IOSG | 去中心化数据层：AI时代的新基础设施

## TL/ DR

我们曾讨论过AI和Web3能够如何各取所长，在计算网络、代理平台和消费应用等各个垂直产业上相辅相成。当聚焦在数据资源这一垂直领域，Web新兴代表项目为数据的获取，共享和利用提供了新的可能性。

传统数据提供商难以满足AI和其他数据驱动产业对高质量、实时可验证数据的需求，尤其在透明度、用户控制和隐私保护方面存在局限

Web3方案正致力重塑数据生态。MPC、零知识证明和TLS Notary 等技术确保数据在多个来源之间流通时的真实性和隐私保护，分布式存储和边缘计算则为数据的实时处理提供了更高的灵活性和效率。

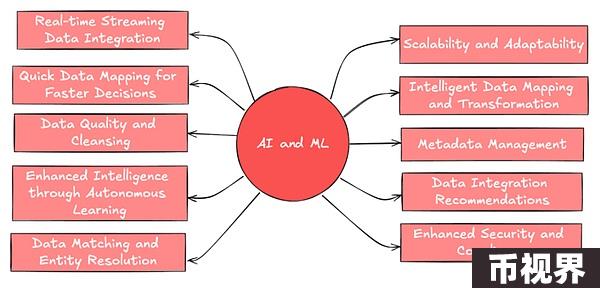
其中去中心化数据网络这一新兴基础设施萌生了几个代表性项目OpenLayer（模块化的真实数据层），Grass（利用用户闲置带宽和去中心化的爬虫节点网络）和Vana（用户数据主权Layer 1网络），以不同的技术路径为AI训练和应用等领域开辟新的前景。

通过众包的容量、无信任的抽象层和基于代币的激励机制，去中心化数据基础设施能够提供比Web2超大规模服务商更私密、安全、高效且经济的解决方案，并且赋予用户对其数据和其相关资源的控制权，构建一个更加开放、安全和互通的数字生态系统。

## 1. 数据需求浪潮

数据已成为各行业创新和决策的关键驱动。UBS预测全球数据量预计将在2020年至2030年间增长超过十倍达到660 ZB，到2025年，全球每人每天将产生463 EB（Exabytes，1EB=10亿GB）的数据。数据即服务（DaaS）市场在快速扩张，根据Grand View Research的报告，全球DaaS市场在2023年的估值为143.6亿美元，预计到2030年将以28.1%的复合年增长率增长，最终达到768亿美元。这些高增长的数字背后是多个产业领域对高质量、实时可信赖数据的需求。

AI模型训练依赖大量数据输入，用于识别模式和调整参数。训练后也需要数据集测试模型的性能和泛化能力。此外，AI agent作为未来可预见的新兴智能应用形式， 需要实时可靠的数据源，以确保准确的决策和任务执行。



(Source: Leewayhertz)

商业分析的需求也在变得多样和广泛，并成为驱动企业创新的核心工具。比如社交媒体平台和市场研究公司需要可靠的用户行为数据来制定策略和洞察趋势， 整合多个社交平台的多元数据, 构建更全面的画像。

对于Web3生态，链上也需要可靠真实数据来支持一些新型金融产品。随着越来越多新型资产在被通证化，需要灵活且可靠的数据接口，以支持创新产品的开发和风险管理，让智能合约可以基于可验证的实时数据执行。

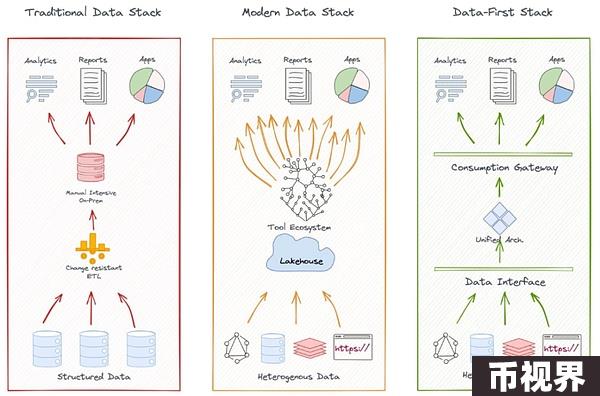
除了以上，还有科研，物联网(IoT)等等。新用例表面各行业对于多样、真实、实时的数据需求激增，而传统系统可能难以应对快速增长的数据量和不断变化的需求。

