Vana: An Open Protocol for Data Sovereignty

Anna Kazlauskas
Open Data Labs
annakaz@opendatalabs.xyz

Abstract. The rise of artificial intelligence (AI) has made user data one of the most important resources in the digital economy, yet users who generate this data do not benefit from its commercial use. We present Vana, a protocol for programmable data ownership that enables collective creation of AI while maintaining individual data sovereignty. By combining personal compute environments with secure enclaves and tokenized data rights, Vana allows users to pool their data for AI training while maintaining cryptographic control. This enables training on datasets of unprecedented scale and diversity by unlocking data previously siloed in separate platforms and services. The protocol introduces Data Liquidity Pools as a novel coordination mechanism, transforming personal data from a static asset into a productive resource that creates ongoing benefits for its creators. This architecture enables new categories of applications previously impossible, from user-owned foundation models trained on cross-platform data to sophisticated data markets, providing a practical path toward democratizing the economic opportunity of artificial intelligence.

1. Introduction and Historical Context

Vitalik Buterin's development of Ethereum in 2015 showed that blockchains could extend beyond simple value transfer to enable programmable state transitions. We propose that the next frontier is programmable data ownership: a protocol that maintains individual data sovereignty while enabling collective creation. By combining personal compute environments, secure enclaves, and tokenized data rights, we create the first truly sovereign network for private user data, enabling new markets and coordination mechanisms that were previously impossible.

1.1 The Original Vision of User Data in Applications

The evolution of personal data on the internet represents a dramatic deviation from its original vision. In 1997, researchers at the MIT Media Lab pioneered wearable computing devices that would store and process personal data locally. The underlying principle was that as more and more personal data was collected, it would have to be stored directly with the individual. This aligned with the internet's early promise of individual sovereignty - users would maintain direct custody of their information rather than entrusting it to third parties.

1.2 The Rise of Centralized Data

Market forces drove a fundamentally different architecture from the original vision of users maintaining direct custody of their data. The emergence of cloud computing platforms offered compelling advantages in scale and efficiency, pushing applications away from local devices and into centralized data centers. As applications migrated to the cloud, user data naturally followed – emails moved from local clients to web browsers, photos from hard drives to cloud storage, and documents from desktop software to web applications. And as companies realized the value of this data through advertising, they further entrenched their monopoly power on data. If you ask most users today whether they own a given platform's data - say their Twitter data or their Reddit posts - most users would actually believe it belongs to the platform, even though it is legally theirs.

This centralization has become particularly problematic in the era of artificial intelligence. User data has become the crucial training resource for foundation models worth billions of dollars, yet the benefits flow entirely to the platforms that aggregate this data rather than the users who create it. Companies like OpenAI and Anthropic must spend enormous sums acquiring training data from these centralized repositories, while the original creators of that data do not own any of the resulting products. The misalignment is clear: users generate the data but have no stake in the AI systems their data helps create.

1.3 Previous Approaches

The idea of users owning their data is not new. Early attempts to restore user data sovereignty through projects like Urbit [1] and Solid [2] introduced the concept of personal servers - individual compute environments where users could store and process their own data. However, these projects failed to achieve widespread adoption due to limited incentive alignment and high technical barriers.

Subsequent blockchain-based approaches enabled data portability but struggled with privacy preservation. Onchain data is inherently public, making these systems unsuitable for personal information. The fundamental challenge became clear: how to combine the coordination capabilities of blockchains with the privacy requirements of personal data.

1.4 The Vana Approach

We propose a system that solves this challenge by combining:

- The sovereignty of personal servers
- The coordination capabilities of blockchain networks
- The privacy guarantees of modern cryptography
- The economic incentives of tokenized markets

This combination enables an entirely new design space, where users can pool their data collectively while maintaining individual control and AI models can be trained on private data without exposing the underlying information. Rewards generated from data can then flow back to its creators through tokenized rights.

The key insight is that data sovereignty and collective creation from data are not mutually exclusive. By providing the right technical primitives and economic incentives, we can create a system that preserves individual rights while enabling unprecedented coordination and the next frontier of AI progress.

2. Technical Architecture

Just as Bitcoin enabled trustless value transfer and Ethereum enabled programmable state transitions, Vana enables programmable data ownership, grounded in the principle of individual data sovereignty.

2.1 The Double Spend Problem for Data

The core challenge in financializing data is that, unlike other digital assets, data's economic value depends on controlled access - once data becomes public, it loses its market value. Traditional blockchains, with their emphasis on public verifiability, are not well suited for working with private data. Vana solves this through an architecture that combines private data custody with public ownership rights.

The blockchain maintains a global state consisting of:

- Data ownership records: Cryptographic proofs of data possession
- Access permissions: Who can access what data, under what conditions
- Validation proofs: Attestations of data quality and authenticity
- Onchain data collective contracts and token balances: Economic rights and governance

While the data itself remains encrypted in personal servers or secure enclaves, the blockchain enables programmatic control over who can access the data, under what conditions, and how value flows back to data creators.

2.2 Core Components

Personal Servers: Personal servers provide the secure foundation for data sovereignty. Each user maintains a personal server P_u defined by a primary encryption key k_p, storage provider credentials k_s, a collection of encrypted data objects D, and associated metadata M. The server implements core functions for data encryption, access control, secure DLP communication, and enclave integration. These personal servers can be run by the user on their machine, by a trusted provider, or in a lightweight way client-side on a user's device. Their core purpose is to run operations on a user's unencrypted data. Users also bring their own storage, and can use existing cloud providers or use a trusted provider. We expect infrastructure providers to emerge around this ecosystem, similar to Infura for Ethereum.

