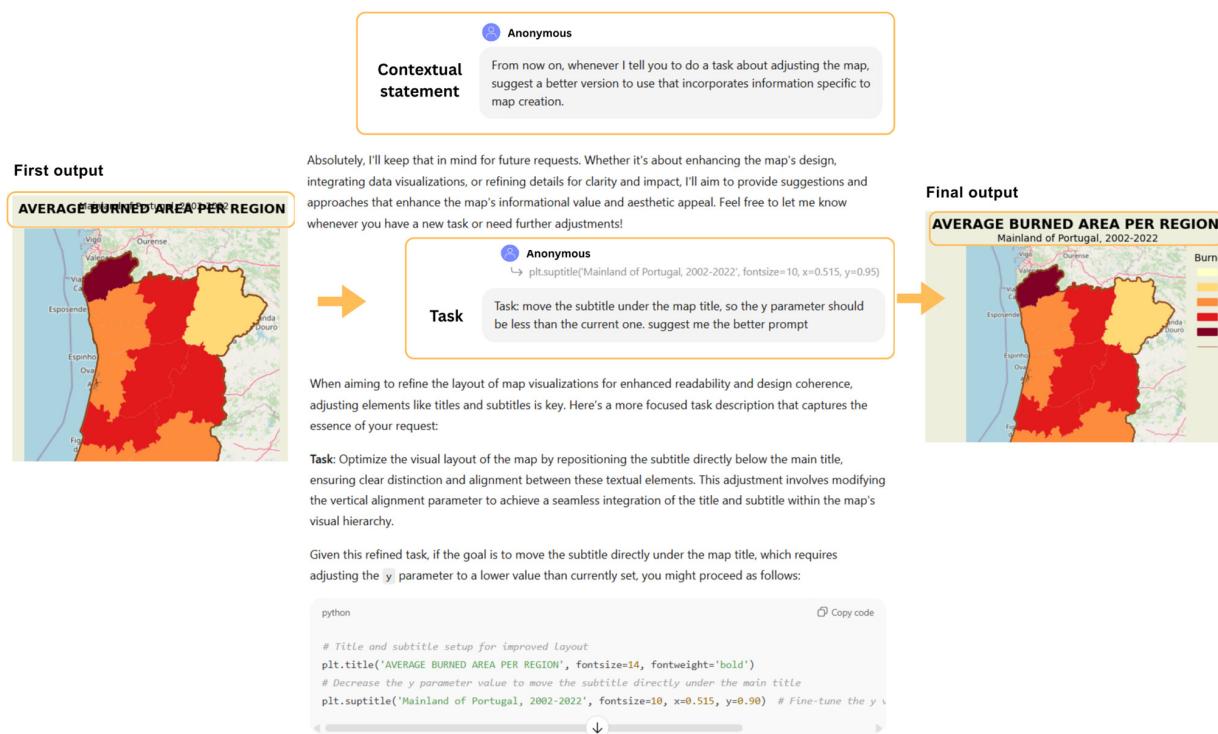
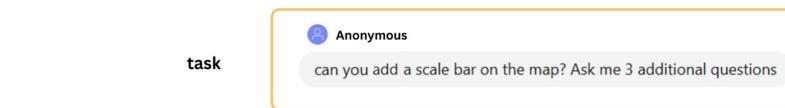


Figure 14. Using Question Refinement to resolve an overlapping issue between the main title and the subtitle.

5.5.2. Scale Bar

Creating a scale bar for a static map with GeoPandas needs to consider the following aspects (see Figure 15):

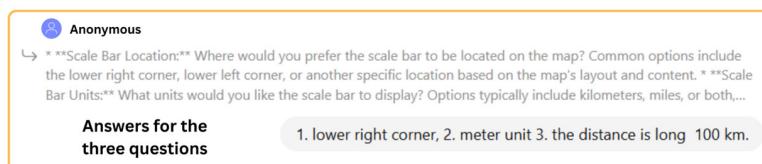
- Scale Bar Placement: The scale bar could be placed in an inappropriate location that overlaps with the other elements. Common options for prompts include the lower right corner, lower left corner, and upper left corner, which do not obscure map details.
- Scale Bar Unit of Measurement: The scale bar utilizes ‘*Matplotlib-scalebar*’, which provides metric unit options (e.g., kilometers, miles, etc.). The prompt also includes the ‘*Scale Bar Length*’, which calculates the length of the scale bar in pixels to accurately represent the unit in kilometers.
- Scale Bar Style: To make the scale bar less dominant than the map field, the prompt can customize the scale bar color for both numeric text and bar.



Adding a scale bar to your map can significantly enhance its usefulness, providing viewers with a reference for distance. To integrate a scale bar effectively, here are three questions that will help tailor its inclusion:

1. **Scale Bar Location:** Where would you prefer the scale bar to be located on the map? Common options include the lower right corner, lower left corner, or another specific location based on the map's layout and content.
2. **Scale Bar Units:** What units would you like the scale bar to display? Options typically include kilometers, miles, or both, depending on the map's audience and purpose.
3. **Scale Bar Size:** How long should the scale bar be, relative to the map's scale? This can be a specific distance (e.g., 10 km, 5 miles) or a proportion of the map's width (e.g., one-fifth of the map width).

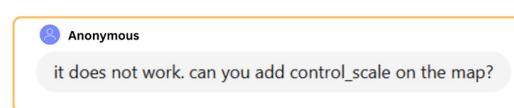
Your answers will guide the process of adding a scale bar that meets your needs, enhancing the map's readability and informational value.



100 km

Figure 15. Using Cognitive Verifier to create a scale bar in a static map (symbol * is used as a new sub-question within the prompt; symbol ** is used as a sub-question title).

The scale bar for the interactive map can be added to the map using the built-in ‘*Scale control*’ in Folium. This control automatically adds a scale bar to the map, showing distances in both kilometers and miles, which is dynamically adjusted according to the zoom level (see Figure 16).



Certainly! To add a scale to your Folium map, you can use the `control_scale` parameter directly within the `folium.Map` creation. This parameter automatically adds a scale to the map when set to `True`. It looks like there was some confusion regarding how to implement this feature directly with a simple parameter.

Here's the corrected version of your script with the `control_scale` parameter properly set:



Figure 16. A solution for resolving a scale bar error in an interactive map by adjusting Folium’s parameters.

5.5.3. Credits

To add a proper credit that includes data source, author's name, and date when the map is created, the information is assigned directly in a prompt to generate a text string (see Figure 17). The prompt also indicates that the credits should not overlap with other elements and ensure that they are less dominant in size and color. In general, the credits maintain small fonts and proper spacing between lines. For an interactive version, the prompt identifies the hyperlinks of the data sources and embeds them in the credits.

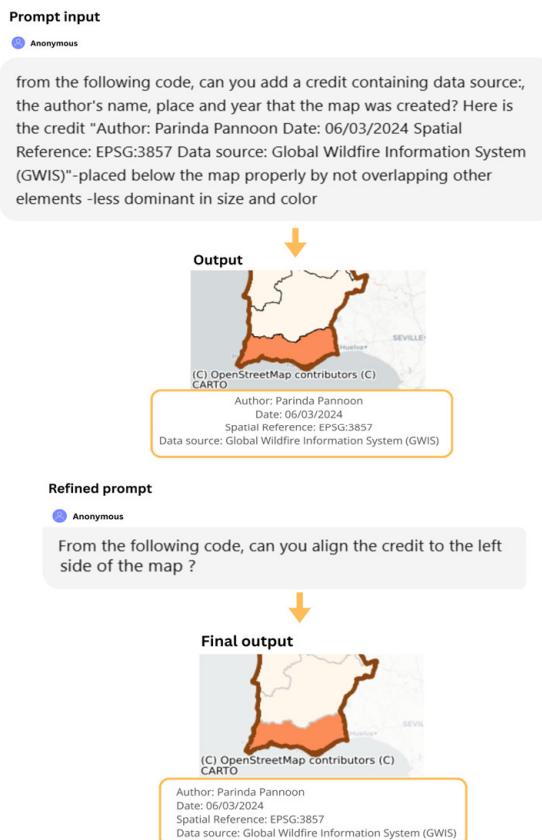


Figure 17. Basic prompts for generating credit information and adjusting credit alignment in a static map.