2.传统数据生态的局限性和问题

典型的数据生态系统包括数据收集、存储、处理、分析和应用。中心化模式的特点是数据集中收集并存储、由核心企业IT团队管理运维, 并实施严格的访问控制。

比如谷歌的数据生态系统涵盖了从搜索引擎、Gmail到Android操作系统等多个数据源，通过这些平台收集用户数据, 存储在其全球分布的数据中心, 然后使用算法处理和分析，以支撑各种产品和服务的开发与优化。

在金融市场里举例，数据和基础设施LSEG（前Refinitiv）则通过从全球交易所、银行和其他主要金融机构获取实时和历史数据，同时利用自有的Reuters News网络收集市场相关新闻，运用专有算法和模型生成分析数据和风险评估作为附加产品。



(Source: kdnuggets.com)

传统数据架构在专业服务方面有效，但集中化模式的局限性日益明显。特别是在新兴数据源的覆盖、透明度和用户隐私保护方面，传统数据生态系统正面临挑战。这里例举几个方面：

数据覆盖不足：传统数据提供商在快速捕捉和分析如社交媒体情绪、物联网设备数据等新兴数据源方面存在挑战。中心化系统难以高效地获取和整合来自众多小规模或非主流来源的"长尾"数据。

比如2021年GameStop事件就揭示了传统金融数据提供商在分析社交媒体情绪时的局限性。Reddit等平台上的投资者情绪迅速改变了市场走势，但像Bloomberg和Reuters这样的数据终端未能及时捕捉到这些动态，导致市场预测滞后。

数据可访问性受限：垄断限制了可访问性。许多传统提供商通过API/云服务开放部分数据，但高昂的访问费用和复杂授权流程仍然增加了数据整合的难度。

链上开发者难以快速接入可靠的链下数据，高质量数据被少数巨头垄断，访问成本高。

数据透明度和可信度问题：许多中心化数据提供商对其数据收集和处理方法缺乏透明度，且缺乏有效的机制来验证大规模数据的真实性和完整性。大规模实时数据的验证仍然是一个复杂的问题, 中心化的本质也增加了数据被篡改或操纵的风险。

隐私保护和数据所有权：大型科技公司大规模商用了用户数据。用户作为私人数据的创造者，很难从中获得应有的价值回报。用户通常无法了解他们的数据如何被收集、处理和使用，也难以决定数据的使用范围和方式。过度收集和使用也导致严重的隐私风险。

例如，Facebook的Cambridge Analytica事件就暴露了传统数据提供商如何在数据使用透明度和隐私保护方面存在巨大漏洞。

数据孤岛：此外，不同来源、格式的实时数据难以快速整合，影响了全面分析的可能性。很多数据往往被锁在组织内部，限制了跨行业和跨组织的数据共享和创新，数据孤岛效应阻碍了跨域的数据整合和分析。

比如在消费行业，品牌需要整合来自电商平台、实体店、社交媒体和市场研究的数据，但这些数据可能由于平台形式不统一或被隔离，难以整合。再例如，像Uber和Lyft这样的共享出行公司，虽然它们都收集大量来自用户的，关于交通、乘客需求和地理位置的实时数据，但由于竞争关系，这些数据无法提出并共享整合。

除此以外，还有成本效率、灵活性等问题。传统数据商正在积极应对这些挑战, 但异军突起的Web3技术为解决这些问题提供了新的思路和可能性。

## 3.Web3数据生态

## 3.Web3数据生态

自2014年IPFS（InterPlanetary File System）等去中心化存储方案发布以来，业界涌现出一系列新兴项目，致力于解决传统数据生态的局限性。我们看到去中心化数据解决方案已经形成了一个多层次、相互连接的生态系统，涵盖了数据生命周期的各个阶段，包括数据生成、存储、交换、处理与分析、验证与安全，以及隐私与所有权。

数据存储：Filecoin和Arweave的快速发展证明了去中心化存储（DCS）正在成为存储领域的范式转变。DCS方案通过分布式架构减少了单点故障风险，同时以更具竞争力的成本效益吸引参与者。随着一系列规模化应用案例的涌现，DCS的存储容量呈现爆发式增长（例如Filecoin网络的总存储容量在2024年已达到22 exabytes）。

