Pantera合伙人：全面解析Sahara AI

作者：Paul Veradittakit ，Pantera Capital合伙人；翻译：本站xiaozou

Sahara AI的使命是创造一个更加开放、公平和协同的人工智能经济，让人们尽可能容易地参与其中。利用区块链，Sahara确保所有贡献者（数据贡献者、labelers、模型开发人员等）得到公平的补偿，数据和模型保持主权，人工智能资产具有安全性，权限可以创建、共享和交易。

## 1、AI堆栈现状

当前的AI堆栈可分为以下几层：

数据收集和标注

数据是从各种来源收集的（例如网络抓取、公共数据集、用户生成的数据），必须遵守许可要求以避免法律问题。数据根据手头的任务进行标注（例如分类、对象识别）。

模型训练及服务

数据被输入到模型中，模型调整其内部参数（权重）以最小化误差。这需要相当昂贵且耗时的计算。

AI智能体的创建和部署

创建AI智能体的用户体验通常涉及使用TensorFlow等工具，需要有技术专长。

计算资源

模型训练需要昂贵的处理。

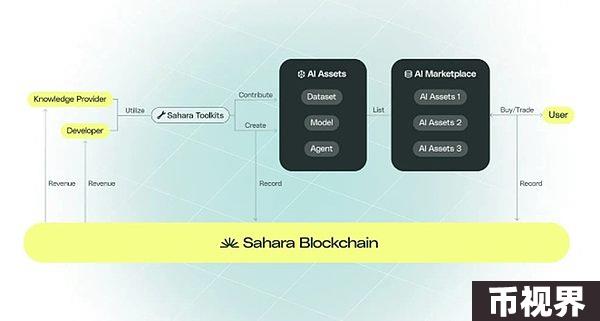
每一层都竞争激烈且多样化，在很大程度上，有一种执行方式被证明是最有效的。例如，数据收集最好使用大型公共数据集（如书籍），并使用专门数据（研究论文）进行微调。模型训练最好在专门的硬件上完成，AI智能体应该很容易使用即插即用的资源来建立一个开发者社区，计算资源应是分布式的，以便精准奖励计算资源提供者。这些结合在一起将带来更好的AI模型和更强大的社区。

Web2公司正朝着这个方向努力，但由于他们的设计师集中式的，所以面临着严重的限制。从企业和技术的角度来看，这些公司旨在限制访问并隔离堆栈的各个不同部分，从而导致不同的安全标准、数据库设计、后端集成和货币化策略。实际上，这样的设计很糟糕，是无法应对人工智能经济模式的转变的。

举例来说，OpenAI已经构建了一个非常强大的基础模型，并开始通过其无需许可的GPT wrapper市场吸引社区builders，但只允许表面的提示定制，不支持底层模型的重构。该公司所有的计算资源都是用投资者的钱购买的，预计今年年底将亏损50亿美元。

## 2、AI协同经济

Sahara平台在整个人工智能生命周期中为所有人工智能开发需求提供一站式服务：从数据收集和标注，到模型训练和服务，AI智能体的创建和部署，多智能体通信，人工智能资产交易，以及人工智能资源的众包等。通过使人工智能开发过程民主化以及降低现有系统的进入壁垒，Sahara AI为个人、企业和社区提供了平等的访问权限，共同建设人工智能的未来。



上图概括了用户旅程，描绘了在Sahara AI生态系统中，人工智能资产是如何从创建到使用再到实现用户粘性的过程。值得注意的是，平台内的所有交易都是不可变且可追踪的，所有权受到保护，资产来源也被记录下来。这支持透明和公平的收益分享模式，确保开发者和数据提供者都能因产生收益而获得适当的补偿。

Sahara的目标是让人们更加轻松地参与到人工智能经济中来。开发者和用户可以这样使用Sahara：

经验丰富的AI开发人员：

开发人员可以使用Sahara SDK和API与Sahara区块链及其AI堆栈的任意层进行交互，例如个性化算力、数据存储和激励结构，以形成他们自己的Sahara AI智能体，可被授权和货币化，以供其他人使用。

