地理空间人工智能基础模型的机遇与挑战

GENGCHEN MAI，美国佐治亚大学地理系SEAI实验室WEIMING HUANG，新加坡南洋理工大学计算机科学与工程学院JIN SUN, 美国佐治亚大学计算学院SUHANG SONG，美国佐治亚大学公共卫生学院DEEPAK MISHRA，美国佐治亚大学地理系NINGHAO LIU，美国佐治亚大学计算学院SONG GAO，美国威斯康星大学麦迪逊分校地理系地理空间数据科学实验室TIANMING LIU，美国佐治亚大学计算学院GAO CONG， 新加坡南洋理工大学计算机科学与工程学院YINGJIE HU，美国布法罗大学地理系GeoAI实验室CHRIS CUNDY，美国斯坦福大学计算机科学系ZIYUAN LI，美国康涅狄格大学商学院RUI ZHU，英国布里斯托大学地理科学学院NI LAO，美国谷歌

arXiv:2304.06798v1 [cs.AI] 13 Apr 2023

大型预训练模型也称为基础模型 （FMs），通过在大规模数据上以任务无关的方式进行训练，可通过微调、少样本学习甚至零样本学习适配到广泛的下游任务。尽管它们在语言和视觉任务中取得了成功，但我们尚未看到开发地理空间人工智能（GeoAI）基础模型的尝试。在本研究中，我们探讨开发用于GeoAI的多模态基础模型的前景与挑战。首先，我们通过测试许多现有基础模型在多个地理空间子领域中七项任务上的表现，考察它们的潜力。这些子领域包括地理空间语义、健康地理学、城市地理学和遥感。结果表明，对于若干仅涉及文本模态的地理空间任务，例如地名识别、位置描述识别以及美国州级/县级痴呆症时间序列预测，这些任务无关的大语言模型在零样本或少样本学习设置下能够胜过特定任务的全监督模型。然而，在其他地理空间任务中，尤其是涉及多种数据模态的任务（例如基于兴趣点的城市功能分类、基于街景图像的城市噪声强度分类以及遥感图像场景分类），现有基础模型的表现仍不及特定任务模型。基于这些观察，我们认为，开发GeoAI基础模型的一项主要挑战，是应对地理空间任务的多模态特性。在讨论每种地理空间数据模态所面临的独特挑战后，我们提出一种可能性：构建能够通过地理空间对齐对各类地理空间数据进行推理的多模态基础模型。最后，我们讨论开发此类GeoAI模型所面临的独特风险与挑战。

允许为个人或课堂使用而免费制作本作品全部或部分内容的数字或纸质副本，前提是这些副本不得用于营利或商业利益目的，不得为此进行制作或分发，并且副本首页必须注明本声明及完整引文。必须尊重本作品中归其他方而非 ACM 所有的组件的版权。允许在注明出处的情况下摘录。若要以其他方式复制、重新出版、发布到服务器或分发至列表，则必须事先获得明确许可和/或支付费用。请通过 permissions@acm.org 申请许可。© 2018年 计算机协会。稿件已提交至 ACM。

CCS概念： • **计算方法** → **自然语言处理**；无监督学习；**计算机视觉**；**知识表示与推理**；**神经网络**； • **应用计算** → **地球与大气科学**。

附加关键词和短语：基础模型、地理空间人工智能、多模态学习

ACM 参考格式：
Gengchen Mai、WEIMING HUANG、JIN SUN、SUHANG SONG、DEEPAK MISHRA、Ninghao Liu、Song Gao、TIANMING LIU、GAO CONG、YINGJIE HU、CHRIS CUNDY、ZIYUAN LI、Rui Zhu 和 Ni Lao。2018 年。关于地理空间人工智能基础模型的机遇与挑战。载于。 ACM，纽约州纽约，美国，37 页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

机器学习（ML）和人工智能（AI）的最新发展彰显了数据和计算资源的巨大力量。在互联网规模数据集上训练的超大型模型，已在多种学习任务中取得最先进水平（SOTA）的表现。尤其是，这些模型前所未有的成功推动了当代机器学习模型训练方式的范式转变。如今，人们不再从头开始训练特定任务模型，而是对这类预训练模型（所谓的“基础模型（FMs）”[13]）进行微调，或采用少样本／零样本学习策略加以适配，随后将其部署于广泛的领域 [15, 112]。这类基础模型能够促进跨领域的知识迁移与共享，并减少对特定任务训练数据的需求。基础模型的例子包括：1）大型语言模型（*LLM*），例如 PaLM [140], LLAMA [132], GPT-3 [15], InstrucGPT [106],和 ChatGPT [104]；2）大型视觉基础模型，例如 Imagen [126], Stable Diffusion [124], DALL·E2 [117], 和 SAM [72]；3）大型多模态基础模型，例如 CLIP [112], OpenCLIP [54], BLIP [82], OpenFlamingo [11],KOSMOS-1 [49], 和 GPT-4 [105]；以及 4）大型强化学习基础模型，例如 Gato [121]。

尽管取得了这些成功，但目前鲜有研究探索开发一种类似的地理空间人工智能（GeoAI）基础模型，而地理空间人工智能处于地理空间科学发现与人工智能技术的交汇处 [31, 55, 93]。这里的核心技术挑战在于 GeoAI 本质上具有多模态特性。GeoAI 的核心数据模态包括文本、图像（例如遥感图像或街景图像）、轨迹数据、知识图谱以及地理空间矢量数据（例如来自 OpenStreetMap 的地图图层），所有这些数据均包含重要的地理空间信息（例如几何信息和语义信息）。每种模态都具有特殊的结构，需要采用各自独特的表示方式——要在单一模型中有效地结合所有这些表示，并融入适当的归纳偏置，需要经过审慎的设计。GeoAI 的多模态特性阻碍了现有预训练基础模型在所有 GeoAI 任务中的直接应用。

本文为开发地理空间人工智能基础模型奠定基础。我们首先在第2节简要概述现有基础模型。随后，在第3节中，我们通过系统比较若干热门基础模型与多种最先进的全监督任务专用机器学习或深度学习模型在不同地理空间领域多项任务上的表现，探究现有基础模型在地理空间人工智能中的潜力：1）**地理空间语义**：地名识别和位置描述识别任务；2）**健康地理学**：美国州级和县级痴呆症死亡人数时间序列预测任务；3）**城市地理学**：基于兴趣点（POI）的城市功能分类任务和基于街景图像的噪声强度分类任务；4）**遥感**：遥感图像场景分类任务。我们相应地讨论基础模型在不同地理空间任务中的优势和问题。接下来，在第4节中，我们详细阐述开发地理空间人工智能基础模型所面临的挑战。为所有地理空间人工智能数据模态创建一个单一基础模型可能是一项艰巨的任务。为此，我们首先考察地理空间人工智能任务中使用的各类数据模态，并由此展开讨论。随后，我们提出构想，设计一种能够应对上述挑战的地理空间人工智能新型多模态基础模型框架。最后，我们指出在开发此类通用模型时可能存在的一些风险和挑战，这些问题

应在第5 节开发地理空间人工智能通用模型时予以考虑。最后，我们在第6节总结本文。

我们的贡献可概括如下：

- 据我们所知，这是首项系统性研究，旨在考察多种现有尖端基础模型在多个地球科学领域的不同地理空间任务中的有效性及其存在的问题[1]。我们针对七项地理空间任务建立了多种基础模型基线方法，以支持未来的地理空间通用人工智能（GeoAGI）研究。

- 我们探讨了开发用于地理空间人工智能的多模态基础模型所面临的挑战，并提出了一个有望实现这一目标的框架。
- 我们讨论了在多模态地理空间基础模型的开发和评估过程中需要考虑的风险与挑战。

2 相关工作

2.1 语言基础模型

不到十年间，大规模语言建模（大语言模型）已经彻底革新了计算自然语言能力 [15, 69, 109, 116]。语言建模 [62] 是一项简单的任务：根据前面的词元预测序列中的下一个词元[2]；从某种意义上说，它对应一种自监督目标，因为除了自然文本语料库之外，无需人工标注。当应用于海量语料库（例如来自互联网的各种主题文档）时，大语言模型能够显著提升语言理解与生成能力。各种迁移学习和规模扩展研究 [40, 43, 66] 表明，下游任务质量与自监督模型规模和数据规模的对数之间几乎呈线性关系。随着数据和计算资源日益丰富，语言建模已成为开发日益强大模型的一种可靠方法。

这类大语言模型的代表性示例是 OpenAI GPT [15, 105, 114, 115]。通过使用海量网络数据进行预训练，GPT模型掌握了网络上几乎所有领域的知识，可将其用于解决不同垂直领域的问题 [15]。获取这类知识的接口已变得越来越简单、直观，涵盖从使用带标签的数据进行监督微调 [114, 115], ，到少样本学习 [15] 、指令 [106], 、对话 [104] 和多模态性 [105]。本研究全面分析了 GPT 和其他大语言模型应用于不同地理空间领域时的潜力与局限性。

2.2 视觉基础模型

长期以来，计算机视觉一直由特定任务模型主导：例如，用于目标检测的 YOLO [120] 、用于实例分割的 Detectron [143] ，以及用于图像超分辨率的 SRGAN [78] 。最新的例子是 Meta AI 的分割一切模型（SAM） [72], ，该模型专为交互式目标分割而设计。在 ImageNet [24] 上训练的 ResNet [36]已被用作许多此类任务的骨干特征提取器。它可以被视为视觉基础模型的早期形式。

受语言基础模型巨大成功的启发，计算机视觉领域构建了大规模视觉基础模型，使其能够适配任何视觉任务。计算机视觉领域对语言模型理念最直接的借鉴，体现在图像生成模型上。自生成对抗网络

[1]这项工作是在我们此前发表于 ACM SIGPATIAL 的 4 页视觉论文基础上的重要扩展，新增了健康地理学、城市地理学和遥感领域的五项任务。2022 [92] [2]此外，还有一种不同的变体，旨在预测文本中的掩码跨度 [69, 116]。

[32, 67], 以来，图像生成模型的质量随着基于扩散的模型的发展取得了重大突破 [41]。Imagen [126] 基于大型基于 Transformer 的语言模型来理解文本提示，并使用扩散模型生成高保真图像。DALL-E·2 [117] 训练扩散解码器，以反向还原来自 CLIP 等视觉语言模型的图像编码器。经过预训练后，它能够生成各种风格和特征的图像。Stable Diffusion [124] 使用变分自编码器（VAE） [71]将原始图像从像素空间转换到潜在空间，使扩散过程更易于管理且更加稳定。它在以文本、姿态、边缘图、语义图和场景深度 [156]为条件进行生成方面展现出极强的灵活性。另一方面，GigaGAN [64], 是近期扩展生成对抗网络模型规模的一次尝试。

视觉 Transformer（ViT） [25] 是视觉基础模型中广泛使用的一种架构。为了扩展该模型，研究人员开发了大规模 ViT [153]。Swin Transformer [88]模型旨在应对将常规 Transformer 模型适配到具有不同空间分辨率的图像时所面临的独特挑战。研究人员还开发了其他大规模非 Transformer 模型，以达到相同的性能水平：ConvNext [89] 是卷积神经网络的“现代化”版本，参数量庞大，性能与 Swin Transformer 相近。MLP-Mixer [131] 是一种仅在图像数据上使用多层感知机的架构，在图像分类数据集上取得了具有竞争力的成绩。

2.3 多模态基础模型

开发能够对复杂数据进行多模态推理和理解的人工智能模型，是一个很有前景的思路。在日常生活中，人类自然会进行多模态推理 [108]；例如，当一个人想到“狗”这个概念时，不仅会想到这个英文单词及其含义，也会想到与之相关的视觉图像和声音。在地理空间任务中，多模态数据无处不在。一般而言，不同模态的数据提供不同的“视角”，这些视角相互补充，提供更多信息，从而帮助全面理解数据。

近年来，在构建能够对不同领域（尤其是视觉与语言）进行联合推理的大规模多模态基础模型方面取得了显著进展。CLIP [54, 112] 是最早广泛采用的视觉语言联合训练框架之一。它利用自监督对比学习，学习视觉与文本特征的联合嵌入。BLIP [82]通过使用从互联网收集的图像所合成生成的图像描述进行训练，进一步改进了 CLIP。它旨在同时处理视觉语言理解和生成任务。BEiT-3 [138]是一种通用多模态基础模型，在视觉任务和视觉语言任务上均实现了最先进的性能。它结合了多模态专家网络的特征。Florence [151]是一种视觉语言基础模型，能够为物体、场景、图像、视频以及图像描述学习通用的视觉语言表征。同样，KOSMOS-1 [49] 从网络规模的多模态数据中学习，其中包括文本与图像对。它能够将知识从一种模态迁移到另一种模态。Flamingo [6] 是一系列视觉语言模型，仅需少量标注样例即可适配新任务，即少样本学习。它将图像或视频与文本词元一同编码为输入，以对视觉任务进行联合推理。最新版本的 GPT 模型 GPT-4 [105], 也声称能够进行多模态分析，涵盖文本、音频、图像和视频。

3 对现有基础模型在各种地理空间领域有效性的探索

我们想提出的第一个问题是：与最先进的全监督任务特定模型相比，现有的尖端基础模型在各种地理空间任务中的表现如何。地理学是一门非常广泛的学科，包括地理空间语义等多个子领域， [46, 57, 60, 75, 97], 健康地理学

[19, 68, 125], 城市地理学、 [17, 52, 65, 154, 165],遥感 [16, 28, 79, 100, 123],等。为回答上述问题，下面我们将在前文提到的四个地理空间子领域中，使用各种基础模型针对不同任务开展实验。我们将详细讨论现有基础模型的优势与不足。

3.1 地理空间语义

```
[说明]  . . .
段落:   阿拉巴马州 州警称，一名格林维尔男子 has died 因其
        ↪ 受伤后 被一辆皮卡车 在65号州际公路上撞倒
        ↪ 位于朗兹县。
Q: 哪些 词语 在本 段落 表示 命名 地点?
A: 阿拉巴马州; 格林维尔; 朗兹
. . .
--
段落:   华盛顿镇之于威廉斯堡，正如威廉斯堡之于弗吉尼亚州。
Q: 这段文字中的哪些词 表示命名 地点?
A: 华盛顿州；威廉斯堡；弗吉尼亚州
```

清单 1. 使用大语言模型（例如 GPT-3）进行地名识别。黄色方框：待标注的文本片段。橙色方框：GPT-3 的输出。此提示中使用了 8 个少样本样例。为节省空间，这里仅展示 1 个，其余以“...”省略。

```
[指令]  . . .
段落:   爸爸被困在 住宅里。水位涨至腰部以上。 HELP
        ↪ 伍德林路 8111 号, 77028 #休斯敦洪灾
Q: 哪些 词语 在这个 段落 表示地点 描述?
A: 8111 伍德林 Rd, 77028
. . .
--
段落:
        ↪ 哈维飓风救援需求，地址：7506 Jackrabbit Rd, Houston, TX 77095。
Q: 哪些 词语 in this 段落 表示 位置 描述?
A: 德克萨斯州休斯敦市杰克兔路 7506 号，邮编 77095
```

代码清单 2. 使用大语言模型进行位置描述识别，例如 GPT-3. 黄色方框：输入文本片段。橙色方框：GPT-3 的输出。使用了 11 个少样本样例，其中展示了 1 个。

作为讨论的起点，我们首先通过实证展示利用大语言模型解决地理空间语义任务的潜力。我们希望研究结果不仅能展现此类通用少样本学习器在地理空间语义领域的有效性，也能对当前将训练特定任务模型作为地理空间人工智能研究惯例的范式提出挑战。

我们比较了 Huggingface 提供的 4 种不同规模的预训练 GPT-2 [115] 模型，以及最新的 GPT-3 [15] (即 text-davinci-002)、InstructGPT [106] (即 text-davinci-003) 和 ChatGPT [104] (即 gpt-3.5-turbo) 模型（由 OpenAI 开发）与多个 有监督的特定任务 基线方法在两项具有代表性的地理空间语义任务上的表现：(1) 地名识别 [33, 134], 和 (2) 位置描述识别 [45]。

3.1.1 地名识别。 地名识别是命名实体识别（NER）的一个子任务，目标是从给定的文本片段中识别命名地点。我们使用 Hu2014 [47] 和 Ju2016 [61] 基准数据集作为该任务的两个代表性数据集。我们通过使用包含少样本训练示例的适当提示词，使 7 个预训练 GPT 模型执行地名识别任务。在提示中，我们以自然语言指令的形式提供若干训练样本作为少样本学习样本。此类提示的一个示例见列表 1，完整提示可见附录 7 中的列表 A.1。值得注意的是，作为基础模型，ChatGPT 针对聊天机器人用途进行了优化，并期望接收对话式输入，而不是单个大型提示。为进行受控实验，我们首先使用列表1 中所示的相同提示，指导全部 7 个预训练 GPT 模型执行地名识别。我们还将少样本示例转换为对话列表，并将其用作 ChatGPT 的输入，此版本记为 ChatGPT（对话版）；而使用原始提示的 ChatGPT 则记为 ChatGPT（原始版）。

表1 比较了全部 8 种 GPT 模型与 15 种基线方法在两个数据集 Hu2014 [47] 和 Ju2016 [61]上的表现。所有模型均使用相同的测试集进行评估。关于这两个数据集的详细说明，请参阅 Hu 等人 [47]和 Ju 等人 [61]。如表1所示，这 15 种基线方法分为三组：(A) 通用命名实体识别模型；(B) 非神经网络（NN）地理解析器；(C) 全监督的特定任务

基于神经网络的地理解析器。C 组中的所有模型均采用监督方式，在相同的划分训练数据集上进行训练。除最小的 GPT2 模型外，其他所有大语言模型在 Hu2014 数据集上的表现始终优于全监督基线方法，尽管它们只需要一小组自然语言指令，无须额外训练。尤其是，GPT-3 的表现比此前的最先进水平（TopoCluster[23]）提高了 8.7%。有趣的是，InstructGPT 和 ChatGPT 等新型 GPT 模型在 Hu2014 数据集上的表现并未更高。InstructGPT 的表现降幅较小，可以接受；而两个 ChatGPT 模型的表现则下降得更为明显。一种合理的假设是，ChatGPT 在 InstructGPT 的基础上进一步优化，用于聊天机器人应用，但这类应用可能不够“灵活”，无法适应地名识别等新任务。

根据以往研究，[134, 135], Ju2016 数据集是一项非常困难的任务。在该数据集上，我们发现 GPT2-XL 的表现比先前的最先进水平（NeuroTPR [135]）高出 2.5% 以上，而提示中仅使用了*8* 个少样本示例。相比之下，NeuroTPR 这类特定任务模型需要在 599 条带标签的推文以及从 3000 篇维基百科文章中生成的带标签句子上进行监督训练。GPT-3 和 InstructGPT 在 Ju2016 数据集上的表现并未优于 GPT2-XL。与 Hu2014 数据集上的发现类似，ChatGPT 在 Ju2016 数据集上的表现显著下降。与现有的实证发现一致，[15, 115], 我们还发现，这些大语言模型的表现往往会随着可学习参数数量的增加而提升。

3.1.2 位置描述识别。位置描述识别任务稍具挑战性——给定一段文本（例如推文），目标是识别更细粒度的位置描述，例如家庭地址、高速公路、道路和行政区域。HaveyTweet2017 [45] 被用作该任务的代表性基准数据集之一。该任务采用相同的一组预训练 GPT 模型和 15 种基线方法。列表2 展示了该任务中使用的一个提示示例，完整提示见附录8 。A.1。

