面向自然历史博物馆大规模数字化藏品的对话式人工智能增强探索系统

YIYUAN WANG, 悉尼科技大学计算机科学学院，澳大利亚ANDREW JOHNSTON,悉尼科技大学计算机科学学院，澳大利亚ZOË SADOKIERSKI, 悉尼科技大学设计学院，澳大利亚RHIANNON STEPHENS, 澳大利亚博物馆研究所，澳大利亚博物馆，澳大利亚SHANE T. AHYONG,澳大利亚博物馆研究所，澳大利亚博物馆，澳大利亚

arXiv:2603.10285v1 [cs.HC] 11 Mar 2026

近年来，自然历史博物馆的数字化工作产生了大量藏品数据，但其规模和科学复杂性往往妨碍公众访问和理解。传统的数据管理工具（如数据库）将探索局限于基于关键词的搜索，或要求用户具备专门的模式知识。本文介绍了一种系统设计，利用对话式人工智能查询澳大利亚博物馆生命科学藏品中近170万条数字化标本记录。该系统通过以人为中心的设计流程进行设计和开发，包含用于视觉空间探索的交互式地图，以及一个能够检索详细标本数据并回答藏品相关问题的自然语言对话代理。系统利用当代大型语言模型的函数调用能力，通过外部 API 动态检索结构化数据，从而实现与庞大且频繁更新的数据集进行快速实时交互。本研究提供了一种将大型博物馆藏品与基于自然语言的查询相连接的新方法，并为未来面向自然历史博物馆的科学人工智能代理设计提供参考。

CCS 概念： • **以人为中心的计算** → **交互式系统与工具**； **交互设计实证研究**； •**应用计算** →数字图书馆与档案馆。

附加关键词和短语：生物多样性信息学、聊天机器人、对话式人工智能、数据探索、数字化藏品、以人为中心的设计、人类-数据交互、大型语言模型、自然历史博物馆、自然语言界面

1 引言

自然历史博物馆通常收藏着数量庞大的藏品，不可能全部在面向公众的展览中展出。例如，澳大利亚博物馆收藏了超过 2100 万件自然科学标本。作为科学机构的责任之一，保护和研究这些标本至关重要，因为它们对发现澳大拉西亚生物多样性贡献巨大，并为生物学、生态学和人文学科研究奠定基础。数字化工作日益助力标本保护 [10, 28]。在博物馆语境中，数字化可概括为“创建物件或物件某些方面的数字版本”。这通常包括将标本图像、地点和采集日期、采集者、目录编号及其他元数据录入数据库。此项工作量巨大；经过十多年由指定博物馆工作人员和志愿者在线上及现场持续开展的努力，澳大利亚博物馆如今已拥有超过 400 万条数字化标本记录，并将其存储和管理在数据库中。

如何传播这些数字化记录已成为新的重点。传统上，库房藏品通过标本借出向外界传播，而借用者通常是其他科研机构的研究人员，而非普通公众。

作者联系方式：Yiyuan Wang，yiyuan.wang@uts.edu.au，悉尼科技大学计算机科学学院，澳大利亚悉尼；Andrew Johnston，andrew.johnston@uts.edu.au，悉尼科技大学计算机科学学院，澳大利亚悉尼；Zoë Sadokierski，zoe.sadokierski@uts.edu.au，悉尼科技大学设计学院，澳大利亚悉尼；Rhiannon Stephens，rhiannon.stephens@australian. museum，澳大利亚博物馆研究所，澳大利亚博物馆，澳大利亚悉尼；Shane T. Ahyong，shane.ahyong@australian.museum，澳大利亚博物馆研究所，澳大利亚博物馆，澳大利亚悉尼。

面向普通公众的传播也一直通过常规学校教育项目和馆藏导览开展，但每次涉及的内容都很有限。然而，这些方式表明，大量未展出的馆藏存在由来已久的局限：大多数普通民众和远程受众难以接触到它们。为应对这些挑战，本文介绍了一种交互式系统的设计与开发，该系统支持通过网络探索澳大利亚博物馆的大规模数字化生命科学藏品。

博物馆已开始尝试运用新兴技术，扩大公众对其馆藏的访问。例如，生成式人工智能（人工智能）已应用于文化和艺术馆藏，以提升访客参与度 [3, 41], ；扩展现实（扩展现实）技术则通过实体博物馆的数字孪生支持远程参观 [15, 24]。近年来，博物馆越来越多地采用对话式人工智能，改善访客与馆藏之间的互动，从入口处担任向导的聊天机器人 [37] ，到嵌入展出展品中的聊天机器人 [7]。例如，剑桥博物馆将基于标本数据训练的大型语言模型（大语言模型）嵌入 13 件配有音频扬声器的展出标本中，让访客能够与“复活”的灭绝动物进行双向对话 [26]。

对话式界面非常适合用于探索大型科学数据集 [33, 40] ，因为用户可以用自然语言提出查询，相较于传统数据库界面或必须具备科学知识，这降低了使用门槛。然而，大语言模型的生成特性可能会对回答的真实性构成挑战，尤其是在询问具体记录或统计数据时 [16]。要确保回答以实际数据为依据，而非模型的预训练知识，就需要采用能够将大语言模型连接到权威数据源的技术。本文介绍的对话式人工智能系统利用大语言模型服务的函数调用机制，从外部数据基础设施中实现有依据的数据检索，以应对这一挑战。

澳大利亚博物馆藏品探索器这一系统，是通过以人为中心的设计流程开发的。我们与澳大利亚博物馆的工作人员开展了焦点小组讨论，以收集需求；随后进行了两轮迭代设计，通过用户测试评估并完善原型。基于这一流程，我们开发了最终应用程序，其中包含一张交互式地图，可用于对近170万条数字化标本记录进行视觉与空间探索；还包含一个对话代理，支持自然语言查询并回答与馆藏相关的问题。我们介绍了该系统的设计流程和对话式人工智能的技术架构，提出了一种支持公众参与数字化博物馆藏品的新方法。

2 相关工作

2.1 聊天机器人辅助的人与数据交互 on

近年来，随着大语言模型的迅速发展，过去三年间，科学和分析领域越来越多地探索借助聊天机器人进行交互，以帮助用户浏览规模庞大、结构复杂的数据集。聊天机器人已被集成到大规模地震模拟数据的交互式探索中，用户通过自然语言对话即可进行元数据查询、空间搜索和基于邻近性的搜索，以及波形相似性检索；结果则通过地图可视化和逐行呈现的 Jupyter 笔记本界面中的文本回复来展示 [33]。多模态聊天机器人已应用于风能研究，通过处理大量科学研究、环境数据和监管文件，简化数据发现；与传统的关键词搜索相比，它能提供融合文本与图像的丰富回复 [40]。对话式界面与可视化界面的结合也已在医疗保健领域得到研究。将聊天机器人交互与数据可视化相结合，有助于理解

由人工智能生成的风险评估，并通过基于证据的解释建立适当的用户信任 [30],；此外，这类系统还能接受文本和图像查询，进行初步医学诊断 [1]。聊天机器人也已被用于改善公众获取复杂行政信息的途径，例如政府计划和保险单 [19]。大语言模型的加入也让可视分析系统受益；例如，系统可以通过协同图表、时间线和筛选界面，对比大语言模型提取的信息与人工生成的数据，从而揭示原本不易察觉的人类偏见 [6]；还可以在大型数据可视化集合中，无需人工标注即可整理可视化类型 [42]。

在文化遗产语境中，人们正在探索对话式界面，以增强访客对博物馆藏品的参与度并提高其可及性。对文化展品中屏幕式聊天机器人的实地研究表明，用户体验和个性化均有所改善 [37]。将艺术品检测与生成式人工智能相结合的智能手机聊天机器人，其访客参与度也高于传统博物馆应用程序 [41]。除现场参观之外，研究者还探讨了由大语言模型驱动的机器人导览员，为实时穿行于博物馆空间的远程访客提供导航和基于位置的信息 [12]。一项近期的设计研究进一步表明，嵌入历史人物模型中的大语言模型原型可作为博物馆体验的设计材料，揭示了参与度与准确性之间的张力，并提出了相应策略：要么聚焦于虚构叙事，要么突出人工智能的局限性以引发反思 [7]。

尽管科学数据探索与文化遗产领域的相关研究日益增多，但在将聊天机器人辅助的数据交互应用于生物多样性数据集和自然科学传播方面仍存在空白。虽然对话式界面已在其他科学领域与基于地图的探索相结合 [14, 33] ，并用于艺术或文化藏品 [25, 41], ，但此前尚无研究将这些方法应用于大规模数字化生命科学藏品；这类藏品包含丰富的分类学、地理和时间信息，有望从对话式访问中受益。

2.2 配置大语言模型以实现有依据的数据检索

早期的聊天机器人系统依赖基于规则的方法或循环神经网络来生成回复，而较新的系统则越来越多地采用大语言模型，以提供更灵活、更能感知上下文的对话能力 [23, 27]。在数据密集型应用中部署大语言模型的一项核心挑战，是确保生成的回复以准确且可检索的信息为依据，而非依赖估算内容。开发者采用了多种技术方法来应对这一挑战，从调整预训练模型的权重，到通过提示和外部工具集成配置基于 API 的大语言模型。微调方法会调整预训练模型的权重，使模型适用于特定应用场景；例如，使用医疗交互数据微调 T5 模型，将用户查询解析为用于糖尿病风险预测的后端函数调用 [30], ；或使用文化遗产数据微调 LLaVA/Vicuna，根据对象级视觉特征生成情境描述 [27]。医疗数据数字化流程已部署专用模型，例如用于印刷文本处理的 LLAMA-3 和用于手写处方识别的 miniCPM [2]。越来越多的应用通过商业 API 使用预训练大语言模型，包括 OpenAI 的 GPT 模型（GPT-3.5、GPT-4、GPT-4o） [12, 33, 37, 40–42], Google 的 Gemini 模型（ Gemini-2.0- Flash） [1, 39], 以及 Anthropic 的 Claude [30]。基于 API 的方法之所以如此普遍，反映出无需承担模型重新训练所需基础设施成本，即可使用先进语言能力的实际优势。

