Variant：去中心数据标注网络Sapien解读

作者：Jesse Walden，Variant Fund联合创始人；编译：0xjs@本站

“数据是新时代的石油。”这个类比很恰当，因为数据是为最具变革性的AI模型提供动力的资源，而且就像原油一样，数据必须经过处理和提炼才能发挥作用。这个主要由人工主导的过程，即数据标注，在 2022 年已经是一个规模达 22 亿美元的产业，并且在未来几年有望呈数量级增长。

然而，随着AI的快速且持续发展，标注需求已经超过了供给。许多客户无法获得他们所需的数据，因为标注公司缺乏提供专业服务的熟练劳动力。还有些客户则因价格过高而无法承受。这个高增长但供应受限的行业似乎已经到了亟待创新的成熟阶段 —— 我们认为加密技术可以成为一个突破口。

Variant很高兴地宣布，Variant领投了Sapien的1050万美元种子轮融资，以拓展其去中心化的数据标注网络。

## Sapien简介

Sapien 背后的主导理念是，可以利用加密技术更快地构建一个更大的标注网络，同时保持较高的数据质量。该团队正在构建一个协议，以便比中心化公司更高效地组织和补偿全球范围内的人工标注员（或 “AI工作者”）—— 包括兼职人员、熟练工人，甚至是小型企业。以下是我们认为加密技术能让 Sapien 在这一领域获得所需优势的几个方面：

1、全球网络：数据标注是一个全球性行业，大型公司（主要位于旧金山湾区）通常会在亚洲设立分支机构，招募当地工人，然后将标注好的数据出售给全球各地的客户 —— 这在不同的银行、法律和劳动体系之间转换时会产生摩擦。一个基于互联网原生的协议是这些不同参与方进行互动的最顺畅平台。Sapien 可以轻松地为客户或标注员开拓任何新的地域，包括召集全球各地的熟练工人来完成小众和专业的任务。

2、优化成本结构：Sapien 将使用加密轨道和稳定币来处理所有支付事宜，并尽可能依靠协议进行协调，而非人工管理。这些创新举措使其能够实现部分运营冗余的自动化处理，而这些运营冗余目前正困扰着该行业，如此一来便能降低成本。

3、代币激励：许多新的标注网络在初期发展阶段都举步维艰，但作为一个代币化的市场，Sapien 可以利用代币在更大规模上激励供给。Sapien 还可以利用奖励及其链上声誉系统来设计更好的激励措施，以提高标注数据的准确性 —— 到目前为止，他们所报告的准确率已达 99%。最后，分配网络价值和所有权应该会激励更多熟练工人加入，因为这能让他们在网络发展过程中获得潜在收益。

凭借这些创新，我们相信 Sapien 能够比其竞争对手发展得更快，同时提供更具性价比的服务。这个模式还处于试验阶段，但如果执行得当，它可能会给整个行业带来革命性的变革。

尽管 Sapien 还是一家年轻的初创公司，但它已经取得了长足进展并在迅速成长。它的早期客户包括一些全球最大的企业，如阿里巴巴、百度和好未来，以及知名的AI研究机构，如 Midjourney。其网络涵盖了来自 70 多个国家的参与者，这是通过与诸如 Worldcoin 和 Yield Guild Games 等大型组织建立合作关系实现的。该网络已经标注了 200 多万个数据点，活跃标注员人数刚刚突破 1 万，且每月以 50% 的速度增长。

## Sapien团队

Sapien的发展进程彰显了其团队的实力 —— 这是一个精简且精英的行业资深人士团队。

他们由首席执行官罗文·斯通（Rowan Stone）领导，他是 Base 的创始团队成员之一，曾助力推出 Layer 2；

还有特雷弗·科弗科（Trevor Koverko），他自 2017 年起就是以太坊社区的资深成员，也是当时最大的现实世界资产（RWA）协议之一 Polymath 的创始人。

团队中还有其他成员，他们曾在 Kraken 负责业务发展，将 Flow 区块链推向市场，以及打造传统人工智能业务。

在我们看来，对于这个领域而言，他们具备加密原生的市场推广（GTM）和企业背景的理想组合。