applied sciences

MDPI

Article

GeoLocator: A Location-Integrated Large Multimodal Model (LMM) for Inferring Geo-Privacy

Yifan Yang [1,*], Siqin Wang [1,*], Daoyang Li [1], Shuju Sun [1] and Qingyang Wu [2]

[1] Spatial Sciences Institute, University of Southern California, Los Angeles, CA 90089, USA; daoyangl@usc.edu (D.L.); shujusun@usc.edu (S.S.)
[2] Department of Environmental Health Sciences, University of California—Los Angeles, Los Angeles, CA 90095, USA; qw@ucla.edu
* Correspondence: yyang295@usc.edu (Y.Y.); siqinwan@usc.edu (S.W.)

Abstract: To ensure the sustainable development of artificial intelligence (AI) application in urban and geospatial science, it is important to protect the geographic privacy, or geo-privacy, which refers to an individual's geographic location details. As a crucial aspect of personal security, geo-privacy plays a key role not only in individual protection but also in maintaining ethical standards in geoscientific practices. Despite its importance, geo-privacy is often not sufficiently addressed in daily activities. With the increasing use of large multimodal models (LMMs) such as GPT-4 for open-source intelligence (OSINT), the risks related to geo-privacy breaches have significantly escalated. This study introduces a novel GPT-4-based model, GeoLocator, integrated with location capabilities, and conducts four experiments to evaluate its ability to accurately infer location information from images and social media content. The results demonstrate that GeoLocator can generate specific geographic details with high precision, thereby increasing the potential for inadvertent exposure of sensitive geospatial information. This highlights the dual challenges posed by online data-sharing and information-gathering technologies in the context of geo-privacy. We conclude with a discussion on the broader impacts of GeoLocator and our findings on individuals and communities, emphasizing the urgent need for increased awareness and protective measures against geo-privacy breaches in the era of advancing AI and widespread social media usage. This contribution thus advocates for sustainable and responsible geoscientific practices.

Keywords: geo-privacy; GPT-4; image comprehension; large multimodal model (LMM); open-source intelligence (OSINT)

Citation: Yang, Y.; Wang, S.; Li, D.; Sun, S.; Wu, Q. GeoLocator: A Location-Integrated Large Multimodal Model (LMM) for Inferring Geo-Privacy. *Appl. Sci.* **2024**, *14*, 7091. https://doi.org/10.3390/app14167091

Academic Editor: Andrea Prati

Received: 5 July 2024
Revised: 3 August 2024
Accepted: 5 August 2024
Published: 13 August 2024

1. Introduction

The advent of the internet, big data, and generative AI has significantly transformed the field of geosciences, offering unprecedented opportunities for collecting, analyzing, disseminating, and generating geographic information [1,2]. In today's digital era, the risk of personal information being quietly leaked is a growing concern, with geographic privacy becoming a focal point. The integrity of geographic data is not only a matter of personal privacy but also a cornerstone of sustainable geoscience practices. Ensuring the privacy of geographic information prevents misuse and distortion, which is crucial for maintaining the trust and accuracy necessary for sustainable geoscience endeavors. Geographic privacy involves the protection and confidentiality of geographic information related to individuals. It primarily includes safeguarding data that disclose personal geographic locations, such as real-time whereabouts, historical movement patterns, or any location-specific information traceable to an individual [3]. Geographic privacy is critical; however, in an era where smartphones and social media platforms are ubiquitous, maintaining this privacy is a significant challenge. While services like navigation, travel booking websites, and social media offer convenience, they also pose potential risks to our geographic privacy through

potential surveillance, unauthorized data mining, and third-party misuse [4]. The existing legal frameworks struggle to keep pace with the rapidly evolving technologies that threaten geographic privacy, further exacerbating these concerns [5].

In the current era of rapid advancements in generative AI, we have identified a commonly overlooked method of geographic privacy leakage: our photos often contain substantial geographic information. In the past, the information embedded in these photos might not have been easily captured [3]. However, today, consider a simple case on social media: you post a photo of yourself visiting a baseball stadium. From this photo, one can infer your geographic location, specifically which stadium it is and its exact address. A single photo can compromise your geographic privacy. With the rapid development of large multimodal models like GPT-4, these models can extract, interpret, and infer geographic information from your posted images. GPT-4's ability to infer details from photos poses a significant threat to geographic privacy [6]. These models can directly reveal precise location details from geotagged images or indirectly through contextual analysis. The potential harm and implications are profound and multifaceted, including identity theft, personal security breaches, and serious intrusions into personal privacy.

Recognizing the potential threat posed by GPT-4 to geographic privacy, we developed GeoLocator, a tool that integrates GPT-4 with geolocation capabilities to demonstrate its ability to infer location information from input images and social media content. To evaluate and compare the privacy attack capabilities of conventional search engines, GPT-4, and GeoLocator, we designed a series of experiments based on inputs from various datasets, including Google Maps images, day/night images, and social media posts. Our experiments demonstrate that GeoLocator can generate specific geographic details with high accuracy, thereby embedding the risk of users unintentionally exposing geospatial information to the public. This underscores the themes of online data sharing, information collection technologies, and geographic privacy jurisprudence. Our findings emphasize the urgency for enhanced awareness and protective measures against geographic privacy leakage in the era of advanced AI and widespread social media usage. By combining these insights with robust privacy protection practices, we aim to ensure that the benefits of geoscience advancements do not come at the expense of individual rights or societal trust. These conclusions highlight the double-edged nature of modern geoscience tools: while they can enhance our understanding and management of geographic resources, if mismanaged, they also pose significant privacy risks.