使用现成的大语言模型时，开发者可以通过多种机制配置模型行为。系统提示工程涉及为模型编写指令，以设定其角色设定、行为约束和输出格式。还可以结合少样本学习，即在提示中提供示例输入输出对，引导模型采用期望的响应模式 [23]。开发者还可以调整推理

参数，例如温度（用于控制响应的随机性）和最大令牌限制，不过与提示工程相比，这些参数提供的控制粒度相对较粗。

其他技术使大语言模型能够利用外部知识源，而不是仅依赖预训练知识中的信息。检索增强生成（RAG）系统从向量数据库中检索相关信息，并在响应生成之前将其输入提示上下文。数据架构的实现方式可以有所不同；例如，将精选知识库分割为数千个可搜索单元以进行基于文本的检索 [1],使用 ChromaDB 和 LlamaIndex 等框架将异构文档类型统一到共享嵌入空间中，从而通过共同的查询嵌入实现对文本和图像的语义搜索 [40], 或使用图像到文本到嵌入流水线，先为视觉输入生成文本描述，再将这些描述嵌入以进行相似度匹配 [1, 27]。

函数调用提供了另一种使大语言模型输出有据可依的技术，使模型能够调用已定义的函数与外部数据源通信。配置了函数调用的大语言模型可以输出发送至外部系统的结构化请求，由外部系统执行这些请求并将结果返回，供大语言模型纳入其响应。这一方法已通过 Text2SQL 方法在数据库交互中得到研究，即大语言模型将自然语言查询转换为可执行的 SQL [16, 23, 31, 39]。此类功能可在多智能体查询处理架构中使用专门的 SQL 代理来实现 [31] ，并可通过提供数据库模式 [39]或通过在对话中存储先前的 SQL 查询和结果来增加上下文感知而得到增强 [23]。对未配置函数调用的基线基于 API 的大语言模型的评估显示其数据检索不准确，进一步凸显了恰当工具定义的重要性 [16]。

较少受到探讨的是通过函数调用将大语言模型直接连接到外部 API 模式，从而无需本地或云端数据存储即可实时检索数据。这一研究空白对于需要与大型、持续更新的数据集交互的应用程序尤为重要，因为开发者一方存储数据可能并不现实，例如本文所述国家数据基础设施汇总的生物多样性收藏记录。

2.3 博物馆参与的交互设计

科学、技术与艺术在博物馆中的交汇可追溯至 20 世纪 50 年代，Ben Laposky 的 Oscillons 是最早的计算机生成展品之一 [20]。近年来，博物馆数字化日益采用人工智能、扩展现实和三维建模等新兴技术，以提升藏品可及性、学习体验和参与度 [4, 9, 13, 20, 43]。近期博物馆展览采用的人工智能技术多种多样，从用于访客与展出标本对话的大语言模型 [26] ，到扩散模型生成的可视化内容，例如 Refik Anadol 在现代艺术博物馆对数百年来艺术作品的重新诠释 [3], ，以及鼓励观众自我表达和共同创作的生成式“绘画” [8, 18]。随着人工智能让展品探索更加个性化、更加以用户为主导，博物馆中的交互设计日益被视为创造力的促进者，而非强制执行预设规则的系统 [34]。大英博物馆也提供了线上范例，通过对话式人工智能界面展示其文化 artefacts [25], ，让用户能够通过自行提出的问题进行学习，并获得即时、自然且信息丰富的回答。这些数字应用共同促进主动探索，而非被动学习。

通过以人为中心的设计流程开发的应用程序，通常会先从利益相关者处收集需求。在文化遗产机构的多项研究中，这种做法已得到实践，例如在实施新技术之前开展探索性访谈和实践型研讨会 [11,22]。焦点小组仍然尤其有价值，是一种成熟的访谈方法，可用于激发集体创意，

捕捉多元观点，并深入探讨选定主题。在交互设计项目中，焦点小组也可用于收集对初步设计原型的反馈 [22]。这也是收集专家见解的一种方法；例如，在数字化框架下收集有关文化遗产管理当前方法和实践的数据 [32], ，并谨慎筛选参与者概况，以确保获得有意义且相关的结果 [32]。基于这些成熟的方法，本文采用以人为中心的设计流程，为开发一个用于探索数字化自然史藏品的对话式人工智能增强系统提供依据。我们与博物馆专业人员和专家开展了焦点小组，以收集需求和领域专业知识，随后进行了两轮迭代式设计改进，最终实施该应用程序。

3 以人为中心的设计流程

本项目是澳大利亚博物馆与悉尼科技大学长期合作的一部分，重点关注新兴技术如何改善科学传播以及公众与博物馆标本藏品之间的互动。项目开始时，研究人员试图了解博物馆数字化工作的当前状况，并确定工作人员对数字化成果的优先事项。为此，大学研究人员在博物馆的工作场所与数字化工作人员并肩工作，并在每周进度会议上向博物馆研究所的高级经理汇报。早期阶段的一项核心成果是了解利益相关者需求，这些需求通过前四个月内开展的两次焦点小组研讨会得以确定。在完成对研讨会成果的分析后，我们[1] 开始了本文所介绍应用程序的迭代设计流程。该流程包括开发初始原型，将利益相关者的优先事项转化为设计目标；咨询数字与信息通信技术部门等相关博物馆部门；以及与博物馆的数字志愿者[2]一起测试概念验证版本。通过不断完善，我们确定了哪些功能是理想的，以及哪些功能可以添加或移除。经过四个月的迭代设计，我们明确了实施目标，并开始开发最终应用程序——澳大利亚博物馆藏品探索器。在介绍该应用程序时，我们描述

其技术实现、当前支持的数据集和用户界面，并通过示例交互场景展示若干核心功能。图 1 概述了上述三个阶段。

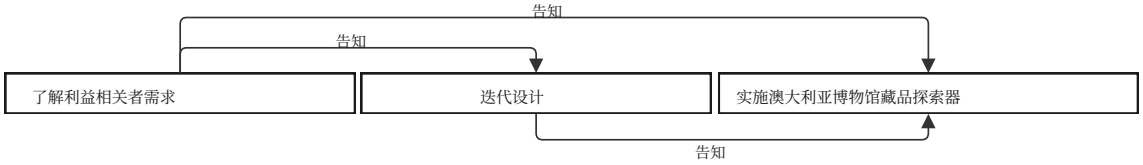


图 1. 澳大利亚博物馆藏品探索器的开发流程。

3.1 澳大利亚博物馆的焦点小组

我们在澳大利亚博物馆举办了两场研讨会，共有30名博物馆工作人员参加。第一场有17名来自自然科学领域的管理人员和科学家参加，第二场则有13名公众参与工作人员参加。两场研讨会遵循相同的实验流程，主要由焦点小组活动组成。焦点小组 [17],是一种半结构化的小组访谈，由主持人引导参与者进行讨论，通常

[1]核心项目团队由三位大学研究人员组成：一位博士后人机交互研究人员、一位高级交互设计研究人员、一位高级视觉传达研究人员；以及两位博物馆管理人员：一位负责馆藏数字化的高级经理和一位生命科学首席科学家。[2]https://australian.museum/get-involved/citizen-science/volunteer-with-us/



参与者围坐在桌旁进行焦点小组讨论



Museum X 研讨会室的全景

研讨会 I：来自自然科学研究领域的 17 名参会者8

- collecti 5 h scienti 员，4 i
- 馆藏管理人员，研究人员，资深 深管理人员，4 个同步焦点小组，2 小时

第 1 部分（第 1 小时）：“第一部分将询问您在工作中如何使用博物馆的数字化藏品，以及为了实现这一目的，哪些藏品内容需要进行可视化。”

- 问题 1：作为馆藏管理员／科研人员，您如何使用我们的数字化藏品？为了以这种方式使用藏品，需要将哪些内容进行可视化？问题 2：您认为我们的数字化藏品有哪些机遇和挑战？您对在博物馆中使用人工智能、虚拟现实/增强现实等可视化技术有何看法？

第二部分（第2小时）：“第二部分请你推测其他研究人员和受众可能如何访问和使用我们的数字化藏品，尤其是其视觉元素。”

- Q1：您认为外部研究人员会如何访问和使用我们的数字化藏品？这些研究人员可能是谁？他们希望可视化馆藏的哪些部分？
- Q2：博物馆参观者可以如何访问和使用我们的数字化藏品？这些参观者可能是谁？他们希望可视化馆藏的哪些部分？

第二场研讨会：来自公众参与的13名与会者7

- exhibit d 1名项目制作人， 3 4 2d i gallat 制作人， ucers， e 2小时
-

第一部分（第1小时）：“第一部分将探讨博物馆当前开展公众推广的情况，以及您认为现有工作取得了哪些成效、有哪些经验教训。”

- Q1：回顾博物馆过去开展的推广项目，哪些项目最为成功？为什么？是否有数字工具助力这些项目取得成功？Q2：在过去向公众介绍博物馆藏品的工作中，您观察到了哪些挑战或局限？是否有具体例子表明数字工具未能达到预期效果？Q3：博物馆现有项目是否未能有效触达某些参观者群体或人口统计群体？在博物馆与更广泛的社区互动的方式上，您认为存在哪些不足？

第二部分（第2小时）：“第二部分请您畅想如何通过不同主题和方式，提升博物馆自然科学藏品及相关研究的推广效果与影响力。”

- Q1：公众可能希望了解博物馆藏品或自然科学主题的哪些方面？相关受众可能是哪些人？Q2：您认为可以如何运用可视化和新兴技术，将这些方面或主题与目标受众联系起来？

图 2. 研讨会现场照片（上）以及焦点小组活动中的访谈问题（下）。

围绕一组预先准备的问题展开。研讨会旨在服务于这一跨机构合作的更广泛目标：形成有关人机交互的见解，例如人机交互系统和范式，以“借助新兴技术改造数字化自然科学馆藏”；研讨会的目的并非仅聚焦于本文讨论的对话式人工智能。相反，本文介绍的项目是根据这些研讨会的发现而产生的主要成果项目。