5.5.4. Basemaps

The AI provides several basemap providers. The common sources include OpenStreetMap, Stamen Terrain, CartoDB, and Esri imagery. However, the AI is not able to visualize some basemaps because of provider limitations. In this research, CartoDB is used as the map background because its low-saturation colors allow the map content to be effectively emphasized. Using the Cognitive Verifier prompt could reduce the number of errors, since it provides additional questions to ensure map projection handling (see Figure 18). Basemaps typically use the Web Mercator projection (EPSG:3857), and the data are reprojected to align with the tiles effectively. Moreover, the Question Refinement also enables basemap details to be rewritten into a more comprehensive and detailed prompt in the suggested version. For a static map, the '*contextily*' library adds tiles as the basemap background. It is recommended to indicate a basemap library name in the prompt to avoid errors.

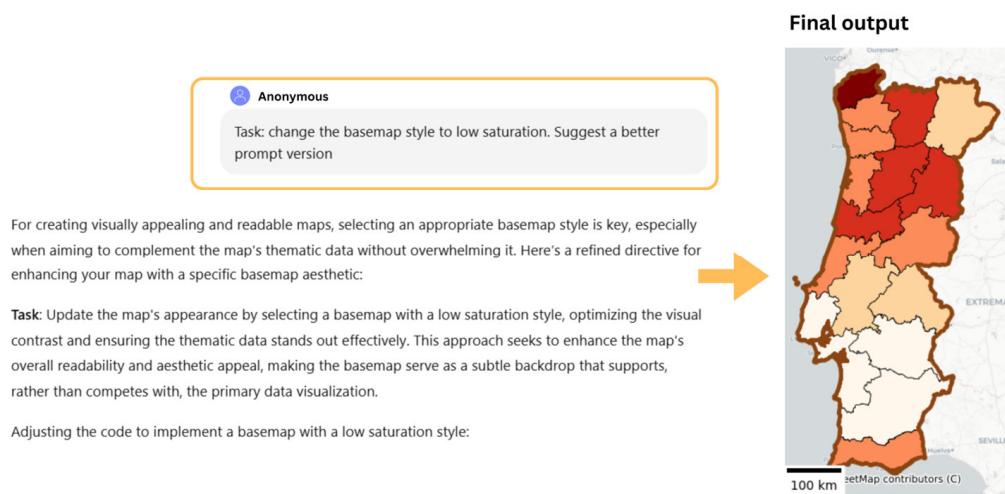


Figure 18. The Question Refinement prompt changes the default basemap style to the CartoDB basemap.

5.5.5. Data Visualization

ChatGPT-4 is capable of plotting data visualizations as subplots within a map field without using an external tool. When creating a well-designed chart or table in GIS software, it is necessary to design them independently using graphic design tools.

The charts are generated based on the prepared dataset in CSV format. For a static map (see Figure 19), the charts given by AI often overlap with other map compositions. A chart layout can be adjusted by the 'Gridspec' variable, a subplot parameter that requires specific dimensions in a prompt. This leads to charts being unsuccessfully generated by the AI in some sessions. The interactive charts are created using Plotly and converted to HTML (see Figure 20). Custom HTML elements are embedded into 'IFrame' and added to the map's root HTML. The common issue is that a chart usually does not appear on a map because of a wrong plotting method in the code. Advanced prompts can be used for customizing several small details at once, such as color, content, position, width, height, and background transparency.

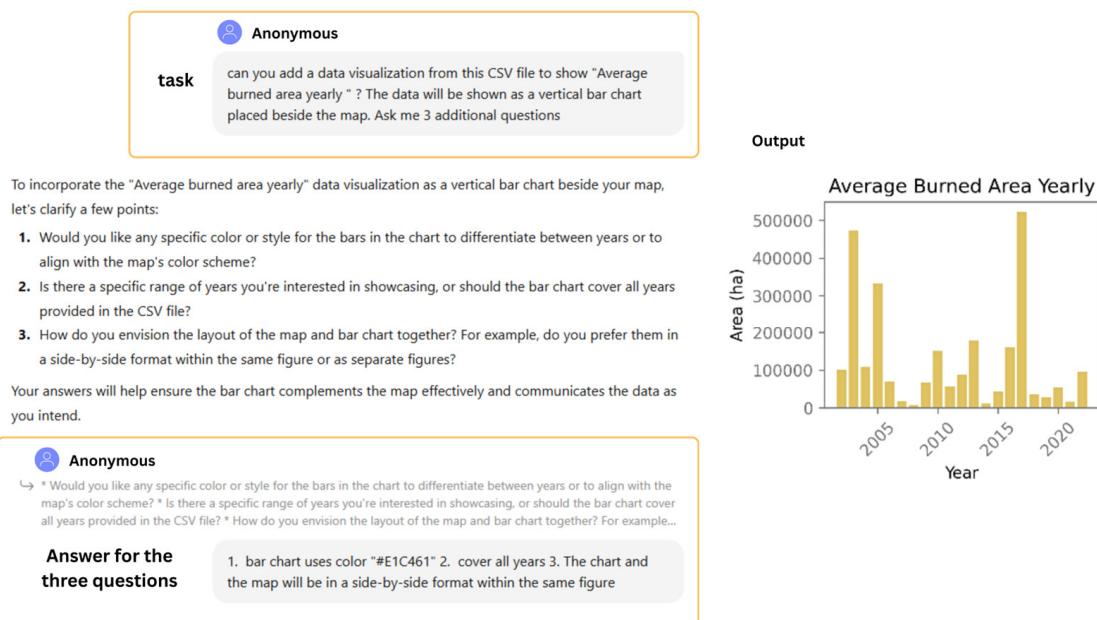


Figure 19. Using Cognitive Verifier to create a bar chart in a static map (symbol * is used as a new sub-question within the prompt; symbol # is used as hexcode number within the prompt).

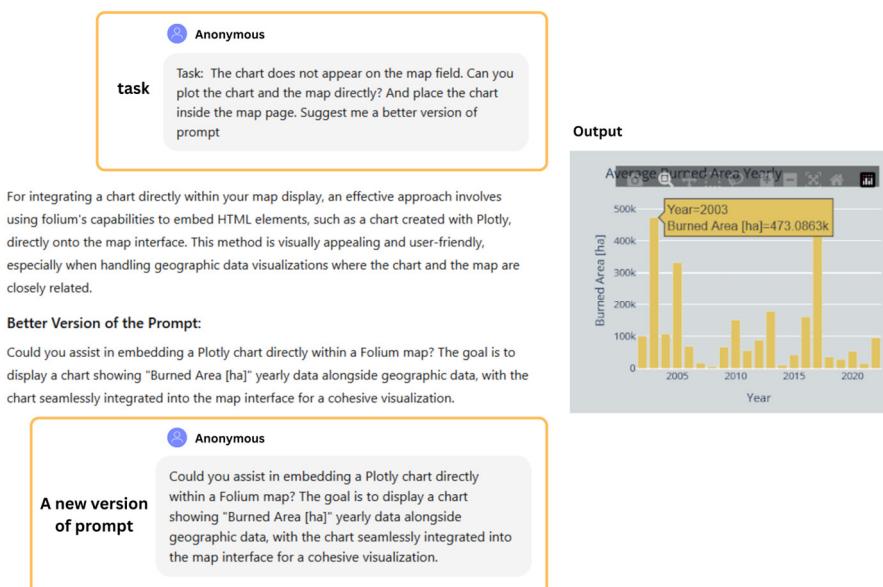


Figure 20. Using Question Refinement to create a bar chart in an interactive map.

