The Graph 如何扩展为 AI 驱动的 Web3 基础设施？

作者：LINDABELL 来源：ChainFeeds

2022 年，OpenAI 推出了 GPT-3.5 模型驱动的 ChatGPT，自此开启了一波又一波的 AI叙事浪潮。然而，尽管 ChatGPT 在多数情况下能能有效处理问题，但其在需要特定领域知识或实时数据时的表现仍可能受限。例如，当询问关于 Vitalik Buterin 过去 18 个月内的代币交易记录时，它无法提供可靠且详尽的信息。为此，The Graph 核心开发团队 Semiotic Labs 结合 The Graph 索引软件栈和 OpenAI 推出了Agentc的项目，能够为用户提供加密货币市场趋势分析和交易数据查询服务。

### The Graph 去中心化 AI 路线图

The Graph 于 2018 年 7 月推出，是一个用于索引和查询区块链数据的去中心化的协议。通过该协议，开发者可以使用开放 API 创建和发布名为子图（Subgraph）的数据索引，使应用程序能够高效检索链上数据。截至目前，The Graph 已经支持超 50 条链、托管了超 7.5 万个项目，并处理了超过 1.26 万亿次查询。

The Graph 能够处理如此庞大数据，离不开其背后的核心团队的支持，包括 Edge & Node、Streamingfast、Semiotic、The Guild、GraphOps、Messari 和 Pinax。其中 Streamingfast 主要提供区块链数据流的跨链架构技术，Semiotic AI 则专注于将 AI 和密码学应用于 The Graph 中。The Guild、GraphOps、Messari 和 Pinax 则各自专注于 GraphQL 开发、索引服务、子图开发和数据流解决方案等领域。



The Graph 布局 AI 并非新想法。早在去年 3 月，The Graph Blog 就发布了一篇文章，概述了利用其数据索引功能进行人工智能应用的潜力。去年 12 月，The Graph 发布了名为「New Era」的新路线图，计划添加大型语言模型的 AI 辅助查询。随着近期白皮书的发布，其 AI 路线图进一步清晰。白皮书中介绍了两项 AI 服务：Inference（推理）和 Agent（代理）Service，允许开发者直接将 AI 功能集成到应用前端，整个过程均由 The Graph 支持。

#### Inference Service：支持多种开源 AI 模型

在传统推理服务中，模型通过中心化的云计算资源对输入数据进行预测。例如，当你向 ChatGPT 提出问题时，它会进行推理并返回答案。然而，这种中心化的方式不仅增加了成本，还存在审查风险。The Graph 希望通过构建一个去中心化的模型托管市场来解决这个问题，使得 dApp 开发者在部署和托管 AI 模型时更加灵活。

The Graph 在白皮书中举了一个例子，展示了如何通过其创建一个应用程序来帮助 Farcaster 用户了解其帖子是否会获得大量点赞。首先，使用 The Graph 的子图数据服务，对 Farcaster 帖子的评论和点赞数量进行索引。接着，训练神经网络预测新的 Farcaster 评论是否会被点赞，并将神经网络部署至 The Graph 的推理服务中。最终开发的 dApp 可以帮助用户撰写能够获得更多点赞的帖子。

这种方式使开发者能够轻松利用 The Graph 的基础设施，将预训练的模型托管在 The Graph 网络上，并通过 API 接口集成到应用中，从而用户在使用 dApp 时就能直接体验这些功能。

而为了给开发者提供更多选择和灵活性，The Graph 的 Inference Service 支持大部分现有的流行模型。其在白皮书中写到，「在 MVP 阶段，The Graph 的 Inference Service 将支持一组经过筛选的流行开源 AI 模型，包括 Stable Diffusion、Stable Video Diffusion、LLaMA、Mixtral、Grok 和 Whisper 等。」未来，任何经过足够测试和索引器操作的开放模型都可以在 The Graph Inference Service 中部署。此外，为了减少部署 AI 模型的技术复杂性，The Graph 提供了用户友好的接口，简化了整个过程，使开发者可以轻松上传和管理他们的 AI 模型，无需担心基础设施维护。

而为了进一步增强模型在特定应用场景下的表现，The Graph 还支持模型进行特定数据集的微调（fine-tuning）。但需要注意的是，微调通常不在 The Graph 上进行。开发者需要在外部对模型进行微调，随后再利用 The Graph 的推理服务部署这些模型。而为了鼓励开发者公开微调后的模型，The Graph 正在开发激励机制，例如在模型创建者和提供模型的索引器之间合理分配查询费用。