我们根据两个不同的参会者群体设计了焦点小组问题。第一场研讨会面向自然科学参与者的问题，围绕自然科学的发展与传播展开，包括工作人员目前如何使用数字化藏品、他们面临的挑战，以及他们对未来外部访问与使用的期望。第二场研讨会面向公众参与领域参会者的问题，则着重探讨如何

加强公众与自然科学的互动，并提升博物馆馆藏的影响力，包括过去公众推广项目的成效与局限，以及对未来数字化方法的看法。问题之间的另一个区别是，与自然科学参与者交流时，更加侧重视觉用途，因为可视化是标本数字化的关键组成部分，有助于科学家识别、分析和解读物种。研讨会的照片和两组焦点小组问题见图2。研讨会已获悉尼科技大学人体研究伦理委员会批准。

3.1.1 将研究发现转化为设计目标。 焦点小组讨论揭示了两个关键成果：一是了解博物馆数字化流程所产生的数据和数字资产；二是提出一系列设想中的未来项目，将新兴技术应用于数字化标本藏品。在这些设想项目中，本文介绍的项目在研讨会结束后被确定为优先开发项目。为了指导其开发，我们分析了与会者表达的需求以及他们提出的当前方法所面临的挑战，据此制定了四项设计目标，并以此指导后续的设计迭代。表1总结了从研讨会发现到设计目标的转化过程。

表1. 利益相关者需求与设计目标的对应关系。

表达的需求	当前挑战	设计目标
传达其庞大的规模与多样性：库房藏品	馆藏对公众而言不可见；数据的数量和丰富程度难以传达	**规模的可理解性**：打造馆藏规模立刻可感知并得到理解
降低获取科学与研究信息的门槛	非专业受众两者皆缺数据库技术技能和自然科学领域知识	**查询可及性**：启用信息检索无需技术专长即可进行检索
自动化受众与专家之间的信息交流流程（即“咨询专家”服务）	人工处理网络提交的问题无法规模化处理且需要多步骤的工作人员转接	**回复可靠性**：准确提供特定馆藏信息并提高效率
触达多元受众，包括偏远地区和服务不足的社区	实体博物馆参观和工作人员咨询受到地理和时间限制	**泛在访问**：支持探索随时随地获取馆藏随时

反复出现的一种观点是，人们希望传达博物馆库房藏品的规模和多样性。大家也认识到，通过传统方式展现这一规模一直很困难，正如与会者所说：

> “我认为，对公众来说，重要的是帮助他们了解我们幕后拥有如此庞大的馆藏，而且我们一直在对这些藏品开展工作。这不仅是个体动物的数量规模，更是馆藏的规模，以及这些类别内部的多样性。以某种方式接触到相关研究——这会在一定程度上改变公众对我们的印象。”

> “我们一直很难向公众传达博物馆藏品的规模——我们有数百万件标本，而且并非所有标本都已数字化……这就是我认为可视化确实能帮助我们向受众传达这一点的原因。”

这些见解促成了我们的第一个设计目标：规模的可理解性，即让用户无需解读数值数据或

浏览复杂的数据库结构，就能立即感知数字化藏品的规模与丰富程度。与会者还指出了访问藏品信息时面临的技术与知识障碍，其中一位与会者将其与自然语言人工智能技术所带来的、更直观的潜在方法进行了对比：

> “与其琢磨如何查询数据，不如直接用自然语言提问，比如‘哪类自然科学藏品数量最多？’，它就能直接给出一个数字……我通常得请人来查询数据。人工智能 *[让我们]* 能够以自然得多的方式查询馆藏数据。”

这一需求促成了我们的第二个设计目标：查询可及性，让用户无需具备技术专长或科学知识即可获取信息；例如，可以使用自由格式输入或自然语言交互。与会者还对现有专家介入式服务的可扩展性和效率表示担忧，并以官方网站上的一项服务为例：

> “我们运营着一项名为‘咨询专家’的服务——目前，用户上传一张照片并输入一段简短留言，然后由博物馆的工作人员回复。如果他们不知道该如何回答，或者无法辨认照片里的东西，就会把问题转给相关团队……之后问题会返回给最初的工作人员，由其回复：‘嘿，这是红点笛鲷。我们网站上有一些关于它的信息。’如果整个流程都能自动化，并且直接从馆藏中提取信息，那就太好了。”

这一洞见促成了我们的第三项设计目标，即回复可靠性：对于所开发的应用程序而言，这不仅是提升日常问答效率的必要条件，也是提供以藏品数据为依据的准确信息的关键。最后，我们发现博物馆一直致力于触达远程受众，由此提出了第四项设计目标，即泛在访问，使用户不受地理位置或时间限制，均可探索馆藏。

3.2 设计迭代

3.2.1 设计迭代 *I*：初始原型与专家咨询。焦点小组讨论之后，我们开发了一个初始交互式原型，用于展示用户通过自然语言对话代理查询标本收藏时可能获得的使用体验，并收集有关其设计的早期反馈。该原型包含一个简洁的聊天界面，以及一个侧边栏，用于按地区和学科显示藏品项目的文字摘要（见图 3）。该原型仅使用三个合成标本记录来演示查询—回复交互。我们使用 Llama 3 实现了该原型；这是一个通过 Ollama Python 软件包在本地运行的开源大语言模型。我们使用*OllamaEmbedding* 对标本记录进行嵌入，并将其编入 *VectorStoreIndex*，从而通过*LlamaIndex* 基于与用户查询的语义相似度检索相关记录，实现检索增强生成。我们使用 Streamlit 实现了用户界面；Streamlit 是一个用于构建网络应用程序的开源 Python 库。

我们咨询了三位具备相关专业知识的博物馆资深员工：两位数字媒体制作人员和一位信息通信技术经理。在一场为时两小时的会议中，研究人员演示了原型，并讨论了它的潜力与局限。会议提出了两项主要建议：第一，将基于文本的馆藏摘要改为基于地图的可视化，因为地点信息和地理参考标本记录对于自然科学藏品的研究方式十分重要；第二，提出两个现有界面作为参考（见图 3）：ALA（澳大利亚生物图谱）的 Lens，它可视化展示了澳大利亚各地理区域的标本密度；以及澳大利亚博物馆的动物资料卡，其中通过包含图像和事实描述的可视化卡片呈现物种。

根据这些反馈，我们将馆藏摘要重新设计为交互式地图，按照采集地坐标标绘标本记录。用户可以点击单条记录，查看弹出卡片，其中显示

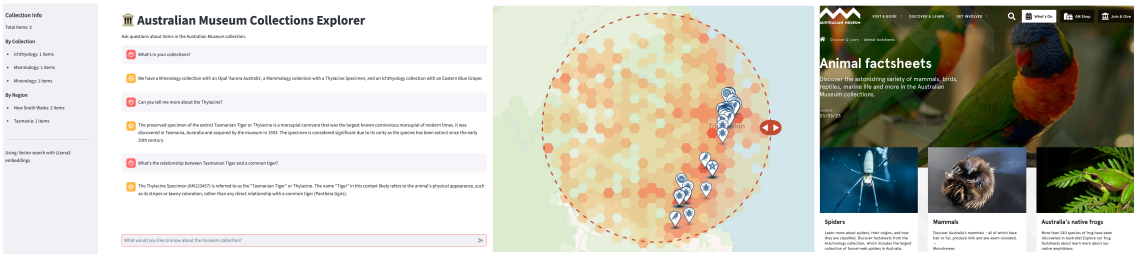


图 3. 我们开发的第一个交互式原型（左）；ALA Lens（中）；博物馆动物资料表（右）。

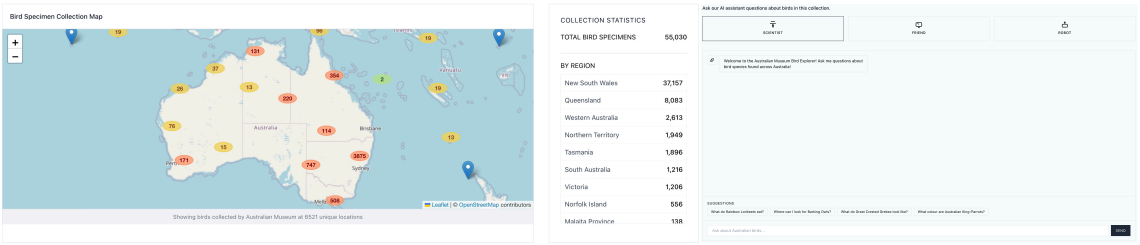


图 4. 我们开发的第二个交互式原型：一个鸟类藏品探索器，由交互式地图的初始版本（左）和对话代理（右）组成。

标本图像（如果有）及其来自藏品数据库的属性。这一改版旨在利用生物多样性数据的空间特性，同时以更丰富、更吸引人的方式呈现单个标本。

3.2.2 设计迭代 *II*：测试概念验证原型。 基于设计迭代 I 的反馈，我们开发了一个概念验证原型，其中纳入交互式地图，以便对标本记录进行视觉和地理探索（见图 4）。该原型被重建为客户端-服务器 Web 应用程序，并以博物馆数字化的鸟类学藏品作为代表性数据集（“鸟类藏品探索器”）。地图使用 React Leaflet.js 实现，并采用标记聚类来处理密集的标本分布；点击聚类可显示按地点分组的各个标本记录，点击标本记录则会弹出从维基百科检索到的物种图像和描述。后端使用 Python 和 Flask 实现。共计 59,295 条鸟类学标本记录存储在托管于亚马逊云科技（AWS）关系数据库服务（RDS）上的 PostgreSQL 数据库中。对话代理由 OpenAI 的 gpt-4-nano 模型驱动。该模型判断用户问题是需要数据库查询，还是可以根据一般知识回答。对于依赖数据库的查询，模型会动态生成 SQL 查询以检索相关记录。我们使用不同的系统提示词实现了三种响应角色——科学家、朋友和机器人，以提供不同程度的正式性和详细程度。随后，我们与博物馆志愿者开展了用户研究，以评估该原型的可用性及改进方向。