表1 总结了不同模型在 HaveyTweet2017 数据集上的评估结果。所有 GPT 模型和 15 种基线模型均使用 HaveyTweet2017 的同一测试集进行评估。在 HaveyTweet2017 数据集上，GPT-3 的召回率在所有方法中最高。然而，所有大语言模型的精确率都相当低（因此 F1 分数也较低）。这是因为大语言模型会隐式地将位置描述识别问题转化为自然语言生成问题（见列表2），这意味着无法保证它们生成的词元会出现在输入文本中。根据表1中的实验结果，我们可以清楚地看到，只需使用少量少样本样例，大语言模型在定义明确的地理空间语义任务上便能胜过全监督的特定任务模型。这表明，大语言模型有望大幅减少地理空间任务对定制架构或大型带标签数据集的需求。然而，如何设计恰当的提示词来指导大语言模型完成给定的地理空间语义任务，仍有待进一步研究。

3.2 健康地理学

下一组实验聚焦于一个重要的健康地理学问题——针对特定地理区域（如城市、县、州等）进行痴呆症死亡人数时间序列预测。随着人口中老年人占比不断上升，据估计，2020年美国65岁及以上的成年人中，超过700万人患有痴呆症；到2030年，这一数字可能增至900多万，并在 2040 [167]前接近1200万。据报道，阿尔茨海默病是最常见的痴呆症类型，也是美国主要死因之一；到 2019 [9]，每三位老年人中就有一位死于阿尔茨海默病或其他痴呆症。值得注意的是，痴呆症死亡率存在显著且长期的地理差异 [4, 8]。地方层面的规划和确定痴呆症预防工作的优先事项

表 1 1 2014 。各种 [47] 2016 [61] 2 2017 [45]

GPT 模型和基线方法在两项地理空间语义任务上的评估结果：( ) 地名识别（Hu和 Ju）以及 ( ) 位置描述识别（HaveyTweet）。我们将所有模型分为四组：(A) 通用命名实体识别；(B) 非神经网络（NN）地理解析器；(C) 全监督的基于神经网络的地理解析器；(D) 使用大语言模型进行少样本学习。“(#Param)”表示大语言模型的可学习参数数量。“(nar. loc.)”和“(bor. loc.)”分别表示窄范围位置模型和宽范围位置模型，定义见 [135]。所有基线方法（即 A、B、C 组中的模型）的结果均取自 [134] 和 [135]，但“0.675†”除外，该结果是通过重新运行 [135]的官方代码获得的。不同 GPT 模型（D 组）的评估结果是通过使用预训练的 GPT2/GPT-3/InstructGPT/ChatGPT 模型并配合适当提示词得出的。四种 GPT2 模型的结果来自使用不同模型规模的 Huggingface 预训练 GPT2 模型。最后四种模型的结果来自使用 OpenAI 的不同 GPT 模型——text-davinci-002、text-davinci-003 和 gpt-3.5-turbo，分别记作 GPT-3、InstructGPT 和 ChatGPT。由于 ChatGPT 接受的是对话输入，而不是单个长提示词，我们对 ChatGPT 的两个版本进行了实验。ChatGPT（原始版）表示使用与其他 GPT 模型相同的提示词；ChatGPT（对话版）表示将提示词中的少样本示例转换为一系列对话。 *由于 OpenAI 应用程序编程接口的限制，我们在随机抽取的 544 个 Ju2016 示例（占数据集的 10%）上评估 GPT-3、InstructGPT 和 ChatGPT。

	模型	#参数	地名识别 位置描述识别				
			Hu2014	Ju2016	HaveyTweet2017		
			准确率 ↓	准确率 ↓	精确率 ↓	召回率 ↓	F 分数 ↓
(A)	斯坦福命名实体识别（自然地名） [30]	-	0.787	0.010	**0.828**	0.399	0.539
	斯坦福命名实体识别（bro. loc.） [30]	-	-	0.012	0.729	0.44	0.548
	重新训练的斯坦福命名实体识别 [30]	-	-	0.078	0.604	0.410	0.489
	不区分大小写的斯坦福命名实体识别（窄地名） [30]	-	-	0.460	0.803	0.320	0.458
	不区分大小写的斯坦福命名实体识别（宽地名） [30]	-	-	0.514	0.721	0.336	0.460
	spaCy 命名实体识别（窄地名） [44]	-	0.681	0.000	0.575	0.024	0.046
	spaCy 命名实体识别（宽地名） [44]	-	-	0.006	0.461	0.304	0.366
	DBpedia Spotlight[99]	-	0.688	0.447	-	-	-
(B)	爱丁堡 [7]	-	0.656	0.000	-	-	-
	CLAVIN [134]	-	0.650	0.000	-	-	-
	TopoCluster [23]	-	0.794	0.158	-	-	-
(C)	CamCoder [33]	-	0.637	0.004	-	-	-
	基础 BiLSTM+CRF [77]	-	-	0.595	0.703	0.600	0.649
	DM NLP（热门推荐） [139]	-	-	0.723	0.729	0.680	0.703
	NeuroTPR [135]	-	0.675†	0.821	0.787	0.678	**0.728**
(D)	GPT2 [115]	117M	0.556	0.650	0.540	0.413	0.468
	GPT2-Medium [115]	345M	0.806	0.802	0.529	0.503	0.515
	GPT2-Large [115]	774M	0.813	0.779	0.598	0.458	0.518
	GPT2-XL [115]	1558M	0.869	**0.846**	0.492	0.470	0.481
	GPT-3 [15]	175B	**0.881**	0.811*	0.603	**0.724**	0.658
	InstructGPT [106]	175B	0.863	0.817*	0.567	0.688	0.622
	ChatGPT（原始版） [104]	176B	0.800	0.696*	0.516	0.654	0.577
	ChatGPT（对话版） [104]	176B	0.806	0.656*	0.548	0.665	0.601

相关策略需要地方死亡率数据。预测次国家级的痴呆症死亡人数，将有助于制定未来有针对性的卫生政策，以消除痴呆症方面的地理差异并实现国家卫生目标。

在这项研究中，我们针对两个地理区域层级——州级和县级的痴呆症死亡人数进行时间序列预测。痴呆症数据来自美国疾病控制与预防中心的流行病学研究广泛在线数据（CDC WONDER[3]），这是一个公开数据集。痴呆症死亡按照《国际疾病分类》第十次修订本（ICD-10）进行分类，

[3]https://wonder.cdc.gov/ucd-icd10.html

包括未特指的痴呆症（F03）、阿尔茨海默病（G30）、血管性痴呆（F01）以及未另行分类的其他神经系统退行性疾病（G31） [73]。

3.2.1 美国州级痴呆症时间序列预测。我们收集了1999年至2020年美国全部51个州的年度痴呆症死亡人数时间序列。1999年至2019年的时间序列用作训练数据，2020年的数据用作真实标签。我们在本任务中使用的预训练GPT模型与第3.1 节所用模型相同。与地理空间语义实验不同，我们在零样本设置下使用所有GPT模型，因为我们认为，历史时间序列数据足以让大语言模型进行预测。列表3 展示了本实验所用的一个提示示例，其中以加利福尼亚州为例。

只有51条时间序列，每条包含22个数据点，因此许多序列深度学习模型（如循环神经网络（RNN）和Transformer） [133] 都有在该数据集上过拟合的风险。因此，我们选择当前最先进的基于机器学习的时间序列预测模型——ARIMA（自回归积分滑动平均）作为全监督任务专用基线模型。我们使用1999年至2019年的数据，针对各州的时间序列分别训练ARIMA模型，并对2020年的数据进行预测。我们对所有ARIMA超参数进行调优，以获得最佳结果。此外，我们还使用持久性模型 [103, 107] 作为参照。持久性模型假设时间序列的未来值在当前时点与预测时点之间保持不变。在本研究中，我们将各州2019年的痴呆症死亡人数作为2020年预测值。

表格 2 展示了不同 GPT 模型与两种基线方法之间的模型性能比较。有趣的是，所有 GPT2 模型在所有评估指标上的表现都很差，甚至不如简单的持久性模型。这表明 GPT2 可能难以进行零样本时间序列预测。另一方面，GPT-3、InstructGPT 和两个 ChatGPT 模型表现尚可。尤为值得关注的是，尽管 InstructGPT 并未针对这一特定任务进行微调，但它在所有评估指标上的表现都优于最佳 ARIMA 模型。我们提出了两个假设，用以解释 InstructGPT 在时间序列预测任务中的出色表现：1）在大型文本语料库上训练后，InstructGPT 可能已具备进行零样本时间序列预测所需的智能，而这本质上是一个自回归问题。2）在大型文本语料库上训练期间，InstructGPT 和 GPT-3 可能接触过美国州级痴呆症时间序列数据。

虽然我们无法确定这些原因中哪一个是 InstructGPT 成功的主要因素，但这些结果令人十分鼓舞。与表1中的结果类似，两个 ChatGPT 模型的表现都不如 InstructGPT。更多实验分析请参见县级实验。

表 2. 各类 GPT 模型和基线方法在美国州级痴呆症时间序列预测任务上的评估结果。我们将所有模型分为四组：(A) 简单持久性模型；(B) 完全监督的机器学习模型，例如 ARIMA；(C) 使用大语言模型进行零样本学习。“(#Param)”表示大语言模型的可学习参数数量。各 GPT 模型的标记方式与表 1 相同。我们使用四种评估指标：MSE（均方误差）、MAE（平均绝对误差）、MAPE（平均绝对百分比误差）和 R^2。↑和↓表示各项指标中模型表现更优的方向。对于所有 GPT 模型，我们将1999年至2019年的时间序列信息编码到提示中，并让模型生成2020年的数据。

	模型	#参数	均方误差 ↓	平均绝对误差 ↓	平均绝对百分比误差 ↓	R^2 ↑
(A) 简单	持久性 [103, 107]	-	985,179	630	0.096	0.971
(B) 监督式机器学习	ARIMA [58]	-	562,768	462	0.067	0.984
(C) 零样本语言模型	GPT2 [115]	117M	44,635,055	4,898	0.955	-0.271
	GPT2-Medium [115]	345M	42,315,630	4,616	0.745	-0.209
	GPT2-Large [115]	774M	39,039,733	4,250	0.779	-0.132
	GPT2-XL [115]	1558M	35,355,840	3,912	0.709	-0.026
	GPT-3 [15]	175B	587,263	474	0.070	0.983
	InstructGPT [106]	175B	**387,413**	**365**	**0.055**	**0.989**
	ChatGPT（原始版） [104]	176B	1,143,675	623	0.121	0.967
	ChatGPT（对话版） [104]	176B	4,224,811	1,131	0.240	0.890

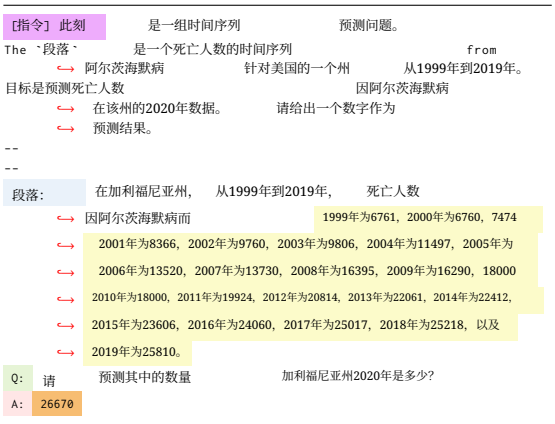


图3. 通过零样本学习使用大语言模型进行美国州级阿尔茨海默病时间序列预测。黄色区块：美国某州的历史时间序列数据。橙色框：InstructGPT 的输出。此处以加利福尼亚州为例，正确答案为29400。

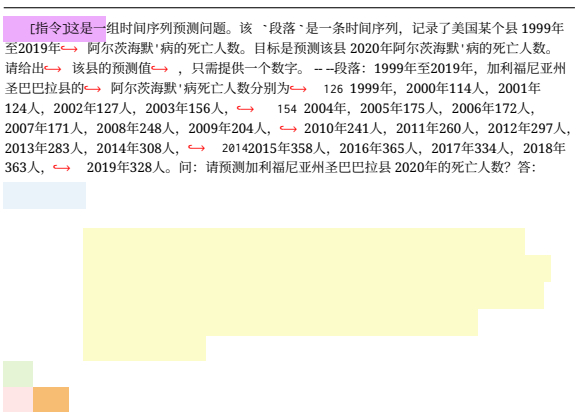


图4. 通过零样本学习使用大语言模型进行美国县级阿尔茨海默病时间序列预测。黄色区块：美国某县的历史时间序列数据。橙色框：InstructGPT 的输出。此处以加利福尼亚州圣巴巴拉县为例，正确答案为373。

3.2.2 美国县级痴呆症时间序列预测。在县级数据方面，我们使用了所有有可用数据的美国县的痴呆症死亡人数时间序列，共选取2447个美国县进行分析。我们仅考虑1999年至2020年间痴呆症年度死亡记录跨度超过四年的县。与第3.2.1节类似，我们利用截至给定年份的所有可用数据训练ARIMA模型并生成GPT提示词，然后预测下一年。我们采用与州级实验相同的一组GPT模型和基线方法进行县级实验。列表4 以加利福尼亚州圣巴巴拉县为例，展示了本实验中使用的一个提示词。

表3. 各种GPT模型和基线方法在美国县级痴呆症时间序列预测任务中的评估结果。我们使用与表2相同的模型集合和评估指标。

	模型	#Param	均方误差 ↓	平均绝对误差 ↓	平均绝对百分比误差 ↓	R^2 ↑
(A) 简单	持久性 [103, 107]	-	1,648	16.9	0.189	0.979
(B) 监督式机器学习	ARIMA [58]	-	1,133	15.1	0.193	0.986
（C）零样本大语言模型	GPT2 [115]	117M	77,529	92.0	0.587	-0.018
	GPT2-Medium [115]	345M	226,259	108.1	0.611	-2.824
	GPT2-Large [115]	774M	211,881	94.3	0.581	-1.706
	GPT2-XL [115]	1558M	162,778	99.8	0.627	-1.082
	GPT-3 [15]	175B	1,105	14.5	0.180	0.986
	InstructGPT [106]	175B	**831**	**13.3**	**0.179**	**0.989**
	ChatGPT（原始版） [104]	176B	4,115	23.2	0.217	0.955
	ChatGPT（对话版） [104]	176B	3,402	20.7	0.231	0.944

表3 比较了不同模型的结果。从这些结果中可以看出类似的发现。所有GPT2模型的表现都很差。然而，GPT-3和InstructGPT在所有评估指标上均优于表现最佳的ARIMA模型，而两个ChatGPT模型的表现则不如它们。在两个ChatGPT模型中，除平均绝对百分比误差外，ChatGPT（对话版）在所有指标上的表现都略优于ChatGPT（原始版）。

为了进一步了解每个模型预测误差的地理分布，我们在图 1 中展示了各模型在美国每个县的预测误差。在图中，红色表示相应模型的高估，而蓝色表示低估。此外，颜色的深浅表示预测误差的大小，颜色越深，误差越大。我们可以看到，持久性模型、ARIMA、GPT-3 和 InstructGPT 总体上表现出较好的预测性能。然而，预测百分比误差在不同美国县之间的分布并不均匀。由于持久性模型使用前一年的数据作为预测值，图1a表明，不同县的痴呆症死亡人数增长率存在不均衡现象。美国西南部近期痴呆症死亡人数有所增加，导致持久性模型低估了真实数据。目前的预测误差地图显示，GPT-3 和 InstructGPT 的误差分布在美国各县之间并不均匀，但尚不清楚这种分布不均是否源于模型中编码的地理偏差，还是痴呆症死亡人数增长率的空间异质性。仍需进一步分析以确定这些分布不均的原因。

从图1 中可以明显看出，所有 GPT2 模型都严重低估了痴呆症数据。为了解这一现象的原因，以及 GPT-3 和 InstructGPT 的优势，我们在表4中展示了不同 GPT 模型对美国四个县生成的答案。从表4可以看出，GPT2 经常重复提示中提供的信息，而不是生成新的预测。例如，在佐治亚州克拉克县和加利福尼亚州圣巴巴拉县的案例中，三个 GPT2 模型（即 GPT2-Medium、GPT2-Large 和 GPT2-XL）预测的数值都与 1999 年的数据相同。这表明这些模型严重依赖提示信息，而不是从时间序列数据中学习，这或许可以解释它们为何表现不如 GPT-3 和 InstructGPT 等其他模型。在另外两个县的案例中，GPT2 模型的预测结果差异很大。在大多数情况下，InstructGPT 和 ChatGPT（原始版）都生成一个数字作为预测，表明它们理解了预期执行的任务。唯一的例外是圣巴巴拉县的案例，ChatGPT（原始版）生成了一个包含合理预测的简短句子。然而，根据我们的评估，ChatGPT（原始版）的预测不如 GPT-3。有趣的是，在使用 ChatGPT 的

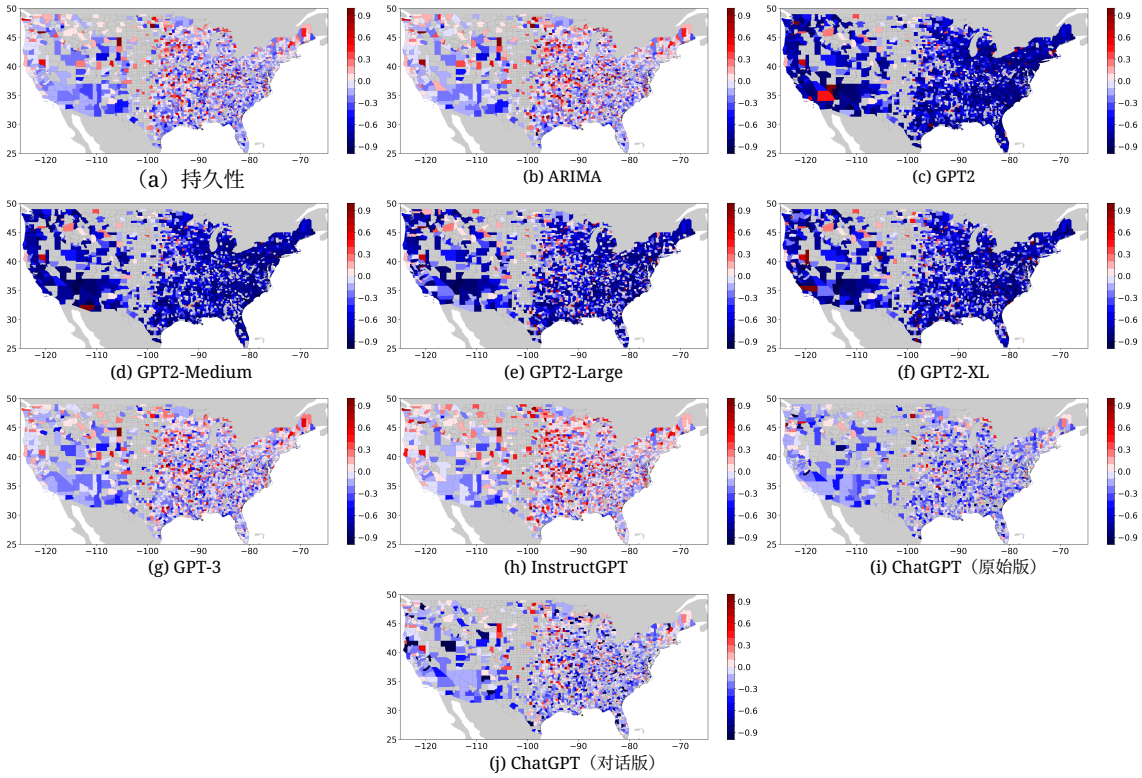


图1. 在美国县级痴呆症死亡人数时间序列预测任务中，各基线模型和 GPT 模型的预测误差地图。每个美国县的颜色表示各模型在该县的预测百分比误差 $PE = (Prediction - Label)/Label$。灰色县表示其1999年和2020年的痴呆症数据不可用。