2. Related Work

We commence with providing a comprehensive overview of the capabilities of large multimodal models (LMMs) in inferring geographic details, and the key milestones and innovative techniques that have shaped the evolution of LMMs.

2.1. LMMs Introduction

The transformative emergence of the transformer architecture [7] set a new precedent in the field, laying a robust foundation for contemporary large language models. This breakthrough was followed by the development of pivotal models in natural language processing, notably GPT [8] and bidirectional encoder representations from transformers [9]. More recently, with the development of computing power and advanced training techniques such as instruction tuning and reinforcement learning from human feedback [10,11], LLMs such as ChatGPT [12] can achieve superior results in various downstream applications without the need for task-specific tuning. For example, LLMs excel in abstract summarization, producing meaningful overviews of text passages. This capability can be particularly beneficial in fields with vast amounts of text, like legal practice, academic research, and medicine, aiding in the efficient navigation of dense information repositories [13,14]. Furthermore, the ability of LLMs to understand context and user intent has led to applications in customer service, personal assistance, and interactive educational tools [15].

Concurrently, there is a notable trend towards integrating LLMs with vision-based models, heralding a new era of large multimodal models. This integration expands the range of tasks they can perform and aligns more closely with the multimodal nature of human cognition. LMMs differ from LLMs by processing and interpreting both textual and other types of data such as images. This advancement led to groundbreaking advancements in visual understanding and reasoning. For instance, the proprietary GPT-4 model [12], renowned for its illustrative abilities, and open-source models like Large Language and Vision Assistant [16], have demonstrated exceptional skill in blending textual and visual information. These models have shown proficiency in tasks ranging from generating website code from visual prompts [17] to recognizing complex details in image-rich contexts [18]. Their success illustrates not only the versatility of LMMs in handling multimodal data but also their potential in transforming tasks that require an intricate understanding of both visual and textual elements. In the next section, we continue to explore existing work integrating large multimodal models with a variety of tasks and examine the application of artificial intelligence in geography.

2.2. LMMs Applications

LMMs have a very wide range of application capabilities. LMMs are the most cutting-edge technology and have been widely employed in diverse domains. In the medical field, Hou et al. found that the current multimodal models, GPT-4 and Bard, could handle visual assignments. For example, GPT-4 successfully solved 96.7% of the visual problems, facing minimal difficulty only with Parsons problems [15]. Parsons problems require precise logical ordering of code snippets, which requires a deep understanding of programming languages. Solving Parsons problems also requires a detailed understanding of the context and a prediction of the results of code execution. GPT-4 can generate and understand natural language, but for this task, which requires rigorous logical reasoning and programming knowledge, this is beyond the main design scope of GPT-4. Yuan et al. found that LMMs can be applied to enhance various aspects of healthcare. Particularly, they highlighted the crucial role of LMMs, investigating their ability to process diverse data types like medical imaging and electronic health records to augment diagnostic accuracy [14]. Fabian et al. proposed a novel zero-shot species classification framework that leverages multimodal foundation models. This framework involves instruction-tuning vision-language models to generate detailed visual descriptions of camera-trap images, using terminology similar to that of experts [19]. Picard et al. evaluated GPT-4V, a vision language model, in engineering design tasks, demonstrating its capabilities and limitations. Their study provides foundational insights for the application of vision language models in engineering [20]. Warner et al. explored the shift in medical AI systems towards deep learning models, focusing on LMM's impact on medical image analysis and clinical decision support systems [21]. Oh et al. introduced an LMM for radiation therapy, integrating clinical text with images, demonstrating enhanced performance in breast cancer treatment, a first in such clinical text integration for oncology [22]. Microsoft delved into the capabilities of GPT-4 Vision, highlighting its proficiency in video understanding, visual reasoning, and other areas. They underscored the substantial potential applications of this technology in various sectors, including industry, medical fields, auto insurance, and image generation. In summary, LMMs have showcased their strong and diverse application capabilities in solving visual problems in the aforementioned domains. Following these existing studies, we take LMMs as the baseline model to further develop our location-integrated model, GeoLocator.

2.3. Geography-Related LLMs

The use of LLMs in the spatial science domain has been relatively limited until quite recently. Earlier this year, Roberts et al. explored the geographical knowledge and reasoning skills of GPT-4 through a series of experiments, ranging from basic tasks like location estimation to complex applications like route planning and itinerary creation. Their study highlights GPT-4's capability in geospatial reasoning and its potential for diverse appli-

cations in geography-related fields [23]. Then Deng et al. developed K2, a specialized language model for geoscience, trained on a tailored corpus, showing enhanced performance in geoscience-specific tasks like question answering and knowledge reasoning, setting a new standard for domain-specific language models [24]. Li et al. introduced GeoLM, a language model integrating geospatial data with linguistic information, using geo-entity anchors and spatial coordinate embeddings for enhanced geo-entity understanding [25]. Hu et al. developed a method that combines geo-knowledge with GPT models for improved extraction of location descriptions from social media messages during disasters. This approach, using only 22 training examples, achieved over 40% improvement in accuracy compared to standard named-entity recognition methods, significantly aiding in the rapid and efficient response to disaster scenarios [26]. Recently, Bhandari et al. assessed the geospatial knowledge and reasoning capabilities of LLMs, using experiments on geocoordinate prediction, geospatial preposition analysis, and multidimensional scaling, revealing their potential in geospatial reasoning tasks [27]. Extending from the existing research, we reformulated the regular LMMs to create our location-integrated model, GeoLocator, and tested out its capacity to infer geospatial information and geo-privacy.