参与者。 我们从澳大利亚博物馆招募了12名数字志愿者（DigiVol 志愿者）参与用户研究。DigiVol 是该博物馆的公民科学项目，公众志愿者在其中协助将实体标本数字化。参与者通过电子邮件和口耳相传的方式招募。本用户研究已获悉尼科技大学人体研究伦理委员会批准。表2总结了参与者人口统计特征及其背景经历。

表2。设计迭代 II 中参与者概况总结（$N = 12$）。

Item	总结
年龄（$M \pm SD$）	51.6 ± 24.9 年
性别	7 女性 , 5 男性
既往人工智能经验	4 人经常使用，5 人使用有限，3 人从未使用
线下博物馆参与	每周均参与（志愿者工作）
展厅参观	每月2次，每两个月3次，每年几次7次
远程博物馆参与	每周3次，每月2次，每年几次5次，从不2次

实验流程。 每次实验均以单独形式进行，持续约30–45分钟。参与者阅读研究说明书并签署知情同意书。随后，他们通过笔记本电脑访问原型，并在出声思考的同时自由探索界面。实验未设定具体任务；参与者被要求按照自然状态与原型互动，并在认为自己已充分探索后停止。探索阶段由一名研究人员进行观察，但除非参与者请求帮助，否则不会介入。探索结束后，我们开展了一次半结构式访谈，内容涵盖初步印象、感知到的实用性、地图和对话代理组件的可用性，以及改进建议（完整的访谈问题见附录A）。实验过程均进行了录音，并手动转录以便进行数据分析。

结果。 参与者平均向对话代理提出10个问题（$M = 9.7$, $SD = 7.9$）。我们还使用主题分析法对访谈数据进行了进一步分析 [5]。分析得出了四个主要主题：(1) 视觉空间探索的价值（$N = 10$）；(2) 对标本图像的需求（$N = 9$）；(3) 无依据人工智能回复的不可靠性（$N = 5$）；(4) 多种响应角色的复杂性（$N = 4$）。根据主题分析结果，我们进一步提出了设计启示，以指导最终系统的设计。

主题*1*：视觉空间探索的价值。 大多数参与者都认可基于地图的交互，因为这种视觉呈现让地理特征更易于理解，正如P2所说：“我是视觉型学习者。如果能看到它们分布在哪里、来自哪里，我就能更好地理解一些。”这种交互有时会带来偶然发现，参与者在熟悉的地区发现了不熟悉的物种，正如 P3 所解释的：“ 它提供了一些鸟类名称，是我在探索该地区时没想到会出现的。”

设计启示*1*：保留并增强地图交互。用户对地图界面的积极反馈，验证了它是直观呈现馆藏的有效方式。最终系统旨在处理规模更大的数据集，涵盖鸟类学以外的馆藏，并改进对同地标本的展示。

主题*2*：需要标本图像。第二个主要主题是用户希望在地图界面和对话代理的回复中看到标本图像。P4表示：“我不太了解鸟类……找到鸟的时候，我希望能看到鸟的图像。”而P12则建议：“或许能够查看某个区域，并看到那里有些什么东西的照片，我觉得这样会很不错。”还有几位参与者建议提供鸟鸣音频，以帮助辨认鸟类。

设计启示*2*：整合馆藏专属图像。最终系统并未依赖维基百科等外部来源获取物种图像，而是设计为直接从数据存储库检索标本图像，在地图弹窗中显示，并根据用户请求在对话代理的回复中提供。

主题*3*：无依据人工智能回复的不可靠性。 一些参与者觉得智能代理的回答缺乏细节，正如 P6 对含糊的位置解析所评论的那样：“显然，有些内容对我想要的东西来说太笼统了……如果我想了解更具体的地点，就没那么有用了。” 参与者有时会测试一些自己已经知道答案的问题，并发现人工智能的回答并不准确。P9 指出，对系统的信任“取决于它们有多准确，究竟是假新闻还是不是。” 智能代理回答笼统且不可靠，主要是由以下原因造成的

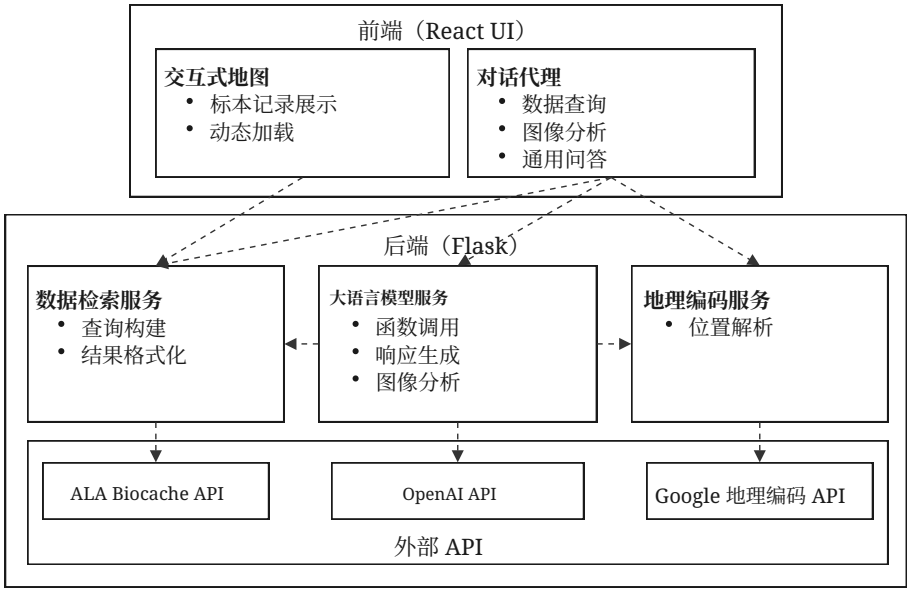


图5. Australian Museum CollectionExplorer 的系统架构概览。

原型的数据管道设计：当大语言模型生成错误的 SQL 查询并且无法检索数据时，系统便会退而使用模型的一般知识，这有时会导致错误。

设计启示*3*：通过函数调用实现基于数据的检索。最终系统没有通过生成 SQL 查询来查询云数据库，而是经过修订，采用函数调用机制，从可通过应用程序接口访问的数据基础设施中动态请求结构化数据。这种方法能够扩展系统可访问的数据量并提高数据的时效性，同时提升数据检索的准确性，因为模型调用的是带有经过验证的参数的预定义函数，而不是生成不受架构约束的 SQL 查询。

主题*4*：多个响应角色的复杂性。三种响应模式（科学家、朋友、机器人）令部分参与者感到困惑，正如 P2 所说：“我一开始不太清楚它们之间有什么区别……能获得各种各样的回答固然不错，但我大概只需要科学家模式的回答就够了。”P7 也指出：“我觉得如果你是第一次使用这个界面，看到不同的人在回答，可能会有点不知所措……我认为只保留两种模式就可以了。”参与者表达偏好时，始终倾向于科学家模式，因为它能提供更丰富的信息。

设计启示*4*：提供单一响应模式。三个角色设定选项被精简为一个，因为参与者认为它们之间的差异微乎其微，并且更倾向于科学家模式。

4 实现澳大利亚博物馆藏品探索器

4.1 系统架构

图5展示了澳大利亚博物馆藏品探索器的系统架构。该应用程序由基于 React 的前端、Flask 后端服务器，以及用于藏品数据检索、自然语言交互和地理解析的外部 API 集成构成。该应用程序使用 JavaScript（React）开发前端

界面，并使用 Python 与 Flask 框架开发后端服务器。前端网页界面和后端服务器部署并托管在云平台（Vercel[3] 和 Render[4]）上。前端提供两个主要界面组件：用于标本记录视觉空间探索的交互式地图，以及用于自然语言查询的对话代理。后端在用户界面与三项外部 API 服务之间编排请求：用于标本采集数据的澳大利亚生物图谱（ALA）Biocache API、用于将地点名称解析为地理坐标的 Google 地理编码 API，以及用于语言模型和函数调用以从 ALA Biocache 与 Google 地理编码 API 检索数据的 OpenAI API。

4.2 数据来源

系统从澳大利亚博物馆在线动物学藏品数据集中检索并查询标本记录。该数据集通过澳大利亚生物图谱[5] ——澳大利亚的国家生物多样性数据聚合平台与数据基础设施——进行访问。截至2026年2月20日，共有1,685,922条具有有效地点信息的标本记录被动态检索并显示在交互式地图上，可供大语言模型服务查询。该数据集包含澳大利亚博物馆生命科学藏品中的部分数字化标本记录，涵盖昆虫学、海洋无脊椎动物、鸟类学、鱼类学、软体动物学、哺乳动物学、爬虫两栖动物学和蛛形动物学。ALA Biocache API提供了对标本记录属性的访问，包括分类学信息（学名、俗名以及从界到物种的分类阶元）、地理数据（地点描述、州和坐标）、采集日期、采集者信息、目录编号和标本图像。作为数据提供方，澳大利亚博物馆会频繁更新该数据集。

4.3 对话式智能体实现

我们采用 OpenAI 的 gpt-5-mini 模型作为自然语言理解与生成引擎。为利用函数调用能力，我们配置了模型的工具定义组件，使我们能够向大语言模型提供描述可用函数的 JSON 架构规范。我们编写了这些函数，用于向外部 API 发送包含相关参数的结构化 HTTP 请求，并从响应中获取数据。主要有三个函数：`搜索_标本`，用于根据分类学、地理、时间范围、采集者信息和图像可用性等筛选条件查询标本记录；`获取_标本_统计数据`，用于返回汇总计数和分面分布；以及`按_标本_编号_查询`，用于通过目录编号检索特定标本的详细信息。详细的 JSON 架构和函数实现见附录 B.1 和 B.2。