对话情境中，即 ChatGPT（对话版），ChatGPT 通常会返回一个很长的句子。在纽约县的案例中，ChatGPT（对话版）无法给出预测，这表明 ChatGPT 在聊天机器人情境中很有用，但对于时间序列预测等其他任务而言，可能并非最佳选择。

3.3 城市地理学

第三组基础模型实验聚焦于城市地理学领域的研究问题。我们选取了两个具有代表性的任务：1）**城市功能任务**旨在根据某个地理区域内的兴趣点（POIs）预测该区域的城市功能 [51, 52, 102, 147, 152]；2）**城市感知任务**旨在根据街景图像（SVI）预测城市邻里特征（例如房价、安全性、噪声强度等级） [65, 154, 162]。由于这些任务涉及点数据、文本和图像等不同的数据模态，因此我们使用不同的基础模型来处理各项任务。

3.3.1 基于兴趣点的城市功能分类。第一项实验聚焦于根据地理区域内的兴趣点（POI）预测该区域的城市功能。这是城市地理学中的一项常见任务，旨在了解城市空间的结构 [51, 52, 102, 147, 152]。

表4。不同GPT模型针对美国四个县的县级痴呆症时间序列预测问题生成的答案。“County”和“Label”列分别表示美国县及其在2020年的真实痴呆症死亡人数。“Generated Answer”表示不同GPT模型生成的答案，其中加粗的数字是我们从这些答案中提取的预测值。

县	标签	模型	生成答案
佐治亚州克拉克县	79	GPT2	阿尔茨海默病的死亡人数在1999年为 **21** 。 , 2000年中的37 , 32 于 2001年 , 2002年的25个 , 2003年时29 国产成人 , 2004年为44 , 38英寸 2005, 2006年为41 , 2007年为52 , 2008年为46
		GPT2-Medium	2020年，佐治亚州克拉克县阿尔茨海默病死亡人数为**21**。
		GPT2-Large	阿尔茨海默病死亡人数：**21** 2020年，2021年37人，2022年32人，2023年25人，2024年29人，44人， 2025, 2026年38人 , 2027年42人, 42
		GPT2-XL	2020年的数字是：**21**。
		GPT-3	2020年佐治亚州克拉克县的阿尔茨海默病死亡人数预计为**81**。
		InstructGPT	**83**
		ChatGPT（原始版）	**78**
		ChatGPT（对话版）	根据现有历史数据，预测乔治亚州克拉克县阿尔茨海默病死亡人数 在2020年将为**81**。
纽约县，纽约州	835	GPT2	1999年阿尔茨海默病死亡人数为**91** , 2000年为121 , 2001年为118 , 2002年为155 , 214 于2003年 , 2004年中的243 , 2005年中的219 , 2006年中的294 , 2007年的296 , 2008年为364彩在线
		GPT2-Medium	纽约州纽约县阿尔茨海默病导致的死亡人数为**9,965**。
		GPT2-Large	2020年纽约州纽约县阿尔茨海默病的死亡人数为**91**。
		GPT2-XL	2020年中的数字是：**901**。
		GPT-3	2020年纽约州纽约县的阿尔茨海默病死亡人数为**932**。
		InstructGPT	**902**
		ChatGPT（原始版）	**855**
		ChatGPT（对话版）	根据所提供的历史数据，时间序列预测的一种常见方法是使用指数平滑法等方法， 或 ARIMA。然而，如果没有关于数据中特定模式或趋势的进一步信息，就很难 以准确预测2020年纽约州纽约县阿尔茨海默病死亡人数。 因此，我无法给出一个单一的预测数值。
加利福尼亚州旧金山县	592	GPT2	阿尔茨海默病导致的死亡人数为**1 ,000**。阿尔茨海默病的死亡人数为 2 ,000. 数量 阿尔茨海默病死亡人数为 3 ,000. 死亡人数来自
		GPT2-Medium	2020年阿尔茨海默病的死亡人数为**201** 。
		GPT2-Large	2020年阿尔茨海默病死亡人数为**574**。
		GPT2-XL	2020年中的数字是**201**。
		GPT-3	**571**
		InstructGPT	**590**
		ChatGPT（原始版）	**579**
		ChatGPT（对话版）	根据提供的历史数据，一种可能的预测方法是使用时间序列模型，例如 ARIMA 或 指数平滑法。如果不拟合这样的模型，一种简单的方法是假设趋势的幅度与之相近。 到近年来，并呈现类似水平的年度变化。采用这种方法，可以对死亡人数作出一种可能的预测。 2020年加利福尼亚州旧金山县阿尔茨海默病患者人数为**583**。
加利福尼亚州圣巴巴拉县	373	GPT2	阿尔茨海默病的死亡人数在1999年为 **126** 。 , 2000年中的114 , 2001年为124 , 2002年为127 , 2003年为156 , 2004年为154 , 2005年为175 , 2006年为172 , 171 于 2007年 , 248（2008年）
		GPT2-Medium	2020年，加利福尼亚州圣巴巴拉县阿尔茨海默病导致的死亡人数为**126** 。
		GPT2-Large	加利福尼亚州圣巴巴拉县阿尔茨海默病导致的死亡人数为：**126** 人（2020年）。
		GPT2-XL	2020年的数字是：**126**。
		GPT-3	预计2020年加利福尼亚州圣巴巴拉县阿尔茨海默病死亡人数约为**350**。
		InstructGPT	**345**
		ChatGPT（原始版）	。我预测2020年加利福尼亚州圣巴巴拉县阿尔茨海默病死亡人数为**356**。
		ChatGPT（续）	。根据所提供的历史数据，预测2020年圣塔阿尔茨海默病死亡人数为 加利福尼亚州芭芭拉县是**327**。

为定量评估大语言模型在这项城市功能预测任务中的表现，我们使用了来自中国深圳的兴趣点（POI）数据集，其中包含 303,428 个兴趣点和 5,461 个含有兴趣点的城市街区[26, 27, 159, 160]。我们将该数据集记为 $UrbanPOI5K$。图2 展示了兴趣点和区域的地理分布。真实数据来自北京大学城市景观基础数据集。该数据集提供了研究区域内十种城市功能类型的详细空间分布：森林、水域、未利用地、交通、绿地、工业、教育与政府、商业、居住和农业。为简化任务，我们将较少见的城市功能类型——森林、水域、未利用地、绿地和农业——合并为户外与自然这一功能类型。由此得到六种城市功能类型：(1) 居住；(2) 商业；(3) 工业；(4) 教育、医疗保健、市政、政府和文化；(5) 交通设施；(6) 户外与自然。总共有 5,344 个区域具有真实数据。我们将该数据集随机划分为训练集、验证集和测试集，比例为 60%:20%:20%。测试数据集用于评估不同模型的表现，而验证集仅用于监督基线方法。

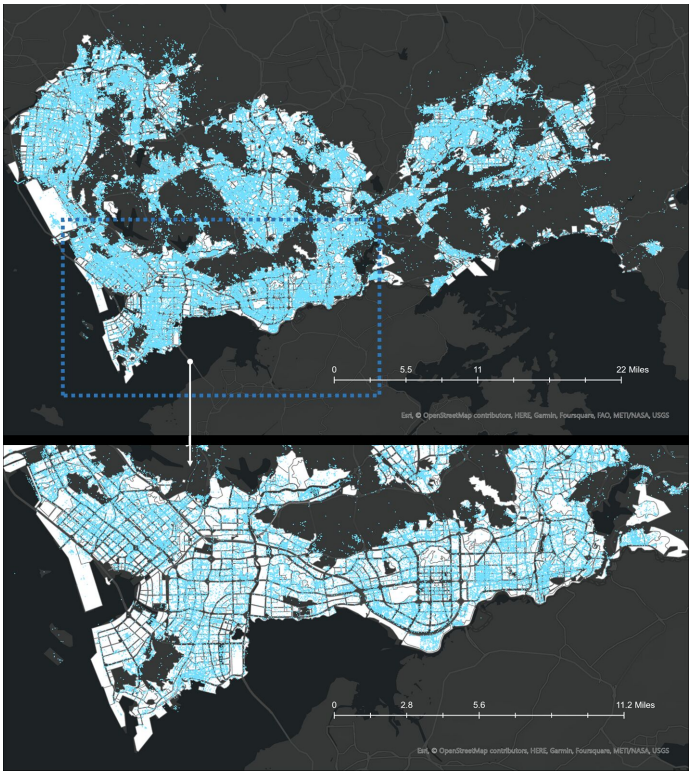


图2. *UrbanPOI5K*数据集中的兴趣点数据空间分布。

[说明] 共有六种 land use 类型：（1）居住，（2）商业，（3）工业，（4）教育，医疗 保健，市政，政府
↪ and 文化，（5）交通 设施，and （6）户外区域和 自然区域。
段落： 在这个城市区域中，有 128 个 兴趣点，包括 2 中餐馆，1 food 餐馆，2 酒店，2 公寓
↪ 酒店，1 日常 生活服务，1 移动通信商店，24家公司、1家物流公司、1家 real 房地产中介机构，1种彩票 零售商，
↪ 3 美容店，1 美甲，2 家理发店，4 网吧，3 浴缸按摩，2 体育场，4 训练 机构，1 药店，4
↪ 汽车销售，6家汽车服务店，2家汽车维修店，1家汽车租赁店，1家汽车零部件店，3家购物场所，5家商店，5 停车 lot，5 个停车位 lot
↪ 入口，2 个交通设施 设施，1 港口，1 道路交叉口，1 自动取款机，2 办公室 建筑物，2 居住 区域，7
↪ 建筑物，1 房地产，1 公园，1 工厂，7 行政机构，1 出入口，3 大门，6 便利店，4 住宅
↪ 建筑材料。
Q: 什么是 主要 land 用途类别 这个城市区域的?
A: 户外区域和 自然区域

段落： 在此城市区域中，有 17 个兴趣点，包括 1 家餐厅、3 个公共厕所、3 家殡葬服务机构，以及 2 个供步行和骑行使用的道路驿站，
↪ 1 海滩，2 停车场，2 道路交叉口，1 企业，2 行政机构。
Q: 什么是 the 主要 land 使用类别 这个城市区域的?
A: 户外与自然

列表 5. POI-街区 基于大语言模型的城市功能分类，例如 ChatGPT（原始版）。黄色方块：一个新城市的兴趣点统计。n 待分类。橙色框：ChatGPT（原始版）的输出。

为了使大语言模型能够处理此类任务，我们将城市区域内的兴趣点集合转换为一段文本，描述不同地方类型的兴趣点出现频率。然后，我们要求大语言模型根据该段落预测该区域的城市功能（尽管混合功能的城市区域很常见，我们在此要求预测占主导地位的功能）。清单5 展示了此任务的一个提示示例，其中包含一个段落—问题—答案三元组作为示范。通过此类提示调整的大语言模型正在进行预测

表5. 各种GPT模型和监督基线模型在 *UrbanPOI5K* 数据集上执行基于兴趣点的城市功能分类任务的评估结果。我们将模型分为三组：（A）基于监督学习的神经网络模型；（B）使用大语言模型进行零样本学习；（C）使用大语言模型进行单样本学习。我们使用准确率、加权精确率和加权召回率作为评估指标。由于加权F1 分数与准确率相同，因此不予列出。每组中的最佳模型均已突出显示。

	模型	准确率	精确率	召回率
(A) 监督式神经网络	Place2Vec [145, 152]	0.540	0.512	0.516
	HGI [52]	**0.584**	**0.568**	**0.563**
(B) 零样本大语言模型	GPT2 [115]	**0.318**	0.105	**0.158**
	GPT2-Medium [115]	0.025	0.102	0.040
	GPT2-Large [115]	0.005	0.001	0.002
	GPT2-XL [115]	0.001	0.108	0.002
	GPT-3 [15]	0.144	**0.448**	0.141
	ChatGPT（原始版） [104]	0.075	0.376	0.106
	ChatGPT（对话版） [104]	0.051	0.232	0.046
（C）单次提示大语言模型	GPT2 [115]	0.149	0.079	0.085
	GPT2-Medium [115]	0.317	0.104	0.156
	GPT2-Large [115]	0.057	0.083	0.021
	GPT2-XL [115]	**0.324**	0.105	0.159
	GPT-3 [15]	0.176	0.486	0.190
	ChatGPT（原始版） [104]	0.195	**0.524**	**0.245**
	ChatGPT（对话版） [104]	0.093	0.451	0.085

在单样本设置下。对于零样本设置，我们只需从提示中移除这一段落—问题—答案三元组。我们使用不同规模的 GPT2、GPT-3 和两个 ChatGPT 模型，在零样本和单样本设置下执行此任务。为了进行比较，我们采用了两种有监督学习神经网络基线方法：

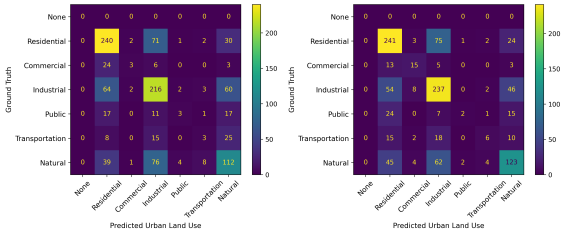
- **Place2Vec**：我们首先按照 Place2Vec 方法 [145]学习兴趣点类别嵌入。然后，给定包含 K 个兴趣点的城市区域，我们将每个兴趣点转换为相应的 Place2Vec 嵌入，并像 Zhai 等人 [152] 那样进行均值池化，获得区域嵌入。随后，将得到的街区嵌入输入一个单隐藏层多层感知机（MLP），在 *UrbanPOI5K*训练数据集上监督学习其城市功能。

- **HGI**：这是一种基于兴趣点类别语义表征的监督城市区域嵌入方法 [52]。随后，将 HGI 区域嵌入输入采用相同配置的多层感知机，以预测主要城市功能。

表5 展示了所有模型在测试数据集上的评估结果。此外，我们在图3、4和5中展示了两个基线模型、7 个零样本 GPT模型和 7 个单样本 GPT 模型的混淆矩阵。我们可以看到：

- 在零样本设置下，GPT-3 在所有 GPT模型中取得最高的精确率，但仍不如 HGI 模型。

- 有趣的是，在零样本设置中，最小的 GPT2 实现了最佳准确率和召回率，这有悖直觉。原因见图4a。GPT2 几乎将所有街区都预测为“居住”，这占真实数据的 30+%。



(a) Place2Vec (b) HGI

图 3. Place2Vec 和 HGI 在表 5中 A 组的混淆矩阵（针对 $UrbanPOI5K$ 数据集）。

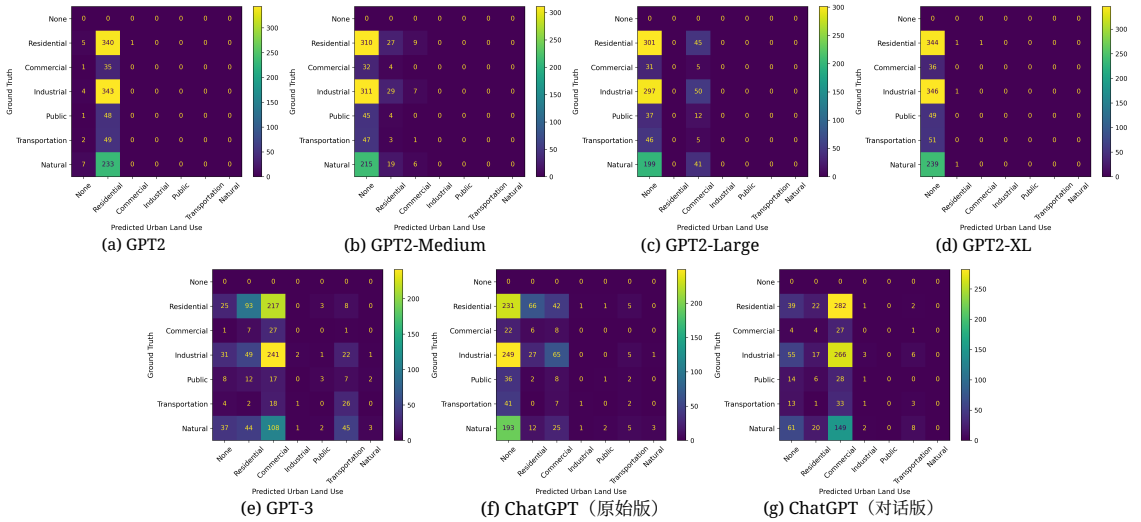
• 在单样本设置下，ChatGPT（原始版）在精确率和召回率两方面均优于所有 GPT模型。其精确率达到 52.4%，仅比 HGI 低 4.4%。图 5f 中的混淆矩阵也表明，ChatGPT（原始版）在所有城市功能类别上的表现都相当不错。
• 在单样本设置下，GPT2-XL 的准确率最高。然而，图5d 显示，GPT2-XL 对“居住”类别存在严重偏向。

这些实验结果凸显了使用大语言模型进行城市功能分类所面临的挑战。其表现不佳主要有两个原因:

• 兴趣点最初用于在线地图服务中的搜索，其本身就包含了丰富的商业场所信息，例如餐厅和酒店。相反，与日常生活联系不太紧密的场所，例如工厂，往往会缺失。在这方面，深圳是一座以工业为主导的城市，而真实数据显示，工业区域的数量远远多于商业区域。然而，鉴于输入给大语言模型的是与商业相关的兴趣点，大语言模型往往会预测大量区域属于商业区域。

• 此外，大语言模型无法获取兴趣点的空间分布，而兴趣点的空间分布会显著影响基于兴趣点的城市功能预测，因为不同的兴趣点空间分布会产生不同的空间交互模式，进而形成不同的城市功能。尽管监督基线方法 Place2Vec 和 HGI 在地点类型嵌入的无监督训练阶段均是基于兴趣点分布进行学习的，但我们无法将兴趣点的空间分布信息提供给大语言模型。将兴趣点集合转换为图像同样行不通。这是因为许多兴趣点会聚集在市中心区域，较大的像素尺寸会导致大量兴趣点落入同一个像素中。另一方面，较小的像素尺寸又会使城市空间图像过于庞大，无法由其他深度图像编码器处理。此外，采用较小像素尺寸的城市空间图像所包含的信息会非常稀疏，难以供图像编码器学习。换言之，我们需要采用专门的神经网络架构来直接处理点数据（以及折线数据和多边形数据）。因此，**有必要将各种地理空间矢量数据的编码架构纳入其中，例如位置编码 [91, 95], 折线编码 [118, 149], 和多边形编码技术[98]， 以推进 GeoAI基础模型开发**。我们将在第 4.6 节中对此进行详细讨论。

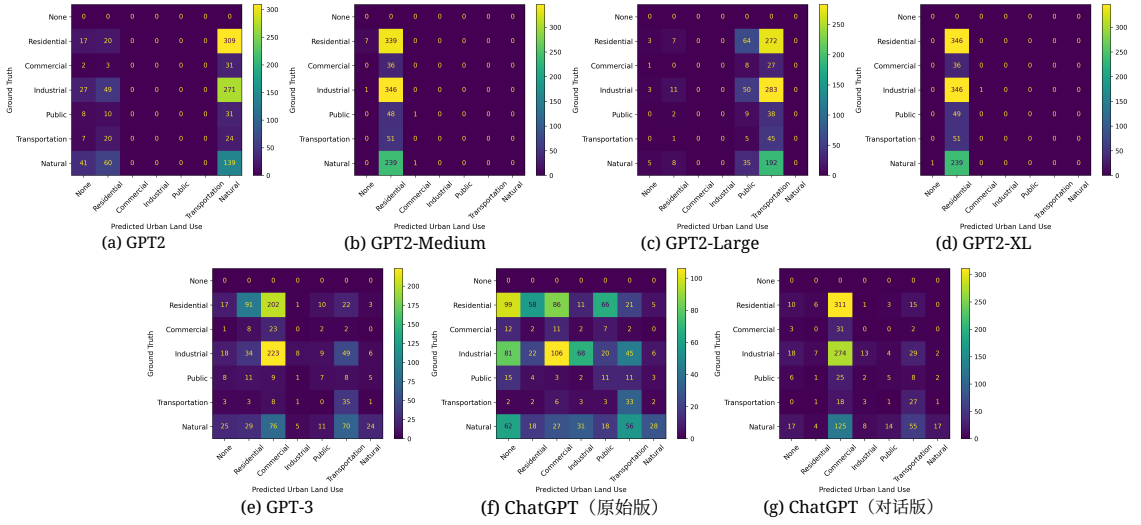
3.3.2 基于街景图像的城市噪声强度分类。街景图像（SVI）广泛用于许多城市地理学研究，以了解城市街区的不同特征，例如安全性 [154], 美观度、富裕程度 [80], 令人沮丧的程度 [154], 房价 [65], 噪声强度等级 [162], 可达性 [35], 等。它已成为补充遥感图像的重要数据来源。



(a) GPT2 (b) GPT2-Medium (c) GPT2-Large (d) GPT2-XL

(e) GPT-3 (f) ChatGPT（原始版） (g) ChatGPT（对话版）

Fig<escape>. 4. 所有 GPT 模型（表 5 中的 B 组）在 *UrbanPOI5K* 数据集上的零样本设置下的混淆矩阵</escape> ng.