处理和分析：Fluence等去中心化数据运算平台通过边缘计算（Edge Computing）技术提高了数据处理的实时性和效率，特别适用于物联网（IoT）和AI推理等对实时性要求较高的应用场景。Web3项目利用联邦学习、差分隐私、可信执行环境、全同态加密等技术在计算层上提供灵活的隐私保护和权衡。

数据市场/交换平台：为了促进数据的价值量化和流通，Ocean Protocol通过代币化和DEX机制, 创建了高效且开放的数据交换渠道， 例如帮助传统制造公司（奔驰母公司Daimler）合作开发数据交换市场，以帮助其供应链管理中的数据分享。另一方面，Streamr则创造了适用于IoT和实时分析场景的无许可、订阅式数据流网络，在交通、物流项目中显示了出色的潜力（例如与芬兰智能城市项目合作）。

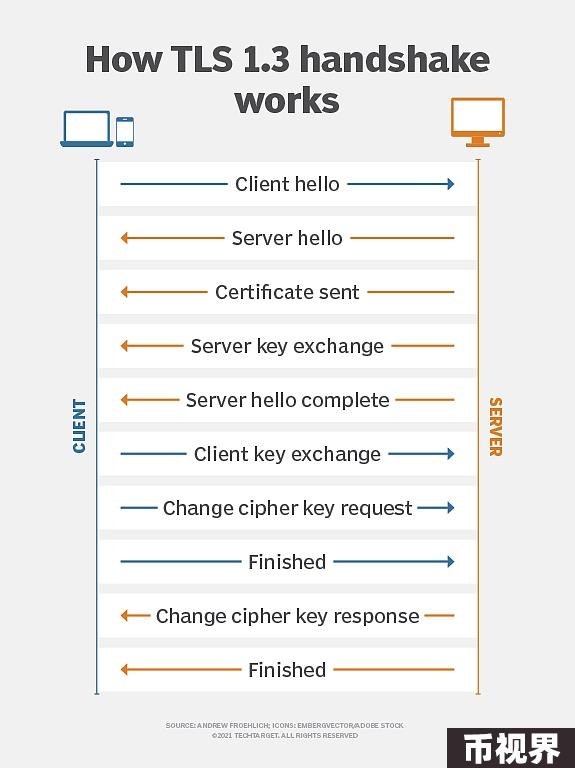
随着数据交换和利用的日益频繁，数据的真实性、可信度和隐私保护成为了不可忽视的关键问题。这促使Web3生态系统将创新延伸到了数据验证和隐私保护领域，催生了一系列突破性的解决方案。

### 3.1 数据验证与隐私保护的革新

许多web3技术及原生项目正致力于解决数据真实性和私有数据保护问题。除了ZK，MPC等技术发展被广泛应用，其中传输层安全协议公证（TLS Notary）作为一种新兴的验证方法尤其值得关注。

TLS Notary简介

传输层安全协议（TLS）是一种广泛用于网络通信的加密协议，旨在确保客户端和服务器之间的数据传输的安全性、完整性和保密性。它是现代网络通信中常见的加密标准，被用于HTTPS、电子邮件、即时通讯等多个场景。



（TLS加密原理， Source：TechTarget）

十年前诞生之际，TLS Notary的最初目标是通过在客户端(Prover)、服务器以外引入第三方“公证人”来验证TLS会话的真实性。

使用密钥分割技术，TLS会话的主密钥被分为两部分，分别由客户端和公证人持有。这种设计允许公证人作为可信第三方参与验证过程，但不能访问实际通信内容。这种公证机制旨在检测中间人攻击、防止欺诈性证书，确保通信数据在传输过程中没有被篡改，并允许受信任的第三方确认通信的合法性，同时保护通信隐私。

由此，TLS Notary提供了安全的数据验证，并有效平衡了验证需求和隐私保护。

在2022年, TLS Notary项目由以太坊基金会的隐私和扩展探索(PSE)研究实验室重新构建。新版本的TLS Notary协议从头开始用Rust语言重写, 融入了更多先进的加密协议（如MPC）, 新的协议功能允许用户向第三方证明他们从服务器接收到的数据的真实性，同时不泄露数据内容。在保持原有TLS Notary核心验证功能的同时，大幅提升了隐私保护能力，使其更适合当前和未来的数据隐私需求。