下面列出了用户自然语言查询的典型处理流程。附录 B.3 中提供了一个包含技术细节的数据查询与响应流程示例。

(1) **意图识别：**语言模型判断该查询是需要藏品数据，还是可以依据一般知识回答。关于特定标本、采集统计以及日期或地理分布的查询会调用函数调用机制。(2) **参数提取：**对于数据查询，模型从自然语言输入中提取参数。例如，查询“显示20世纪80年代来自新南威尔士州的袋鼠标本”得到参数`俗名_名称`：“袋鼠”，`州_省`：“新南威尔士州”，`年份_范围`：`起始_年份`：*1980*，`结束_年份`：*1989*，并触发对函数`搜索_标本`的调用。

[3] https://vercel.com/ [4] https://render.com/ [5] https://www.ala.org.au/

(3) 分类学解析：当使用俗名的初始查询未返回结果时，系统会尝试通过 ALA BIEAPI 自动转换名称。该服务将俗名解析为相应的学名，从而使查询能够使用分类标识符重试。反向转换（学名到俗名）同样受支持。**(4) 地理解析：**用户查询中的地点指称（例如郊区或城市名称）通过 Google 地理编码 API 解析为地理坐标。如果同一名称返回多个匹配项（例如，“Castle Hill”同时存在于新南威尔士州和昆士兰州），系统会筛选到用户指定的州，或查询所有匹配地点。**(5) API 执行：**后端构建查询并针对 ALA Biocache API 执行，同时应用所有已解析的筛选条件。响应可包括匹配的标本记录、分面计数与分布、标本图像，以及指向 ALA 网页界面的直接 URL。**(6) 响应生成与清理：**函数结果会返回给语言模型，由其基于检索数据综合生成自然语言响应。在将生成的响应发回用户之前，会应用后处理模块，移除偶尔出现的技术性叙述，并修正格式错误的 URL。

4.4 交互式地图集成

交互式地图组件使用 Leaflet.js 实现，并为每条标本记录设置了可点击标记，以支持可视化探索。为管理海量数据，地图采用了基于视口的动态加载方式：仅获取当前地图边界内的标本记录，并在用户缩放或拖动地图时重新加载更新后的记录。点击标记会打开弹出窗口，显示标本详细信息，例如俗名和学名、目录编号、采集者姓名、日期、地点，以及（如有）标本图像。对于包含多个同地标本的地点，弹出窗口采用轮播设计，方便用户在不同记录间切换。用户还可以通过切换选项，仅显示带有图像的记录，或显示所有记录（不论图像是否可用）。

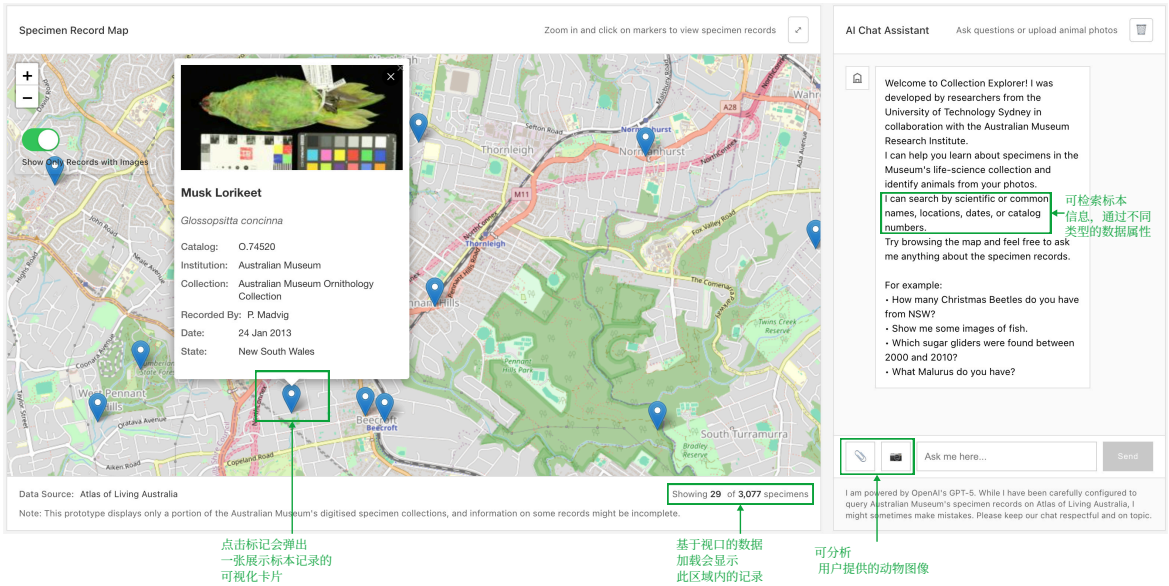


图 6. 澳大利亚博物馆藏品探索器界面，展示了带有弹出窗口的交互式地图，窗口中显示一件虹彩吸蜜鹦鹉标本的详细信息（左），以及显示介绍消息的对话代理（右）。

4.5 用户界面与使用场景

本节介绍澳大利亚博物馆藏品探索器的用户界面，并通过典型示例展示主要交互场景。

4.5.1 界面概述。该应用程序采用分屏布局，左侧为交互式地图，右侧为对话代理。图 6 展示了用户界面的一个示例概览：地图面板在当前视口中以可点击标记的形式显示标本记录；放大后可以看到更多单个标本，点击标记会打开弹出窗口，显示标本的图像（如有）、俗名和学名、目录编号及其他元数据。对话代理面板允许用户发送自由文本并上传图像进行分析，同时显示一条介绍消息，其中列出示例问题以引导新用户，并附有免责声明，指出人工智能模型的局限性。

4.5.2 场景 *1*：查询藏品数据。对话代理的一项核心能力，是通过实时检索相关数据来回答有关博物馆标本藏品的问题。图 7 展示了两个具有代表性的查询。在第一个示例中，用户询问：“2000 年至 2010 年间发现了哪些蜜袋鼯？”系统识别出该查询需要进行标本检索，提取参数（物种：蜜袋鼯；日期范围：2000–2010），并调用 ALA Biocache API 执行检索_标本功能。响应列出了符合条件的记录数量，并展示了具有代表性的标本及其目录编号、日期、地点和采集者。在第二个示例中，用户询问：“你们有多少来自新南威尔士州的圣诞甲虫？”系统调用获取_标本_统计功能，并返回新南威尔士州金龟科标本的准确数量。重要的是，每条响应都附有指向 ALA 的直接 URL，用户可据此访问数据源、核实结果并查阅更多记录。这种透明度旨在增强用户对系统响应的信任，并为希望在界面之外开展更深入调查的用户提供途径。

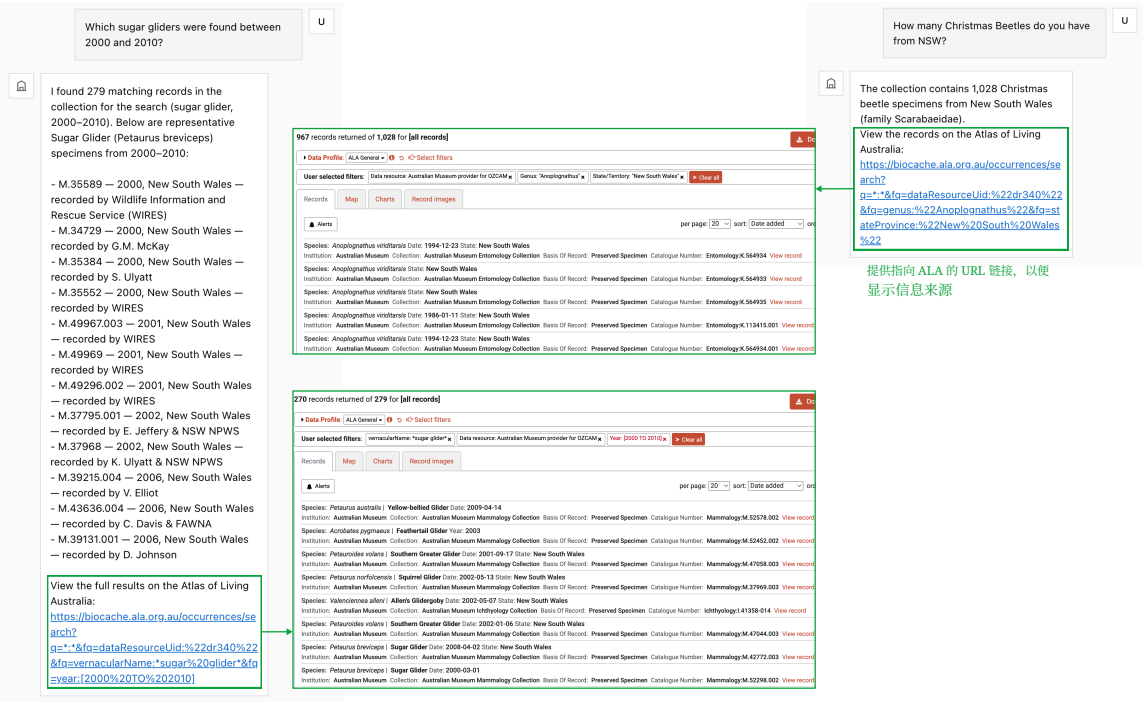


图 7. 藏品数据查询示例：搜索 2000–2010 年采集的蜜袋鼯（左）；查询新南威尔士州的圣诞甲虫数量（右）。两条回复均包含指向 ALA 的直接链接（信息来源），以供核实并进一步探索。

4.5.3 场景 *2*：多模态对话。系统支持双向的基于图像的交互：根据查询从馆藏中检索标本图像，以及分析用户上传的图像以识别动物。

图像输出。当用户请求查看标本图像时，对话代理会通过 ALA Biocache API，直接从博物馆的数字化馆藏中检索照片并提供链接。图 8 和图 9 展示了两个示例：“向我展示麦克利燕凤蝶的图像”会检索该物种的标本照片和馆藏元数据；“向我展示来自昆士兰州的鱼类图像”则会返回该地点最多五条排名靠前的鱼类标本图像结果，以及它们的学名、俗名和元数据。此功能支持用户探索基于图像的馆藏数据，并通过视觉模态加深对物种的了解。