(a) GPT2 (b) GPT2-Medium (c) GPT2-Large (d) GPT2-XL

(e) GPT-3 (f) ChatGPT（原始版） (g) ChatGPT（对话版）

图 5。 所有GPT模型（表中的C组 5）在 *UrbanPOI5K* 数据集上的单样本设置混淆矩阵 .

在本研究中，我们使用了由 Zhao 等人近期开发的街景图像噪声强度数据集，将其作为具有代表性的城市感知任务。该数据集包含从新加坡采集的 579 张街景图像。噪声强度分数（介于 0 和 1 之间）基于人工调查收集。有关该数据集的详细说明，请参阅他们的 Github[4]。由于声音强度分数并非一个普遍认可的指标，而是由 Zhao 等人定义的指标 [162],，因此，对于在通用网络数据上训练的视觉基础模型（如 OpenCLIP [54] 和 BLIP [82] ）而言，直接预测此类分数颇具挑战性。因此，我们将每张街景图像原始的噪声强度分数离散化为四个类别：非常安静（0 - 0.25）、安静（0.25 - 0.50）、嘈杂（0.50 - 0.75）和非常

[4]https://github.com/ualsg/Visual-soundscapes

嘈杂（0.75 - 1.00）。我们将此数据集记为 *SingaporeSVI*579。图6 展示了每个噪声强度类别中的一些街景图像示例。我们将 *SingaporeSVI*579 随机划分为 50% 的训练集和 50% 的测试集，其中测试集用于评估不同的卷积神经网络和基础模型。

由于此前实验中使用的所有 GPT 模型（GPT-4 除外）均为纯语言模型，无法处理图像等数据模态。因此，对于基于街景图像的噪声强度预测任务，我们选择了最新的高性能开放式视觉语言基础模型（VLFM），包括 OpenCLIP [54], 、 BLIP [82], 和 OpenFlamingo-9B [11]。尽管还存在性能更强大的视觉语言基础模型，例如 DeepMind 的 Flamingo-9B [6], 、KOSMOS-1 [49], 和 GPT-4 [105], ，但这些模型尚未开放获取，也尚未提供 API访问[5]。我们对各个 VLFM 的设置描述如下：

- **OpenCLIP-L**：我们使用以 LAION-5B 的 LAION-2B 英语子集预训练的 OpenCLIP [54] ViT L/14 模型，作为较小规模的 OpenCLIP 模型。我们从 Huggingface[7]下载预训练模型。
- **OpenCLIP-B**：我们使用以 LAION-5B 的 LAION-2B 英语子集训练的 OpenCLIP [54] ViT-bigG/14 模型，作为较大规模的 OpenCLIP 模型。预训练模型来自 Huggingface[8]。
- **BLIP**：我们使用 Huggingface[9] 提供的预训练 BLIP-2 模型 [81] ，该模型由类CLIP图像编码器、查询变换器（Q-Former）和大型语言模型（Flan T5-xl）组成。
- **OpenFlamingo-9B**：我们使用 Huggingface[10] 提供的预训练 OpenFlamingo-9B 模型 [11] ，该模型由图像编码器（CLIP ViT-L/14 [54]）和大型语言模型（LLaMA-7B [132]）组成。

所有视觉语言基础模型均在 *SingaporeSVI*579 的测试集上以零样本设置进行评估。由于不同的视觉语言基础模型需要不同的图像输入格式，并且期望使用不同风格的文本提示词，我们将在下文介绍每个视觉语言基础模型的零样本流程：

- **OpenCLIP-L** 和 **OpenCLIP-B**：我们首先使用如下形式的文本模板，将四个噪声强度类别名称编码为四个文本嵌入：“一个噪声强度为 *[NOISE_INTENSITY_CLASS]*”。随后，给定一张街景图像，我们使用 OpenCLIP ViT 图像编码器将其编码为图像嵌入。计算该图像嵌入与四个类别文本嵌入之间的余弦相似度，并选取相似度最高的类别作为预测结果。

- **BLIP**：给定一张街景图像，我们使用如下形式的提示：“该区域的噪声强度是多少，是 *1.*非常安静、*2.* 安静、*3.*嘈杂，还是 *4.*非常嘈杂？”，以指示 BLIP 的语言编码器预测其噪声强度类别。

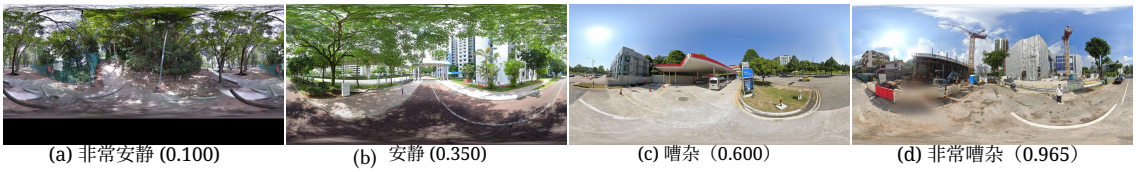
- **OpenFlamingo-9B**：我们使用如下形式的提示：“噪声强度分为四个等级：*1.* 非常安静，*2.* 安静，*3.* 嘈杂，或 *4.* 非常嘈杂。 <图像>该区域的噪声强度为”，以指示 OpenFlamingo-9B 预测给定图像的噪声强度。这里，“<image>”表示一个图像标记，编码器采用 CLIP ViT-L/14。

我们选择四种卷积神经网络模型（卷积神经网络）作为替代基线方法，与这些视觉语言基础模型进行比较：AlexNet [74], ResNet18 [37], ResNet50 [37], 和 DenseNet161 [48]。所有卷积神经网络模型的权重首先使用 Place365 预训练权重进行初始化 [164], ，然后仅在 *SingaporeSVI*579 训练数据集上通过全监督微调其最终的 softmax 层。由于训练数据量非常有限，我们选择这种线性探测方法，而不是对整个卷积神经网络架构进行完全微调。

[5]请注意，在我们提交本文时，GPT-4 API 仍不支持视觉问答。 [6]https://laion.ai/blog/laion-5b/ [7] https://huggingface.co/laion/CLIP-ViT-L-14-laion2B-s32B-b82K [8] https://huggingface.co/laion/CLIP-ViT-bigG-14-laion2B-39B-b160k [9] https://huggingface.co/Salesforce/blip2-flan-t5-xl [10] https://huggingface.co/openflamingo/OpenFlamingo-9B

表 6. 各种视觉语言基础模型和基线方法在基于城市街景图像的新加坡街景图像噪声强度分类数据集上的评估结果579 [162]。我们将模型分为两组：(A) 经过监督微调的卷积神经网络（卷积神经网络）；(B) 使用视觉语言基础模型进行零样本学习。我们使用准确率和加权 F1 分数作为评估指标。每组的最佳分数以突出显示。

	模型	#Param	准确率	F1
(A) 监督微调卷积神经网络	AlexNet [74]	58M	0.452	0.405
	ResNet18 [37]	11M	0.493	**0.442**
	ResNet50 [37]	24M	**0.500**	0.436
	DenseNet161 [48]	27M	0.486	0.382
（B）零样本基础模型	OpenCLIP-L [54, 113, 127]	427M	0.128	0.089
	OpenCLIP-B [54, 113, 127]	2.5B	0.169	0.178
	BLIP [81, 82]	3.9B	**0.452**	**0.405**
	OpenFlamingo-9B [11]	8.3B	0.262	0.127

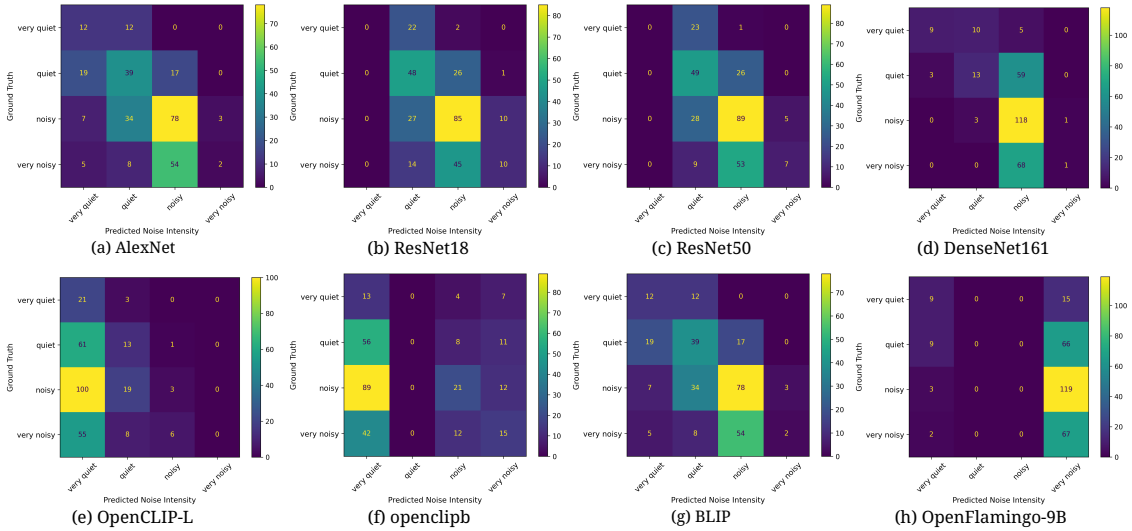


(a) 非常安静 (0.100) (b) 安静 (0.350) (c) 嘈杂（0.600） (d) 非常嘈杂（0.965）

图 6. 一些街景图像示例 *SingaporeSVI579* 数据集中的街景图像示例。图像说明指出了该图像所属的噪声强度类别 s
图像所属的 类别，括号中的数字表示 Zhao 等人 [162] 给出的原始噪声强度分数。

表6 比较了不同微调后的 CNN 模型与四种零样本视觉语言基础模型的表现。结果表明，在零样本学习设置下，BLIP 在所有视觉语言基础模型中取得了最高的准确率和加权 F1 分数。BLIP 的表现与 AlexNet 相当，但仍略逊于表现最佳的模型 ResNet18 和 ResNet50。为了进一步了解不同模型对各噪声强度类别的分类准确率，我们在图中展示了所有模型的混淆矩阵 7。可以看出，OpenCLIP-L、OpenCLIP-B 和 OpenFlamingo-9B 的预测存在明显偏差。OpenCLIP-L 和 OpenCLIP-B 倾向于将大多数街景图像归类为“非常安静”，而 OpenFlamingo-9B 则将大多数图像归类为“非常嘈杂”。另一方面，只有 BLIP 对全部四种噪声强度类别的预测较为均衡且合理，与这些微调后的 CNN 模型类似。

这些结果令人鼓舞：零样本 BLIP 的表现可与经过微调的模型相媲美。从图7g 中可以看出，BLIP 对目标城市区域的噪声强度等级有大致的判断，例如，它会将大多数“非常嘈杂”的区域误分类为“嘈杂”。这表明，BLIP 对噪声强度等级的理解采用了不同的尺度。例如，新加坡的人类受访者标注为“非常嘈杂”的地方，在 BLIP 看来可能还达不到“非常嘈杂”的程度，因为它可能见过许多噪声大得多的城市区域。就此而言，BLIP 总体上能够胜任这项城市感知任务。与此同时，我们也认识到，大多数开放式视觉语言基础模型仍不够强大，无法将视觉特征与城市研究中那些重要却微妙的语义和概念联系起来。例如，当模型看到图6d中的施工现场时，我们希望视觉语言基础模型能够预测这是一个非常嘈杂的街区。看到图6d中的高植被覆盖率时，视觉语言基础模型应将这一视觉特征与语言空间中的“安静”概念联系起来。本研究强调，当前的视觉语言基础模型具备一定能力，能够根据视觉输入理解城市街区的特征。然而，在纯语言任务上，其能力总体上仍不及当前的大语言模型。此外，我们认为，城市感知任务作为城市地理学中的经典任务，比当前常见的视觉问答任务



(a) AlexNet (b) ResNet18 (c) ResNet50 (d) DenseNet161

(e) OpenCLIP-L (f) openclipb (g) BLIP (h) OpenFlamingo-9B

图 7. 所有基线方法和视觉语言基础模型在 *SingaporeSVI579* 数据集上的混淆矩阵。

在视觉语言基础模型研究中常用 [49, 112]，部分原因是这类任务具有一定的主观性，且标注数据集稀缺。这进一步凸显了地理空间人工智能领域的基础模型研究所面临的独特挑战。

3.4 遥感

我们的最终实验聚焦于一项典型的遥感（RS）任务——遥感图像场景分类。我们选择了一个广泛使用的航空图像场景分类数据集，*AID* [144]，该数据集包含 1 万个场景和 30 种航空场景类型。这些数据采集自 Google Earth 影像。有关该数据集的详细说明，请参阅 Xia 等人 [144] 。*AID* 未提供官方数据集划分，因此我们采用分层抽样将数据集划分为训练集和测试集，训练集占 80%，测试集占 20%，确保两者具有相似的场景类型标签分布。

与第 3.3.1节中的街景图像分类任务类似，我们使用四种卷积神经网络模型（即 AlexNet、ResNet18、ResNet50 和 DenseNet161）以及四种视觉语言基础模型（即 OpenCLIP-L、OpenCLIP-B、BLIP 和 OpenFlamingo-9B）。对于所有卷积神经网络模型，其权重首先使用 ImageNet-V1 预训练权重进行初始化，并在 *AID* 训练数据集上通过全监督微调其最终 softmax 层。对于视觉语言基础模型，其模型性能很大程度上取决于语言模型组件能否正确理解每种遥感图像场景类型的语义。然而，*AID* 中的许多遥感图像场景类型名称都含糊不清，例如“中心”和“商业”。我们发现，如果保留原始场景类型名称，OpenCLIP 等模型不会将任何遥感图像归入这两种类型。因此，我们将“中心”改为“剧院”（尽管这只能部分涵盖该类别的语义），并将“商业”改为“商业区”，在提示中使用这些名称。使用此类提示的模型标记为“(*Updated*)”，而“(*Origin*)”表示提示中使用了 *AID* 中的原始遥感图像场景类型名称。我们在零样本学习设置下评估所有视觉语言基础模型。遵循第 3.3.1节中的街景图像分类任务，我们在 *AID* 数据集上使用类似的提示格式。

表7 总结了四种微调后的 CNN 模型和零样本视觉语言基础模型的实验结果。可以看出，AlexNet 在所有 CNN 模型中取得了最高的准确率和 F1 分数。令人惊讶的是，OpenCLIP-L (*Updated*) 取得了

表 7. 遥感图像场景分类数据集上各种视觉-语言基础模型和基线方法的评估结果， *AID* [144]。我们使用与表 6 中相同的模型集合。“(*Origin*)”表示我们使用 *AID* 中原始的遥感图像场景类别名称来填充提示，而“(*Updated*)”则表示我们对部分类别名称进行了更新，以改善其对基础模型的语义解释。我们使用准确率和 F1 分数作为评估指标。

	模型	#Param	准确率	F1
监督微调卷积神经网络	AlexNet [74]	58M	**0.831**	**0.827**
	ResNet18 [37]	11M	0.752	0.730
	ResNet50 [37]	24M	0.757	0.738
	DenseNet161 [48]	27M	0.818	0.807
零样本基础模型	OpenCLIP-L (*Origin*) [54, 113, 127]	427M	0.708	0.688
	OpenCLIP-L (*Updated*) [54, 113, 127]	427M	**0.710**	**0.698**
	OpenCLIP-B (*Origin*) [54, 113, 127]	2.5B	0.699	0.668
	OpenCLIP-B (*Updated*) [54, 113, 127]	2.5B	0.705	0.686
	BLIP (*Origin*) [82]	2.5B	0.500	0.473
	BLIP (*Updated*) [82]	2.5B	0.520	0.494
	OpenFlamingo-9B [11]	8.3B	0.206	0.154

在所有视觉语言基础模型中取得了最高的准确率和 F1 分数。我们发现，在这项任务中，模型更大并不一定会带来更好的结果。例如，最大的模型 OpenFlamingo-9B 的准确率仅为 0.206。一个可能的原因是，这些较大的视觉语言基础模型可能没有在训练数据中接触过遥感图像，而训练数据通常包含从网络爬取的一般图像和文本。另一方面，OpenCLIP 的预训练数据明确包含卫星图像[54]。然而，BLIP 和 OpenFlamingo-9B 均未说明它们在预训练阶段是否使用了遥感图像。请注意，街景图像与互联网图像十分相似，而互联网图像被广泛用于视觉语言基础模型的预训练。另一方面，遥感图像（例如卫星图像和无人机图像）在视觉上与互联网照片截然不同；互联网照片大多由消费者使用数码相机在地面拍摄。如果 BLIP 和 OpenFlamingo-9B 的视觉编码器没有在遥感图像上进行预训练，那么它们提取的特征就无法与具有相似语义的文本特征很好地对齐，导致其在*AID* 数据集上的表现不佳。我们的研究强调，对视觉语言基础模型使用多样化的视觉输入（包括遥感图像）进行预训练，对于提升其在遥感任务中的表现至关重要。

另一个重要的观察是，提示词中蕴含的语义在决定模型性能方面发挥着关键作用。例如，当使用原始场景类型名称“center”时，通常没有任何模型能够理解其背后模糊的含义。然而，只需将“center”改为“theater”，就能帮助 OpenCLIP 正确检索相关的遥感场景，尽管“theater”并不是描述这一类别的完美名称。尽管如此，这一简单的修改充分体现了在使用基础模型执行地理空间任务时，选择具有明确表达力的提示词的重要性。