5.5.6. Tooltips and Layer Control

Tooltips and Layer control are additional map elements for the interactive version. The '*folium.GeoJsonTooltip*' function displays more informative details. Additionally, map effects can be added, along with the tooltips, when the mouse hovers over each region. The State boundary and the Mainland area are added in the layer control widget, which uses Folium's '*LayerControl*' function (see Figure 21). These interactive widgets do not require complex requirements in a prompt. Among five iterations, ChatGPT-4 often returns desired results with few attempts.

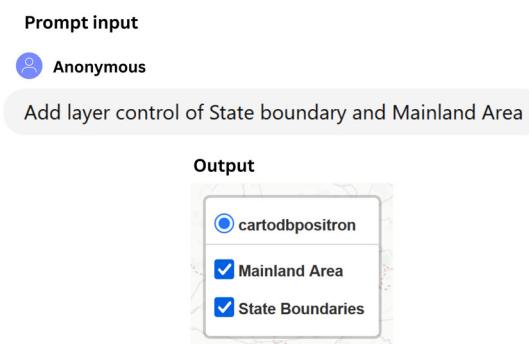


Figure 21. Basic prompt for creating a Layer control function in an interactive map.

6. Quality Assessment

This chapter aims to answer the research question of how well the quality of AI-generated maps compares to that of human-generated maps. The AI-generated map has several limitations; this evaluation can measure whether the ability of ChatGPT-4 achieves as many correct cartographic principles as the traditional method. This stage considers two main steps for assessing map quality. The first one is setting map references based on cartography principles. The second step is considering the map readability, which consists of size, color, lettering, and interpretation.

According to [37], a predefined set of cartographic principles that apply to a particular map is referred to as a map specification. Setting map references according to predefined specifications ensures map quality. Each map composition in both AI-generated and human-generated maps is created according to the same specifications, adhering to

cartographic rules. Such a method helps identify the strengths and weaknesses of a map. The more the map composition complies with cartographic rules, the better the quality and communication efficiency will be. Map readability focuses on the ease of reading and interpreting a map. The first issue of the AI-generated map is placement; the AI does not consider overlapping with the other map elements, leading to illegible and less esthetic appeal. In the typographical hierarchy, a subtitle was generated with the same characteristics as the main title, which is not supposed to attract so much attention, so a prompt needed to be customized to meet the specification. The map quality is divided into three levels of suitability, most suitable, intermediate, and least suitable, based on how much the maps meet cartographic and readability criteria. The map criteria have been established according to cartographic design and rules, as outlined in the book by [14].

Evaluating a choropleth map for cartographic suitability involves several key considerations, as set in Table 1. The suitability levels were set according to the thematic method, data classification, sequential colors, legend colors, and description (see the full list of map composition criteria in the Appendix A). The most important characteristics of the choropleth map are data normalization and discrete classes with a sequential color scheme. Creating a legend in ChatGPT-4 requires users to specify a legend color and description that corresponds to the map, as ChatGPT-4 is likely to provide inconsistent results. For example, the legend represents a continuous legend, or legend labels describe the severity levels instead of showing one value in a class.

Table 1. Suitability criteria for evaluating a choropleth map.

| | Suitability Levels | | |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Map Compositions | Least Suitable | Intermediate | Most Suitable |
| 1. legend | <ul style="list-style-type: none"> - Does not represent burned area values in each class. - Uses a legend of a continuous bar. - Colors do not correspond on the map. | <ul style="list-style-type: none"> - Represents ranges of the value in each class. - Discrete legend from high to low, with graduated colors. - Colors do not correspond on the map, the tones are slightly different but the data can still be understood. - Upper and lower values of each class are duplicated. | <ul style="list-style-type: none"> - Represents ranges of the value in each class. - Colors correspond with the color on the map. - The number of classes is the same as the map. - The data intervals are classified properly according to statistical methods. |
| 2. map field | <ul style="list-style-type: none"> - Does not use a sequential scheme on areas. - Does not use a warm color scheme. - Data is organized into more than seven or less than four classes. - Colors do not associate with the actual data. - The colors on the map do not correspond to the legend. | <ul style="list-style-type: none"> - Uses a sequential scheme on areas. - Single data value is represented by each area as a ratio. - Uses a warm color scheme - Data is organized into more than seven or less than four classes. - The colors on the map are slightly different from the legend. | <ul style="list-style-type: none"> - Uses graduated-color: sequential scheme on areas. - The data values are ratios (normalized) represented by each area. - Uses a warm color scheme. - The color symbols support the reader in making comparisons between high and low. - Data is organized into more than seven or less than four classes. - The colors on the map correspond to the legend. |

7. Discussion

This section summarizes the results of the two stages and answers all the research questions. In the first research question, the cartographic completeness was assessed to

reveal the influence of different prompt patterns. The AI's capability and learning ability were assessed through several attempts, errors, and hallucinated results for the second research question. The map suitability levels, as set in Section 5, were also evaluated for comparison with traditional methods for the third research question.

7.1. Cartographic Completeness (RQ1)

The map compositions created in Section 5 were evaluated to assess how many map elements were created successfully. The results reveal that most static and interactive map fields can be successfully generated for all five iterations (see Figure 22). There are two legends associated with the static maps that are inconsistent with the map colors. Scale bar, subtitle, and labels of the static maps are also unsuccessfully created by the AI. Comparing the number of static map compositions that cannot be achieved with the interactive version, there are 15 map elements that fail in interactive maps, including credits, scale bar, labels, data visualization, tooltips, and layer control; most of these are from advanced prompts. Using the advanced prompt patterns for creating static choropleth maps can achieve more elements than interactive maps.

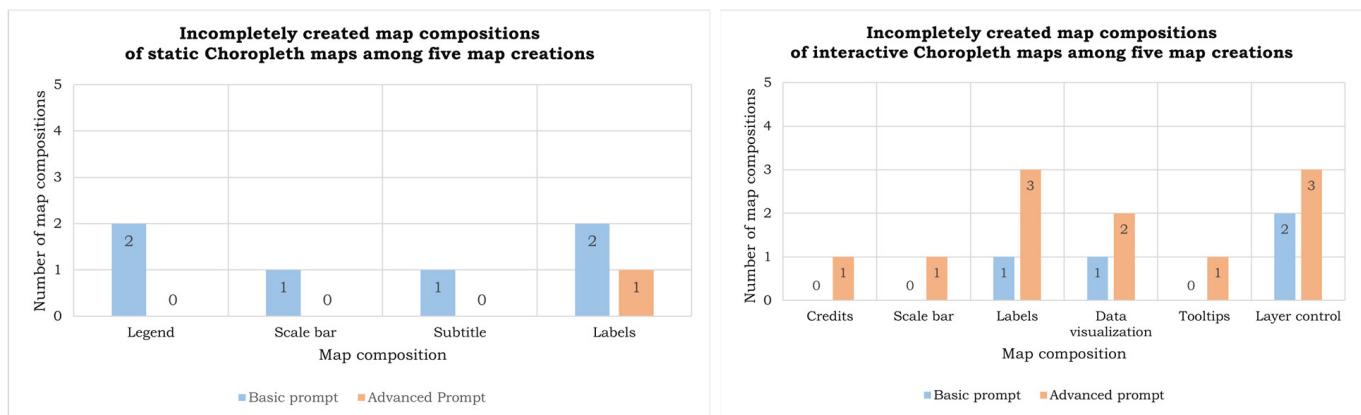


Figure 22. The number of map compositions of the static (left) and interactive (right) maps that were unsuccessfully created among five iterations.