在验证推理任务执行方面，The Graph 提供了多种方法，如可信权威、M-of-N 共识、交互式欺诈证明和 zk-SNARKs。这四种方式各有优缺点，其中可信权威依赖于可信实体；M-of-N 共识需要多个索引器验证，增加作弊难度的同时也增加了计算和协调成本；交互式欺诈证明安全性较强，但不适用于需要快速响应的应用；而 zk-SNARKs 则是技术实现较为复杂，不适合大型模型。

The Graph 认为开发者和用户应有权根据自己的需求选择合适的安全级别。因此，The Graph 计划在其推理服务中计划支持多种验证方法，以适应不同安全需求和应用场景。例如，在涉及到财务交易或重要业务逻辑的场合，可能需要使用更高安全性的验证方法，如 zk-SNARKs 或 M-of-N 共识。而对于一些低风险或以娱乐为目的的应用，则可以选择成本较低、实施较简单的验证方法，如可信权威或交互式欺诈证明。此外，The Graph 还计划探索隐私增强技术，以改善模型和用户隐私问题。

#### Agent Service：帮助开发者构建自主 AI 驱动应用

相较于 Inference Service 主要是运行训练好的 AI 模型进行推理，Agent Service 更为复杂，其需要多个组件协同工作，以使得这些 Agent 能够执行一系列复杂和自动化的任务。The Graph 的 Agent Service 价值主张是将 Agent 的构建、托管和执行都集成至 The Graph，并由索引器网络提供服务。

具体而言，The Graph 会提供一个去中心化的网络，支持 Agent 的构建和托管。当 Agent 被部署在 The Graph 网络上后，The Graph 索引器便会提供必要的执行支持，包括索引数据、响应链上事件等各种交互请求。



正如上文提到的，The Graph 核心开发团队 Semiotic Labs 已经推出了一个早期 Agent 实验产品Agentc，结合了 The Graph 的索引软件栈和 OpenAI ，其主要功能是将自然语言输入转换为 SQL 查询，方便用户直接查询区块链上的实时数据，并将查询结果以易于理解的形式呈现给用户。简单理解，Agentc 专注于为用户提供便捷的加密货币市场趋势分析和交易数据查询，其所有数据均来自以太坊上 Uniswap V2、Uniswap V3、Uniswap X及其分叉，价格每小时更新一次。



此外，The Graph 还表示 The Graph 使用的 LLM 模型准确率仅有 63.41%，因此存在错误响应的问题。为了解决这一问题，The Graph 正在开发一种名为 KGLLM（Knowledge Graph-enabled Large Language Models）的新型大语言模型。

KGLLM 通过使用 Geo 提供的结构化知识图谱数据，能够显著减少生成错误信息的概率。Geo 系统中每个声明都由链上时间戳和投票验证支持。集成 Geo 的知识图谱后，代理可以应用于多种场景，包括医疗法规、政治发展、市场分析等，从而提升代理服务的多样性和准确性。例如，KGLLM 可以利用政治数据为去中心化自治组织（DAO）提供政策变更建议，并确保基于当前且准确的信息。

KGLLM 优势还包括：

结构化数据的使用：KGLLM 使用的是结构化的外部知识库。信息在知识图谱中以图形形式建模，使得数据之间的关系一目了然，因此查询和理解数据也变的更加直观；

关系数据处理能力：KGLLM 尤其适合处理关系数据，例如它可以理解人和人之间的关系、人和事件之间的关系等。且其使用图遍历算法，通过在知识图谱中跳跃多个节点（类似于地图上移动）找到相关信息。通过这种方式，KGLLM 可以找到最相关的信息来回答问题；

高效的信息检索和生成：通过图遍历算法，KGLLM 提取的关系会以自然语言转换成模型可以理解的提示，通过这些清晰的指示，KGLLM 模型能够生成更加准确和相关的回答。

### 展望

The Graph 作为「Web3 的 Google」，利用其优势弥补了当前 AI 服务的数据短缺问题，并通过引入 AI 服务简化了开发者的项目开发流程。随着更多 AI 应用的开发和使用，用户体验将有望得到进一步提升。未来，The Graph 开发团队将继续探索人工智能与 Web3 结合的可能性。此外，其生态系统中的其他团队，如 Playgrounds Analytics 和 DappLooker，也在设计与代理服务相关的解决方案。