与表5中的结果相比，表7 中的实验结果突显了遥感图像面临的独特挑战。我们将在第4.4节详细讨论基础模型在遥感领域的改进。

4 面向地理空间人工智能的多模态基础模型

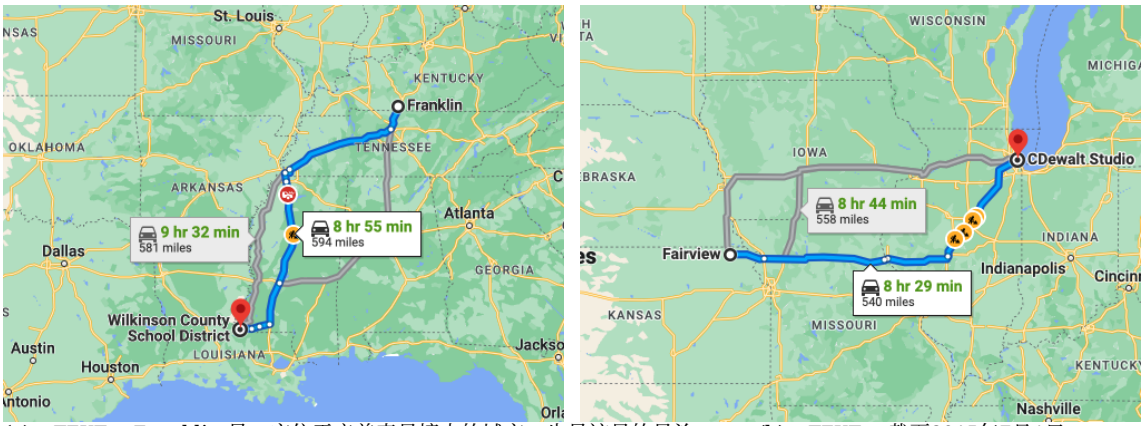
第3 节探讨了将现有基础模型应用于不同地理空间领域的各类任务时的效果。我们可以看到，许多大语言模型的表现优于全监督任务专用机器学习/深度学习模型，并在若干地理空间任务上取得了出人意料的良好效果，例如地名识别、位置描述识别以及痴呆症时间序列预测。然而，在其他地理空间任务（即两个经测试的城市地理学任务和一个遥感任务）中，尤其是在涉及多种数据模态（例如点数据、街景图像、遥感

图像等）的任务中，现有基础模型的表现仍不如特定任务模型。事实上，许多地理空间任务的一个独特特点是涉及多种数据模态，例如文本数据、知识图谱、遥感图像、街景图像、轨迹以及其他地理空间矢量数据。这给地理空间人工智能基础模型的开发带来了重大挑战。因此，本节将讨论各类数据模态特有的挑战，并提出一种面向未来地理空间人工智能的潜在框架，该框架将利用多模态基础模型。

4.1 地理文本数据

尽管表1中的结果颇具前景，大语言模型在更复杂的地理空间语义任务上仍然表现吃力，例如地名消歧/地理解析 [7, 33, 134]和地理问答（地理空间问答） [94, 97],，因为大语言模型无法以扎根于现实世界的方式进行（隐式）空间推理。作为一个具体示例，我们展示 GPT-3 在地理解析任务中的不足。我们从 Ju2016 数据集中选取两个示例，要求 GPT-3 同时完成：1）识别地名；以及 2）预测其地理坐标。提示见列表 6，地理解析结果则在图 8中可视化。我们发现，在这两种情况下，GPT-3 都能正确识别地名，但预测的坐标却距离真实值 500+ 英里。此外，我们注意到，只要生成的地理坐标发生轻微的空间偏移，GPT-3 对这组新坐标的对数概率就会显著降低。换言之，大语言模型生成坐标的概率并不遵循托布勒地理学第一定律 [130]。GPT-3 还会生成无效的纬度/经度坐标，这表明现有大语言模型仍远不能妥善处理复杂的数值推理和空间推理任务。

图 9 展示了大语言模型在回答涉及空间关系的地理问题时，另一个令人不满意的例子。在这个例子中，ChatGPT 生成的答案称 Monore 位于佐治亚州雅典的北部，但实际上它位于雅典的西南部。这个例子表明，大语言模型并未充分理解空间关系的语义。出现这一错误的原因可能是，ChatGPT 回答这个空间关系问题时，依据的是对其内部文本知识记忆的搜索，而非进行空间推理。解决这一问题的一个潜在方法是使用地理空间知识图谱[18, 166],，以引导大语言模型进行显式的空间关系计算。我们将在下一节进一步讨论这一点。



(a) [TEXT]: Franklin 是一座位于辛普森县境内的城市，也是该县的县治，..... (b) [TEXT]: 截至2015年7月1日，Fairview 市的人口为 260 人。..... 图 8. GPT-3 在 Ju2016 数据集上的地理解析示例，对比预测坐标（图钉标记）与真实值坐标（起始点）。文本中识别出的地名已加下划线。

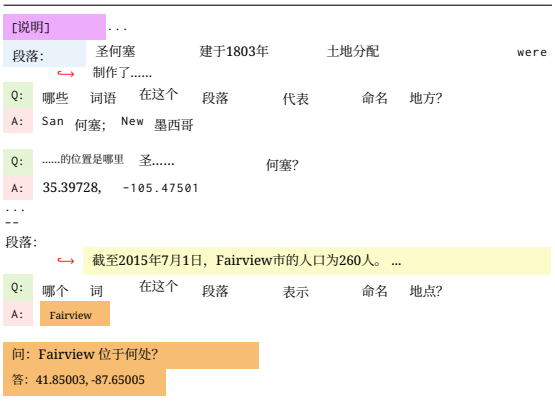
4.2 地理空间知识图谱

尽管大语言模型具备出色的端到端预测和生成能力，但它们生成的内容可能未能充分涵盖事实知识，甚至可能包含不符合事实的信息。为解决这一问题，知识图谱（KG）可以作为有效的信息来源，补充大语言模型。知识图谱具有事实性，因为其中的信息通常提取自可靠来源，并由人工编辑人员进行后处理，以进一步确保移除错误内容。作为一种重要的领域知识图谱，地理空间知识图谱（GeoKG），例如 *GeoNames* [2], *LinkedGeoData* [10], *YAGO2* [42], *GNIS-LD* [122], *KnowWhereGraph* [56], *EVKG* [111],等，通常由权威数据源和空间数据库生成。例如，GNIS-LD 是基于美国地质调查局的地理名称信息系统（GNIS）构建的。这确保了这些地理空间数据的真实性。

具体而言，开发能够联合考虑文本数据和地理空间知识图谱的地理空间人工智能多模态基础模型，可以带来多方面优势。首先，从模型角度看，可以将（地理空间）知识图谱整合到大语言模型的预训练或微调过程中，例如，检索知识实体的嵌入以进行上下文表示学习 [110], 、融合知识实体与文本信息 [38, 158], 、设计侧重于重构知识实体 [161] 和三元组 [148]的学习目标。其次，从数据角度看，地理空间知识图谱可以提供带有上下文的语义和时空知识，以助力提示工程或数据生成，例如，利用知识图谱中的上下文信息丰富提示词 [14, 142] ，并将知识图谱三元组转换为特定领域的自然语言文本语料库 [1]。第三，从应用角度看，可以将地理空间知识图谱中的事实转换为自然语言，以增强文本生成 [150], ，用于（地理）问答 [29, 96]和对话系统 [141]等场景。最后，从推理角度看，地理空间知识图谱通常提供地理实体的空间足迹，使大语言模型能够像神经符号机那样执行显式空间推理 [84]。这有助于避免图9中所示的错误。

4.3 街景图像

部分 3.3.1 展示了现有视觉语言基础模型在基于街景的地理空间任务上的有效性。然而，表格 6 表明



列表 6：使用大语言模型进行地理解析，例如 GPT-3。黄色方框：待进行地理解析的文本片段。橙色方框：GPT-3 的输出。

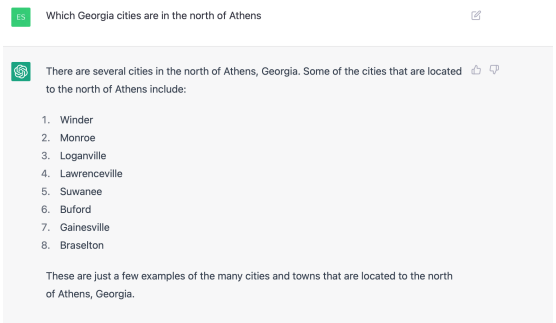


图 9：ChatGPT 在回答有关拓扑关系的地理问题时出错的一个例子。在此例中，Monore 不在佐治亚州雅典的北部，而是在其西南部。

这表明，如果我们希望开发用于地理空间人工智能的基础模型，就需要考虑城市感知任务的一些独特特征。

虽然街景图像与常见视觉语言任务中使用的自然图像相似，但两者有一个主要区别：常见视觉语言任务通常关注图像中的事实知识（例如“这张图像里有多少辆汽车”），而城市感知任务通常涉及人类对图像的高层感知，例如根据街景图像判断一个街区的安全性、贫困程度、美观度和声音强度。与事实知识相比，这类高层感知知识很难估计，而且相应的标签也相当稀少。此外，许多感知概念都模糊且主观，这增加了此类任务的难度。因此，为了开发能够在各种城市感知任务中达到最先进性能的地理空间人工智能基础模型，我们需要开展一些领域研究，为每个城市感知概念提供明确的定义，并开发一些标注数据集，用于地理空间人工智能基础模型的预训练。

4.4 遥感

随着计算机视觉技术的进步，深度视觉模型已成功应用于不同类型的遥感（RS）任务，包括图像分类／回归 [12, 123], 土地覆盖分类 [12],和目标检测[76]。与通常基于 RGB 图像开展的常规视觉任务不同，遥感任务基于来自不同传感器的多光谱／高光谱图像。现有的大多数遥感研究侧重于使用来自特定传感器的数据，为特定遥感任务训练单一模型 [76]。研究人员通常使用相同的训练数据集比较不同模型的性能，并根据准确率统计数据决定模型的实现方案。然而，我们看到，计算机视觉领域中的基础模型（FM）呈现出发展趋势，例如 CLIP [112], Flamingo-9B [6]，这些模型需要进一步发展，以应对遥感任务所面临的独特挑战。第 3.4 节中的遥感实验表明，当前视觉语言基础模型与特定任务深度模型之间仍存在性能差距。为了弥合这一差距，并开发能够在各种遥感任务上实现最先进性能的地理空间人工智能基础模型，我们需要考虑遥感图像和任务的独特性。

除了具备**任务无关**特性外，遥感基础模型还应满足以下要求：1）**传感器无关**：能够无缝推理来自不同传感器、具有不同空间或光谱分辨率的遥感图像；2）**时空感知**：能够处理遥感图像的时空元数据，并针对图像地理定位和目标跟踪等任务执行地理空间推理；3）**环境不变**：能够在各种背景环境条件和景观结构下，分解并分离感兴趣物体的光谱特征。近期相关进展包括地理感知遥感模型 [12] 或自监督／无监督遥感模型 [12, 123],，这些模型均与任务无关。然而，我们尚未开发出能够满足所有这些特性的遥感任务基础模型。

总之，应致力于利用遥感技术开发地理空间人工智能基础模型，以应对气候变化带来的紧迫环境挑战。这需要构建复杂模型，不再局限于图像分类，而是转向对生态系统功能进行建模，例如森林结构、碳封存、城市热环境、沿海洪涝和湿地健康状况。传统上，遥感被广泛用于研究这些现象，但研究方式局限于特定地点和特定传感器。传感器无关、时空感知且环境不变的基础模型，有望改变我们对这些复杂环境现象趋势和行为的理解。

4.5 轨迹与人类移动性

轨迹各种传感器序列智能位置元序数设备和地球空间人工智能中的另一种重要数据类型 .

设备）生成的数字轨迹数据日益增多，加之深度学习方法的进步，使得新型地理空间人工智能模型得以发展，用于建模人类出行模式。这些模式对于城市管理、交通服务等至关重要。利用深度学习建模人类动态，通常涉及四项典型任务： [90], 轨迹生成 [22, 118], 起讫点（OD）流量生成 [129, 146], 人口流入/流出流量预测 [59, 83], 下一位置预测 [85, 119]。

为了开发支持人类移动分析的地理空间人工智能基础模型，我们需要考虑以下几个方面：1) 预训练并生成任务无关的轨迹嵌入 [101, 136],，以表示来自各种轨迹 [85]的高层次移动语义（例如时空感知、路线和位置序列）；2) 轨迹的上下文感知对比学习：人类移动会受到其工作类型、周围建成环境、交通基础设施以及许多其他时空和环境因素的约束 [90,128, 137]；地理空间人工智能基础模型应能够将轨迹与各种上下文表示关联起来，例如道路网络（例如 Road2Vec [86], [21]）、兴趣点构成或土地利用类型 [155], 、城市形态 [20], 和人口分布 [53]；3) 在训练此类地理空间人工智能基础模型时，应保护用户地理隐私 [70]，因为轨迹数据可能会揭示个人的敏感位置，例如住宅和个人出行。利用密码学或差分隐私 [5] 以及联邦学习框架的隐私保护技术，可以纳入轨迹 [119]的地理空间人工智能基础模型训练过程。

4.6 地理空间矢量数据

开发用于地理空间人工智能的基础模型面临的另一项重大挑战，是地理空间矢量数据的复杂性。这类数据几乎应用于所有地理信息系统和制图平台中。例如，第 3.2 节讨论的美国州级和县级痴呆症数据（多边形数据）、第 3.3.1 节介绍的城市 POI 数据（点和多边形数据）、制图折线数据 [149], 建筑物轮廓数据 [98], 道路网络（由点和折线构成）以及许多其他数据。与自然语言处理和计算机视觉中的文本（一维）或图像（二维）结构良好、更适合常见神经网络架构不同，矢量数据呈现出更复杂的数据结构，包括点、折线、多边形和网络 [95]。因此，开发一种能够无缝编码或解码不同类型矢量数据的基础模型尤其具有挑战性。

值得注意的是，近期开发的位置编码 [91, 95], 、折线编码 [118, 149], 和多边形编码技术[98] 可以被视为此类模型的基础构建模块。此外，由于对地理空间矢量数据进行编码（例如，具有地理空间感知能力的图像分类[91]）或解码（例如，地理解析 [134]），以及执行空间推理（例如，地理空间问答 [97]），是大多数地理空间人工智能任务不可或缺的组成部分，因此，开发面向矢量数据的基础模型是迈向地理空间人工智能多模态基础模型的关键一步。这一点也使地理空间人工智能基础模型区别于其他领域现有的基础模型。

4.7 面向地理空间人工智能的多模态基础模型

除了这些数据模态之外，地理空间人工智能领域还经常研究其他数据集，例如带有地理标签的视频、空间社交网络、传感器网络等。鉴于这些多样化的数据模态，如何开发一种能够最佳整合所有这些数据的地理空间人工智能多模态基础模型，仍然是一个值得探讨的问题。

当我们考察现有的多模态基础模型，例如 CLIP [112],DALL·E2 [117], MDETR [63],VATT [3],BLIP [82], DeepMind Flamingo [6], KOSMOS-1 [49], 时，可以看到以下通用架构：1) **首先使用独立的嵌入模块对不同模态的数据进行编码**（例如，用于文本的 Transformer 和用于图像的 ViT [112]）；2)（可选）通过拼接**融合不同模态的表征**；3)（可选）使用**更多 Transformer 层**进行跨模态推理，从而实现一定程度的语义对齐，例如让“医院”一词关注医院的图片；4) 使用**生成式或判别式预测模块**对不同模态进行预测，以实现自监督训练。

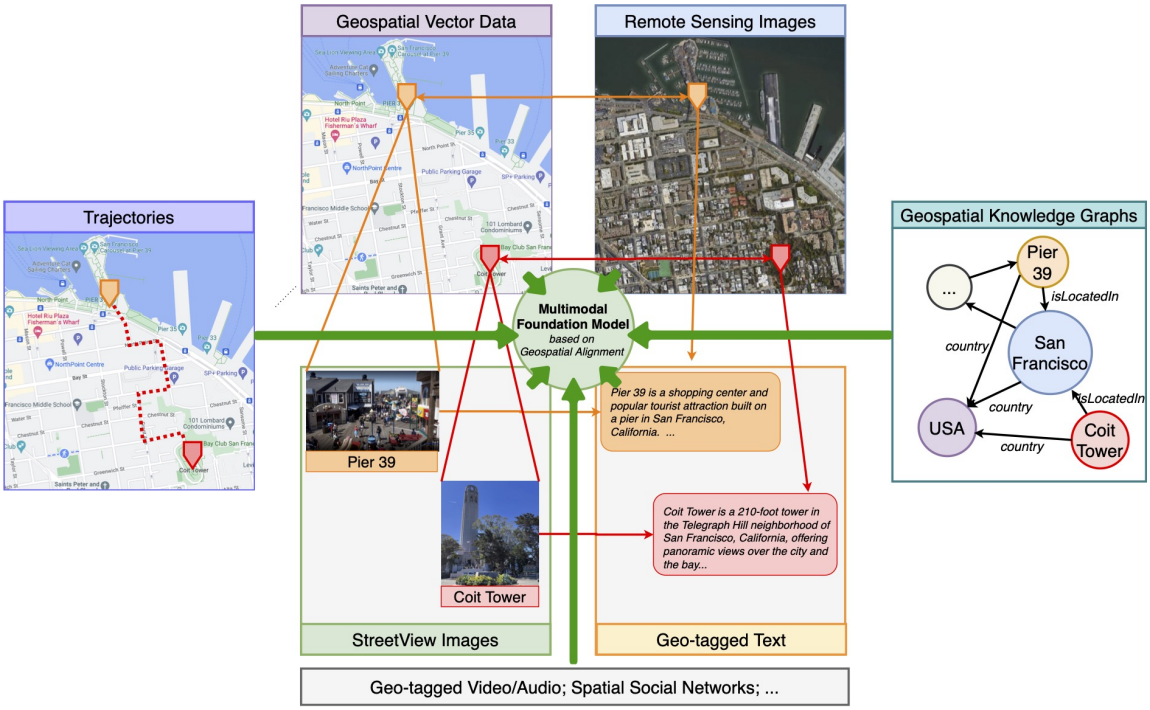


图10. 一种通过地理空间关系实现不同数据源间对齐的多模态基础模型

这些架构的一个薄弱之处是缺乏与地理空间矢量数据的整合，而地理空间矢量数据是空间推理的基础，有助于地理空间人工智能中多种模态之间的对齐。这被认为是地理空间人工智能任务的核心和关键。因此，我们提议将步骤 2 替换为**对齐**不同模态的表征（例如，带有地理标签的文本和遥感图像），方法是为其表征添加位置编码[91]，然后再进行融合。图10 阐释了这一思路。带有地理标签的文本数据、街景图像、遥感图像、轨迹和地理空间知识图谱，都可以通过其地理足迹（矢量数据）轻松对齐。这种模型的关键优势在于能够实现跨模态的空间推理和知识迁移。