3. Workflow to Develop a New Tool—GeoLocator

Based on the GPT-4, we have developed a tool capable of inferring location information from images, which we named GeoLocator (https://chat.openai.com/g/g-qxqvMb6YJ-geolocator, accessed on 10 August 2024). GPT-4 is a large multimodal model acceptable for the input of images and texts and emitting text outputs; although it is not capable as a human being in many real-world scenarios, GPT-4 has demonstrated human-level performance on a variety of professional and academic benchmarks. GeoLocator we developed is a customized version of ChatGPT (https://chat.openai.com/gpts/editor, accessed on 10 August 2024). GeoLocator's strength is to use the powerful feature extraction and linguistic inference capabilities of large multimodal models to infer location information from images. At the same time, we developed GeoLocator with a large number of model commands built in to avoid transferring lengthy contexts each time. As shown in Figure 1, forming the most important core of GeoLocator are the instructions for model input and features for model output.

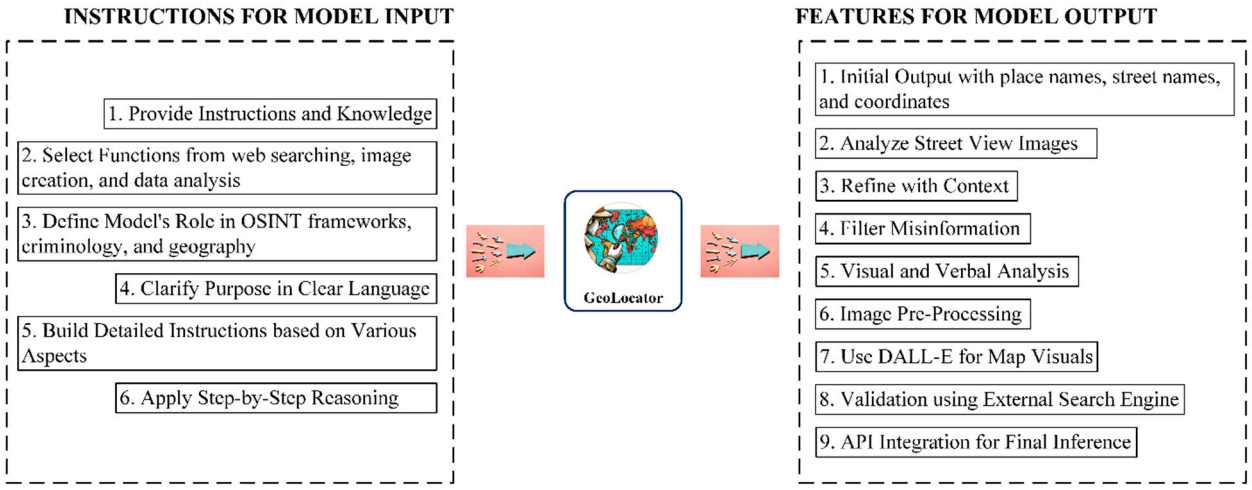

Figure 1. GeoLocator instructions and features.

GeoLocator was created based on ChatGPT with customized functionalities. Creating GeoLocator is easy following the below procedure: starting a conversation, giving it instructions and additional knowledge, and then choosing what it can do, such as searching the web, making images, or analyzing data. Instructions are the key to GeoLocator. The process of creating instructions involves engineering skills, such as defining the model's

roles as being proficient with OSINT frameworks, criminology, and geography. The purpose of the model is described in clear language, and its task is to extract every detail from the photographs and come up with a sound analysis. In building the instructions for GeoLocator, we emphasized the need to focus on image details, exchangeable image file format (EXIF) data, traffic rules, human and physical geography, and unique regional clues. Step-by-step reasoning was used to improve accuracy. GeoLocator was built on GPT-4, which has the inherent advantage that GPT-4 can infer geographic information on its own, although the extent to which it can do so remains unknown. Instructions are key to the success of building GeoLocator. After many attempts and enriched with instructions, GeoLocator as developed has the potential to significantly improve its ability to infer geographic information.

When built with enriched instructions, our GeoLocator offers a variety of features. It provides detailed place names, street names, and coordinates in the final location result. It performs an initial analysis using only street-view images and refines its inferences using contextual information provided by the user. Users may provide misleading information that needs to be recognized and effectively differentiated. GeoLocator expresses itself visually and verbally, for example by recognizing and highlighting road signs and landmarks, and then performs a reasonable verbal analysis. It uses a code interpreter to preprocess images to deal with issues of low resolution, including noise reduction, resolution enhancement, zooming, or cropping. GeoLocator can utilize DALL-E image generation to obtain rich visual information, such as mapping spatial relationships (geographic manuscript style), topographic maps, or street maps. Finally, we merge the external search engine validation step into GeoLocator to confirm its conjectures. Based on GeoLocator's conclusions, we run external APIs, such as the Google Earth API, which provides features like detailed map drawing, geocoding, and street view to provide feedback on street location information.