### 3.2 TLS Notary的变体与扩展

近年TLS Notary技术也在持续演进，在基础上发展产生了多个变体，进一步增强了隐私和验证功能：

zkTLS：TLS Notary的隐私增强版本，结合了ZKP技术，允许用户生成网页数据的加密证明，而无需暴露任何敏感信息。它适用于需要极高隐私保护的通信场景。

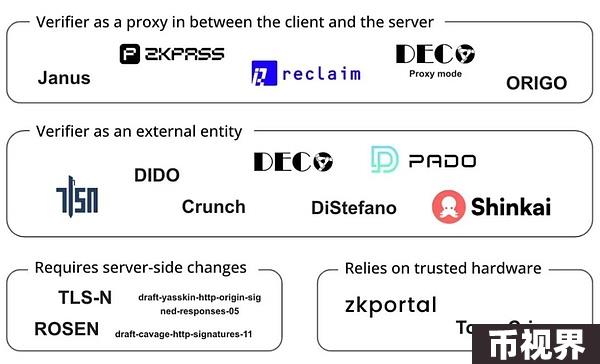
3P-TLS (Three-Party TLS)：引入了客户端、服务器和审计者三方，在不泄露通信内容的情况下，允许审计者验证通信的安全性。这一协议在需要透明度但同时要求隐私保护的场景中非常有用，如合规审查或金融交易的审计。

Web3项目们使用这些加密技术来增强数据验证和隐私保护，打破数据垄断，解决数据孤岛和可信传输问题，让用户得以不泄露隐私地证明如社媒账号拥有权、用以金融借贷的购物记录，银行信用记录、职业背景和学历认证等信息，比如：

Reclaim Protocol 使用zkTLS技术，生成HTTPS流量的零知识证明，允许用户从外部网站安全导入活动、声誉和身份数据，而无需暴露敏感信息。

zkPass 结合3P-TLS技术，允许用户无泄露地验证真实世界的私有数据，广泛应用于KYC、信用服务等场景，并且与HTTPS网络兼容。

Opacity Network 基于zkTLS，允许用户安全地证明他们在各个平台（如Uber、Spotify、Netflix等）的活动，而无需直接访问这些平台的API。实现跨平台活动证明。



(Projects working on TLS Oracles, Source: Bastian Wetzel)

Web3数据验证作为数据生态链条上的一个重要环节，应用前景十分广阔，其生态的兴荣正引导着一个更开放、动态和以用户为中心的数字经济。然而，真实性验证技术的发展仅仅是构建新一代数据基础设施的开始。

## 4. 去中心化数据网络

一些项目则结合上述的数据验证技术，在数据生态的上游，即数据溯源、数据的分布式采集和可信传输上做出更深入的探索。下面重点讨论几个代表性项目：OpenLayer，Grass和Vana，它们在构建新一代数据基础设施方面展现出独特的潜力。

### 4.1 OpenLayer

OpenLayer是a16z Crypto 2024春季加密创业加速器项目之一，作为首个模块化的真实数据层，致力于提供一个创新的模块化解决方案，用于协调数据的收集、验证和转换，以同时满足Web2和Web3公司的需求。OpenLayer已吸引了包括Geometry Ventures、LongHash Ventures在内的知名基金和天使投资者的支持。

传统数据层存在多重挑战：缺乏可信验证机制，依赖中心化架构导致访问性受限，不同系统间的数据缺乏互操作性和流动性，同时也没有公平的数据价值分配机制。

一个更加具象化的问题是，当今AI训练数据正变得日益稀缺。在公共互联网上，许多网站开始通过反爬虫限制措施来防止AI公司大规模抓取数据。

而在私密专有数据方面，情况则更为复杂，许多有价值的数据由于其敏感性质而以隐私保护的方式存储，缺乏有效的激励机制。在这种现状下，用户无法安全地通过提供私人数据获得直接收益，因此不愿意共享这些敏感数据。