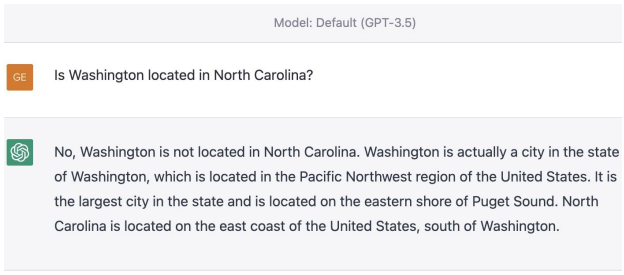
5 风险与挑战

尽管近期取得了进展，但随着更多先进基础模型的发布，若干挑战也逐渐显现 [163]。首先，随着基础模型规模不断增大，需要提高训练和微调这些模型的计算效率。其次，由于越来越多的大语言模型并未开源，在无法访问其内部参数的情况下，将知识融入这些模型变得颇具挑战。第三，随着大语言模型越来越多地部署在远程第三方环境中，保护用户隐私变得愈发重要。

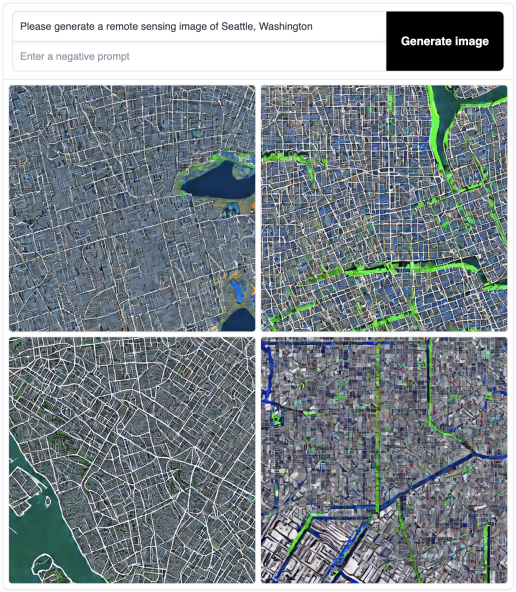
除了基础模型普遍面临的这些挑战之外，在地理空间人工智能基础模型的开发过程中，也存在许多独特的挑战和风险。

5.1 地理保真度

许多基础模型因生成不准确且具有误导性的结果而受到批评 [50, 132]。在地理空间场景中，生成忠实反映地理信息的结果对于几乎所有地理空间人工智能任务都尤为重要。除第9 节中的图4.1外，图11 展示了 ChatGPT 和 Stable Diffusion 生成的两个地理信息不准确的结果。图 11a 中，



(a) ChatGPT 生成的地理位置不准确结果



(b) Stable Diffusion 生成的地理位置不准确结果

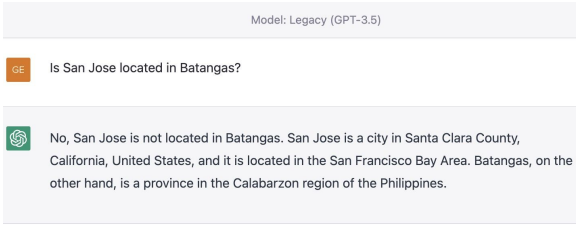
图 11. 不同语言与视觉基础模型生成的一些地理位置不准确的结果。(a) 未能正确生成预期答案“北卡罗来纳州华盛顿”。此外，华盛顿州没有名为华盛顿的城市。华盛顿州最大的城市应为西雅图。(b) Stable Diffusion 生成的遥感图像未能呈现正确的地理布局，例如道路网络、水体等。

预期答案应为“北卡罗来纳州华盛顿”[11]。然而，ChatGPT 表示北卡罗来纳州没有华盛顿。此外，华盛顿州最大的城市应为西雅图，而且该州并没有名为华盛顿的城市。图11b 展示了 Stable Diffusion 生成的 4 幅遥感图像[12]。尽管这些图像看起来与卫星图像相似，但很容易看出它们是伪造的遥感图像，因为图中地理要素的布局显然不属于世界上任何一座城市。事实上，生成忠实的遥感图像是一项热门且重要的遥感任务 [34, 39]，其中几何精度对于下游任务非常重要。

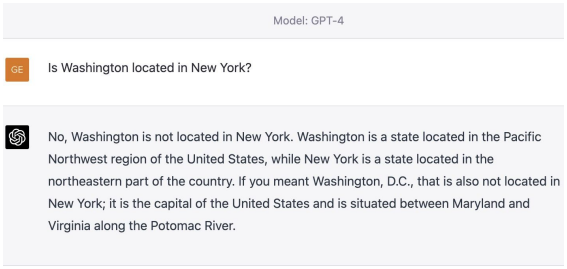
5.2 地理偏差

众所周知，基础模型可能会放大数据中已有的社会不平等和偏差 [13, 132, 157]。对于地理空间人工智能而言，一个尤为重要但常被人工智能研究忽视的因素是地理偏差 [87], 。例如，Zilong等人 [87] 指出，目前所有地理解析器都存在严重的地理偏差，倾向于数据丰富地区。当前的大语言模型也存在同样的问题。图12 展示了两个例子：ChatGPT 和 GPT-4 都因继承自这些模型的地理偏差而生成不准确的结果。与美国加利福尼亚州圣何塞相比，八打雁省圣何塞[13] 在许多文本语料库中是一个知名度较低的地名。同样，与美国华盛顿州和美国华盛顿哥伦比亚特区相比，纽约州华盛顿[14] 也是一个知名度较低的地名。因此，ChatGPT 和 GPT-4 都会错误地解读这些地名。与特定任务模型相比，基础模型受到的地理偏差影响更大，原因有：1）训练数据以大规模方式收集，因此很可能由

[11] https://en.wikipedia.org/wiki/Washington,_North_卡罗来纳州
[12] https://huggingface.co/spaces/stabilityai/stable-diffusion [13] https://en.wikipedia.org/wiki/San_Jose,_八打雁省 [14] https://en.wikipedia.org/wiki/Washington,_New_York

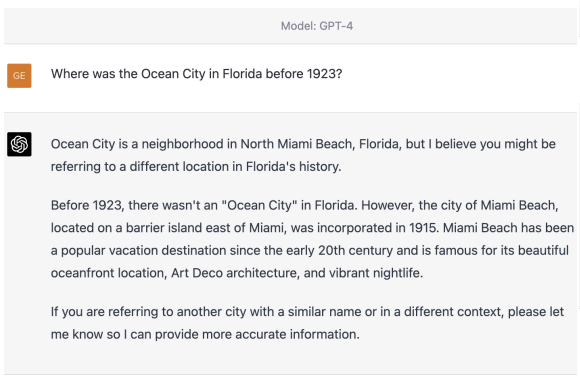


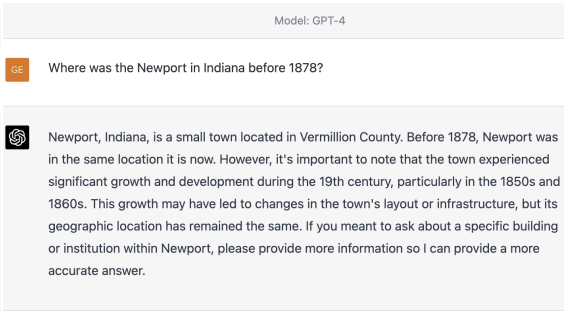
(a) ChatGPT因地理bias



(b) GPT-4因地理偏差而生成的不准确结果

图 12. 由于地理偏差，ChatGPT 和 GPT-4 生成的一些不准确结果。(a) 与八打雁省圣何塞相比，美国加利福尼亚州圣何塞是一个更知名的地名。因此，ChatGPT 对名称“圣何塞”的解读有误，导致答案错误。(b) 美国华盛顿州和美国华盛顿哥伦比亚特区都是两个以“华盛顿”命名的知名地点。正确答案“纽约州华盛顿”知名度较低，因此导致答案不准确。





(a) 不准确 ChatGPT因时间偏差而生成的结果 (b) GPT-4因时间偏差而生成的不准确结果

图 13. GPT-4 因时间偏差而生成的一些不准确结果。(a) 佛罗里达州的弗拉格勒海滩 在 1913 年至 1923 年期间曾被命名为 大洋城，而如今 佛罗里达州的大洋城 被用来称呼佛罗里达州的另一个地方。GPT-4 未能识别这一情况，因此返回了错误答案。(b) 印第安纳州的喷泉城 在 1834 年至 1878 年期间曾被命名为 纽波特，而如今 纽波特 被用来称呼另一个城市，即位于维米利恩县的印第安纳州纽波特。GPT-4 未能正确回答这一问题。

代表性过高的社区或地区；2）大量的可学习参数和复杂的模型结构使模型解释和去偏变得更加困难；3）基础模型的地理偏差很容易被下游所有适配模型继承 [13]，，从而对社会造成更大的危害。这表明亟需设计适当的（地理）去偏框架。

5.3 时间偏差

与地理偏差类似，基础模型也会受到时间偏差的影响，因为与历史地理实体相比，当前地理实体的训练数据要丰富得多。时间偏差也会导致结果不准确。图13展示了两个例子。在这两种情况下，历史地点的名称都被用于指代附近的其他地点。由于过度依赖偏向当前地理知识的预训练数据，GPT-4 无法回答这两个问题。时间偏差和地理偏差是地理空间人工智能基础模型发展过程中亟待解决的关键挑战。

5.4 空间尺度

地理信息可以用不同的空间尺度来表示，这意味着在不同的地理空间人工智能任务中，同一地理现象／对象可能具有截然不同的空间表示（点与多边形）。例如，城市交通预测模型必须将旧金山（SF）表示为复杂的多边形，而地理解析器通常将 SF 表示为单个点。由于基础模型面向各种不同的下游任务开发，因此需要能够处理具有不同空间尺度的地理空间信息，并根据给定的下游任务推断应采用的空间尺度。开发这样的模块是构建有效地理空间人工智能基础模型的关键组成部分。

5.5 泛化能力与空间异质性

地理空间人工智能面临的一个尚未解决的问题是，如何在空间上实现模型的泛化能力（“可复现性” [31]），同时仍允许模型捕捉空间异质性。面对具有不同空间尺度的地理空间数据，我们希望基础模型能够学习一般的空间趋势，同时记住位置特定细节。这种泛化能力是否会在下游地理空间人工智能任务中引入不可避免的内在模型偏差？这些记忆的局部信息是否会导致全球预测问题的预测面过于复杂？随着大规模训练数据的使用，这一问题可能会进一步加剧，因此需要谨慎处理。

6 结论

本文探讨了为地理空间人工智能开发多模态基础模型（基础模型）的前景与挑战。我们将现有大语言模型和视觉语言基础模型作为零样本或少样本学习器的表现，与特定任务的全监督最先进模型进行比较，展示基础模型的潜力。比较涵盖地理空间语义、健康地理学、城市地理学和遥感等多个地理空间子领域中的七项任务。在一些纯语言地理空间任务中，大语言模型作为零样本或少样本学习器，能够胜过特定任务的全监督模型；但在其他地理空间任务上，现有基础模型的表现仍不及特定任务的全监督模型，尤其是涉及多种数据模态的任务（例如基于兴趣点的城市功能分类、基于街景图像的城市噪声强度分类，以及遥感图像场景分类）。我们认识到，为地理空间人工智能开发基础模型的主要挑战在于地理空间任务具有多模态性。在讨论各类地理空间数据模态所面临的独特挑战后，我们提出了对新型地理空间人工智能多模态基础模型的设想：该模型应基于不同数据模态之间通过地理空间关系实现的对齐进行预训练。最后，我们讨论了此类模型所面临的一些独特挑战与风险。

参考文献

[1] Oshin Agarwal、Heming Ge、Siamak Shakeri 和 Rami Al-Rfou。2021年。用于知识增强语言模型预训练的基于知识图谱的合成语料生成。载于*2021*年计算语言学协会北美分会年会论文集：人类语言技术。计算语言学协会，3554–3565。[2] Dirk Ahlers。2013年。GeoNames 地名录数据准确率评估。载于第七届地理信息检索研讨会论文集。74–81。[3] Hassan Akbari、Liangzhe Yuan、Rui Qian、Wei-Hong Chuang、Shih-Fu Chang、Yin Cui 和 Boqing Gong。2021年。Vatt：用于原始视频、音频和文本多模态自监督学习的 Transformer。神经信息处理系统进展 34（2021年），24206–24221。[4] Igor Akushevich、Arseniy P Yashkin、Anatoliy I Yashin 和 Julia Kravchenko。2021年。阿尔茨海默病及相关痴呆症死亡率的地理差异。美国老年医学会杂志 69，8（2021年），2306–2315。[5] Mohammad Al-Rubaie 和 J Morris Chang。2019年。隐私保护机器学习：威胁与解决方案。*IEEE*《安全与隐私》17，2（2019年），49–58。[6] Jean-Baptiste Alayrac、Jeff Donahue、Pauline Luc、Antoine Miech、Iain Barr、Yana Hasson、Karel Lenc、Arthur Mensch、Katie Millican、Malcolm Reynolds、Roman Ring、Eliza Rutherford、Serkan Cabi、Tengda Han、Zhitao Gong、Sina Samangooei、Marianne Monteiro、Jacob Menick、Sebastian

Borgeaud、Andy Brock、Aida Nematzadeh、Sahand Sharifzadeh、Mikolaj Binkowski、Ricardo Barreira、Oriol Vinyals、Andrew Zisserman 和 Karen Simonyan。2022年。Flamingo：一种用于少样本学习的视觉语言模型。arXiv abs/2204.14198（2022年）。[7] Beatrice Alex 等。2015年。调整爱丁堡地理解析器以适用于历史地理定位。《人文与艺术计算国际期刊》9，1（2015年）。[8] 阿尔茨海默病协会等。2021年。改变阿尔茨海默病的发展轨迹：2025年之前的治疗如何挽救生命并节省资金。2015年。网址：https://www.alz. org/media/Documents/changing-the-trajectory-r. pdf *[访问日期：2018年-07-18][*WebCite 缓存 ID 710WNv2LM*]*（2021年）。[9] 阿尔茨海默病协会等。2022年。阿尔茨海默病事实与数据。《超越正常衰老：了解轻度认知障碍》。阿尔茨海默病协会。[10] Sören Auer、Jens Lehmann 和 Sebastian Hellmann。2009年。LinkedGeoData：为数据网络添加空间维度。载于《语义网—ISWC 2009：第八届国际语义网会议，ISWC 2009》，2009年10月25日至29日，美国弗吉尼亚州尚蒂伊。会议论文集第8卷。施普林格，731–746。[11] Anas Awadalla、Irena Gao、Joshua Gardner、Jack Hessel、Yusuf Hanafy、Wanrong Zhu、Kalyani Marathe、Yonatan Bitton、Samir Gadre、Jenia Jitsev、Simon Kornblith、Pang Wei Koh、Gabriel Ilharco、Mitchell Wortsman 和 Ludwig Schmidt。2023年。OpenFlamingo。https://doi.org/10.5281/zenodo. 7733589[12] Kumar Ayush 等。2021年。地理感知自监督学习。载于 CVPR 2021。2021年 IEEE/CVF 国际计算机视觉大会（ICCV），10181–10190。[13] Rishi Bommasani、Drew A Hudson、Ehsan Adeli、Russ Altman、Simran Arora、Sydney von Arx、Michael S Bernstein、Jeannette Bohg、Antoine Bosselut、Emma Brunskill 等。2021年。基础模型的机遇与风险。arXiv 预印本 arXiv:2108.07258（2021年）。[14] Ryan Brate、Minh-Hoang Dang、Fabian Hoppe、Yuan He、Albert Meroño-Peñuela 和 Vijay Sadashivaiah。2022年。通过融入知识图谱的提示词改进语言模型预测。载于知识图谱深度学习研讨会（DL4KG@ISWC 2022）。[15] Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等。2020年。语言模型是少样本学习器。《神经信息处理系统进展》33（2020年），1877–1901。[16] Marshall Burke、Anne Driscoll、David B Lobell 和 Stefano Ermon。2021年。利用卫星影像理解并促进可持续发展。《科学》371，6535（2021年），eabe8628。[17] Ling Cai、Krzysztof Janowicz、Gengchen Mai、Bo Yan 和 Rui Zhu。2020年。交通Transformer：捕捉时间序列的连续性与周期性以进行交通预测。《地理信息系统汇刊》24，3（2020年），736–755。[18] Ling Cai、Krzysztof Janowicz、Rui Zhu、Gengchen Mai、Bo Yan 和 Zhangyu Wang。2022年。HyperQuaternionE：一种用于定性时空推理的双曲嵌入模型。《地理信息学》（2022年），1–39。[19] Serina Chang、Emma Pierson、Pang Wei Koh、Jaline Gerardin、Beth Redbird、David Grusky 和 Jure Leskovec。2021年。新冠肺炎移动网络模型揭示不平等现象并为重新开放提供参考。《自然》589，7840（2021年），82–87。[20] Wangyang Chen、Abraham Noah Wu 和 Filip Biljecki。2021年。利用深度学习对城市形态进行分类：城市活力应用。《计算机、环境与城市系统》90（2021年），101706。[21] Yile Chen、Xiucheng Li、Gao Cong、Zhifeng Bao、Cheng Long、Yiding Liu、Arun Kumar Chandran 和 Richard Ellison。2021年。稳健的道路网络表征学习：交通模式与出行语义相结合。载于第30届 ACM信息与知识管理国际会议论文集。211–220。[22] Seongjin Choi、Jiwon Kim 和 Hwasoo Yeo。2021年。Trajgail：利用生成对抗模仿学习生成城市车辆轨迹。《交通研究C辑：新兴技术》128（2021年），103091。[23] Grant DeLozier、Benjamin Wing、Jason Baldridge 和 Scott Nesbit。2016年。从历史文本中创建新型地理定位语料库。载于 LAW-X 2016。188–198。[24] Jia Deng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei。2009年。ImageNet：一个大规模分层图像数据库。载于2009年 IEEE计算机视觉与模式识别会议。IEEE，248–255。[25] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly、Jakob Uszkoreit 和 Neil Houlsby。2021年。一张图像胜过16×16个词：用于大规模图像识别的Transformer。ICLR（2021年）。[26] Shouji Du、Shihong Du、Bo Liu 和 Xiuyuan Zhang。2019年。利用上下文信息从超高分辨率图像中提取大规模城市功能区：一种多尺度分割方法。《遥感》11，16（2019年），1902。[27] Shouji Du、Shihong Du、Bo Liu、Xiuyuan Zhang 和 Zhijia Zheng。2020年。融合遥感图像与开放社会数据的大规模城市功能区制图。《地理信息科学与遥感》57，3（2020年），411–430。[28] Amna Elmustafa、Erik Rozi、Yutong He、Gengchen Mai、Stefano Ermon、Marshall Burke 和 David Lobell。2022年。利用卫星影像、建筑物轮廓和深度模型理解非洲农村地区的经济发展。载于第30届国际地理信息系统进展会议论文集。ACM，1–4。[29] Angela Fan、Claire Gardent、Chloé Braud 和 Antoine Bordes。2019年。利用局部知识图谱构建将序列到序列模型扩展至多文档输入。载于2019年自然语言处理实证方法会议暨第9届自然语言处理国际联合会议。[30] Jenny Rose Finkel、Trond Grenager 和 Christopher D Manning。2005年。通过吉布斯采样将非局部信息融入信息抽取系统。载于第43届计算语言学协会年会论文集（ACL’05）。363–370。