First, GeoLocator employs step-by-step reasoning throughout its process. It provides a detailed explanation of its reasoning, outlining how it arrives at specific conclusions, including identifying key visual cues, analyzing EXIF data, and cross-referencing contextual information. Second, GeoLocator delivers clear and explicit intermediate outputs. During the inference process, it produces intermediate results, such as identified landmarks, road signs, and coordinates, allowing users to track the progress of its analysis. Additionally, GeoLocator uses visual annotations to highlight specific elements in images used for its reasoning. For example, it may circle landmarks or annotate road signs, clearly indicating the factors influencing its conclusions. External validation is another critical feature, as GeoLocator integrates steps to cross-reference its findings with search engines and external APIs, such as the Google Earth API, helping to verify the accuracy of its inferences and adding a layer of reliability. Moreover, GeoLocator includes a user feedback loop, enabling users to provide feedback on its inferences, which the model can use to improve its accuracy over time. By incorporating the transparency and explainable features, we aim to demystify GeoLocator's operations and ensure that users have a clear understanding of its reasoning process, thereby addressing concerns about it functioning as a black-box algorithm.

4. Experimental Design to Test GeoLocator

We designated a series of experiments to test the capacity of GeoLocator in inferring location information and evaluating its modelling performance. We compared the results of geospatial/location information identified by three tools and/or platforms: Google search engine, GPT-4, and GeoLocator, based on images and texts. Such a comparison was also implemented in different languages. We observed that GeoLocator has the strongest location inference ability across the three tools mentioned above and can even deduce exact street addresses.

Given the size of our study, for all experiments, we judged the performance of the model solely based on whether it can correctly infer the location of the information given.

To give the reader a better understanding of GeoLocator's analysis process for the following experiments, we designed the flowchart shown in Figure 2.

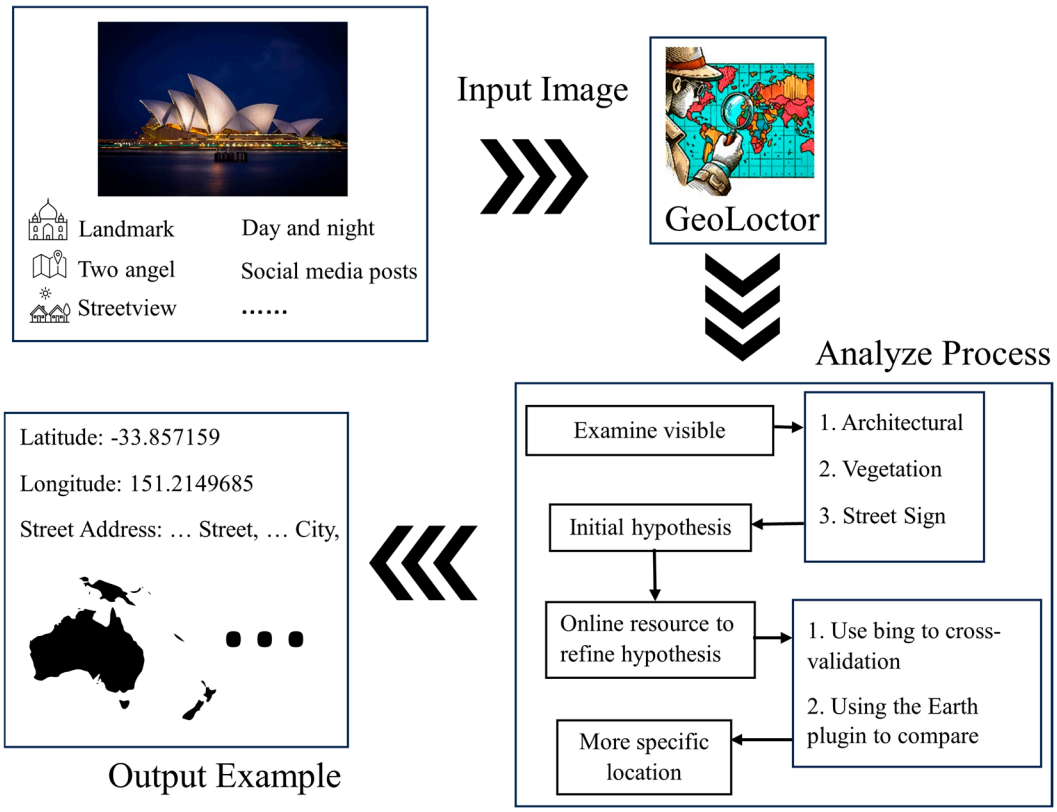

Figure 2. GeoLocator working flowchart.

Step 1: Prepare the input of data

To evaluate the effectiveness of the Google search engine, GPT-4, and GeoLocator across various image types, a diverse set of data sources was gathered. This included images from Google Maps, photographs taken by our research team, Google Images, and posts from social media. Google Maps served as a resource for geographically diverse and detailed images, offering both reliability and recency. Our dataset (Table 1) comprised 100 locations from Google Maps, including 50 iconic landmarks (e.g., the Statue of Liberty in New York) and 50 street views without obvious landmarks. Moreover, with the help of Google Maps, it was able to access different angles of street views. Forty images of 20 different locations from two different angles were selected. Additionally, we captured 40 images of 20 specific locations at separate times (day and night) to assess changes in environmental conditions. Ten images from Google Images were also chosen to assess the impact of language input on the results. Three social media posts were selected, combined with text and images from the authors' personal accounts. The inclusion of social media images aimed to mimic real-life scenarios. While the dataset may appear small, it has been carefully curated to provide diverse and representative scenarios for the initial study. The variety in image sources, angles, lighting conditions, and contexts allowed us to thoroughly evaluate the capabilities and limitations of each tool under different circumstances. Furthermore, the core of the research focused on the enhanced capabilities of the large multimodal model (GeoLocator) to infer geographic details more effectively.