AI开发新手：

通过无代码/低代码环境，开发人员可以通过直观的界面和预构建模板创建和部署AI资产。

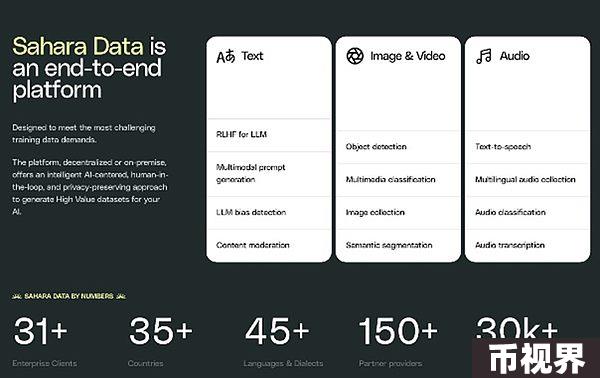
AI训练：

要参与人工智能模型训练，用户只需访问一个网站，在那里他们可以完成人工智能训练任务然后获得可交易代币的补偿，任务范围从解决基本数学问题到短视频描述。

AI用户：

用户可以通过直观的UI轻松使用AI智能体。用户可以灵活购买访问及进一步开发的许可权限，甚至可以交易AI资产shares。

用户将能够创建自己的个性化数据“知识库”，并使用用户自己的数据创建专门的人工智能。就像其他人工智能一样，这将允许他人访问，而训练数据仍然是完全私密安全的。



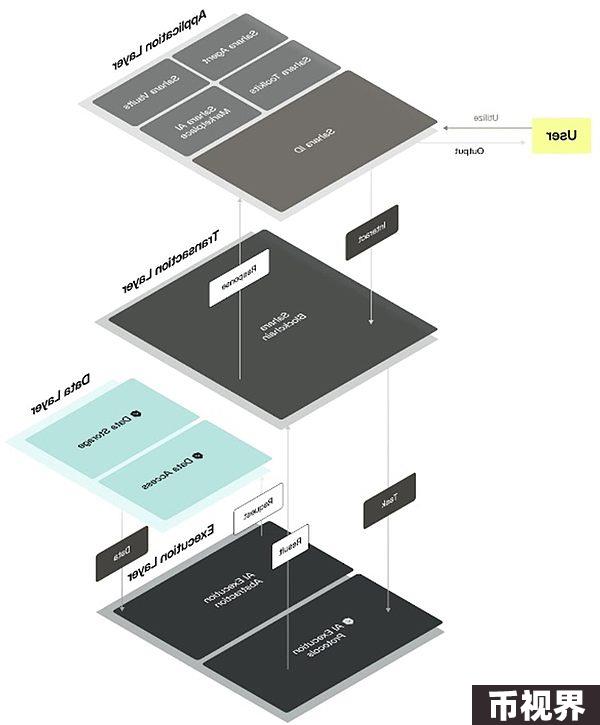
公司：

公司也可以创建AI智能体（或“商业代理”），训练他们自己的专有数据，由于系统运行在Sahara区块链上，得益于去中心化AI智能体生成和服务，成本要低得多。

企业也可以付费生成Sahara数据，它融合了AI自动标注和人工标注，有效地创建了高质量、保护隐私的多模型数据集。

除了面向企业的产品已经被一些知名客户使用之外，其他所有功能都还没有发布，但都有了发布计划。

## 3、技术概况



Sahara团队将系统设计得尽可能简单易用，抽象出了确保AI堆栈各部分的兼容性、盈利性和安全性所需的复杂性。在幕后，Sahara团队开发了无数创新来实现这一目标。举几个例子：

Sahara区块链最大限度地降低了gas费，完全兼容EVM，Sahara跨链通信（SCC）协议实现了跨区块链的安全、无需许可的数据传输，促进了无需信任的互操作性。

Sahara AI-Native Precompiles（SAPs）是预编译的智能合约，用于优化AI任务的性能，以减少计算开销，包括训练执行SAPs和推理执行SAPs。

Sahara区块链协议（SBPs）管理人工智能资产以确保会计责任，如AI Attribution跟踪贡献和分配奖励，AI Asset Registry来管理AI资产、AI许可和AI所有权的注册和来源。

数据管理在链上和链下进行，AI资产元数据、承诺和证明在链上，而重要的数据集、AI模型和补充信息在链下进行，以优化数据检索、安全性和数据可用性。

Collaborative Execution Protocols（协作执行协议）支持跨AI训练、聚合和服务的联合AI模型开发和部署。像PEFT这样的其他模型允许技术微调，Privacy Preserving Compute（隐私保护计算）支持差分隐私、同态加密和秘密共享，Fraud Proofs（欺诈证明）的功能如其名所示。

## 4、完全集成的AI堆栈

该团队由南加州大学终身教授Sean Ren和加州大学伯克利分校校友Tyler Z掌舵，前者曾被在《麻省理工科技评论》科技创新35人之列，并荣获2023年度三星研究员，后者曾担任Binance Labs的投资总监。团队的其他成员具有斯坦福大学、加州大学伯克利分校、AI2、Toloka、Stability AI、微软、币安、谷歌、Chainlink、LinkedIn、Avalanche等公司的背景或经验，贡献了宝贵的专业知识。

Sahara还有顶级AI原生研究人员和企业客户提供建议：

Laksh Vaaman Sehgal（Motherson Group副主席）

Rohan Taori（人类研究科学家）

Teknium（Nous Research联合创始人）

Vipul Prakash（Together AI首席执行官）

Elvis Zhang（Midjourney创始成员）

Sahara AI目前被超过35家领先的技术创新项目和研究机构使用，包括微软、亚马逊、麻省理工学院、Motherson集团和Snap，用于各种人工智能服务，如Shara Data用于数据收集/标注，Sahara Agents用于个性化领域智能体。

生成式AI在技术和市场规模方面仍处于起步阶段；由于难以将整个AI堆栈集成到一个产品中，因此今天的集中式聊天和视频工具的覆盖范围有限。Sahara AI是唯一一家通过模块化设计解决这一瓶颈的公司，该设计使用区块链作为无需许可访问、代币分配和安全性的支柱。为了让所有人都能参与其中，人工智能的未来必须是易于访问且公平的，而Sahara AI是唯一一家朝着这一愿景前进的公司。