A common issue is that map labels cannot be completed using advanced prompts on the interactive maps. The static version has only one label that is unsuccessfully achieved by an advanced prompt. Additionally, incomplete layer control is also commonly found, because it generally cannot organize the map layers to be toggled effectively.

7.2. AI's Capability in Map Generation Across Two Prompt Patterns (RQ2)

This section aims to answer how well AI can create appropriate choropleth maps across different prompt patterns. To understand AI's capability in generating outputs, this paper focuses on evaluating the number of attempts, error messages, and incorrect results for each prompt pattern. Analyzing these factors allows users to identify best practices for successful map creation and to understand the limitations of AI in the cartographic domain.

7.2.1. Number of Attempts

An attempt indicates the number of prompts used during map creation until a final map output is achieved. Creating a thematic map requires several attempts, as ChatGPT-4 can sometimes provide hallucinated or incorrect results. Moreover, map elements can be degraded and disappear during the process, and then those elements need to be regenerated again, with the use of more refined prompts. Figure 23 indicates that the basic prompt

pattern in the choropleth maps uses more attempts in all five iterations (362 attempts). On the other hand, the advanced prompt requires fewer attempts (261 attempts). All the map compositions in the basic prompt have a wider range of attempts, indicating a high variability of attempts. When using the advanced prompt, the number of attempts is decreased, which means that the advanced prompt facilitates creating a thematic map and reduces the number of iterations to achieve the desired outputs.

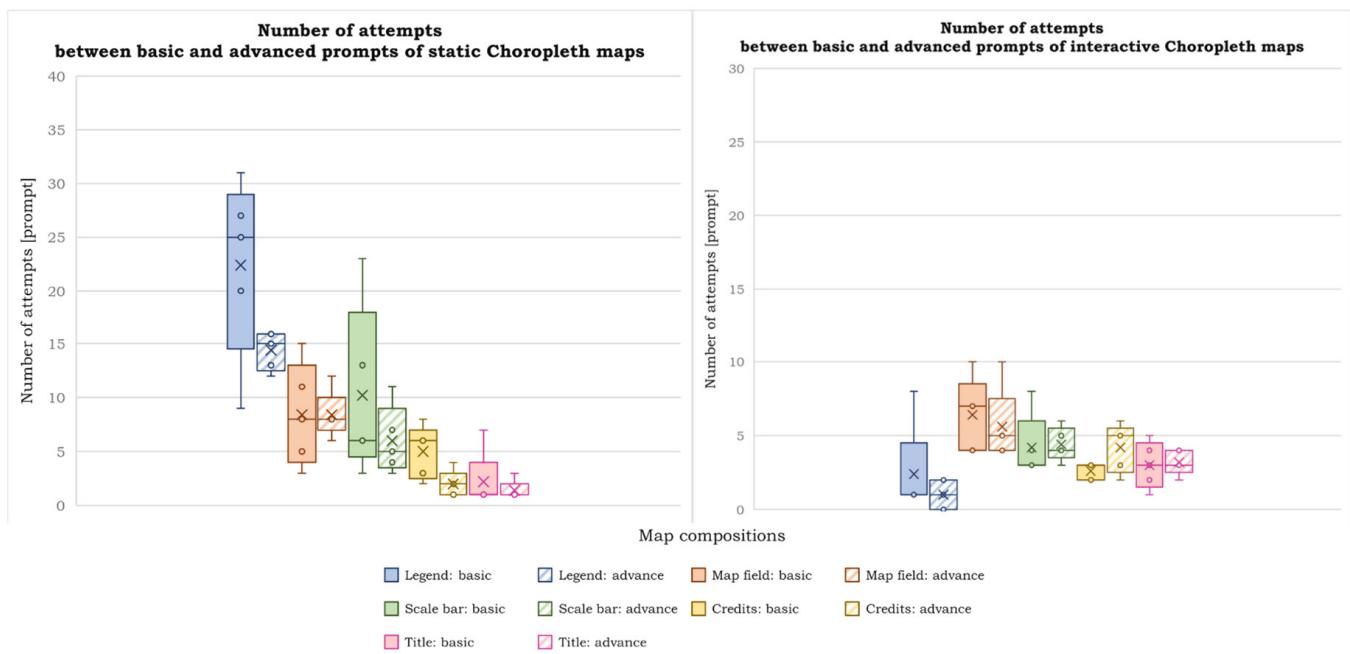


Figure 23. Comparison of the number of attempts: basic vs. advanced prompts on the static maps (left) and interactive maps (right).

For the interactive map, the map field of the basic prompt has a similar range to the advanced pattern, with a slightly higher distribution. The number of attempts used in the advanced prompt for the legends shows a small range from zero to two attempts, which means that some legends in the interactive map are automatically created by the predefined function of the Folium library. For the credits and title in an interactive map, the advanced prompt can lead to more attempts, as the map library does not have a predefined text box for adding text elements. Therefore, it requires more attempts to create HTML elements overlaid on the layout.

7.2.2. Number of Incorrect Results

The results reveal (see Figure 24) how the two types of prompts return incorrect results caused by the AI. ChatGPT-4 can cause hallucinations in the code results, which means generating facts with reasoning errors [39]. Such results indicate that the model's learning ability is limited by its cartographic knowledge, resulting in incorrect outputs. For the static maps, the incorrect results vary highly across map elements of legend, map field, scale bar, and data visualization, with some elements showing a significant reduction in errors when using advanced prompts. In particular, legends are made more accurately with the advanced prompts than the basic prompt. The advanced prompt provides a total number of incorrect results that is lower than that of the basic prompt.

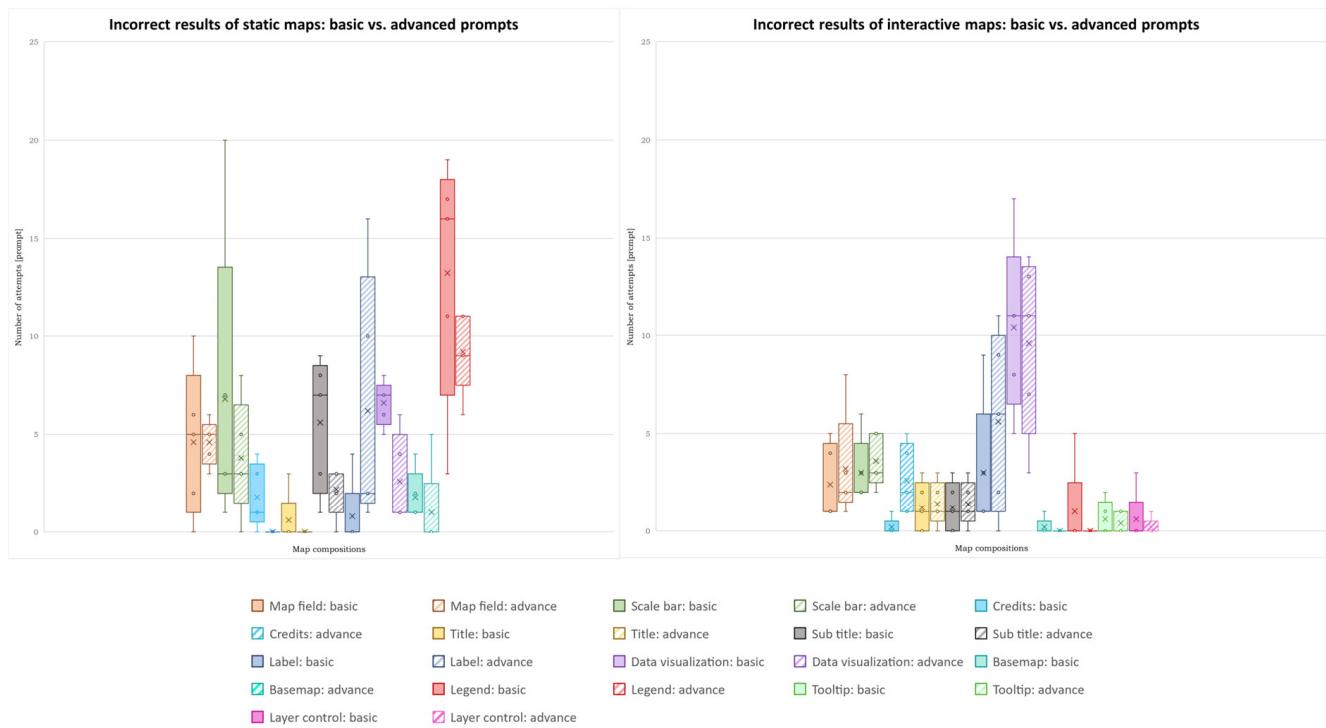


Figure 24. Comparison of the number of incorrect results: basic vs. advanced prompts on the static maps (**left**) and interactive maps (**right**).

