The inference of geographic location in images through GPT-4 and customized GPTs.

realize the ability limits, shortcomings and expansion space of large models in the process of location recognition or inference.

The GPT large model can infer geographic location through text, pictures, audio and video (multimedia) (in the form of social media). The large model has a strong distinction from the traditional recognition of geographic location, and you can tell you the steps of GPT recognition of geographic location through its own logic. However, this process brings two problems, the credibility of the large model and the violation of the user's geographic privacy by the large model. This also made me realize that the following work can be carried out, whether the large model can be used to complete the user profile with geographic attributes for the user's social media content. For the reasoning logic of the large model (which I think is stronger than the understanding of ordinary people), summarize the key steps, build a model by yourself to reproduce and compare with the large model. The purpose of building the model by yourself is to avoid black boxes as much as possible. And discuss the relationship between the large model and geographic privacy and location privacy, and propose prevention strategies.

Research Proposal: Research on Geographic Attribute Inference and Privacy Protection Strategy of Social Media Content Using GPT

Background and Motivation:

GPT has been widely used in multimedia content analysis, including text, images, audio and video. These models have strong geolocation inference capabilities and can extract geographic information from social media content. However, this capability also brings potential threats to user privacy, especially concerns about geolocation privacy and location privacy. Therefore, this study aims to explore the application, efficiency, limitations and impact of GPT in geolocation inference on user privacy, and develop corresponding privacy protection strategies.

Research objectives:

Evaluate and define the capabilities and limitations of large generative models in geolocation inference: Through literature review and experimental analysis, clarify the geolocation inference accuracy and limitations of the model when processing social media content.

Build and verify autonomous models: In order to avoid relying on black box models, develop an open source, transparent geolocation inference model and compare its performance with existing GPT models.

Explore potential risks of geolocation privacy: Analyze the impact that geolocation information inferred by the model may have on user privacy.

Develop privacy protection strategies: Based on the model analysis results, propose effective geolocation privacy protection measures to prevent the abuse of geolocation data.

Methodology:

Data collection: Collect public data on social media platforms as analysis samples. Model development and training: Use open source machine learning frameworks to develop geolocation inference models, and train and optimize them. Performance evaluation: Compare the performance differences between the autonomous model and the GPT model in the geolocation inference task. Privacy risk analysis: Analyze the specific impact of geographic information inference on personal privacy through case studies. Strategy formulation: Design and test a series of geo-privacy protection measures, such as data desensitization, access control, etc.

Expected results:

Provide a comprehensive evaluation report describing the performance and limitations of large-scale generative models in inferring the geographic attributes of social media content. Develop a transparent and verifiable geolocation inference model to provide the community with a reliable alternative tool. Develop a set of practical geo-privacy protection strategies to help alleviate the public's concerns about privacy violations that may be caused by big data analysis

Liu, Y., Kuai, C., Ma, H., Liao, X., He, B. Y., & Ma, J. (2024). Semantic Trajectory Data Mining with LLM-Informed POI Classification. arXiv preprint arXiv:2405.11715.

这篇论文核心是解释一种新的人类出行轨迹挖掘方法。这种方法的核心是用于优化交通系统、改善路线优化、交通管理以及研究人类出行模式。该方法首先使用大型语言模型（LLMs）来标注兴趣点（POI）的活动类型，然后采用基于贝叶斯的算法来推断轨迹中每个停留点的活动类型。这种方法在传统的基于规则的轨迹挖掘方法中整合了语义信息，比如通过POI数据推断出的活动类型，从而大幅提高挖掘质量。传统方法未能有效整合这类语义信息，限制了效率和准确性。而在这篇文章中，通过使用LLMs来处理POI数据的不完整特征问题，并以此进行分类，从而克服了现有学习算法对数据完整性的高依赖性。使用OpenStreetMap (OSM) POI数据集进行的评估，该方法在POI分类中达到了93.4%的准确率和96.1%的F-1得分，在活动推断中达到了91.7%的准确率和92.3%的F-1得分，显示出了较高的效果。

POI 是“兴趣点”（Point of Interest）的缩写，通常指在地图上能够代表某种特定信息或服务的位置点。这些点可能是商业地点、旅游景点、公共设施或任何其他对人们有特定兴趣的地方。例如，餐馆、博物馥、学校、公园、商店等都可以是POI。在交通和导航系统中，POI 用于帮助用户找到所需的地点，并在路径规划中提供有用的停靠点。在数据分析和地理信息系统（GIS）中，POI 数据用于研究地理位置的特性，分析商业趋势，城市规划，人口流动等方面。通过这些兴趣点的数据，可以提高轨迹挖掘的准确性和效率，使得交通管理和路线优化更加科学和精确。