为了解决这些问题，OpenLayer结合数据验证技术搭建了一个模块化真实数据层（Modular Authentic Data Layer），并以去中心化+经济激励的方式来协调数据收集、验证和转换过程，为Web2和Web3公司提供一个更安全、高效率、灵活的数据基础设施。

### 4.1.1OpenLayer模块化设计的核心组件

OpenLayer提供了一个模块化的平台以简化数据的收集、可信验证和转换过程流程：

a) OpenNodes

OpenNodes是OpenLayer生态系统中负责去中心化数据收集的核心组件，通过用户的移动应用、浏览器扩展等渠道收集数据，不同的运营商/节点可以根据其硬件规格执行最适合的任务而优化回报。

OpenNodes支持三种主要的数据类型，以满足不同类型任务的需求：

公开可用的互联网数据（如金融数据、天气数据、体育数据和社交媒体流）

用户私人数据（如Netflix观看历史、Amazon订单记录等）

来自安全来源的自报告数据（如由专有所有者签名或特定可信硬件验证的数据）。

开发者可以轻松添加新的数据类型,指定新的数据源，需求和数据检索方法, 用户可以选择提供去识别化的数据以换取奖励。这种设计使得系统可以不断扩展以适应新的数据需求，多样化的数据源使得OpenLayer能够为各种应用场景提供全面的数据支持，也降低了数据提供的门槛。

b) OpenValidators

OpenValidators负责收集之后的数据验证，允许数据消费者确认用户提供的数据与数据源的完全匹配。所有提供的验证方法可以进行加密证明的, 验证结果可以在事后被证实。同一类型的证明,有多个不同的提供商提供服务。开发者可以根据自己的需求选择最适合的验证提供商。

在初始用例中,特别是针对来自互联网API的公共或私有数据, OpenLayer以TLSNotary作为验证解决方案，从任何Web应用程序导出数据,并在不损害隐私的情况下证明数据的真实性。

不局限于TLSNotary，得益于其模块化设计, 验证系统可以轻松接入其他验证方法，以适应不同类型的数据和验证需求包括但不限于:

Attested TLS connections: 利用可信执行环境(TEE)建立经过认证的TLS连接,确保数据在传输过程中的完整性和真实性。

Secure Enclaves: 使用硬件级别的安全隔离环境(如Intel SGX)来处理和验证敏感数据,提供更高级别的数据保护。

ZK Proof Generators: 集成ZKP, 允许在不泄露原始数据的情况下验证数据的属性或计算结果。

c) OpenConnect

OpenConnect是OpenLayer生态系统中负责数据转换，实现可用性的核心模块，处理来自各种来源的数据, 确保数据在不同系统间的互操作性,以满足不同应用的需求。例如:

将数据转换为链上预言机（Oracle）格式,便于智能合约直接使用。

将非结构化原始数据转换为结构化数据,为AI训练等目的进行预处理。

对于来自用户私人账户的数据,OpenConnect提供了数据脱敏功能以保护隐私，也提供了组件来增强数据共享过程中的安全性,减少数据泄露和滥用。为了满足AI和区块链等应用对实时数据的需求, OpenConnect支持高效的实时数据转换。

当下，通过和Eigenlayer的集成，OpenLayer AVS运营商监听数据请求任务,负责抓取数据并进行验证,然后将结果报告回系统，通过EigenLayer质押或重质押资产,为其行为提供经济担保。如恶意行为被证实,将面临质押资产被罚没的风险。作为EigenLayer主网上最早的的AVS(主动验证服务)之一,OpenLayer已经吸引了超过50个运营商和40亿美元的再质押资产。

总的来说，OpenLayer所构建的去中心化数据层在不牺牲实用性和效率的前提下，扩展了可用数据的范围和多样性， 同时通过加密技术和经济激励，确保了数据的真实性和完整性。其技术对于寻求获取链下信息的Web3 Dapp、需要用真实输入来训练和推断的AI模型，以及希望根据现有身份和声誉来细分和定位用户的公司都有广泛的实际用例。用户也得以价值化他们的私有数据。

### 4.2 Grass

Grass是由Wynd Network开发的旗舰项目，旨在创建一个去中心化的网络爬虫和AI训练数据平台。在2023年末，Grass项目完成了由Polychain Capital和Tribe Capital领投的350万美元种子轮融资。紧接着，在2024年9月，项目又迎来了由HackVC领投的A轮融资，Polychain、Delphi、Lattice和Brevan Howard等知名投资机构也参与其中。