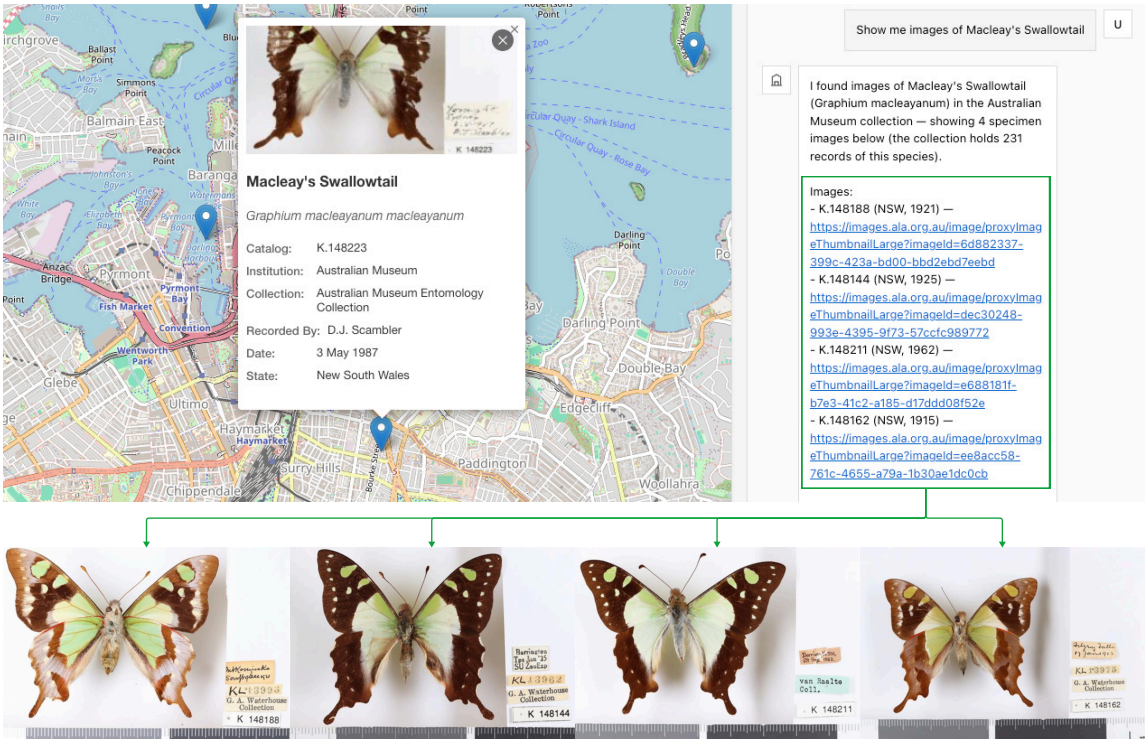


图 8. 图像输出示例 1：在地图上发现一条麦克利燕凤蝶记录，并请求查看该物种的更多图像。

图像输入。用户还可以上传图像，既可以从文件中选择，也可以使用设备摄像头实时拍摄，以进行图像分析。系统利用语言模型内置的视觉能力分析上传的图像，并返回物种鉴定结果及相关描述信息。图 9 展示了一个示例：系统对在花园中拍摄的冠鸠进行了鉴定，并描述了其鉴别特征、栖息地、食性和行为。回复还包含已鉴定物种的博物馆标本记录链接，将现实世界观察与数字化藏品联系起来。



图 9. 图像输出示例 2（左）：请求昆士兰州的鱼类标本。图像输入示例（右）：将一张在野外拍摄的鸟类照片识别为冠鸠，并提供描述信息以及相关博物馆记录的链接。

5 讨论

5.1 从泛在访问到泛在能动性

本文开发了一个系统，供用户探索澳大利亚博物馆数字化馆藏中的近 170 万条生命科学标本记录。基于地图的交互通过可视化呈现和对生物多样性的空间探索，有效展现了馆藏的规模与丰富程度。本项目的核心工作还包括开发能够提供特定馆藏信息的对话式人工智能代理，使用户不仅可以浏览界面上呈现的内容，还能提出关于馆藏的问题。

对话式人工智能的兴起，或许标志着数字化藏品的探索方式正发生更广泛的转变。在过去二十年里，随着普适计算技术和移动电话等个人设备带来**泛在访问**，借助这些技术的数字博物馆应用不断提升博物馆馆藏的可及性，并促进文化与科学教育。相关实例包括基于移动应用的公民科学项目，例如记录本地环境中的蛙鸣，以帮助发现物种分布 [29],；也包括在家中通过虚拟现实远程参观博物馆 [15], ，突破博物馆与公众交流的物理和时间界限。然而，尽管访问日益普及，应用中的查询方式和展示内容仍主要由机构主导并预先设定。大多数应用仍规定着反馈循环的结构和用户学习的路径，从而限制了用户自主塑造探索过程的能力。

随着工业5.0强调人机协作和以人为中心， [36, 38], 我们有机会超越泛在访问，迈向**泛在能动性** ——用户不仅可以访问数字化藏品，还能主导对其进行的探索。过去两年，生成式人工智能技术展现了新的

可能性；例如，大语言模型能够围绕用户的兴趣展开对话， [26], 扩散模型则可以实时生成定制化可视化 [18]。这些能力让我们得以设想这样一个未来：知识系统分布式发展，由用户驱动，个人能够实时以自己的方式探索馆藏。本项目开发的对话代理正是朝这一方向迈出的一步：用户不再受限于预先设定的查询或导航结构，而是可以按照自己的思路进行探究，并直接从藏品数据中获取回复。

在这一新范式下，博物馆应当拥抱个体的好奇心，转向**好奇心驱动的交互模型**，使数字化藏品不再只是静态的知识库，而成为一个灵活、响应迅速的系统，以及驱动其交互性的技术“燃料”。博物馆不再仅仅扮演信息策展者的角色，而应成为个性化、开放式探索的引导者；技术则用于连接充满好奇心的个体与自然史。

5.2 将馆藏重新融入现实世界环境

长期以来，博物馆一直是人们前来欣赏展出展品的场所。尽管亲临现场参观仍是人们参与博物馆活动的主要方式，数字技术却带来了体验馆藏的新方式，从基于手势的桌面交互和触摸屏交互 [21] ，到借助沉浸式虚拟现实或增强现实设备 [35]，通过感官与动作带来沉浸体验的扩展现实应用。尽管这类交互让知识变得更加具体、更易于探索，但它们往往局限于博物馆空间的实体围墙之内，或缺乏与参观者日常环境更深层次的融合。

焦点小组的一项发现是，当展览与人们的日常生活产生联系时，会变得更具吸引力。同样，在用户测试中，参与者常常会先探索地图上自己居住或工作的地点，或者询问有关自己曾去过的地点的标本的问题。这种行为表明，人们天生倾向于将馆藏与亲身经历联系起来——这为个人叙事和故事提供了自然的起点，也赋予了数字探索之旅以意义。对于自然史藏品而言，**情境化交互** 在传达科学知识和文化知识方面尤其有价值。通过将数字化藏品嵌入生活环境，它们有望成为交织于人们日常空间和活动中的“隐形层”。本项目开发的基于地图的藏品探索器也有助于实现这一目标：它按地理位置呈现标本，使用户能够将生物多样性知识与对自己有意义的地点联系起来。由于用户可以通过设备上的任何浏览器访问该应用程序，因此有望将系统拓展为现场移动体验。我们也正在规划未来项目，以拓展当前系统并支持这一方向。

5.3 局限性与未来工作

本研究存在若干值得指出的局限性。首先，设计迭代 II 期间开展的用户评估涉及数字志愿者参与者，他们此前已通过每周的现场数字化工作与澳大利亚博物馆建立联系。这一群体基于对馆藏的熟悉程度提供了宝贵反馈，但未必能代表更广泛的受众，尤其是普通公众或此前很少接触自然史藏品的远程用户。我们已部署最终版澳大利亚博物馆藏品探索器，可通过 URL 链接访问，并在用户界面中集成了在线调查，以收集用户体验数据。为评估系统在普通公众用户中的表现，我们正开始为期六个月的纵向数据收集。

其次，系统的技术架构依赖外部 API 服务，并因此受到第三方可用性、运营成本和速率限制等因素的影响。此外，尽管函数调用机制能够使对话代理的回答以检索数据为依据，但当查询超出已定义的函数模式范围，或 API 调用失败时，大语言模型仍可能生成不准确的回答。正在进行的纵向部署将使我们能够大规模收集有关这些技术层面的实证数据，包括运营成本和响应延迟的精确估算。这也将使我们能够收集用户提出的更全面的问题类型，并评估系统对用户体验的影响。

6 结论

本文介绍了一个名为澳大利亚博物馆藏品探索器的交互式系统。该应用程序通过交互式地图支持对近170万条数字化标本记录进行视觉—空间探索，并使用户能够通过基于最新大语言模型服务的函数调用机制实现的对话代理提出开放式问题。该人工智能代理基于检索数据提供针对馆藏的回答。

我们着重介绍了藏品探索器的设计与开发历程。对以人为中心的设计过程的记录使我们的研究可被其他博物馆复现，并为自然历史机构设计未来的科学人工智能代理提供参考。该系统还展示了一种将基于大语言模型的对话代理纳入领域专用应用的快速部署方法。

所提供的技术实现不仅可应用于自然科学藏品，也可应用于其他类型的数字化博物馆馆藏，例如文化 artefacts 和历史艺术品。因此，本研究为考虑采用类似数字技术的机构提供了实用指导。