在处理兴趣点（POI）数据中的不完整特征问题时，使用大型语言模型（LLMs）的方法主要涉及以下几个步骤：

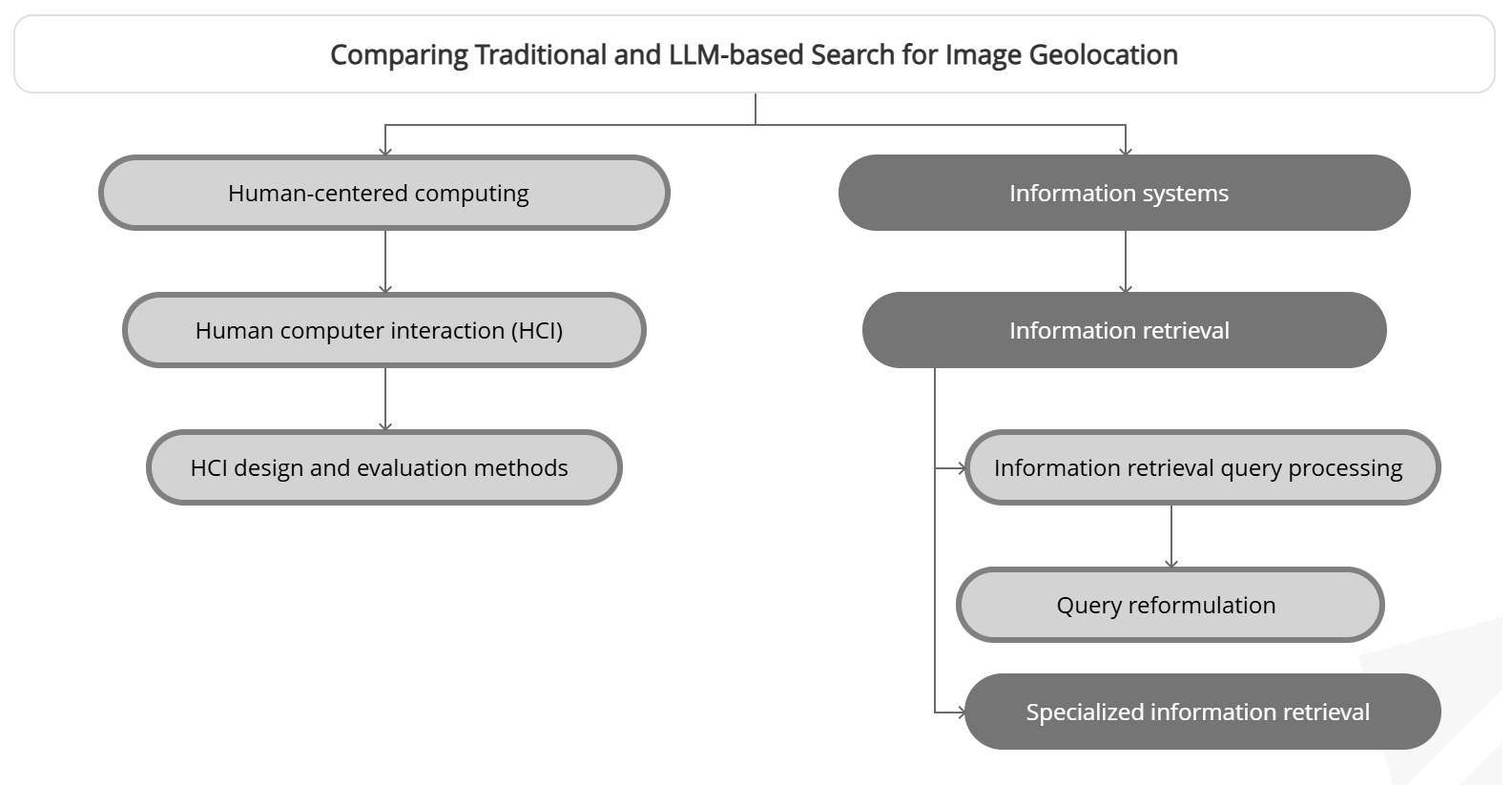
1. 首先对POI数据进行初步的清理和格式化，例如去除噪音、填充缺失的标签等。然后，从现有的数据中提取可用的特征，如位置信息、类别、用户评价等。