The advanced prompts for the interactive maps resulted in more incorrect outputs in the choropleth maps across all five iterations. The map field generated by both prompt types shows minimal differences and is even slightly higher in the advanced prompts. However, the advanced prompts substantially reduce the number of code errors when embedding charts or data visualizations. Labels, titles, and credits do not show significant differences between the two prompt types, indicating that the AI performs similarly for both in these map elements.

In conclusion, ChatGPT-4 produces more errors or hallucinated results in complex components such as map fields, legends, and data visualizations. The use of advanced prompts consistently reduces incorrect results in some map elements, particularly in data visualization and legend design. However, for map fields, the differences between the two prompt patterns are small, and advanced prompts produce even more incorrect results for interactive maps.

7.2.3. Number of Error Messages

Error messages can occur during code generation by ChatGPT; the cause may be due to several factors. The main error that usually happens in the experiments is the Error Analyzing issue. The error is potentially caused by model bias in the training data, complex datasets, and inadequate data handling capabilities. The issue interrupts the code generation because ChatGPT-4 cannot process and read the shapefile datasets properly. This can occur with certain types of spatial data and issues related to processing geometries. Assessing the technical issue helps to understand the limitations of the ChatGPT-4 model in processing spatial data and its performance in creating maps.

In conclusion, the Error Analyzing issue consistently occurs when creating map fields and data visualizations. Shapefiles are processed for plotting maps, and CSV files are used to generate charts or other visualizations within the map layout. ChatGPT-4 processes the provided data by unzipping the files and preparing them as data frames. Errors may arise when the AI is unable to process, plot, or load these files. Most map

compositions generated from advanced prompts result in fewer errors, but they remain inconsistent, as more errors appear in the interactive versions using advanced prompts, as shown in Table 2. This suggests that errors are more likely to occur when creating complex interactive elements.

Table 2. The number of error messages occurred during code generation.

| Map Compositions | Error Messages Occurred During Code Generation | | | |
|--------------------|------------------------------------------------|-----------------|-----------------|-----------------|
| | Static Map | | Interactive Map | |
| | Basic Prompt | Advanced Prompt | Basic Prompt | Advanced Prompt |
| Legend | 1 | 0 | 0 | 0 |
| Map field | 12 | 0 | 3 | 6 |
| Scale bar | 4 | 0 | 1 | 0 |
| Credits | 0 | 0 | 1 | 0 |
| Data visualization | 3 | 3 | 8 | 15 |

According to the findings, there are key principles that can be engineered for map tasks. Clear and precise instructions help the LLM focus on the task and reduce ambiguity in the output. Including prompt details such as symbol libraries, types of thematic methods, color schemes, and expected outcomes leads to better results. Without clear context, the model may produce generic or less relevant responses.

Assigning roles to the LLM (e.g., “Act as a cartographer”) helps in tailoring the responses to domain-specific requirements. When the LLM hallucinates outputs, specifying explicit parameters in the libraries, such as legend intervals and scale bar parameters, can minimize errors and improve alignment with cartographic rules.

However, there are limitations and challenges in refining prompts. Challenges arise when the data on the choropleth map is not normalized, because ChatGPT-4 usually returns the choropleth maps without normalization due to the lack of library-specific functionality and the inability to handle spatial attributes. Therefore, refining prompts, either on basic or advanced prompts, could not be achieved in this case.

The findings reveal deterministic characteristics in AI-driven map creation. ChatGPT-4 always provides different results on each execution due to the nature of LLM systems, as probabilistic models and reasoning are used to make decisions. The map designed by ChatGPT-4 has a specific layout style; the given symbol and color scheme do not differ much because of predefined library functions and the constrained knowledge of large language models (LLMs). On the other hand, a heuristic for generating maps learns from training data to make decisions, and it can also learn to refine new responses based on previous errors. ChatGPT is not guaranteed to provide an accurate choropleth map on the first attempt, but it is valuable when dealing with large spatial datasets.

7.3. Map Quality: AI vs. Traditional Method (RQ3)

Evaluating the quality of AI-generated maps can help indicate how far their quality differs from that of traditional methods (see Figure 25). The results from both prompt types are combined and compared with human-generated maps, as shown in Figure 26.

Among the 10 maps created using the two prompt patterns, the choropleth maps show a similar number of suitable and intermediate outputs, particularly in aspects such as legends, map fields, and charts. Some maps show the lowest suitability in elements like scales, labels, and legends due to code generation errors. For instance, slight differences between legend colors and map fields can lead to an intermediate rating. Five out of

ten map fields are rated as most suitable, as is the case for the legends. However, the AI-generated maps lack flexibility in the placement of elements. Most subtitles receive an intermediate rating because they overlap with the main title.

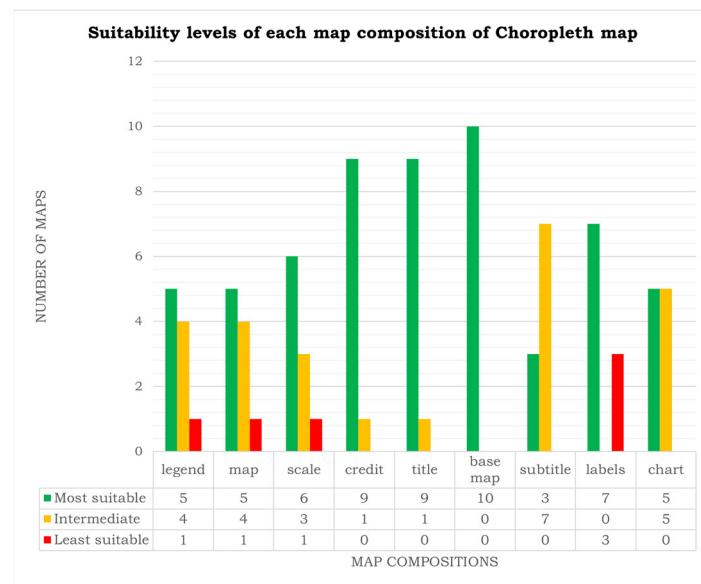


Figure 25. Suitability levels of each map composition.