我们提到AI训练需要新的数据敞口，而其中一个解决方案是使用多 IP 来突破数据访问的权限，为AI进行数据喂养。Grass由此出发，创造了一个分布式爬虫节点网络， 专门致力于以去中心化物理基础设施的方式，利用用户的闲置带宽为AI训练收集并提供可验证数据集。节点通过用户的互联网连接路由web请求, 访问公开网站并编译结构化数据集。它使用边缘计算技术进行初步数据清理和格式化,提高数据质量。

Grass采用了Solana Layer 2 Data Rollup架构,建立在Solana之上以提高处理效率。Grass使用验证器接收、验证和批处理来自节点的web交易,生成ZK证明以确保数据真实性。验证后的数据存储在数据账本(L2)中,并链接到相应的L1链上证明。

### 4.2.1 Grass主要组件

a)Grass节点

与OpenNodes类似，C端用户安装Grass应用或浏览器扩展并运行, 利用闲置带宽进行网络爬虫操作, 节点通过用户的互联网连接路由web请求, 访问公开网站并编译结构化数据集，使用边缘计算技术进行初步数据清理和格式化。用户根据贡献的带宽和数据量获得GRASS代币奖励。

b) 路由器(Routers)

连接Grass节点和验证器,管理节点网络并中继带宽。Routers被激励运营并获得奖励,奖励比例与通过其中继的总验证带宽成正比。

c) 验证器(Validators)

接收、验证和批处理来自路由器的web交易, 生成ZK证明，使用独特的密钥集来建立TLS连接,为与目标web服务器的通信选择适当的密码套件。Grass目前采用中心化验证器,未来计划转向验证器委员会。

d) ZK处理器(ZK Processor)

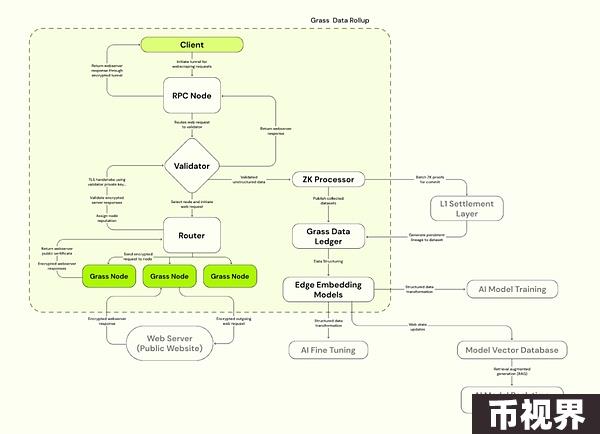
接收来自验证者的生成每个节点会话数据的证明，批处理所有web请求的有效性证明并提交到Layer 1（Solana）。

e) Grass数据账本（Grass L2）

存储完整的数据集，并链接到相应的L1链（Solana）上证明。

f) 边缘嵌入模型

负责将非结构化web数据转换为可用与AI训练的结构化模型。



Source：Grass

### 分析对比Grass和OpenLayer

OpenLayer和Grass都利用分布式网络为公司提供了访问开放互联网数据和需要身份验证的封闭信息的机会。以激励机制促进了数据共享和高质量数据的生产。两者都致力于创造一个去中心化数据层（Decentralized Data Layer）以解决数据获取访问和验证的问题, 但采用了略有不同的技术路径和商业模式。

技术架构的不同

Grass使用Solana上的Layer 2 Data Rollup架构， 目前采用中心化的验证机制, 使用单一的验证器。而Openlayer作为首批AVS，基于EigenLayer构建, 利用经济激励和罚没机制实现去中心化的验证机制。并采用模块化设计，强调数据验证服务的可扩展性和灵活性。

产品差异

两者都提供了类似的To C产品，允许用户通过节点进行数据的价值变现。在To B用例上，Grass提供了一个有趣的数据市场模型，并使用L2来可验证地存储完整的数据，来为AI公司提供结构化、高质量、可验证的训练集。而OpenLayer并没有暂时专用的数据存储组件，但提供更广泛的实时数据流验证服务（Vaas），除了为AI提供数据，也适用于需要快速响应的场景，比如作为Oracle为RWA/DeFi/预测市场项目喂价，提供实时社交数据等等。