2. 对于那些特征信息不全的POI，LLMs可以通过理解与POI相关的文本描述或上下文信息来推断缺失的属性。例如，如果一个POI被标记为“餐馆”，但没有进一步的风格或菜系信息，模型可以通过分析与该POI相关的评论或描述来推断其可能的菜系类型（如中式、意式等）。

3. LMs可以生成缺失的文本描述或标签，根据已有信息推断并填充数据空缺。例如，通过分析地理位置和相似POI的数据，模型可以预测一个没有明确类别的POI可能属于的业务类型。

4. 使用补全和增强后的数据训练特定的机器学习模型，如分类器或聚类算法。这些模型可以进一步用于自动标注新的POI数据，或用于轨迹挖掘中的活动识别和分类。

Wazzan, A., MacNeil, S., & Souvenir, R. (2024, March). Comparing traditional and LLM-based search for image geolocation. In Proceedings of the 2024 Conference on Human Information Interaction and Retrieval (pp. 291-302).



这篇论文探讨了传统搜索引擎与基于大型语言模型（LLM）的搜索引擎在图像地理定位任务上的表现比较。图像地理定位是指确定一张图片拍摄地点的过程。研究主要分析了用户如何与这两种类型的搜索引擎互动，特别是在查询表达策略方面。

研究中，共有60名参与者参与实验，他们被随机分配使用传统搜索引擎或基于LLM的搜索引擎来帮助确定图片的拍摄地点。结果显示，使用传统搜索引擎的参与者在预测图片位置的准确性上超过了使用LLM-based搜索的参与者。