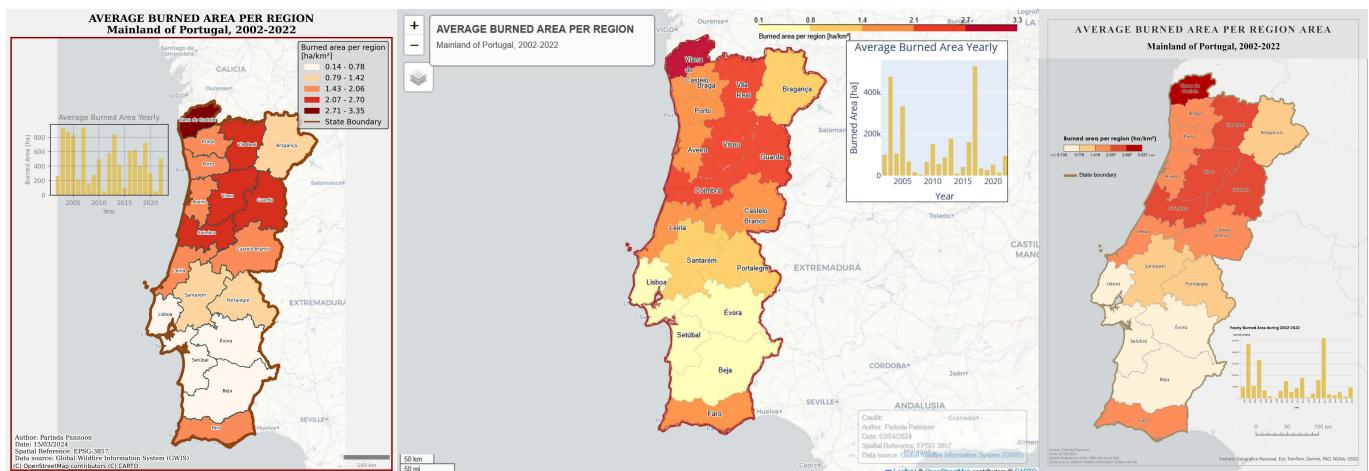


Figure 26. The static (left), interactive (middle), and human-made (right) map outputs.

7.4. Comparison with Recent Work on ChatGPT-4

Recent work by [28] demonstrated that ChatGPT can support map creation but often struggles with spatial reasoning, data interpretation, and adherence to cartographic conventions—issues that closely align with our observations. Their study similarly noted that ChatGPT tends to misinterpret thematic map terminology and requires iterative refinement, which supports our findings regarding the instability of legend generation and prompt sensitivity in choropleth mapping tasks. Likewise, ref. [27] introduced the concept of Autonomous GIS, demonstrating that GPT-4 can automate parts of the GIS workflow but frequently fails in tasks involving code debugging, complex geometry handling, or spatial data loading. This mirrors our results, particularly in the context of shapefile processing errors, geometry-related failures in Folium, and the model's limited ability to maintain consistent map components across iterations. While their system integrates multiple tools to mitigate these issues, our results confirm that standalone LLM prompting remains insufficient for robust geospatial automation. Taken together, these studies support the broader

conclusion that LLMs—despite their promising capabilities—currently require substantial human oversight for cartographic reliability, and that their performance is highly dependent on the prompt specificity. Our findings reinforce this trend while providing a detailed, map-element-level evaluation within the context of choropleth mapping.

7.5. Time Efficiency

Time efficiency is one of the key factors in producing maps. Creating accurate maps within a short time frame can enhance decision-making on projects, especially rapid mapping for emergency responses. The study results reveal that the map created by ChatGPT-4 generally took up to 4–6 h to obtain the final result; a map from the AI's code requires execution in external tools to visualize the result. This indicates that using the AI tool is not an efficient approach for creating a thematic map with comprehensive elements. On the other hand, ArcGIS Pro is a user-friendly tool for creating maps that are quick and easy to use because of its comprehensive functionality and intuitive map view on the interface. However, creating a map with GIS software can be costly and requires a subscription, and the map quality depends on the user's expertise and GIS knowledge. ChatGPT-4 is suitable for a quick visualization of geospatial data, as it has the capability of plotting a map field and manipulating data within a short time. The weakness is that it cannot generate a complex map correctly with all map elements as quickly as humans.

8. Limitations and Future Work

This study demonstrates the potential of GeoAI to support automated map creation, conducted by ChatGPT-4, as state-of-the-art at the time of the research (spring 2024). The authors are aware of the limitations associated with the capabilities and performance of ChatGPT-4, particularly in the context of rapid AI evolution. The model used in this study showed constraints in several areas: inconsistent interpretation of prompts, especially when handling cartographic terminology; occasional hallucinations in code, leading to non-functional or syntactically incorrect outputs; limited ability to handle geospatial data structures, particularly shapefiles with complex geometries; and insufficient contextual understanding of cartographic principles, resulting in issues such as incorrect legend construction, unsuitable color schemes, or suboptimal element placement.

Even with advanced prompt strategies (Cognitive Verifier and Question Refinement), several map elements required repeated prompt adjustments. The LLM was prone to degrading previously correct elements when new instructions were introduced, and some outputs—especially interactive components—could not be completed reliably. This underlines that prompt engineering cannot fully compensate for the model's limited internal representation of cartographic rules, and that AI-generated maps still require post-processing and expert supervision. While ChatGPT-4 can follow predefined specifications, it lacks human-like sensitivity to esthetics, hierarchy, and layout optimization. It cannot reliably detect overlapping elements, evaluate perceptual quality, or judge map composition holistically. As a result, many outputs achieved only intermediate suitability scores when compared to traditional human-made maps.

Findings of this study are based on one thematic method (choropleth), one dataset, specific Python libraries (GeoPandas, Folium), and one LLM version. Thus, the results cannot be directly generalized to other cartographic methods or future models. Nevertheless, the authors of this study believe that preserving the ChatGPT-4 results is an essential contribution to the GeoAI topic: as a “historical snapshot”; as the definition of the baseline of LLM-assisted cartography at the moment of the study; and as a contribution to the rapid progression of GeoAI technologies, leading to its integration into cartographic practice.

This research is built on the author's master thesis [40], which has not been previously published in a peer-reviewed form. As representatives of the academic environment, authors consider it essential to remain actively involved in the progress of GeoAI. Authors are fully aware of the limitations of the present study but would like to continue pursuing further systematic investigations, testing newer AI models, and refining methodological approaches, so that future work can more accurately capture the capabilities and challenges of AI-driven cartography. Further research focuses on these observations, and will be focused on evaluation with new AI models (ChatGPT-5+, Gemini 2.5+, Claude Sonnet 4.5+). With the models capable of interpreting code and spatial data, future studies could examine workflows in which the AI not only generates code but also reads and critiques the visual output, enabling iterative self-correction. Finally, beyond choropleth maps, future work should investigate AI-assisted creation of other cartographic methods such as proportional symbol maps, dot density maps, graduated symbols maps, or flow maps.

Finally, ethical considerations are increasingly important when generating or evaluating AI-based outputs. Issues of accuracy and reliability arise because LLMs may produce plausible-looking but incorrect spatial representations, potentially misleading users who assume algorithmic precision. Moreover, LLMs do not automatically follow cartographic standards and may generate outputs that violate established principles of classification (see Figure 7), symbology or visual hierarchy. This is further complicated by the possibility of hallucinations, in which the model produces fabricated values, incorrect code, or spatially inconsistent elements without signaling uncertainty. These concerns are consistent with the findings of [31], which highlighted the ethical risks associated with AI-generated maps, including opacity, misrepresentation, and the challenge of ensuring trustworthy outputs.

9. Conclusions

This study demonstrates that ChatGPT-4 is capable of generating both static and interactive choropleth maps using GeoPandas and Folium, with advanced prompt patterns leading to fewer errors, greater completeness of map elements, and more consistent outputs compared to basic prompts. The evaluation shows that AI-generated maps can meet several core cartographic requirements, although the quality still varies across map components and iterations. Human-made maps remain more flexible and precise, especially in layout, symbology, and adherence to cartographic rules.