[31] Michael F Goodchild 和 Wenwen Li。2021年。社会科学和环境科学中的跨时空重复验证必须是弱的。《美国国家科学院院刊》118，35（2021年）。[32] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。2020年。生成对抗网络。ACM 通讯 63，11（2020年），139–144。[33] Milan Gritta、Mohammad Taher Pilehvar 和 Nigel Collier。2018年。哪个墨尔本？利用地图增强地理编码。载于 ACL 2018。1285–1296。[34] Xiaolin Han、Huan Zhang、Jing-Hao Xue 和 Weidong Sun。2021年。一种光谱—空间联合光谱超分辨率方法及其在 HJ-1A 卫星图像中的应用。IEEE 地球科学与遥感快报 19（2021年），1–5。[35] Kotaro Hara、Jin Sun、Robert Moore、David Jacobs 和 Jon Froehlich。2014年。Tohme：利用众包、计算机视觉和机器学习检测路缘坡道。载于第27届 ACM 用户界面软件与技术年度研讨会论文集（美国夏威夷州檀香山）（UIST ’14）。计算机协会，美国纽约州纽约，189–204。https://doi.org/10.1145/2642918.2647403[36] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。2016年。用于图像识别的深度残差学习。载于 IEEE 计算机视觉与模式识别会议论文集。770–778。[37] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。2016年。用于图像识别的深度残差学习。载于 IEEE 计算机视觉与模式识别会议论文集。770–778。[38] Lei He、Suncong Zheng、Tao Yang 和 Feng Zhang。2021年。Klmo：通过细粒度关系增强知识图谱的预训练语言模型。载于计算语言学协会：EMNLP 2021论文集。4536–4542。[39] Yutong He、Dingjie Wang、Nicholas Lai、William Zhang、Chenlin Meng、Marshall Burke、David Lobell 和 Stefano Ermon。2021年。通过条件像素合成实现卫星影像的时空超分辨率。神经信息处理系统进展 34（2021年），27903–27915。[40] Danny Hernandez、Jared Kaplan、Tom Henighan 和 Sam McCandlish。2021年。迁移学习的缩放定律。arXiv:2102。01293 [cs.LG][41] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。2020年。去噪扩散概率模型。神经信息处理系统进展 33（2020年），6840–6851。[42] Johannes Hoffart、Fabian M Suchanek、Klaus Berberich 和 Gerhard Weikum。2013 年。YAGO2：一个基于维基百科、经过空间和时间增强的知识库。人工智能 194 （ 2013 年），28–61。[43] Jordan Hoffmann、Sebastian Borgeaud、Arthur Mensch、Elena Buchatskaya、Trevor Cai、Eliza Rutherford、Diego de Las Casas、Lisa Anne Hendricks、Johannes Welbl、Aidan Clark、Tom Hennigan、Eric Noland、Katie Millican、George van den Driessche、Bogdan Damoc、Aurelia Guy、Simon Osindero、Karen Simonyan、Erich Elsen、Jack W. Rae、Oriol Vinyals 和 Laurent Sifre。2022年。训练计算最优的大语言模型。CoRR abs/2203.15556（2022年）。[44] Matthew Honnibal 和 Ines Montani。2017年。spaCy 2：利用 Bloom 嵌入、卷积神经网络和增量解析实现自然语言理解。即将刊出 7，1（2017年），411–420。[45] Yingjie Hu，等。2020年。自然灾害期间人们如何描述地点：对哈维飓风推文的分析。载于 GIScience 2020。[46] Yingjie Hu。2018年。地理文本数据与数据驱动的地理空间语义。《地理学指南针》12，11（2018年），e12404。[47] Yingjie Hu 等。2014年。利用 DBpedia 的结构化数据改进基于维基百科的短文本地名消歧。载于 GIR Workshop 2014。1–8。[48] Gao Huang、Zhuang Liu、Laurens Van Der Maaten 和 Kilian Q Weinberger。2017年。密集连接卷积网络。载于 IEEE 计算机视觉与模式识别会议论文集。4700–4708。[49] Shaohan Huang、Li Dong、Wenhui Wang、Yaru Hao、Saksham Singhal、Shuming Ma、Tengchao Lv、Lei Cui、Owais Khan Mohammed、Qiang Liu 等。2023年。语言并非你所需的一切：使感知与语言模型对齐。arXiv 预印本 arXiv:2302.14045（2023年）。[50] Shaohan Huang、Li Dong、Wenhui Wang、Yaru Hao、Saksham Singhal、Shuming Ma、Tengchao Lv、Lei Cui、Owais Khan Mohammed、Qiang Liu 等。2023年。语言并非你所需的一切：使感知与语言模型对齐。arXiv 预印本 arXiv:2302.14045（2023年）。[51] WEIMING HUANG、Lizhen Cui、Meng Chen、Daokun Zhang 和 Yao Yao。2022年。利用保留语义的兴趣点嵌入估计城市功能分布。《国际地理信息科学杂志》36，10（2022年），1905–1930。[52] WEIMING HUANG、Daokun Zhang、Gengchen Mai、Xu Guo 和 Lizhen Cui。2023年。利用兴趣点和层次图信息最大化学习城市区域表示。《ISPRS摄影测量与遥感杂志》196（2023年），134–145。[53] Xiao Huang、Di Zhu、Fan Zhang、Tao Liu、Xiao Li 和 Lei Zou。2021年。通过深度学习从卫星影像感知人口分布：模型选择、邻近效应与系统性偏差。IEEE 地球观测与遥感应用精选主题期刊 14（2021年），5137–5151。[54] Gabriel Ilharco、Mitchell Wortsman、Ross Wightman、Cade Gordon、Nicholas Carlini、Rohan Taori、Achal Dave、Vaishaal Shankar、Hongseok Namkoong、John Miller、Hannaneh Hajishirzi、Ali Farhadi 和 Ludwig Schmidt。2021年。OpenCLIP。https://doi.org/10.5281/zenodo.5143773 如果您使用此软件，请按以下格式引用。[55] Krzysztof Janowicz、Song Gao、Grant McKenzie、Yingjie Hu 和 Budhendra Bhaduri。2020年。GeoAI：用于地理知识发现及其他领域的显式空间人工智能技术。《国际地理信息科学杂志》34，4（2020年），625–636。[56] Krzysztof Janowicz、Pascal Hitzler、Wenwen Li、Dean Rehberger、Mark Schildhauer、Rui Zhu、Cogan Shimizu、Colby K Fisher、Ling Cai、Gengchen Mai 等。2022年。Know、Know Where、KnowWhereGraph：一个高度互联的跨领域知识图谱和地理信息增强服务栈，面向环境智能应用。《AI Magazine》43，1（2022年），30–39。

[57] Krzysztof Janowicz、Simon Scheider、Todd Pehle 和 Glen Hart。2012年。地理空间语义与关联时空数据——过去、现在与未来。语义网 3，4（2012年），321–332。[58] Gwilym M Jenkins、George EP Box 和 Gregory C Reinsel。2011年。时间序列分析：预测与控制。第734卷。约翰·威立父子公司。[59] Renhe Jiang、Zekun Cai、Zhaonan Wang、Chuang Yang、Zipei Fan、Quanjun Chen、Kota Tsubouchi、Xuan Song 和 Ryosuke Shibasaki。2021年。Deepcrowd：用于大规模全市人群密度和流量预测的深度模型。IEEE 知识与数据工程汇刊 35，1（2021年），276–290。[60] Christopher B Jones 和 Ross S Purves。2008年。地理信息检索。《国际地理信息科学杂志》22，3（2008年），219–228。[61] Yiting Ju 等。2016年。事物与字符串：结合实体共现和主题建模，改进短文本中的地名消歧。载于 EKAW 2016。施普林格，353–367。[62] Dan Jurafsky 和 James H. Martin。2009年。语音与语言处理：自然语言处理、计算语言学和语音识别导论，第2版。普伦蒂斯·霍尔，培生教育国际。[63] Aishwarya Kamath、Mannat Singh、Yann LeCun、Gabriel Synnaeve、Ishan Misra 和 Nicolas Carion。2021年。MDETR——用于端到端多模态理解的调制检测。载于 IEEE/CVF 国际计算机视觉大会论文集。1780–1790。[64] Minguk Kang、Jun-Yan Zhu、Richard Zhang、Jaesik Park、Eli Shechtman、Sylvain Paris 和 Taesung Park。2023年。扩展用于文本生成图像的生成对抗网络。arXiv 预印本 arXiv:2303.05511（2023年）。[65] Yuhao Kang、Fan Zhang、Wenzhe Peng、Song Gao、Jinmeng Rao、Fabio Duarte 和 Carlo Ratti。2021年。利用多源大规模地理数据和机器学习理解房价升值。《土地利用政策》111（2021年），104919。[66] Jared Kaplan、Sam McCandlish、Tom Henighan、Tom B. Brown、Benjamin Chess、Rewon Child、Scott Gray、Alec Radford、Jeffrey Wu 和 Dario Amodei。2020年。神经语言模型的缩放定律。CoRR abs/2001.08361（2020年）。[67] Tero Karras、Samuli Laine 和 Timo Aila。2019年。用于生成对抗网络的基于风格的生成器架构。载于 IEEE/CVF 计算机视觉与模式识别会议论文集。4401–4410。[68] Robin Kearns 和 Graham Moon。2002年。从医学地理学到健康地理学：变革十年后的新颖性、地方与理论。《人文地理学进展》26，5（2002年），605–625。[69] Jacob Devlin、Ming-Wei Chang、Kenton 和 Lee Kristina Toutanova。2019年。BERT：用于语言理解的深度双向 Transformer 预训练。载于 NAACL-HLT 2019。4171–4186。[70] Carsten Keßler 和 Grant McKenzie。2018年。地理隐私宣言。地理信息系统汇刊 22，1（2018年），3–19。[71] Diederik P. Kingma 和 Max Welling。2014年。自编码变分贝叶斯。载于第二届国际学习表征会议（ICLR 2014），加拿大艾伯塔省班夫，2014年4月14–16日，会议论文集。arXiv:http://arxiv.org/abs/1312.6114v10 [stat.ML][72] Alexander Kirillov、Eric Mintun、Nikhila Ravi、Hanzi Mao、Chloe Rolland、Laura Gustafson、Tete Xiao、Spencer Whitehead、Alexander C Berg、Wan-Yen Lo 等。2023年。任意分割。arXiv 预印本 arXiv:2304.02643（2023年）。[73] Ellen A Kramarow 和 Betzaida Tejada-Vera。2019年。美国痴呆症死亡率，2000年–2017年。《国家生命统计报告》：疾病控制与预防中心、国家卫生统计中心、国家生命统计系统 68，2（2019年），1–29。[74] Alex Krizhevsky、Ilya Sutskever 和 Geoffrey E Hinton。2017年。使用深度卷积神经网络进行 ImageNet 分类。ACM 通讯 60，6（2017年），84–90。[75] Werner Kuhn。2005年。地理空间语义：为何、关于什么以及如何？载于《数据语义学杂志 III》。施普林格，1–24。[76] Darius Lam、Richard Kuzma、Kevin McGee、Samuel Dooley、Michael Laielli、Matthew Klaric、Yaroslav Bulatov 和 Brendan McCord。2018年。xView：俯视影像中的情境物体。arXiv 预印本 arXiv:1802.07856（2018年）。[77] Guillaume Lample、Miguel Ballesteros、Sandeep Subramanian、Kazuya Kawakami 和 Chris Dyer。2016年。用于命名实体识别的神经网络架构。arXiv 预印本 arXiv:1603.01360（2016年）。[78] Christian Ledig、Lucas Theis、Ferenc Huszár、Jose Caballero、Andrew Cunningham、Alejandro Acosta、Andrew Aitken、Alykhan Tejani、Johannes Totz、Zehan Wang 等。2017年。利用生成对抗网络实现照片级单图像超分辨率。载于 IEEE 计算机视觉与模式识别会议论文集。4681–4690。[79] Jihyeon Lee、Nina R Brooks、Fahim Tajwar、Marshall Burke、Stefano Ermon、David B Lobell、Debashish Biswas 和 Stephen P Luby。2021年。可扩展深度学习，用于识别砖窑并提升监管能力。《美国国家科学院院刊》118，17（2021年），e2018863118。[80] Jihyeon Lee、Dylan Grosz、Sicheng Zeng、Burak Uzkent、Marshall Burke、David Lobell 和 Stefano Ermon。2021年。根据众包街景图像预测生计指标。载于 AAAI 人工智能会议论文集。[81] Junnan Li、Dongxu Li、Silvio Savarese 和 Steven Hoi。2023年。BLIP-2：利用冻结图像编码器和大语言模型引导语言-图像预训练。arXiv 预印本 arXiv:2301.12597（2023年）。[82] Junnan Li、Dongxu Li、Caiming Xiong 和 Steven Hoi。2022年。BLIP：面向统一视觉-语言理解与生成的语言-图像预训练引导方法。载于国际机器学习会议。PMLR，12888–12900。[83] Mingxiao Li、Song Gao、Feng Lu、Kang Liu、Hengcai Zhang 和 Wei Tu。2021年。利用图卷积网络，通过物理与社会空间中的交互预测人类活动强度。《国际地理信息科学杂志》35，12（2021年），2489–2516。[84] Chen Liang、Jonathan Berant、Quoc Le、Kenneth Forbus 和 Ni Lao。2017年。神经符号机：通过弱监督在 Freebase 上学习语义解析器。载于计算语言学协会第55届年会论文集。23–33。

[85] Yan Lin、Huaiyu Wan、Shengnan Guo 和 Youfang Lin。2021年。基于时空轨迹预训练上下文与时间感知的位置嵌入，用于用户下一位置预测。载于《AAAI 人工智能会议论文集》，第35卷。4241–4248页。[86] Kang Liu、Song Gao、Peiyuan Qiu、Xiliang Liu、Bo Yan 和 Feng Lu。2017年。Road2Vec：利用大规模出行路线衡量城市道路系统中的交通交互。《ISPRS国际地理信息期刊》6卷11期（2017年），321。[87] Zilong Liu、Krzysztof Janowicz、Ling Cai、Rui Zhu、Gengchen Mai 和 Meilin Shi。2022年。地理解析：问题已解决还是存在偏差？对地理解析中地理偏差的评估。《AGILE：地理信息科学系列》3（2022年），9。[88] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。2021年。Swin Transformer：利用移位窗口的层次化视觉Transformer。载于《IEEE/CVF国际计算机视觉大会论文集》。10012–10022页。[89] Zhuang Liu、Hanzi Mao、Chao-Yuan Wu、Christoph Feichtenhofer、Trevor Darrell 和 Saining Xie。2022年。面向2020年代的卷积网络。载于《IEEE/CVF计算机视觉与模式识别会议论文集》（CVPR）（2022年）。[90] Massimiliano Luca、Gianni Barlacchi、Bruno Lepri 和 Luca Pappalardo。2021年。人类移动性深度学习综述。《ACM计算调查》（CSUR）55卷1期（2021年），1–44页。[91] Gengchen Mai 等人。2020年。利用网格单元对空间特征分布进行多尺度表示学习。载于 ICLR 2020。OpenReview。[92] Gengchen Mai、Chris Cundy、Kristy Choi、YINGJIE HU、Ni Lao 和 Stefano Ermon。2022年。迈向地理空间人工智能基础模型（视觉论文）。载于《第30届国际地理信息系统进展会议论文集》。1–4页。[93] Gengchen Mai、YINGJIE HU、Song Gao、Ling Cai、Bruno Martins、Johannes Scholz、Jing Gao 和 Krzysztof Janowicz。2022年。符号与亚符号GeoAI：地理空间知识图谱与空间显式机器学习。《地理信息系统汇刊》26卷8期（2022年），3118–3124页。[94] Gengchen Mai、Krzysztof Janowicz、Ling Cai、Rui Zhu、Blake Regalia、Bo Yan、Meilin Shi 和 Ni Lao。2020年。SE-KGE：一种用于地理问答和空间语义提升的位置感知知识图谱嵌入模型。《地理信息系统汇刊》24卷3期（2020年），623–655页。[95] Gengchen Mai、Krzysztof Janowicz、YINGJIE HU、Song Gao、Bo Yan、Rui Zhu、Ling Cai 和 Ni Lao。2022年。GeoAI位置编码综述：方法与应用。《国际地理信息科学杂志》36卷4期（2022年），639–673页。[96] Gengchen Mai、Krzysztof Janowicz、Bo Yan、Rui Zhu、Ling Cai 和 Ni Lao。2019年。上下文图注意力：在不完整知识图谱上回答逻辑查询。载于《第10届国际知识获取会议论文集》。171–178页。[97] Gengchen Mai、Krzysztof Janowicz、Rui Zhu、Ling Cai 和 Ni Lao。2021年。地理问答：挑战、独特性、分类与未来方向。《AGILE：地理信息科学系列》2（2021年），8。[98] Gengchen Mai、Chiyu Jiang、Weiwei Sun、Rui Zhu、Yao Xuan、Ling Cai、Krzysztof Janowicz、Stefano Ermon 和 Ni Lao。2022年。迈向多边形几何体的通用表示学习。《地理信息学》（2022年），1–52页。[99] Pablo N Mendes 等人。2011年。DBpedia Spotlight：为文档网络揭示信息。载于《I-Semantics 2011》。[100] Deepak R Mishra、Hyun J Cho、Shuvankar Ghosh、Amelia Fox、Christopher Downs、Paul BT Merani、Philemon Kirui、Nick Jackson 和 Sachidananda Mishra。2012年。漏油事件后的盐沼状况：遥感估算墨西哥湾漏油事件对路易斯安那盐沼生态环境的影响。《环境遥感》118（2012年），176–185页。[101] Mashaal Musleh、Mohamed F Mokbel 和 Sofiane Abbar。2022年。让我们用语言描述轨迹。载于《第30届国际地理信息系统进展会议论文集》。1–4页。[102] Haifeng Niu 和 Elisabete A Silva。2021年。利用兴趣点数据和神经网络嵌入划分城市功能用地：以大伦敦为例。《计算机、环境与城市系统》88（2021年），101651。[103] Gilles Notton 和 Cyril Voyant。2018年。间歇性太阳能资源预测。载于《可再生能源与电力技术进展》。Elsevier，77–114页。[104] OpenAI。2022年。ChatGPT 正式发布。https://openai.com/blog/chatgpt[105] OpenAI。2023年。GPT-4技术报告。arXiv 预印本 arXiv:2303.08774（2023年）。[106] Long Ouyang、Jeffrey Wu、Xu Jiang、Diogo Almeida、Carroll Wainwright、Pamela Mishkin、Chong Zhang、Sandhini Agarwal、Katarina Slama、Alex Ray 等人。2022年。通过人类反馈训练语言模型遵循指令。载于《神经信息处理系统进展》第35卷（2022年），27730–27744页。[107] Marius Paulescu、Eugenia Paulescu 和 Viorel Badescu。2021年。太阳辐照度临近预报：实现太阳能发电站运行与智能电网管理的有效性。载于《能源管理与电力系统工程的预测建模》。Elsevier，249–270页。[108] Joel Pearson。2019年。人类想象力：视觉心象的认知神经科学。《自然神经科学评论》20卷10期（2019年），624–634页。[109] Matthew E. Peters、Mark Neumann、Mohit Iyyer、Matt Gardner、Christopher Clark、Kenton Lee 和 Luke Zettlemoyer。2018年。深度上下文化词表示。载于NAACL-HLT 2018。计算语言学协会，路易斯安那州新奥尔良，2227–2237页。[110] Matthew E Peters、Mark Neumann、Robert Logan、Roy Schwartz、Vidur Joshi、Sameer Singh 和 Noah A Smith。2019年。知识增强上下文词表示。载于《2019年自然语言处理实证方法会议暨第9届自然语言处理国际联合会议论文集》（EMNLP-IJCNLP）。43–54页。[111] Yanlin Qi、Gengchen Mai、Rui Zhu 和 Michael Zhang。2023年。EVKG：面向智能交通系统的互联互通电动汽车知识图谱。《地理信息系统汇刊》（2023年）。[112] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark 等人。2021年。从自然语言监督中学习可迁移视觉模型。载于《国际机器学习会议论文集》。PMLR，8748–8763页。