两组用户在搜索策略上表现出明显的不同：

使用LLM-based搜索的参与者倾向于使用更长、更自然的语言查询，但他们的搜索会话较短。

在重新构建查询时，使用传统搜索的参与者倾向于在初始查询中增加更多的关键词，而使用LLM-based搜索的参与者则倾向于重新表述他们的初始查询。

这项研究揭示了不同类型搜索工具对用户交互方式和查询策略的影响，对于理解和优化搜索引擎的用户体验具有重要意义。

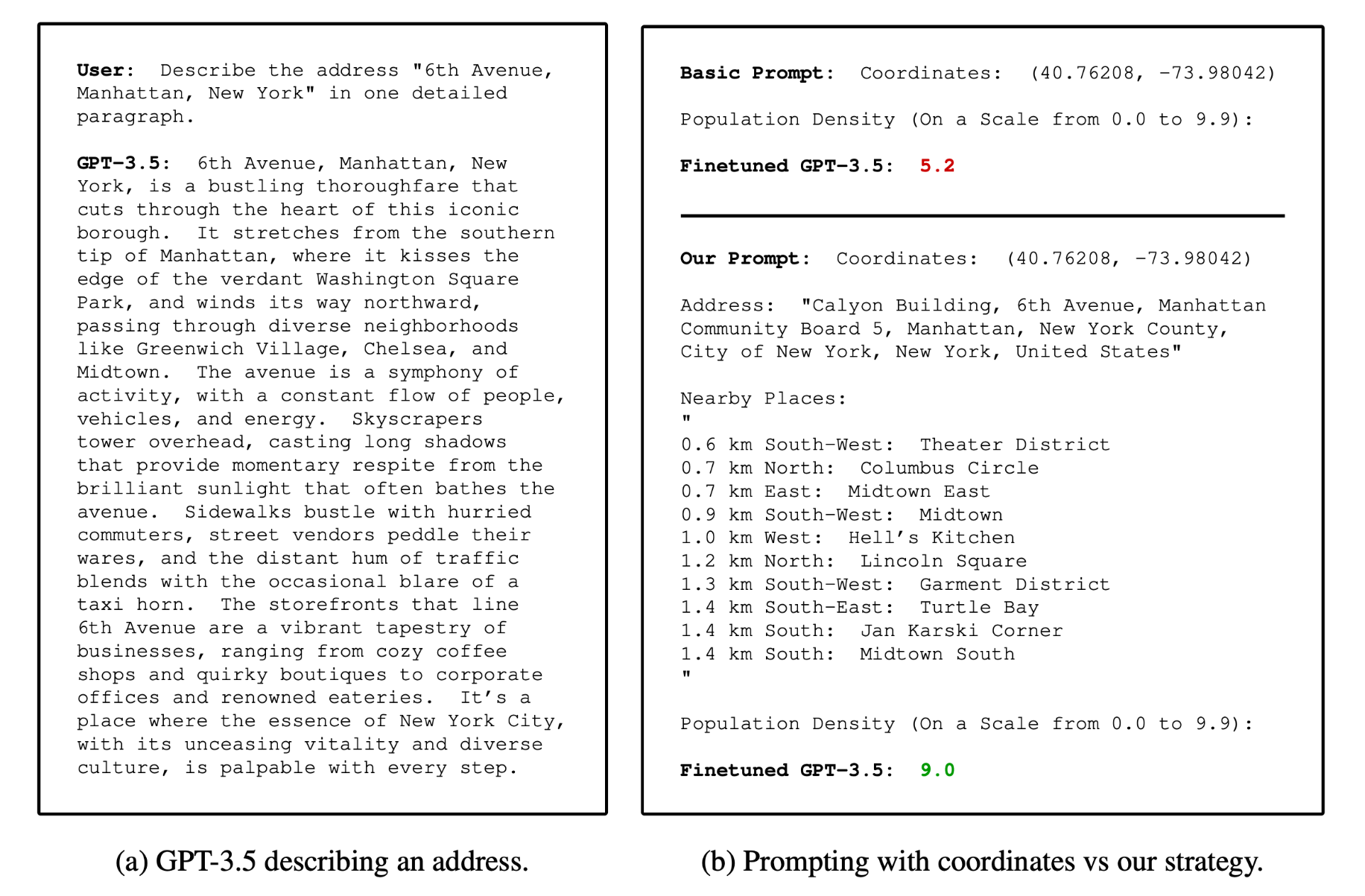
Manvi, R., Khanna, S., Mai, G., Burke, M., Lobell, D., & Ermon, S. (2023). Geollm: Extracting geospatial knowledge from large language models. arXiv preprint arXiv:2310.06213.

这篇论文介绍了名为GeoLLM的新方法，该方法通过结合大型语言模型（LLMs）和辅助地图数据（如OpenStreetMap）来提取地理空间知识，并用于地理空间预测任务。在现代地理空间分析中，虽然常用的全球变量如卫星图像在某些方面表现出色，但也存在成本高和预测能力不足的问题。作者探讨了是否能够利用互联网语言语料库中的丰富知识——这些知识现已被大型语言模型压缩并储存——来解决这些限制。

研究首先证明了LLMs内嵌了关于地点的显著空间信息，但是简单地使用地理坐标来查询LLMs并不有效，不能有效预测如人口密度等关键指标。因此，他们提出了GeoLLM方法，它能够有效地从LLMs中提取地理空间知译，并用于测量人口密度和经济生活状况等国际社会关心的多个任务。

通过实验比较，GeoLLM在这些任务上表现出了较高的性能，相对于使用最近邻或直接从提示信息中获取数据的基线模型，性能提高了70%，甚至达到或超过了现有文疗中基于卫星的基准。实验还显示，GPT-3.5在GeoLLM框架下的表现超过了Llama 2和RoBERTa分别达19%和51%，表明该方法的效果随着模型大小和预训练数据集的增加而提高。此外，LLMs在样本效率、地理空间信息的丰富性和全球范围内的鲁棒性方面表现出色。

总之，GeoLLM不仅展现了利用LLMs进行地理空间预测的潜力，还显示了它在弥补现有地理空间协变量限制和与之良好互补的前景。此外，该研究的代码已在项目网站上公开，以便其他研究者使用和进一步开发。



Wang, Z., Xu, D., Khan, R. M. S., Lin, Y., Fan, Z., & Zhu, X. (2024). LLMGeo: Benchmarking Large Language Models on Image Geolocation In-the-wild. arXiv preprint arXiv:2405.20363.