The findings suggest that large language models can already support certain aspects of the cartographic workflow, particularly for users who lack programming or GIS experience. Integrating LLMs with geospatial libraries offers a viable pathway beyond text-to-image approaches, resulting in spatially accurate maps rather than visually appealing but geographically distorted images. Prompt engineering—especially through the use of refinement and verification patterns—plays a crucial role in enhancing map quality, providing practical guidelines for future GeoAI-assisted cartography. The main advantage of the findings is their ability to interpret and analyze large datasets. This is useful for users who need to visualize geospatial data in thematic maps at a glance. ChatGPT-4 creates a basic map that does not contain complex elements (e.g., scale bar, data visualization) much faster than humans can. It also helps in constructing code for initial visualization tasks, such as web map development. This advantage facilitates users who want to create maps from the code without starting from scratch.

Despite these advances, the method shows clear limitations. ChatGPT-4 struggles with data normalization, geometry handling, optimal placement of map elements, and maintaining consistency when modifying existing code. Some errors stem from inherent model constraints, such as hallucinations and insufficient domain understanding. As the

study was conducted in spring 2024, the results reflect the state of ChatGPT-4 at that time and do not capture improvements introduced in newer models.

The authors are aware that the development of both GIS and GeoAI is progressing very rapidly [41]. Future work should therefore explore more advanced and contemporary LLMs, including ChatGPT-5, Small Language Models trained on cartographic knowledge [42], and workflows integrating retrieval-augmented generation (RAG) to reduce hallucinations [43]. Expanding experiments to additional map types and developing quantitative evaluation metrics will further clarify the generalizability of the approach. Continued research is essential.

Author Contributions: Conceptualization, Rostislav Netek; methodology, Parinda Panno; resources, Parinda Panno; writing—original draft preparation, Parinda Panno; writing—review and editing, Rostislav Netek; visualization, Parinda Panno; supervision, Rostislav Netek. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data, code snippet and materials that support the findings of this study can be found on this GitHub repository: <https://github.com/GeoAI-Map/Materials-for-Creating-maps-by-Artificial-Intelligence> (accessed on 17 November 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Suitability criteria for evaluating the map quality can be found on: <https://github.com/GeoAI-Map/Materials-for-Creating-maps-by-Artificial-Intelligence/blob/main/Appendix.pdf> (accessed on 1 December 2025).

Appendix B. Static Maps

Appendix B.1. Basic Prompt Pattern

Attempt 1: <https://chat.openai.com/share/1763437f-2964-4fba-8e16-8cbf1b0f4d78> (accessed on 1 December 2025) &

<https://chat.openai.com/share/a054ab0b-49fa-4e66-b0cd-207691a0a583> (accessed on 1 December 2025).

Attempt 2: <https://chat.openai.com/share/b5dc860d-4103-4afe-9be4-fb9db2184991> (accessed on 1 December 2025).

Attempt 3: <https://chat.openai.com/share/ec5fa524-7e81-4244-8d5f-00651ca1aced> (accessed on 1 December 2025).

Attempt 4: <https://chat.openai.com/share/703279b3-c8ab-463c-b5ce-ec31a3d013f8> (accessed on 1 December 2025).

Attempt 5: <https://chat.openai.com/share/7971c36d-c642-4508-905c-001b87db6f54> (accessed on 1 December 2025).

Appendix B.2. Advanced Prompt Pattern

Attempt 1: <https://chat.openai.com/share/559838e2-a63d-4389-be7a-10b97786c0c9> (accessed on 1 December 2025).

Attempt 2: <https://chat.openai.com/share/c3b86d41-b9ab-4fdf-945c-152f6de6fc10> (accessed on 1 December 2025).

Attempt 3: <https://chat.openai.com/share/139fcdd1-eb5e-4128-a7c1-fc320bf19662> (accessed on 1 December 2025).

Attempt 4: <https://chat.openai.com/share/a6dc8f0f-dbde-4100-b8f0-daa6989db616> (accessed on 1 December 2025).

Attempt 5: <https://chat.openai.com/share/5ef52675-f0bd-4efe-ac36-b55b504a65fc> (accessed on 1 December 2025).

Appendix C. Interactive Maps

Appendix C.1. Basic Prompt Pattern

Attempt 1: <https://chat.openai.com/share/7504daf8-2683-40ff-8a88-20b58c800bab> (accessed on 1 December 2025).

Attempt 2: <https://chat.openai.com/share/1d0d0d73-d3ca-4e1a-b2b1-ada8a990f06f> (accessed on 1 December 2025).

Attempt 3: <https://chat.openai.com/share/1b9e97da-3936-44f5-bc9b-860fccedb114> (accessed on 1 December 2025).

Attempt 4: <https://chat.openai.com/share/32478f5e-47ca-4c51-b35b-214d3ba79090> (accessed on 1 December 2025).

Attempt 5: <https://chat.openai.com/share/f67d0a17-e253-440b-a430-1daf94c29f72> (accessed on 1 December 2025).

Appendix C.2. Advanced Prompt Pattern

Attempt 1: <https://chat.openai.com/share/12607c2e-5bb4-4dc0-acc0-a5fa3a3e02e4> (accessed on 1 December 2025).

Attempt 2: <https://chat.openai.com/share/40ce52d0-250b-43f4-bd6f-8c1e51c924ab> (accessed on 1 December 2025).

Attempt 3: <https://chat.openai.com/share/e867706a-5658-4d72-842f-a6d6616e23de> (accessed on 1 December 2025).

Attempt 4: <https://chat.openai.com/share/bfa7e0df-a05a-4bfa-b034-223a222a2907> (accessed on 1 December 2025).

Attempt 5: <https://chat.openai.com/share/061694a8-0084-4e65-93b2-e360eafa340f> (accessed on 1 December 2025).

References

1. Rashid, A.B.; Kausik, M.A.K. AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Adv.* **2024**, *7*, 100277. [[CrossRef](#)]
2. Lakshmi Aishwarya, G.; Satyanarayana, V.; Singh, M.K.; Kumar, S. Contemporary Evolution of Artificial Intelligence (AI): An Overview and Applications. In *Advances in Transdisciplinary Engineering*; Singari, R.M., Kankar, P.K., Eds.; IOS Press: Amsterdam, The Netherlands, 2022. [[CrossRef](#)]
3. Youvan, D.C. Parallel Precision: The Role of GPUs in the Acceleration of Artificial Intelligence. 2023. Available online: https://www.researchgate.net/publication/375184141_Parallel_Precision_The_Role_of_GPUs_in_the_Acceleration_of_Artificial_Intelligence (accessed on 17 November 2025).
4. Introducing ChatGPT. 2024. Available online: <https://openai.com/index/chatgpt/> (accessed on 17 November 2025).
5. Kovari, A. Explainable AI chatbots towards XAI ChatGPT: A review. *Heliyon* **2025**, *11*, e42077. [[CrossRef](#)]
6. Acharya, V. Generative AI and the Transformation of Software Development Practices. *arXiv* **2025**, arXiv:2510.10819. [[CrossRef](#)]
7. Hong, M.K.; Hakimi, S.; Chen, Y.-Y.; Toyoda, H.; Wu, C.; Klenk, M. Generative AI for Product Design: Getting the Right Design and the Design Right. *arXiv* **2023**, arXiv:2306.01217. [[CrossRef](#)]
8. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 121–154. [[CrossRef](#)]
9. Li, W.; Hsu, C.-Y. GeoAI for Large-Scale Image Analysis and Machine Vision: Recent Progress of Artificial Intelligence in Geography. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 385. [[CrossRef](#)]
10. Mai, G.; Xie, Y.; Jia, X.; Lao, N.; Rao, J.; Zhu, Q.; Liu, Z.; Chiang, Y.-Y.; Jiao, J. Towards the next generation of Geospatial Artificial Intelligence. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *136*, 104368. [[CrossRef](#)]
11. Lammers, R.; Li, A.; Nag, S.; Ravindra, V. Prediction models for urban flood evolution for satellite remote sensing. *J. Hydrol.* **2021**, *603*, 127175. [[CrossRef](#)]
12. Kaiser, S.K.; Rodrigues, F.; Azevedo, C.L.; Kaack, L.H. Spatio-Temporal Graph Neural Network for Urban Spaces: Interpolating Citywide Traffic Volume. *arXiv* **2025**, arXiv:2505.06292. [[CrossRef](#)]