[113] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、A manda Askell、Pamela Mishkin、Jack Clark 等。2021年。从自然语言监督中学习可迁移视觉模型。载于国际机器学习会议。PMLR，8748–8763。[114] Alec Radford、Karthik Narasimhan、Tim Salimans 和 Ilya Sutskever。2018年。通过生成式预训练提升语言理解。（2018年）。[115] Alec Radford、Jeff Wu、Rewon Child、David Luan、Dario Amodei 和 Ilya Sutskever。2019年。语言模型是无监督多任务学习者。（2019年）。[116] Colin Raffel、Noam Shazeer、Adam Roberts、Katherine Lee、Sharan Narang、Michael Matena、Yanqi Zhou、Wei Li 和 Peter J. Liu。2020年。探索统一文本到文本转换器的迁移学习极限。《机器学习研究期刊》21，140（2020年），1–67。[117] Aditya Ramesh、Prafulla Dhariwal、Alex Nichol、Casey Chu 和 Mark Chen。2022年。利用CLIP 潜变量进行分层文本条件图像生成。arXiv预印本 arXiv:2204.06125（2022年）。[118] Jinmeng Rao、Song Gao、Yuhao Kang 和 Qunying Huang。2020年。LSTM-TrajGAN：一种用于轨迹隐私保护的深度学习方法。载于第11届国际地理信息科学会议（GIScience 2021）——第一部分。施洛斯·达格施图尔·莱布尼茨信息学中心。[119] Jinmeng Rao、Song Gao、Mingxiao Li 和 Qunying Huang。2021年。一种利用去中心化协同机器学习的位置推荐隐私保护框架。《地理信息系统汇刊》25，3（2021年），1153–1175。[120] Joseph Redmon、Santosh Divvala、Ross Girshick 和 Ali Farhadi。2016年。你只需看一眼：统一的实时目标检测。载于IEEE计算机视觉与模式识别会议论文集。779–788。[121] Scott Reed、Konrad Zolna、Emilio Parisotto、Sergio Gomez Colmenarejo、Alexander Novikov、Gabriel Barth-Maron、Mai Gimenez、Yury Sulsky、Jackie Kay、Jost Tobias Springenberg 等。2022年。通用智能体。arXiv预印本 arXiv:2205.06175（2022年）。[122] Blake Regalia、Krzysztof Janowicz、Gengchen Mai、Dalia Varanka 和 E Lynn Usery。2018年。GNIS-LD：将地理名称信息系统地名词典作为关联数据提供和可视化。载于欧洲语义网会议。施普林格，528–540。[123] Esther Rolf、Jonathan Proctor、Tamma Carleton、Ian Bolliger、Vaishaal Shankar、Miyabi Ishihara、Benjamin Recht 和 Solomon Hsiang。2021年。一种利用全球卫星影像开展机器学习的通用且易于使用的方法。《自然·通讯》12，1（2021年），4392。[124] Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。2022年。利用潜扩散模型合成高分辨率图像。载于I EEE/CVF 计算机视觉与模式识别大会论文集。10684–10695。[125] Mark Rosenberg。2014年。健康地理学 I：社会正义、唯心主义理论、健康与医疗保健。《人文地理学进展》38，3（2014年），466–475。[126] Chitwan Saharia、William Chan、Saurabh Saxena、Lala Li、Jay Whang、Emily L Denton、Kamyar Ghasemipour、Raphael Gontijo Lopes、Burcu Karagol Ayan、Tim Salimans 等。2022年。结合深度语言理解的照片级真实感文生图扩散模型。《神经信息处理系统进展》35（2022年），36479–36494。[127] Christoph Schuhmann、Romain Beaumont、Richard Vencu、Cade W Gordon、Ross Wightman、Mehdi Cherti、Theo Coombes、Aarush Katta、Clayton Mullis、Mitchell Wortsman、Patrick Schramowski、Srivatsa R Kundurthy、Katherine Crowson、Ludwig Schmidt、Robert Kaczmarczyk 和 Jenia Jitsev。2022年。LAION-5B：用于训练下一代图文模型的开放大规模数据集。载于神经信息处理系统第三十六届会议数据集与基准赛道。https://openreview.net/forum?id=M3Y74vmsMcY[128] Katarzyna Siła-Nowicka、Jan Vandrol、Taylor Oshan、Jed A Long、Urška Demšar 和 A Stewart Fotheringham。2016年。基于GPS 轨迹和上下文信息分析人类移动模式。《国际地理信息科学杂志》30，5（2016年），881–906。[129] Filippo Simini、Gianni Barlacchi、Massimilano Luca 和 Luca Pappalardo。2021年。一种用于生成移动流的深度引力模型。《自然·通讯》12，1（2021年），6576。[130] Waldo R Tobler。1970年。一部模拟底特律地区城市发展的计算机动画。《经济地理学》46，增刊1（1970年），234–240。[131] Ilya Tolstikhin、Neil Houlsby、Alexander Kolesnikov、Lucas Beyer、Xiaohua Zhai、Thomas Unterthiner、Jessica Yung、Andreas Steiner、Daniel Keysers、Jakob Uszkoreit、Mario Lucic 和 Alexey Dosovitskiy。2021年。MLP-Mixer：一种面向视觉的全MLP架构。arXiv预印本 arXiv:2105.01601（2021年）。[132] Hugo Touvron、Thibaut Lavril、Gautier Izacard、Xavier Martinet、Marie-Anne Lachaux、Timothée Lacroix、Baptiste Rozière、Naman Goyal、Eric Hambro、Faisal Azhar 等。2023年。Llama：开放且高效的基础语言模型。arXiv预印本 arXiv:2302.13971（2023年）。[133] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017年。注意力机制即为所需。《神经信息处理系统进展》30（2017年）。[134] Jimin Wang 和 Yingjie Hu。2019年。利用EUPEG增强空间与文本分析：用于评估地理解析器的可扩展统一平台。《地理信息系统汇刊》23，6（2019年），1393–1419。[135] Jimin Wang、Yingjie Hu 和 Kenneth Joseph。2020年。NeuroTPR：一种用于从社交媒体消息中提取地点的神经网络地名识别模型。《地理信息系统汇刊》24，3（2020年），719–735。[136] Sheng Wang、Zhifeng Bao、J Shane Culpepper 和 Gao Cong。2021年。轨迹数据管理、分析与学习综述。《ACM计算调查》（CSUR）54，2（2021年），1–36。[137] Shaohua Wang、Song Gao、Xin Feng、Alan T Murray 和 Yuan Zeng。2018年。一种基于上下文的地理处理框架，用于优化道路网络上多个移动对象的会合地点。《国际地理信息科学杂志》32，7（2018年），1368–1390。[138] Wenhui Wang、Hangbo Bao、Li Dong、Johan Bjorck、Zhiliang Peng、Qiang Liu、Kriti Aggarwal、Owais Khan Mohammed、Saksham Singhal、Subhojit Som 等。2022年。图像作为一种外语：面向所有视觉与视觉语言任务的BEiT预训练。arXiv预印本 arXiv:2208.10442（2022年）。[139] Xiaobin Wang 等。2019年。DM_NLP在SemEval-2018任务12中的系统：用于地名消歧的流水线系统。载于SemEval 2019。917–923。

[140] Jason Wei、Xuezhi Wang、Dale Schuurmans、Maarten Bosma、Ed Chi、Quoc Le 和 Denny Zhou。2022年。思维链提示能够激发大语言模型的推理能力。arXiv预印本 arXiv:2201.11903（2022年）。[141] Chien-sheng Wu、Richard Socher 和 Caiming Xiong。2019年。面向任务的对话全局到局部记忆指针网络。发表于第七届国际学习表征会议（ICLR 2019）。[142] Xuansheng Wu、Kaixiong Zhou、Mingchen Sun、Xin Wang 和 Ninghao Liu。2023年。图提示方法综述：技术、应用与挑战。arXiv预印本 arXiv:2303.07275（2023年）。[143] Yuxin Wu、Alexander Kirillov、Francisco Massa、Wan-Yen Lo 和 Ross Girshick。2019年。Detectron2。https://github.com/facebookresearch/detectron2。[144] Gui-Song Xia、Jingwen Hu、Fan Hu、Baoguang Shi、Xiang Bai、Yanfei Zhong、Liangpei Zhang 和 Xiaoqiang Lu。2017年。AID：用于航空场景分类性能评估的基准数据集。《IEEE地球科学与遥感汇刊》55卷7期（2017年），3965–3981。[145] Bo Yan、Krzysztof Janowicz、Gengchen Mai 和 Song Gao。2017年。从itdl到Place2Vec：通过从增强空间上下文中学习嵌入，推断地点类型的相似性和相关性。发表于第25届ACM SIGSPATIAL地理信息系统进展国际会议论文集。1–10。[146] Xin Yao、Yong Gao、Di Zhu、Ed Manley、Jiaoe Wang 和 Yu Liu。2020年。利用图卷积网络进行空间起讫流补全。《IEEE智能交通系统汇刊》22卷12期（2020年），7474–7484。[147] Yao Yao、Xia Li、Xiaoping Liu、Penghua Liu、Zhaotang Liang、Jinbao Zhang 和 Ke Mai。2017年。通过整合兴趣点和Google Word2Vec模型感知城市土地利用的空间分布。《国际地理信息科学杂志》31卷4期（2017年），825–848。[148] Michihiro Yasunaga、Antoine Bosselut、Hongyu Ren、Xikun Zhang、Christopher D Manning、Percy S Liang 和 Jure Leskovec。2022年。深度双向语言知识图谱预训练。《神经信息处理系统进展》35（2022年），37309–37323。[149] Wenhao Yu 和 Yujie Chen。2022年。使用编码器-解码器模型填补制图折线中的缺口。《国际地理信息科学杂志》36卷11期（2022年），2296–2321。[150] Wenhao Yu、Chenguang Zhu、Zaitang Li、Zhiting Hu、Qingyun Wang、Heng Ji 和 Meng Jiang。2022年。知识增强文本生成综述。《计算调查》54卷11s期（2022年），1–38。[151] Lu Yuan、Dongdong Chen、Yi-Ling Chen、Noel Codella、Xiyang Dai、Jianfeng Gao、Houdong Hu、Xuedong Huang、Boxin Li、Chunyuan Li等。2021年。Florence：一种用于计算机视觉的新型基础模型。arXiv预印本 arXiv:2111.11432（2021年）。[152] Wei Zhai、Xueyin Bai、Yu Shi、Yu Han、Zhong-Ren Peng 和 Chaolin Gu。2019年。超越Word2vec：结合Place2Vec与兴趣点数据提取和识别城市功能区的方法。《计算机、环境与城市系统》74（2019年），1–12。[153] Xiaohua Zhai、Alexander Kolesnikov、Neil Houlsby 和 Lucas Beyer。2022年。视觉变换器的规模扩展。发表于IEEE/CVF计算机视觉与模式识别会议论文集。12104–12113。[154] Fan Zhang、Bolei Zhou、Liu Liu、Yu Liu、Helene H Fung、Hui Lin 和 Carlo Ratti。2018年。利用机器学习测量人类对大规模城市区域的感知。《景观与城市规划》180（2018年），148–160。[155] Jinbao Zhang、Xia Li、Yao Yao、Ye Hong、Jialyu He、Zhangwei Jiang 和 Jianchao Sun。2021年。Traj2Vec模型：量化居民的空间轨迹并估算城市土地利用类型的比例。《国际地理信息科学杂志》35卷1期（2021年），193–211。[156] Lvmin Zhang 和 Maneesh Agrawala。2023年。为文生图扩散模型添加条件控制。arXiv预印本 arXiv:2302.05543（2023年）。[157] Susan Zhang、Stephen Roller、Naman Goyal、Mikel Artetxe、Moya Chen、Shuohui Chen、Christopher Dewan、Mona Diab、Xian Li、Xi Victoria Lin等。2022年。OPT：开放式预训练Transformer语言模型。arXiv预印本 arXiv:2205.01068（2022年）。[158] X Zhang、A Bosselut、M Yasunaga、H Ren、P Liang、C Manning 和 J Leskovec。2022年。GreaseLM：用于问答的图推理增强语言模型。发表于国际表征学习会议（ICLR）。[159] Xiuyuan Zhang、Shihong Du 和 Qiao Wang。2017年。利用甚高分辨率卫星图像和兴趣点数据对城市功能区进行层次化语义认知。《ISPRS摄影测量与遥感杂志》132（2017年），170–184。[160] Xiuyuan Zhang、Shihong Du 和 Qiao Wang。2018年。整合自下而上分类与自上而下反馈，以改进城市土地覆盖和功能区制图。《环境遥感》212（2018年），231–248。[161] Zhengyan Zhang、Xu Han、Zhiyuan Liu、Xin Jiang、Maosong Sun 和 Qun Liu。2019年。ERNIE：基于信息实体的增强语言表征。发表于第57届计算语言学协会年会论文集。1441–1451。[162] Tianhong Zhao、Xiucheng Liang、Wei Tu、Zhengdong Huang 和 Filip Biljecki。2023年。利用街景图像感知城市声景。《计算机、环境与城市系统》99（2023年），101915。[163] Wayne Xin Zhao、Kun Zhou、Junyi Li、Tianyi Tang、Xiaolei Wang、Yupeng Hou、Yingqian Min、Beichen Zhang、Junjie Zhang、Zican Dong等。2023年。大型语言模型综述。arXiv预印本 arXiv:2303.18223（2023年）。[164] Bolei Zhou、Agata Lapedriza、Aditya Khosla、Aude Oliva 和 Antonio Torralba。2017年。Places：一个用于场景识别的千万级图像数据库。《IEEE模式分析与机器智能汇刊》（2017年）。[165] Di Zhu、Fan Zhang、Shengyin Wang、Yaoli Wang、Ximeng Cheng、Zhou Huang 和 Yu Liu。2020年。通过图卷积神经网络理解地理上下文中的地点特征。《美国地理学家协会年刊》110卷2期（2020年），408–420。[166] Rui Zhu、Krzysztof Janowicz、Ling Cai 和 Gengchen Mai。2022年。通过空间显式神经网络对高阶定性空间关系进行推理。《国际地理信息科学杂志》36卷11期（2022年），2194–2225。[167] Julie M Zissimopoulos、Bryan C Tysinger、Patricia A St. Clair 和 Eileen M Crimmins。2018年。美国人口健康和死亡率变化对未来阿尔茨海默病及其他痴呆症患病率的影响。《老年学期刊：B辑》73卷增刊_1（2018年），S38–S47。

A 附录 A.1 各项实验中使用的完整提示词

```
这是一组 of 地名识别 问题
The `段落 ` 是一组 文本 包含 地方。
目标是 推断 哪些 词语表示命名地点 Places
    ↪ in [g:this] 段落, and 拆分命名 地点
    ↪ with `;`
--
--
段落: 阿拉巴马州警察 称一名 格林维尔男子已
    ↪ died 因其伤势 之后 被 撞上
    ↪ 取件 在朗兹县的65号州际公路上。
Q: 哪些 词语在这 段落 代表命名 地点?
A: 阿拉巴马州; 格林维尔; 朗兹
--
.....
--
段落: 定居 in the 塞尼亚地区 1856年, the year
    ↪ 在布尔本县之后 成立于1855年, were
    ↪ John Van Syckle , Samuel Stephenson 和 Charles
    ↪ 安德森。
Q: 哪种 此处的词语 段落 代表命名 地点?
A: 塞尼亚; 波旁
--
--
段落: [文本]
Q: 哪些 词语在这段落中 表示命名 地点?
A:
```

列表 7. GPT-2 和 GPT-3 模型使用的提示，用于 Hu2014 和 Ju2016 测试集上的地名识别数据集。一个“段落”、“Q”和“A”三元组构成一个语言指令示例。“[TEXT]”将被替换为 the 待标注的文本。我们在此提示中总共使用了 8 个样本。此处仅展示 2 个以节省空间。

```
This is 一组 的位置 描 述 识别
    ↪ 问题
The `段落 ` 是一个集合 文本的 包含 位 置
    ↪ 描 述。
目标是 to [g:infer] 其中 [g:this] 词语表示位置
    ↪ 描述于其中 段落, and 分 割
    ↪ 不同 位 置 描 述 with `; `.
--
--
段落: 爸爸被困在 住宅。 水域正在上升 高过 腰部
    ↪ 。救命 8111 伍德林路, 77028 #休斯敦洪灾
Q: 哪些 这段文字中的 段落 代表位置
    ↪ 描 述?
A: 伍德林路8111号 , 77028
--
.....
--
段落: 严重洪灾 克莱发生 路与Queenston交汇处 west
    ↪ 休斯敦。 正在开展大量人员救援
    ↪ 被困。
Q: 哪些 这些词 段落 表示位置
    ↪ 描述?
A: Clay 路与 Queenston; 休斯敦西部
--
--
段落: [文本]
Q: 哪些 词语在这 段落 表示位置
    ↪ 描述?
A:
```

列表 8. GPT-2 和 GPT-3 模型在 HaveyTweet2017 数据集的测试集上进行地点描述识别时使用的提示。一个“段落”、“Q”和“A”元组构成一个语言指令样本。"[TEXT]" 将替换为待标注的文本。此提示共使用了 11 个样本，为节省空间，此处仅展示其中 2 个。

这是一组地理定位问题。 `段落 ` 是一段包含地点的文本。目标是推断该段落中哪些词语代表命名地点↩，并使用 `;`分隔这些命名地点↩。然后，下一个目标是将每个命名地点定位为具有 5 位小数↩ 精确率的↩地理坐标。----段落：阿拉巴马州警察表示，朗兹县一名来自格林维尔的男子在65号州际公路上被一辆↩ 皮卡车↩ 撞击后，因伤势过重死亡。问：该段落中哪些词语代表命名地点？答：格林维尔；阿拉巴马州；朗兹县。问：格林维尔的地理位置是什么？答：31.83283， −86.63270。问：阿拉巴马州的地理位置是什么？答：32.92040， −86.83519。问：朗兹县的地理位置是什么？答：32.16314， −86.64631----段落：圣何塞于↩ 1803年↩ 成立，当时新墨西哥州的西班牙政府将土地分配给45名男性和两名女性。问：该段落中哪些词语代表命名地点？答：圣何塞；新墨西哥州。问：圣何塞的地理位置是什么？答：35.39728， −105.
新墨西哥州的地理位置是什么？答：34.68965， −106.05006----段落：↩ 截至2015年7月1日，Fairview市的人口为260人。与其他城市、城镇和人口普查指定地区（人口普查指定地区）相比，Fairview的多样性指数排名↩ 处于较低四分位数。问：该段落中哪些词语代表命名地点？答：Fairview。问：Fairview的地理位置是
？

A: 41.85003 , -87.65005

图 9. GPT-3 在 Ju2016 数据集的测试集上进行地理解析时使用的提示。我们使用两个样本作为语言指令。黄色区块表示 Ju2016 数据集中的一个文本片段，橙色区块表示生成答案。最后一行生成的坐标被视为地理解析结果。

GeoAI基础模型 会议缩写“XX”，2018年6月3日至5日，纽约州伍德斯托克

收稿日期：2007年2月20日；修订日期：2009年3月12日；录用日期：2009年6月5日