Table 1. Data source and description.

Data Source	Description	Number of Images
Google Maps	Iconic landmarks	50 images
	Street view without obvious landmarks	50 images
	Images of 20 locations from two different angels	20 sets (40 images)
Taken by a research team	Images of 20 locations at two time slots (i.e., day and nighttime)	20 sets (40 images)
Google Images	Images of 10 locations from China to assess the impact of language input	10 images
Posts from social media	Social media posts sent by research team members	3 posts

Step 2: Compare images' location inference ability among Google search engine, GPT-4, and GeoLocator

This section of the study aimed to evaluate the location inference capabilities of Google search engine, GPT-4, and GeoLocator by sending photos without text prompts, thereby determining whether advanced artificial intelligence tools have better inference performance. This experiment highlights the potential of artificial intelligence tools like GPT-4 and GeoLocator in assisting with geographic location inference. In the experiment, identical images—ranging from iconic landmarks to street views and day/night images of the same place—were uploaded to the Google search engine, GPT-4, and GeoLocator for evaluating the performance of the tools above based on their inference precision.

Step 3: Compare the location inference based on image and text instruction

Although GeoLocator already has reasonable accuracy on location inference, this experiment explored whether it could perform better if additional textual prompts or additional images were provided. To start the evaluation, 40 images taken from 20 different locations with different angles were selected, along with 10 images from Step 3 that had comparatively lower accuracy (accuracy of inference at the country level). GeoLocator's prediction results were then compared before and after applying these additional images or textual instructions.

Step 4: Examine the impact of languages based on inference results

Through further experiments, it was found that GeoLocator's performance varies with different input languages. Therefore, this experiment was designed to evaluate the impact of different languages on the model's inference results. During the experiment, images containing text prompts in different languages were provided to GeoLocator, and its predictions for the geographic location of the images were observed. Finally, we checked whether the model's predictions varied with the change of input language.

Step 5: Evaluate the performance of GeoLocator based on social media posts

The final experiment was focused on a more realistic and complex scenario: social media posts. Unlike structured datasets, social media content often includes a mix of meaningful and seemingly unrelated information, such as emotional expressions, user activities, and various multimedia elements. This complexity makes location inference particularly challenging and provides the best case for simulating attacks on geographic privacy. To test GeoLocator's performance, a set of social media posts containing both text and images was curated. These posts were selected to represent a wide range of scenarios, from everyday updates to more specific location-centric content. The goal was to determine whether GeoLocator, with its finely tuned large multimodal models (LMMs), could accurately infer the locations described or depicted in these posts. Applying GeoLocator to these realistic and varied social media scenarios, we aimed to simulate

real-world conditions where users might unintentionally share geographic information. The results of this experiment are crucial for understanding how effectively GeoLocator can manage the complexities of social media data while ensuring user privacy and providing accurate location inferences. This step highlights GeoLocator's advanced capabilities in navigating the intricate landscape of social media content, showcasing its potential for practical applications in protecting geographic privacy.

5. Results

With the help of GeoLocator, we observed a reasonably high accuracy of location inference in most kinds of images mentioned in the experiment. With the advancement of technology, especially the progress in image inference and geolocation techniques, it is becoming increasingly possible to infer where a photo was taken.

5.1. Compare Images' Location Inference Ability among Google Search Engine, GPT-4, and GeoLocator

This experiment highlights the potential of artificial intelligence tools like GPT-4 and GeoLocator. They not only improve the speed and accuracy of inference but also expand the capabilities of users in conducting such tasks. We calculated the image inferring accuracy by dividing the number of images successfully inferred in the experiment by the total number of images used in that kind of experiment. The outcomes of our experiments are displayed in Table 2. We selected 10 representative images to show in Table 3 in detail. To simplify our results, we grouped the results of our experiments into four geographical categories (by country, state, city/town, and street). These were color-coded for inference accuracy, ranging from dark green for the most precise to light green for the least. We consider inferences accurate to the street level as successful, and this criterion is used to evaluate the accuracy of the three tools' inferences.

Table 2. Results of image inferring accuracy.

Image Type	Sample Size	Google Search Engine	GPT-4	GeoLocator
Iconic landmark	50	88%	60%	94%
Street view	50	16%	18%	54%
Daytime image	20	25%	40%	70%
Nighttime image	20	10%	15%	35%

Because the location inference of Google search engine is based on tags of images, it works well for deducing from pictures of tourist attractions or iconic buildings with rich information, making it easy for search engines to infer these locations. Users only need to upload the photo to the search engine for comparison or inquiry to get the address results. However, it is difficult to infer the location of street views because there is little information about various street views on the internet, making it hard to establish web links and thus infer street views.

For street-view image inference, when a user inputs a picture and asks for the possible address, GPT-4 can utilize its image inference and analysis model to predict the shooting location. It can analyze various factors such as landmarks, natural features, architectural styles, vegetation types, and weather conditions in the photos to complete the inference and prediction of image locations. This enables GPT-4 to provide relatively accurate geographic location predictions based on the photo's content, even in the absence of tag information.