参考文献

[1] Ishita Agarwal、V Sakthivel和P Prakash。2025。迈向包容性医疗保健：用于初步诊断的基于大语言模型的多模态聊天机器人。*IEEEAccess*（2025）。[2] Mahmoud Amgad、Bahaa Shamoon、Hagar Mekhemar、Rodeen Mostafa、Toqa Tamer、Habiba Mohamed和Sahar Fawzi。2025。医疗数据数字化：基于大语言模型的数据录入、清洗与可视化解决方案。载于*2025*年第*22*届国际学习与技术会议（*L&T*），第22卷。IEEE，339–342。[3] Refik Anadol。2024。Unsupervised。https://www.moma.org/calendar/exhibitions/5535 访问日期：2025年2月2日。[4] Fabrizio I Apollonio、Michele Zannoni、Filippo Fantini、Simone Garagnani和Luca Barbieri。2025。博物馆馆藏三维模型的精确可视化与交互：从采集到数字展亭。*ACM* 计算与文化遗产期刊 18, 1（2025），1–25。[5] Virginia Braun和Victoria Clarke。2006。在心理学中运用主题分析。*Qualitative Research in Psychology* 3, 2（2006），77–101。[6] Dany Diaz、Felipe Moreno-Vera、Juanpablo Heredia、Fabrício Venturim和Jorge Poco。2024。FishBiasLens：整合大型语言模型与可视分析以进行偏差检测。载于*2024*年*IEEE* 可视分析科学与技术 *VAST* 挑战赛。IEEE，17–18。[7] Maria Padilla Engstrøm和Anders Sundnes Løvlie。2025。将大语言模型用作交互式博物馆装置的设计材料。载于*2025*年*ACM*交互系统设计会议伴随出版物。386–390。[8] Kexue Fu、Ruishan Wu、Yuying Tang、Yixin Chen、Bowen Liu和RAY LC。2024。“一点点被侵蚀”：通过展示生成式人工智能共创成果，提升文化遗产传播中的参与度与叙事。载于*2024*年*ACM* 交互系统设计大会论文集。2833–2850。[9] Remo Grillo、Caterina Morelli、Vincenzo Rubano和Fabio Vitali。2025。提升科学博物馆馆藏可及性的方法（SCAMP）。*ACM* 计算与文化遗产期刊 18, 2（2025），1–20。[10] Maarten Heerlien、Joost Van Leusen、Stephanie Schnörr、Suzanne De Jong-Kole、Niels Raes和Kirsten Van Hulsen。2015。自然历史生产线：科学收藏数字化的工业化方法。计算与文化遗产期刊（*JOCCH*）8, 1（2015），1–11。[11] Isabella Hodgson、Dennis Hoffmann、Luca Junge、Frederic Theis和Nicolas Dittgen。2024。学会正确倾听：基于迭代式共创的德国海事博物馆参与式数字项目。*ACM* 计算与文化遗产期刊 17, 3（2024），1–19。[12] Yaxin Hu、Arissa J Sato、Jingxin Du、Chenming Ye、Anjun Zhu、Pragathi Praveena和Bilge Mutlu。2025。NarraGuide：用于远程地点探索的基于大语言模型的叙事型移动机器人。载于第*38*届*ACM* 用户界面软件与技术研讨会年会论文集。1–15。

[13] Burhan Rashid Hussein、Owais Ahmed Malik、Wee-Hong Ong和Johan Willem Frederik Slik。2022。计算机视觉与机器学习技术在数字化植物标本中的应用：系统性文献综述。Ecological Informatics 69（2022），101641。[14] Sichen Jin和Shrirang Abhyankar。2024。ChatGrid：由大语言模型赋能的电网可视化。载于2024年IEEE能源数据可视化研讨会（EnergyVis）。IEEE，12–17。[15] Thomas Kersten、Felix Tschirschwitz和Simon Deggim。2017。开发一座虚拟博物馆，其中包括以虚拟现实呈现的建筑史4D呈现。载于TC II与CIPA复杂建筑的三维虚拟重建与可视化，2017年3月1–3日，希腊纳夫普利奥。Copernicus，361–367。[16] Minhyuk Ko、Dibyendu Brinto Bose、Weilu Wang、Mohammed Seyam和Chris Brown。2024。理解大语言模型生成SQL 查询的性能。载于2024年IEEE可视化语言与以人为中心的计算研讨会（VL/HCC）。IEEE，359–361。[17] Richard A Krueger。2014。焦点小组：应用研究实用指南。Sage出版社。[18] Peter Kun、Matthias Freiberger、Sebastian Risi和Anders Sundnes Løvlie。2024。GenFrame：为任意女孩绘制肖像的交互式画框。载于2024年ACM交互系统设计会议伴随出版物。353–358。[19] Babymol Kurian、A Aafreen Fathima、T Afra Fathima和R Shahista Begum。2024。GovInfohub：促进知情参与与可及性的动态政府计划聊天机器人。载于2024年计算、通信与应用信息学进展国际会议（ACCAI）。IEEE，1–6。[20] Jingjing Li、Xiaoyang Zheng、Ikumu Watanabe和Yoichi Ochiai。2024。博物馆展览中数字化转型技术的系统性综述。Computers in Human Behavior（2024），108407。[21] Duri Long、Tom McKlin、Anna Weisling、William Martin、Hannah Guthrie和Brian Magerko。2019。协同创作式博物馆装置中身体参与与表达的轨迹。载于2019年创造力与认知会议论文集。246–257。[22] Zhicong Lu、Peng Tan、Yi Ji和Xiaojuan Ma。2022。工艺+ 制作研讨会：引导学生参与以非物质文化遗产为导向的创意设计。载于2022年ACM 交互系统设计大会论文集。1071–1084。[23] M Maragadhavalli Meenakshi、M Selvi、R Srimathi和A Aravindhan。2025。上下文感知SQL查询生成：增强关系数据库中的自然语言交互与查询关联。载于2025年信息技术创新趋势国际会议（ICITIIT）。IEEE，1–6。[24] 澳大利亚博物馆。2025。展览虚拟导览。https://australian.museum/exhibition/virtual-exhibition-tours/?gad_source=1&gclid=CjwKCAiAqfe8BhBwEiwAsne6gYaJH4ZAgWi3-EUibkAKHD_faJJayl06BnweRnbayP94LqhAUvy3ERoCeJIQAvD_BwE 访问日期：2025年2月2日。[25] 大英博物馆。2024。The Living Museum。https://www.livingmuseum.app/explore 访问日期：2025年2月2日。[26] 剑桥大学。2024。公众受邀在新颖的人工智能实验中与博物馆动物对话。https://www.cam.ac.uk/research/news/public-invited-to-chat-to-museum-animals-in-novel-ai-experiment 访问日期：2025年2月2日。[27] Pavan Kartheek Rachabatuni、Filippo Principi、Paolo Mazzanti和Marco Bertini。2024。使用多模态大语言模型、面向文化遗产的上下文感知聊天机器人。载于第15届ACM多媒体系统会议论文集。459–463。[28] Constance Rinaldo、Diane Rielinger、Joseph Deveer和Danielle Castronovo。2023。连接图书馆、档案馆与博物馆：支持自然历史科学的馆藏。ACM 计算与文化遗产期刊 16, 1（2023），1–24。[29] Jodi JL Rowley、Corey T Callaghan、Timothy Cutajar、Christopher Portway、Kathy Potter、Stephen Mahony、Dane F Trembath、Paul Flemons和Adam Woods。2019。FrogID：公民科学家提供经验证的澳大利亚青蛙生物多样性数据。Herpetological Conservation and Biology 14, 1（2019），155–170。[30] Reza Samimi、Aditya Bhattacharya、Lucija Gosak、Gregor Stiglic和Katrien Verbert。2025。用于糖尿病风险预测之基于证据解释的视觉对话界面。载于第7届ACM对话式用户界面会议论文集。1–18。[31] Sonal Singh、Yadunath Gupta和Soudip Roy Chowdhury。2024。MEQA——使用多智能体大语言模型的多模态交互式企业问答系统。载于第8届数据科学与数据管理国际会议论文集（第12届ACM IKDD CODS与第30届COMAD）。422–426。[32] Irini Stamatoudi和Konstantinos Roussos。2024。面向博物馆与文化遗产机构的可持续文化遗产管理模式。ACM 计算与文化遗产期刊（2024）。[33] Houjun Tang和David McCallen。2025。EQSIM Agent：用于大规模地震模拟数据交互式探索的对话式人工智能。载于高性能计算、网络、存储与分析国际会议SC'25研讨会论文集。582–587。[34] Robyn Taylor、John Bowers、Bettina Nissen、Gavin Wood、Qasim Chaudhry、Peter Wright、Lindsey Bruce、Sarah Glynn、Helen Mallinson和Roy Bearpark。2015。创造魔法：为博物馆环境中的开放式交互而设计。载于2015年ACM SIGCHI创造力与认知会议论文集。313–322。[35] Kihiro Tokuno、Fusako Kusunoki、Shigenori Inagaki和Hiroshi Mizoguchi。2024。Talkative Museum：面向协作式儿童—家长—标本互动的增强现实交互式博物馆导览系统。载于第23届ACM交互设计与儿童会议年会论文集。754–758。[36] Attila Tóth、László Nagy、Roderick Kennedy、Belej Bohuš、János Abonyi和Tamás Ruppert。2023。以人为中心的工业5.0协作架构。MethodsX 11（2023），102260。[37] Georgios Trichopoulos、Kostas Ordoumpozanis和George Caridakis。2025。对基于大语言模型的聊天机器人提升文化展品参观者用户体验的评估。ACM 计算与文化遗产期刊（2025）。

[38] 欧盟。2024。《工业5.0》。
https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry- 50 en 访问日期：2025年2月6日。_[39] V Usha、Nalagarla Chiru Abhinash、Sakhamuri Nitin Chowdary、V Sathya、Eeda Ramakrishna Reddy 等。2024。使用大型语言模型增强数据库交互，以改进数据检索与分析。载于*2024*年智能信息物理系统与物联网第二届国际会议（*ICoICI*）。IEEE，第1302–1306页。[40] Sridevi Wagle、Sai Munikoti、Rounak Meyur、Jonathan Whiting、Hayley Farr、Anurag Acharya、Sameera Horawalavithana、Mahantesh Halappanavar、Jan Strube 和 Laura Fierce。2025。利用多模态人工智能实现风能研究中的高效数据发现。载于*2025*年高级研究计算实践与经验：协作的力量。第1–3页。[41] Huan Wang 和 Andrii Matviienko。2025。借助生成式人工智能聊天机器人体验艺术博物馆。载于*2025*年*ACM*国际交互媒体体验会议论文集。第430–436页。[42] Zhiyang Wang、Devin Bayly 和 Carolina Roe-Raymond。2025。多模态大语言模型辅助的大规模可视化馆藏可视化。载于*2025*年高级研究计算实践与经验：协作的力量。第1–4页。[43] Yuting Zhou、Juanjuan Chen 和 Minhong Wang。2022。关于将虚拟现实和增强现实融入博物馆学习的元分析综述。*Educational Research Review* 36（2022），100454。

访谈方案

The 以下问题用于在设计迭代 II 期间开展的半结构式访谈 .