文章主要探讨了在图像理解应用中的关键任务——图像地理定位，并尝试解决在分析自然环境下（即“野外”）的复杂图像时现有方法常常失败的问题。

研究团队首先从Google Street View收集了来自不同国家的图像，并创建了一个新的图像数据集。随后，他们使用这一数据集系统评估了多模态语言模型在地理定位方面的能力。评估分为两种类型：无需训练的评估和基于训练的评估，这些评估分别针对闭源和开源的多模态语言模型进行。

研究结果表明，闭源模型在地理定位任务上表现出了更优越的能力。然而，通过细致的调整，开源模型也能够达到与闭源模型相媲美的表现。这些发现对于推动图像地理定位技术的发展，特别是在处理自然环境中拍摄的复杂图像时，具有重要的意义和应用前景。

Zelko, Andreea-Cristina, and Matej Kucera. "Location Inference on Twitter." 21st SC@ RUG 2023-2024: 14.

探讨了通过社交媒体帖子推断个人地点信息的技术及其可能的应用。在互联网上，个人地点数据既有研究也有商业用途，但这类数据很少公开可用。由于缺乏地理标记的数据，许多需要这种数据的研究工作往往难以进行。

为了解决这个问题，论文深入探讨了几种地点推断技术。通过分析社交媒体帖子的文本内容和其他上下文信息，这些技术能够推断出准确的位置信息。论文列出了这些算法使用的输入数据类型以及它们能够达到的结果，并强调了社交媒体用户了解这些信息的重要性，以便他们有效控制公开的信息量。

此外，研究还探讨了将大型语言模型（LLMs）应用于基于文本的位置推断的可行性。实证研究结果显示，LLMs能够从文本中提取信息，几乎不需要预处理。因此，论文指出了这类工具的普遍可用性及其与其他现有方法结合时的潜在能力。同时，论文也提到了所介绍算法以及基于LLM的方法的一些局限性。

这篇论文不仅探讨了社交媒体地点推断技术的科技层面，也提醒了公众关注与此相关的隐私问题，强调了在这个领域存在的隐私关切的重要性。

Haris, Erum, Anthony G. Cohn, and John G. Stell. "Exploring Spatial Representations in the Historical Lake District Texts with LLM-based Relation Extraction." GeoExT@ ECIR. 2024.

论文主要研究了如何通过使用大型语言模型（特别是生成预训练转换模型，即Generative Pre-trained Transformer，GPT）来挖掘和理解历史叙事中的空间关系。研究聚焦于英格兰湖区，利用了“湖区文学语料库”。

在处理历史叙述时，揭示过去景观的空间复杂性是一个挑战。该研究通过自然语言处理技术提取文本描述中的空间关系，进一步深入理解历史叙事中固有的空间维度。具体方法是利用语言模型从文本中生成语义三元组（semantic triples），这些三元组捕捉实体和位置之间的微妙联系，并将这些关系以网络图的形式可视化，从而提供一个关于空间叙事的图形表示。

通过这种方法，研究不仅增进了对英格兰湖区空间结构的深入理解，还提供了一种在不同历史背景下揭示空间关系的新途径。这项工作对历史学、地理学乃至文学研究领域都有重要的贡献，尤其是对于那些关注历史文本中空间维度解析的研究者。

Luo, Yuxiao, et al. "Deciphering Human Mobility: Inferring Semantics of Trajectories with Large Language Models." arXiv preprint arXiv:2405.19850 (2024).

文章主要探讨了如何使用大型语言模型（LLM）来深入解析和推断人类移动轨迹的语义信息，这对于城市规划、公共安全等多种应用领域至关重要。

虽然个体轨迹数据（如手机定位数据）富含时空信息，但常常缺乏足够的语义细节，这限制了其在深入移动性分析中的应用。现有方法仅能从这些数据中推断出基本的日常活动序列，而无法深入理解复杂的人类行为和用户特征，且这些方法还依赖于难以获取的辅助数据集，如旅行调查数据。

为了克服这些限制，论文提出了一种新的框架——轨迹语义推断与大型语言模型（TSI-LLM），定义了轨迹语义推断的三个关键维度：用户职业类别、活动序列和轨迹描述。TSI-LLM框架通过采用时空属性增强的数据格式（STFormat）和设计包含上下文的提示（context-inclusive prompt），使得LLMs能更有效地解释和推断轨迹数据的语义。

通过在现实世界轨迹数据集上的实验验证，TSI-LLM框架证明了其在解析复杂人类移动模式方面的有效性。这项研究探索了LLMs在增强轨迹数据语义分析方面的潜力，为更复杂、更易于访问的人类移动性研究铺平了道路。这不仅推动了个体轨迹数据的深入利用，也为相关应用提供了新的方法和视角。

Chen, Guanhua, Wenhan Yu, and Lei Sha. "Unlocking Multi-View Insights in Knowledge-Dense Retrieval-Augmented Generation." arXiv preprint arXiv:2404.12879 (2024).