Outperforming GPT-4, GeoLocator utilizes enhanced instructions to infer specific location information from street view images and performs a more comprehensive inference process. By integrating additional context and analysis steps, GeoLocator achieves superior performance. It can not only predict the country or city of the picture but even infer the name of the street or building where the image was taken, even in the absence of specific road signs or street names.

Table 3. Results of comparing images' location inference ability among Google search engine, GPT-4, and GeoLocator.

Images	Data Type	Google Search Engine	GPT-4	GeoLocator	Distance * (miles)
	Street View	State	City/Town	Street	0.0034
	Street View	City/Town	Unknown **	City/Town	10
	Street View	State	Unknown **	Country	32.49
	Street View	Country	Unknown **	Country	126.42
	Landmark	Street	Unknown **	Street	0.0044
	Landmark	City/Town	Street	Street	0.0019
	Landmark	Unknown **	Country	City/Town	55.22
	Landmark	Street	Street	Street	0.0449
	Street View	City/Town	Unknown **	City/Town	2.62

Table 3. *Cont.*

Images	Data Type	Google Search Engine	GPT-4	GeoLocator	Distance * (miles)
	Nighttime Image	Unknown **	Country	City/Town	3.28

* Distance means the distance difference between the real location and the inferred location. ** This means that the Google Search Engine, GPT-4, or GeoLocator is not able to infer the address. Geographical categories: country, state, city/town, and street. A deeper background color indicates that a more specific address is inferred in terms of the geographical categories mentioned above.

5.2. Compare the Location Inference Based on Images and Text Instructions

We evaluated whether GeoLocator could improve its performance in inferring the location of street-view images by providing additional image perspectives or textual prompts when the inference of location is not at street level. In both experiments, whether an additional image was added or there was an additional textual prompt, the model's inference accuracy increased. To assess GeoLocator's proficiency in inferring locations from images captured at various angles of the same site, we conducted a series of 10 experiments in which each scenario could potentially deduce a street-level location, using Taipei 101 as a representative example. It did not give us an expected answer at first, and we gave the model another image taken from the same place but at a different angle. Then GeoLocator did successfully infer the place where the photo was taken. With additional information, GeoLocator is able to infer the location more precisely.

To evaluate the accuracy of location information identified by inputting additional text prompts, we tested GeoLocator's performance. We chose to highlight the USC University Park Campus case as an example, where each scenario could potentially deduce a street-level location as well. After we uploaded an image representing the campus, GeoLocator first observed the environment in the picture and inferred that it was in a temperate Mediterranean climate area. The people in the picture were dressed lightly and casually, and the activities looked like they were happening on a campus. Based on these clues, GeoLocator concluded that the picture was shot at a university in the United States. Having received this unsatisfactory answer, we gave the model three prompts until it gave us a street-level answer. In this situation, with richer information, GeoLocator showed better performance than relying on image input alone. This multimodal approach enabled GeoLocator to infer and understand the content of images, including specific details like road signs and street names, more effectively. The advantage of this multimodal analysis is that it is not just a simple overpay of different modalities of information, but rather a deep understanding and analysis of the interrelationships between these diverse types of data, leading to more comprehensive and precise conclusions. In this situation, with more information, GeoLocator performed better than when only given input of images, being able to infer more images with specific details like road signs and street names.

5.3. Examine the Impact of Languages on Inference Results

In this experiment, we investigated how different inputs affect GeoLocator's reasoning procedure and subsequent inference results. We provided GeoLocator with images containing text prompts in various languages and observed how these inputs influenced its geographical location references. Specifically, we tested inputs in Chinese and English to determine whether there were significant differences in the model's inference results. Our findings indicated that while there were no substantial differences in the overall inference results between the two languages, the model's reasoning processes differed. For instance, when the input text was in English, the model focused on natural geography, flora, architecture, and infrastructure in its analysis. In contrast, when the input text was in

Chinese, the model emphasized country-specific, regional, and zoning clues, concentrating on more specific details. Moreover, when queried in English, the model identified the image as being from a park, a classification that did not occur with queries in Chinese. This variation may be attributed to several factors. First, the model's training data could contain language-specific biases, leading it to associate images with geographical locations more commonly referenced by speakers of a particular language. Second, different languages may convey distinct cultural and geographic context clues, which the model may use to inform its location predictions.

5.4. Evaluate the GeoLocator's Performance on Social Media Posts

The final experiment assesses GeoLocator's capability to infer the location of a picture taken in a more complex real-world scenario. Social media posts often feature photos of the same location from various angles, accompanied by text describing the time or place, as well as some extraneous information. This complexity necessitates a higher level of synthesis from GeoLocator to accurately determine the location. The experiment comprises three sub-experiments: two involving tourist posts and one featuring a daily-life post. The results indicate that GeoLocator can infer locations down to the city level and provide detailed profiles of individuals in the posts. It accurately pinpointed the exact street or area in two out of the three posts. Furthermore, the distances between GeoLocator's estimations and the actual locations in the images were all within 100 miles.

6. Discussion and Conclusions

Our development of GeoLocator, based on over 200 experimental tests across diverse data sources including Google Maps, author-captured images, Google Images, and social media posts, has led to four key discoveries. Firstly, GeoLocator demonstrated exceptional performance in inferring specific locations, particularly street views, surpassing both Google search engine and GPT-4. Secondly, we observed an enhancement in GeoLocator's performance when it was provided with additional images from different angles or textual prompts. Thirdly, we identified that the language of input text influences GeoLocator's reasoning process. Despite overall similar levels of accuracy, different languages led to variations in the model's focus and thought processes. Lastly, GeoLocator exhibited robust capabilities in inferring locations from social media posts, even amidst complex and diverse data such as varying orientations and accompanying social media text. These findings emphasize the advancements in AI-driven geolocation tools and their potential in various applications, from tourism to protecting geo-privacy, particularly highlighting GeoLocator's exceptional ability to infer locations from complex multimodal data sources.