因此，如今Grass的目标客户群主要面向AI公司和数据科学家，提供大规模、结构化的训练数据集，也服务于需要大量网络数据集的研究机构和企业；而Openlayer则暂时面向需要链下数据源的链上开发者，需要实时、可验证的数据流的AI公司，以及支持创新的用户获取策略，如验证竞品使用历史的Web2公司。

未来的潜在竞争

然而，考虑到行业发展趋势, 两个项目的功能确实有可能在未来趋同。Grass不久后可能也将提供实时的结构化数据。而OpenLayer作为一个模块化平台,未来也有可能扩展到数据集管理拥有自己的data ledger, 因此两者的竞争领域可能会逐渐重叠。

并且，两个项目都可能会考虑加入数据标注(data labelling)这一关键环节。Grass在这方面可能会更快推进, 因为他们拥有庞大的节点网络 - 据报道已超过220万个活跃节点。这一优势使Grass有潜力提供基于人类反馈的强化学习(RLHF)服务,利用大量标注数据来优化AI模型。

然而,OpenLayer凭借其在数据验证和实时处理方面的专长, 其在私人数据的专注，可能会在数据质量和可信度方面保持优势。此外,OpenLayer作为Eigenlayer的AVS之一, 可能在去中心化验证机制上有更深入的发展。

尽管两个项目可能在某些领域展开竞争, 但它们各自的独特优势和技术路线也可能导致它们在数据生态系统中占据不同的利基市场。



（Source：IOSG, David）

4.3 VAVA

作为一个以用户为中心的数据池网络，Vana同样致力于为AI和相关应用提供高质量数据。相比OpenLayer和Grass，Vana采用了更不同的技术路径和商业模式。Vana在2024年9月完成500万美元融资，由Coinbase Ventures领投，此前获得Paradigm领投的1800万美元A轮融资，其他知名投资者包括Polychain, Casey Caruso等。

最初于2018年作为MIT的一个研究项目启动，Vana旨在成为一个专门为用户私有数据设计的Layer 1区块链。其在数据所有权和价值分配上做出的创新使用户能够从基于其数据训练的AI模型中获利。Vana的核心在于通过无需信任、私密且可归因的数据流动性池（Data Liquidity Pool）和创新的 Proof of Contribution 机制来实现私人数据的流通和价值化：

4.3.1. 数据流动性池（Data Liquidity Pool）

Vana 引入了一个独特的数据流动性池（DLP）概念：作为Vana网络的核心组件，每个DLP都是一个独立的点对点网络，用于聚合特定类型的数据资产。用户可以将他们的私人数据（如购物记录、浏览习惯、社交媒体活动等）上传至特定DLP，并灵活选择是否将这些数据授权给特定的第三方使用。数据通过这些流动性池被整合和管理，这些数据经过去识别化处理，确保用户隐私的同时允许数据参与商业应用，例如用于 AI 模型训练或市场研究。

用户向DLP提交数据并获得相应的DLP代币（每一个DLP都有特定的代币）奖励，这些代币不仅代表用户对数据池的贡献，还赋予用户对DLP的治理权和未来利润分配权。用户不仅可以分享数据，还可以从数据的后续调用中获取持续的收益（并提供可视化追踪）。与传统的单次数据售卖不同，Vana 允许数据持续参与经济循环。

4.3.2. Proof of Contribution 机制

Vana 的另一核心创新之一是 Proof of Contribution（贡献证明）机制。这是Vana确保数据质量的关键机制，让每个DLP可以根据其特性定制独特的贡献证明函数，以验证数据的真实性和完整性，并评估数据对AI模型性能提升的贡献。这一机制确保用户的数据贡献得到量化和记录，从而为用户提供奖励。与加密货币中的“工作量证明”（Proof of Work）类似，Proof of Contribution 根据用户贡献的数据质量、数量以及被使用的频次来为用户分配收益。通过智能合约自动执行，确保贡献者获得与其贡献匹配的奖励。