13. Kang, Y.; Wang, C. Envisioning Generative Artificial Intelligence in Cartography and Mapmaking. *arXiv* **2025**, arXiv:2508.09028. [CrossRef]
14. Field, K. *Cartography The Definitive Guide to Making Maps*; Esri Press: Redlands, CA, USA, 2018.
15. Openshaw, S.; Openshaw, C. *Artificial Intelligence in Geography*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1997.
16. Richter, K.-F.; Scheider, S. Current topics and challenges in geoAI. *KI-Künstl. Intell.* **2023**, 37, 11–16. [CrossRef]
17. Esri. History of GIS | Timeline of Early History & the Future of GIS. Available online: <https://www.esri.com/en-us/what-is-gis/history-of-gis> (accessed on 2 January 2024).
18. Dardas, A. GeoAI Series #2: The Birth and Evolution of GeoAI. 2020. Available online: <https://resources.esri.ca/education/geoai-series-2-the-birth-and-evolution-of-geoai> (accessed on 22 December 2023).
19. Li, S.; Dragicevic, S.; Castro, F.A.; Sester, M.; Winter, S.; Coltekin, A.; Pettit, C.; Jiang, B.; Haworth, J.; Stein, A.; et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* **2016**, 115, 119–133. [CrossRef]
20. Netek R (2013): Interconnection of Rich Internet Application and Cloud Computing for Web Map Solutions. In Proceedings of the 13th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria, 16–22 June 2013; Volume 1, pp. 753–760, ISBN 978-954-91818-9-0, ISSN 1314–2704. [CrossRef]
21. Snow, S. Future Impacts of GeoAI on Mapping. Esri. 2020. Available online: <https://www.esri.com/about/newsroom/arcuser/geoai-for-mapping/> (accessed on 2 January 2024).
22. Minaee, S.; Mikolov, T.; Nikzad, N.; Chenaghlu, M.; Socher, R.; Amatriain, X.; Gao, J. Large Language Models: A Survey. *arXiv* **2024**, arXiv:2402.06196. [CrossRef]
23. Zhang, Y.; He, Z.; Li, J.; Lin, J.; Guan, Q.; Yu, W. MapGPT: An autonomous framework for mapping by integrating large language model and cartographic tools. *Cartogr. Geogr. Inf. Sci.* **2024**, 51, 717–743. [CrossRef]
24. Yang, N.; Wang, Y.; Wei, Z.; Wu, F. MapColorAI: Designing contextually relevant choropleth map color schemes using a large language model. *Cartogr. Geogr. Inf. Sci.* **2025**, 1–19. [CrossRef]
25. Shomer, H.; Xu, J. Automated Label Placement on Maps via Large Language Models. *arXiv* **2025**, arXiv:2507.22952. [CrossRef]
26. Wang, C.; Kang, Y.; Gong, Z.; Zhao, P.; Feng, Y.; Zhang, W.; Li, G. CartoAgent: A multimodal large language model-powered multi-agent cartographic framework for map style transfer and evaluation. *Int. J. Geogr. Inf. Sci.* **2025**, 39, 1904–1937. [CrossRef]
27. Li, Z.; Ning, H. Autonomous GIS: The next-generation AI-powered GIS. *Int. J. Digit. Earth* **2023**, 16, 4668–4686. [CrossRef]
28. Tao, R.; Xu, J. Mapping with ChatGPT. *ISPRS Int. J. Geo-Inf.* **2023**, 12, 284. [CrossRef]
29. Memduhoglu, A. Towards AI-Assisted Mapmaking: Assessing the Capabilities of GPT-4o in Cartographic Design. *ISPRS Int. J. Geo-Inf.* **2025**, 14, 35. [CrossRef]
30. Chen, B.; Zhang, Z.; Langrené, N.; Zhu, S. Unleashing the potential of prompt engineering in Large Language Models: A comprehensive review. *arXiv* **2023**, arXiv:2310.14735. [CrossRef]
31. Kang, Y.; Zhang, Q.; Roth, R. The Ethics of AI-Generated Maps: A Study of DALLE 2 and Implications for Cartography. *arXiv* **2023**, arXiv:2304.10743. [CrossRef]
32. Juhász, L.; Mooney, P.; Hochmair, H.; Guan, B. ChatGPT as a Mapping Assistant: A Novel Method to Enrich Maps with Generative AI and Content Derived from Street-Level Photographs. 2023. Available online: <https://eartharxiv.org/repository/view/5480/> (accessed on 17 November 2025).
33. Patel, H.; Parmar, H. Prompt Engineering For Large Language Model. 2024. Available online: https://www.researchgate.net/publication/379048840_Prompt_Engineering_For_Large_Language_Model?channel=doi&linkId=65f8a42d1f0aec67e2a6673e&showFulltext=true (accessed on 17 November 2025).
34. GWIS. Global Wildfire Information System. Available online: <https://gwis.jrc.ec.europa.eu/> (accessed on 25 March 2024).
35. San-Miguel-Ayanz, J.; Artes, T.; Oom, D.; Ferrari, D.; Branco, A.; Pfeiffer, H.; Liberta, G.; De Rigo, D.; Durrant, T.; Grecchi, R.; et al. Global Wildfire Information System Country Profiles. 2020. Available online: <https://gwis-reports.s3-eu-west-1.amazonaws.com/countriesprofile/gwis.countryprofiles.pdf> (accessed on 17 November 2025).
36. Dobesova, Z.; Vavra, A.; Netek, R. Cartographic aspects of creation of plans for botanical garden and conservatories. *Int. Multidiscip. Sci. GeoConfrence SGEM* **2013**, 1, 653. [CrossRef]
37. Vansteenvoort, L.; Maeyer, P.D. An approach to the quality assessment of the cartographic representation of thematic information. In Proceedings of the 22th International Cartographic Conference (ICA), La Coruna, Spain, 9–16 July 2005.
38. White, J.; Fu, Q.; Hays, S.; Sandborn, M.; Olea, C.; Gilbert, H.; Elnashar, A.; Spencer-Smith, J.; Schmidt, D.C. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv* **2023**, arXiv:2302.11382. [CrossRef]
39. OpenAI; Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; et al. GPT-4 Technical Report. *arXiv* **2024**, arXiv:2303.08774. [CrossRef]
40. Panno, P. Automated Map Creation Using Large Language Models: A Case Study with ChatGPT-4. Master’s Thesis, Palacký University Olomouc, Olomouc, Czech Republic, 2024. Available online: https://www.geoinformatics.upol.cz/dprace/magisterske/panno24/docs/Panno_n_thesis.pdf (accessed on 17 November 2025).

41. Konicek, J.; Netek, R.; Burian, T.; Novakova, T.; Kaplan, J. Non-Spatial Data towards Spatially Located News about COVID-19: A Semi-Automated Aggregator of Pandemic Data from (Social) Media within the Olomouc Region, Czechia. *Data* **2020**, *5*, 76. [[CrossRef](#)]
42. Raza, M. LLMs vs. SLMs: The Differences in Large & Small Language Models. Splunk. 2024. Available online: https://www.splunk.com/en_us/blog/learn/language-models-slm-vs-llm.html (accessed on 10 May 2024).
43. Merritt, R. What Is Retrieval-Augmented Generation Aka RAG? NVIDIA Blog. 2024. Available online: <https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/> (accessed on 17 November 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.