Our development of GeoLocator contributes to and extends the existing research on LMMs and their applications, especially in geography-related tasks in multiple aspects. Our GeoLocator represents an innovative integration of geospatial data with LMMs. While existing research, like Li et al.'s work on GeoLM [19], has begun exploring the integration of linguistic information with geospatial data, GeoLocator takes this integration further. It not only processes textual and visual data but also effectively infers detailed geospatial information such as street-level locations, enhancing the granularity of location inference. Furthermore, our GeoLocator extends the application of LMMs in geospatial contexts. Previous research demonstrated LMMs' capabilities in general tasks like geoscience and medical imaging. However, GeoLocator specifically targets the nuanced task of geo-privacy and location inference, demonstrating effectiveness in complex real-world scenarios such as social media analysis involving multiple data types and orientations. Our findings align with and diverge from existing literature, similar to studies by Roberts et al. and Bhandari et al., showing robust geospatial reasoning capabilities in LMMs [17,21]. However, the enhanced performance of GeoLocator in street-level inference and its effectiveness in multimodal analysis (including textual, visual, and geographic data) marks a significant advancement beyond the general capabilities discussed in the current literature. Our study provides new insights into the impact of language on LMMs' inference abilities. While

existing studies have primarily focused on the model's performance in specific domains or tasks, our findings highlight the nuanced ways in which input language can affect a model's reasoning process and outcome, adding a new dimension to the understanding of LMMs' capabilities.

The emergence and application of GeoLocator in the field of geospatial data analysis have significant policy implications, especially regarding privacy protection and governmental usage. Governments could leverage GeoLocator for more effective enforcement of privacy regulations. By comprehending the capabilities and limitations of such advanced geospatial reasoning tools, policymakers can develop more informed guidelines and regulations to prevent and protect user geo-privacy. In urban planning and development, GeoLocator could assess the potential privacy impacts of new projects, such as how new constructions or urban developments might affect an area's geospatial data footprint and its implications for resident privacy. Moreover, GeoLocator holds substantial promise for public safety and natural disaster management. During natural disasters or public safety emergencies, GeoLocator could assist governments in rapidly and accurately assessing on-the-ground conditions. For instance, it could detect real-time locations of natural disasters, such as fires, through social media analysis. This capability to swiftly gather and analyze geospatial information can significantly improve response times and resource allocation, contributing to the development of more sustainable and resilient communities.

Our study is subject to certain limitations in the data and methodology used for developing and implementing GeoLocator. One notable limitation is GPT-4's handling of lengthy instructions. When provided with substantial textual input, GPT-4 often struggled with effectively processing the middle sections of the text. This issue raised concerns about the model's ability to consistently interpret and analyze long-form instructions or data inputs. Additionally, the variability in GPT-4's outputs highlights the need for methods to stabilize its responses. This stabilization is crucial for achieving consistent and reproducible results, particularly in applications where decision-making heavily relies on the model's output. Another concern is GPT-4's permeability, specifically its tendency to inadvertently disclose construction information and custom prompts. This poses a risk of unintentional exposure of sensitive or proprietary data. Moreover, the model's potential to provide download links for private knowledge bases is a significant security concern. Ensuring the confidentiality and integrity of data processed by GPT-4 is paramount, especially when dealing with sensitive geospatial information.

To conclude, the development of GeoLocator represents a significant advancement in integrating language models with geospatial data, highlighting the potential risks of geo-privacy infringement through image and text analysis. This research demonstrates the potential of AI-driven tools to enhance our understanding and interaction with geographic data, while simultaneously emphasizing the need for careful consideration of privacy and ethical implications. GeoLocator's capability to analyze and infer precise locations from complex multimodal data sources marks a breakthrough in the field, showcasing the transformative power of AI in reshaping geospatial analysis. However, this innovation also carries the responsibility of ensuring the ethical and secure use of such technologies. Our study serves not merely as a pilot exploration but also as a cautionary tale, setting a precedent for future research focused on balancing technological advancement and the preservation of geo-privacy. As we continue to push the boundaries of what AI can achieve, it is crucial to remain vigilant about the implications of these advancements, ensuring that they are used to improve society while safeguarding individual privacy and security. Insights into how personal geographic privacy can be inferred through step-by-step analysis using GeoLocator can inspire new methods for privacy protection by understanding and mitigating the reasons behind such inferences. In summary, the development of GeoLocator underscores its immense potential in enhancing the sustainability of geosciences through advanced AI-driven geospatial analysis. By carefully managing the ethical and privacy concerns associated with these technologies, we can harness their benefits to foster a more informed, secure, and sustainable future.

Author Contributions: Y.Y.: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, visualization, writing—original draft. S.W.: conceptualization, validation, investigation, resources, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition. D.L.: data curation, writing—original draft, formal analysis. S.S.: data curation, visualization, formal analysis. Q.W.: writing—original draft, formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Acknowledgments: We would like to express our gratitude to Yixian Zhang for the significant contributions made to the development of GPTs, as well as for the valuable discussions and assistance in the initial ideation and experimental design of this manuscript.