### Vana的技术架构

数据流动性层（Data Liquidity Layer）

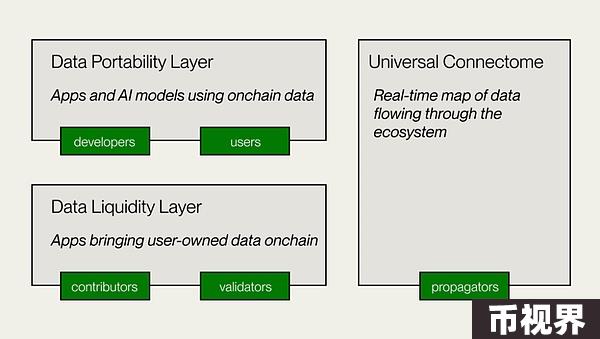
这是Vana的核心层，负责数据的贡献、验证和记录到DLPs，将数据作为可转移的数字资产引入链上。DLP创建者部署DLP智能合约，设定数据贡献目的、验证方法和贡献参数。数据贡献者和托管者提交数据进行验证，贡献证明（PoC）模块会执行数据验证和价值评估，根据参数给予治理权和奖励。

数据可移植层（Data Portability Layer）

这是数据贡献者和开发者的开放数据平台，也是Vana的应用层。Data Portability Layer为数据贡献者和开发者提供一个协作空间，以使用DLPs中积累的数据流动性构建应用。为User-Owned模型分布式训练，AI Dapp开发提供基础设施。

通用连接组（Connectome）

一个去中心化账本，也是一个贯穿整个Vana生态系统的实时数据流图，使用权益证明共识（Proof of Stake）记录Vana生态系统中的实时数据交易。确保DLP代币的有效转移并为应用提供跨DLP数据访问。与EVM兼容，允许与其他网络、协议和DeFi应用程序互操作。



（Source：Vana）

Vana 提供了一条较为不同的路径，专注于用户数据的流动性和价值赋能，这种去中心化的数据交换模式不仅适用于 AI 训练、数据市场等场景，也为 Web3 生态系统中用户数据的跨平台互通与授权提供了一个新的解决方案，最终创建一个开放的互联网生态系统，让用户拥有并管理自己的数据，以及由这些数据创造的智能产品。

5.去中心化数据网络的价值主张

数据科学家克莱夫·哈姆比（Clive Humby）在2006年说过数据是新时代的石油。近20年间，我们见证了"提炼"技术的飞速发展。大数据分析、机器学习等技术使得数据价值得到了空前释放。根据IDC的预测，到2025年，全球数据圈将增长到163 ZB，其中大部分将来自个人用户，随着IoT、可穿戴设备、AI 与个性化服务等新兴科技的普及，未来大量需要商用的数据将也将来源于个人。

### 传统方案的痛点：Web3 的解锁创新

Web3数据解决方案通过分布式节点网络，突破了传统设施的局限，实现了更广泛、更高效的数据采集，同时提升了特定数据的实时获取效率和验证可信度。在此过程中，Web3 技术确保了数据的真实性和完整性，并能有效保护用户隐私，从而实现更公平的数据利用模式。这种去中心化的数据架构，推动了数据获取的民主化。

不管是OpenLayer和Grass的用户节点模式，还是Vana 通过用户私有数据的货币化，除了提高特定数据采集的效率，也让普通用户共享数据经济的红利，创造一种用户与开发者双赢的模式，让用户真正掌控和获益于他们的数据和相关资源。

通过代币经济，Web3 数据方案重新设计了激励模型，创造了一个更加公平的数据价值分配机制。吸引了大量用户、硬件资源与资本的注入，从而协调并优化了整个数据网络的运行。

比起传统数据解决方案，它们也拥有模块化与可扩展性：比如Openlayer的模块化设计，为未来的技术迭代和生态扩展提供了灵活性。得益于技术特性，优化AI模型训练的数据获取方式，提供更丰富、更多样化的数据集。

从数据生成、存储、验证到交换与分析，Web3驱动的解决方案通过独有技术优势解决者传统设施的诸多弊端，同时也赋予用户对个人数据的变现能力，引发数据经济模式的根本转变。随着技术进一步发展演进和应用场景的扩大，去中心化数据层有望和其他Web3数据解决方案一起，成为下一代关键基础设施，为广泛的数据驱动型产业提供支持。