人口统计信息

- 您今年多大?
- 您如何描述自己的性别?

先前经验

- 您使用过生成式人工智能产品吗（例如 ChatGPT、Claude、Copilot、Gemini、Midjourney）？ 如果使用过，频率如何?
- 您多久参观一次澳大利亚博物馆?
- 您多久使用一次澳大利亚博物馆的数字产品（网站、应用程序、虚拟现实项目）？
- 您是否通过任何方式探索过澳大利亚博物馆的馆藏（亲自参观、在线浏览、使用软件或参观展览）？

原型评估

- 您对这个应用程序的初步印象如何?
- 你为什么会提出那些问题?
- 你认为聊天机器人的回答有多大用处？你的评估标准是什么?
- 你认为这个应用程序在哪些方面有用或没有用?
- 你觉得使用地图和聊天机器人探索鸟类藏品的体验如何?
- 你喜欢这个应用程序的哪些方面？你不喜欢这个应用程序的哪些方面?
- 你认为这种应用程序的完整开发版本可以如何协助你的工作?
- 您对如何改进这项技术或使用体验还有其他建议吗?

B 函数调用规范

本附录介绍对话代理中函数调用机制的工具定义（B.1 小节）、函数实现（B.2 小节），以及数据查询与响应流程（B.3 小节）。

B.1 工具定义

工具定义是描述语言模型可用函数的 JSON 架构。清单 1 展示了主要搜索函数的架构。其余两个函数采用类似的架构结构，汇总见表 3。

清单 1. search_specimens 函数的 JSON 架构

```
{
   "type": " function ", " function ": { "name": " search_specimens ", " description ": " 通过 ALA
   Biocache API 搜索 OZCAM 标本数据集", " parameters ": { "type": " object ",
```

" properties ": { " scientific_name ": { "type": " string ", " description ": " 任意分类学层级的学名 " }, " common_name ": { "type": " string ", " description ": " 生物的通用名称 / 地方名称 " }, " state_province ": { "type": " string ", " description ": " 澳大利亚州或领地 " }, " locality ": { "type": " string ", " description ": " 具体地点（郊区、城市或地区）" }, " year_range ": { "type": " object ", " properties ": { " start_year ": {"type": " integer "}, " end_year ": {"type": " integer "} } }, " has_image ": { "type": " boolean ", " description ": " 按图像可用性筛选 " }, " limit ": { "type": " integer ", " description ": " 返回结果的最大数量 " } } } }
}

表 3. 工具定义摘要

函数	**用途**	**关键参数**
搜索_标本	使用筛选条件查询标本记录	学名_，俗名_，州_省， 地点，年份_范围，是否有_图像
获取_标本_统计信息	返回汇总计数和分面分布	学名_，俗名_，包含_分面
获取_标本_根据_id	检索详细的标本信息	标本_id（必填）

B.2 函数实现

我们项目中实现的 Python 函数将模型的结构化参数转换为发送至外部 API 端点的 HTTP 请求。本节以一个示例说明 search_specimens 函数如何连接到 ALA Biocache API 中相应的端点，处理针对新南威尔士州袋鼠标本的请求。

从大语言模型接收参数。语言模型从用户的查询中提取了结构化参数：

```
{
   " common_name ": " 袋鼠 ", " state_province ": "新南威尔士州 ", "
   year_range ": {" start_year ": 1980, " end_year ": 1989}

}
```

构建查询 ry 参数。该函数将这些参数转换为 ALA Biocache API 所需的参数格式 I:

```
参数 ={
    "q": "*:*",
    "fq": [
          ' 数据资源唯一标识符 :" dr368 "' ,
          ' 俗名 :* 袋鼠 *',
          ' 州或省 :" New ␣南部 ␣威尔士 "' ,
          '年份 :[1980␣TO␣1989]'
    ],
    " 页面大小 ": 10,
    " 分面 ": " 州或省 ,年份 , 科 "
}
```

向端点发送 HTTP 请求。 该函数向 ALA Biocache 端点发送了 GET 请求 :

```
响应 = 请求 .get(" https :// biocache -ws.ala.org.au/ws/ occurrences /
search ", params=参数 , timeout =60) 数据 =响应 .json ()
```

接收来自应用程序接口的响应。 ALA 服务器返回了标本记录，其字段名称由该应用程序接口定义：

```
{ " 记录总数 ": 47, " 出现记录 ": [{ "通用唯一标识符": "
a1b2c3d4 -e5f6 -7890" , " 学名 ": " Macropus giganteus ", " 俗
名 ": " 东部灰袋鼠 ", " 十进制度纬度 ": -36.45, " 十进制度经度 ":
148.26 , " 州或省 ": "新南威尔士州 ",
```

```
    " 年份 ": 1985 }]}
```

格式化并返回给大语言模型。该函数将检索数据重新整理为传递给大语言模型的格式：

```
{
  " 总数 _ 记录 ":  47 ,
  " 标本 ":  [
    {
      " 学名_ ":  " Macropus  giganteus ",
      " 通用 _name ":  " 东方  Grey  袋鼠 ",
      " 地点 ": {"state ":  "New  南  威尔士 "} ,
      "日期":  {"年份":  1985 }
    }
  ] ,
  " ala_url ":  " https :// biocache .ala.org.au /..."
}
```

B.3 数据查询与响应流程

本节展示了用户查询“显示 Castle Hill 附近的青蛙”到对话代理给出最终响应的数据流。图 10 展示了数据查询与响应流程。

步骤 1：用户发送查询。前端通过 POST 请求将消息发送到后端。**步骤 2：后端请求 OpenAI**。后端在应用程序接口请求中包含了用户消息和工具定义。**步骤 3：OpenAI 返回函数调用**。模型判断该查询需要进行数据查询：

```
{ " tool_calls ":  [{ " function ": { "name": " search_specimens ", " arguments ": "{\" common_name \":
\" 青蛙 \", \" 地点 \": \" Castle Hill \"}" } } ]}
```

步骤 4：后端解析地点。该函数向 Google 地理编码发送请求。应用程序接口返回坐标：纬度 −33.731，经度 151.004：

```
响应  =  请求.get( " https :// maps. googleapis .com/maps/api/
geocode /json", 参数 ={ " 地址 ": " Castle ␣Hill ,␣ 澳大利亚 ", " 地区 ": "
au", " 组件 ": " 国家 :AU",
```

```
    "密钥": 应用程序接口_密
钥 } )
```

第 5 步：后端检索标本记录。后端使用解析出的坐标向 ALA Biocache 发送了请求。API 返回了 23 条匹配的标本记录。

```
响应  =  请求 .get( " https :// biocache -ws.ala.org.au/ws/ occurrences /
search ", 参数 ={ "q": "*:*", "fq":  ["数据资源唯一标识符 :\" dr368 \"", " 俗名 :
* 青蛙*"] ,"纬度": -33.731 , "经度": 151.004 , " 半径 ": 5, "页面大小 ": 10 } )
```

第 6 步：将结果返回给 OpenAI。格式化后的函数结果被发送回语言模型。**第 7 步：OpenAI 生成响应**。模型利用检索到的数据综合生成了自然语言响应：*“*我找到了 *Castle Hill*、新南威尔士州附近的 *23* 个蛙类标本。该馆藏包括绿树蛙（*Litoriacaerulea*）、佩龙树蛙（*Litoria peronii*）和东部常见小蛙（*Crinia signifera*）。大多数标本采集于 *1985* 年至 *2005* 年之间。*”*

步骤 8：处 理后处理。 后端在必要时清理响应，并将其返回前端以供显示。

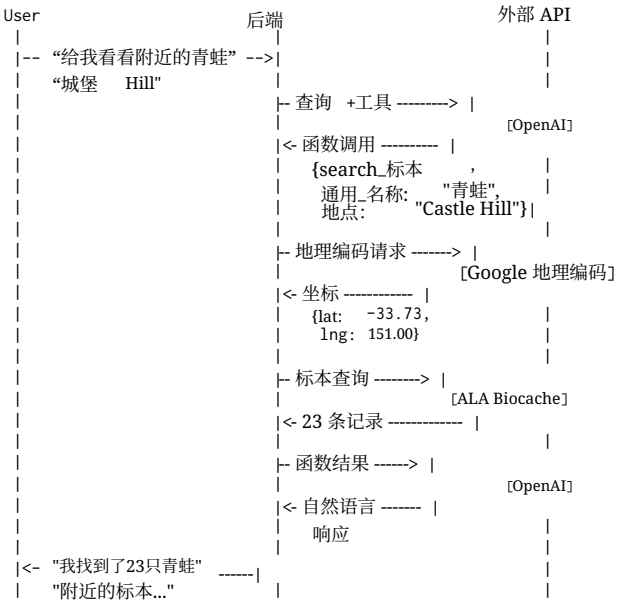


图10。示例用户消息的数据查询和响应流程。