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

References

1. Zhu, T.; Ye, D.; Wang, W.; Zhou, W.; Philip, S.Y. More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 2824–2843. [CrossRef]
2. Janowicz, K.; Gao, S.; McKenzie, G.; Hu, Y.; Bhaduri, B. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 625–636. [CrossRef]
3. Jiang, H.; Li, J.; Zhao, P.; Zeng, F.; Xiao, Z.; Iyengar, A. Location privacy-preserving mechanisms in location-based services: A comprehensive survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]
4. Di Minin, E.; Fink, C.; Hausmann, A.; Kremer, J.; Kulkarni, R. How to address data privacy concerns when using social media data in conservation science. *Conserv. Biol.* **2021**, *35*, 437–446. [CrossRef]
5. Nair, M.M.; Tyagi, A.K. Privacy: History, statistics, policy, laws, preservation and threat analysis. *J. Inf. Assur. Secur.* **2021**, *16*, 24–34.
6. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
7. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
8. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving language understanding by generative pre-training. 2018; *in progress*.
9. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
10. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744.
11. Wang, Y.; Mishra, S.; Alipoormolabashi, P.; Kordi, Y.; Mirzaei, A.; Arunkumar, A.; Ashok, A.; Dhanasekaran, A.S.; Naik, A.; Stap, D.; et al. Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks. *arXiv* **2022**, arXiv:2204.07705.
12. OpenAi. ChatGPT. 2023. Available online: https://openai.com/chatgpt (accessed on 19 December 2023).
13. Holmes, J.; Ye, S.; Li, Y.; Wu, S.-N.; Liu, Z.; Wu, Z.; Zhao, H.; Jiang, X.; Liu, W.; Wei, H.; et al. Evaluating Large Language Models in Ophthalmology. *arXiv* **2023**, arXiv:2311.04933.
14. Yuan, M.; Bao, P.; Yuan, J.; Shen, Y.; Chen, Z.; Xie, Y.; Zhao, J.; Chen, Y.; Zhang, L.; Shen, L.; et al. Large Language Models Illuminate a Progressive Pathway to Artificial Healthcare Assistant: A Review. *arXiv* **2023**, arXiv:2311.01918. [CrossRef]
15. Hou, I.; Man, O.; Mettille, S.; Gutierrez, S.; Angelikas, K.; MacNeil, S. More Robots are Coming: Large Multimodal Models (ChatGPT) can Solve Visually Diverse Images of Parsons Problems. *arXiv* **2023**, arXiv:2311.04926.
16. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. *arXiv* **2023**, arXiv:2304.08485.
17. Zhu, D.; Chen, J.; Shen, X.; Li, X.; Elhoseiny, M. MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models. *arXiv* **2023**, arXiv:2304.10592.
18. Zhang, Y.; Zhang, R.; Gu, J.; Zhou, Y.; Lipka, N.; Yang, D.; Sun, T. LLaVAR: Enhanced Visual Instruction Tuning for Text-Rich Image Understanding. *arXiv* **2023**, arXiv:2306.17107.
19. Fabian, Z.; Miao, Z.; Li, C.; Zhang, Y.; Liu, Z.; Hernández, A.; Montes-Rojas, A.; Escucha, R.; Siabatto, L.; Link, A.; et al. Multimodal Foundation Models for Zero-shot Animal Species Recognition in Camera Trap Images. *arXiv* **2023**, arXiv:2311.01064.
20. Picard, C.; Edwards, K.M.; Doris, A.C.; Man, B.; Giannone, G.; Alam, M.F.; Ahmed, F. From Concept to Manufacturing: Evaluating Vision-Language Models for Engineering Design. *arXiv* **2023**, arXiv:2311.12668.

21. Oh, Y.; Park, S.; Byun, H.K.; Kim, J.S.; Ye, J.C. LLM-driven Multimodal Target Volume Contouring in Radiation Oncology. *arXiv* **2023**, arXiv:2311.01908.
22. Yang, Z.; Li, L.; Lin, K.; Wang, J.; Lin, C.-C.; Liu, Z.; Wang, L. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). *arXiv* **2023**, arXiv:2309.17421.
23. Roberts, J.; Lüddecke, T.; Das, S.; Han, K.; Albanie, S. GPT4GEO: How a Language Model Sees the World's Geography. *arXiv* **2023**, arXiv:2306.00020.
24. Deng, C.; Zhang, T.; He, Z.; Xu, Y.; Chen, Q.; Shi, Y.; Xu, Y.; Fu, L.; Zhang, W.; Wang, X.; et al. K2: A Foundation Language Model for Geoscience Knowledge Understanding and Utilization. *arXiv* **2023**, arXiv:2306.05064.
25. Li, Z.; Zhou, W.; Chiang, Y.-Y.; Chen, M. GeoLM: Empowering Language Models for Geospatially Grounded Language Understanding. *arXiv* **2023**, arXiv:2310.14478.
26. Hu, Y.; Mai, G.; Cundy, C.; Choi, K.; Lao, N.; Liu, W.; Lakhanpal, G.; Zhou, R.Z.; Joseph, K. Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *Int. J. Geogr. Inf. Sci.* **2023**, *37*, 2289–2318. [CrossRef]
27. Bhandari, P.; Anastasopoulos, A.; Pfoser, D. Are Large Language Models Geospatially Knowledgeable? *arXiv* **2023**, arXiv:2310.13002.