文章主要研究了在法律和医学等知识密集领域中，如何通过改进检索增强生成（RAG）来提高大型语言模型（LLMs）的应用效果。现有的检索方法常常因缺乏多角度视野而在解释性和可靠性方面受限，这一点在需要高度专业知识的领域尤为重要。

论文提出了一种新的多视角RAG框架，称为MVRAG，专为知识密集领域设计。这个框架通过从多个领域视角进行意图感知的查询重写，以提高检索的精确度，从而提高最终推断的效果。通过在法律和医疗案例检索上进行的实验，证明了该框架在召回率和精确度方面都有显著提高。

MVRAG框架的多角度检索方法释放了多视角信息在增强RAG任务中的潜力，加速了LLMs在知识密集型领域的进一步应用。这种方法不仅提高了检索的质量和生成内容的相关性，还增强了模型在处理复杂查询时的理解能力，为高度专业化领域的语言模型应用提供了新的可能性。

**检索增强生成（Retrieval-Augmented Generation，简称RAG）**是一种结合了检索（retrieval）和生成（generation）技术的自然语言处理模型。这种模型首先从一个大型的、结构化或非结构化的知识库中检索出与输入查询相关的信息，然后使用这些信息来增强生成模型的回答能力，以产生更准确、更丰富的输出。

RAG的工作流程如下：

输入处理：接收用户的查询或问题作为输入。

检索阶段：根据输入查询，RAG从一个或多个外部数据源（如维基白皮书、专业知识库或互联网）中检索相关信息。这可以通过传统的信息检索技术实现，比如使用向量空间模型、倒排索引等。

信息融合：检索到的信息被送入生成模型。在这一阶段，模型将输入查询和检索到的内容结合起来，共同影响最终的生成结果。

生成阶段：使用语言生成模型（通常是基于Transformer的模型，如GPT系列）根据融合了检索信息的上下文生成文本回答或内容。

使用检索增强生成（RAG）来提高大型语言模型（LLMs）在从文本和照片中进行地理位置推断的应用具体

**检索增强**：通过整合外部知识库，如地理信息系统（GIS）和地标数据库，RAG可以提供更多上下文信息，帮助LLMs在地理位置推断中实现更精确的识别。

**多模态输入**：结合文本和图像信息可以使模型更全面地理解和推断地理位置，因为图像提供了直观的位置线索，而文本可能包含描述或地点的具体信息。

集成高质量地理数据源：

利用开源地理信息数据（如OpenStreetMap）和商业地图服务（如Google Maps API），获取详细的地理位置数据和属性。

集成历史和文化地理数据，提供关于特定地点的背景信息和文化意义。

多模态数据处理：

开发算法来分析图像中的地理标志和特征（如建筑、自然景观），并将这些视觉信息与文本描述相结合。

利用先进的图像识别和自然语言处理技术，提高对文本和图像内容的理解深度。

用户交互与反馈机制：

设计交互界面，允许用户输入文本和上传图片，以及反馈推断结果的准确性。

利用用户反馈来持续优化模型，提高地理位置推断的准确率和可靠性。

应用场景开发：

开发旅游和文化遗产探索应用，利用地理位置推断来提供关于旅游景点的详细信息。

在紧急响应和公共安全领域，开发应用程序来快速定位事件和事故的地点。

使用gpt-4等大多模态模型对于用户社交媒体上传的图片，视频和文本这三个最常见的因素进行地理信息数据的挖掘和推断，构建一个以人为本的，具有地理属性的框架或者说人物画像，来表征人地特点，使用RAG的手法增加准确性。

**地理信息推断**：使用GPT-4等多模态模型，从社交媒体上传的图片、视频和文本中挖掘和推断地理信息。

**人物画像构建**：建立包含地理属性的人物画像，反映用户与其环境的互动和特征。

**隐私保护策略研究**：研究和开发有效的隐私保护措施，确保用户数据处理的安全性和合规性。

**黑盒探索**：讲GPT-4的推断分析过程提取出来，使用开源的模型进行复现和尝试；针对文本，视频和图片三个不同的数据类型来设计三个模型，使用知识图谱的形式来串联三个模型。