Data Liquidity Pools: DLPs serve as the coordination layer for collective data assets, enabling users to pool data while maintaining individual sovereignty. Each DLP enforces data validation rules, manages access rights, and governs token distribution through smart contracts.

Secure Enclaves: Secure enclaves provide Trusted Execution Environments (TEEs) for private computation, enabling complex operations while maintaining data privacy. All data validation and processing occurs within these isolated environments.

2.3 State Transitions and the Data Economy

Just as Bitcoin solved the double-spend problem for digital currency, Vana solves the "double-spend" problem for data. When data is sold or made public, it typically loses its economic power since it can be freely copied and redistributed. Vana preserves the economic power of data through a combination of privacy preservation and programmable access rights.

Data transactions in Vana can be viewed as state transitions, where each transaction T produces a new state S' according to the transition function:

$$\text{APPLY}(S,T) \rightarrow S' \text{ or ERROR}$$

The state transition system handles both data validity and economic transactions in a unified way. The state S tracks:

- Data ownership and delegated rights
- Access permissions and their conditions
- DLP membership and governance rights
- Token balances and trading state
- Financial positions and market states

Transactions T can include both data operations and economic operations, which can be atomically linked within the same transaction. For example, a single transaction could both grant model training access and execute the corresponding payment, or revoke data access rights when a loan defaults. Because Vana is fully EVM-compatible, these atomic operations between data rights and smart contracts enable new economic primitives while maintaining strict data access controls.

Examples of transactions representing data operations include: registering new data ownership, granting or delegating access permissions, creating or modifying DLPs, and executing model training permissions.

2.4 Data Portability

The combination of personal servers and blockchain-managed permissions enables true data portability. Users can grant any application access to their data through permission transactions, maintaining cryptographic control while enabling flexible usage. When a new application requests data access:

- The application specifies required data types and access level
- The user signs a transaction granting access, and decrypts their data in their personal server, for example, client-side in the application
- The application can query the user data as needed

Unlike traditional platforms where data transfer means surrendering control, Vana ensures that users maintain sovereignty while enabling data to flow freely between applications and be used to train AI models. These technical foundations create the basis for the sophisticated coordination mechanisms detailed in Section 3, where we explore how collective data assets are created, governed, and monetized at scale.

3. Data Coordination Layer

Vana enables coordination across groups of users by introducing Data Liquidity Pools (DLPs), which allow users to aggregate and tokenize their data. By aggregating non-fungible data contributions through a framework of metadata schemas, validation functions, token economics, and governance parameters, DLPs enable AI researchers to access comprehensive datasets while ensuring individual users maintain sovereignty over their information.

3.1 Data Liquidity Pools

Data Liquidity Pools (DLPs) serve as the fundamental coordination mechanism for collective data assets in the Vana network. Unlike traditional liquidity pools in DeFi that coordinate fungible token pairs, DLPs coordinate non-fungible personal data contributions while maintaining privacy and sovereignty.

The key innovation in DLPs is their ability to aggregate data access rights while preserving individual sovereignty. Much like how labor unions aggregate worker bargaining power, DLPs aggregate data bargaining power. When an AI researcher wishes to train a model on personal data, rather than negotiating with thousands of individual users and managing thousands of separate data access agreements, they can work directly with DLPs that have already aggregated user data rights.

A DLP can be formally defined as a tuple (M, V, T, G) where:

- M represents the metadata schema that defines the structure of acceptable data
- V is the validation function that verifies data quality and authenticity
- T defines the token economics
- G specifies the governance parameters

When a user contributes data to a DLP, it undergoes a three-stage process:

1. Data Preparation: The user encrypts their data with a derivation of their public key and creates a secondary encryption layer for DLP access. This ensures the user maintains sovereignty while enabling collective usage.
2. Validation: The DLP's validation function V executes within a TEE to verify the contribution's authenticity, quality, and uniqueness without exposing the underlying data. This produces an attestation A = (h, s, m) where h is the data hash, s is the quality score, and m is metadata.
3. Pool Integration: Upon successful validation, the DLP updates its state to include the new contribution and mints tokens according to the contribution value function v(A) -> R+.

When an AI researcher or model developer wants to access the data, they negotiate with the DLP's governance rather than individual users. Since data contributors receive tokens representing their ownership share in the DLP, they directly benefit from the AI models and other products their data creates. This creates natural market incentives for high-quality data contributions, efficient pricing of data access, sustainable revenue sharing models, and ongoing data maintenance and updates.

For example, an AI researcher can put up a proposal to access 10% of the underlying dataset directly for quality control purposes, then 100% of the overall dataset to train an AI model (without directly exposing the data), in exchange for burning some of the DLP token, as agreed upon with the DAO. The AI researcher could then create a new token specific to the trained model, requiring AI model token burns for each inference call, with 30% of these AI model tokens distributed to the DLP token holders as compensation for their data contribution.

The collective structure of DLPs transforms data from a one-time sale into a productive asset, where contributors capture ongoing rewards from their data's utility in AI training and applications. This corrects the misaligned incentives of traditional data marketplaces where users sell their data outright with no stake in the products ultimately created from their data.

3.2 Privacy-Preserving Proof of Contribution

Data verification on Vana is fundamentally privacy-preserving, offering two paths for validation while ensuring data never leaves the user's control unprotected:

- Local Verification: Zero-knowledge proofs generated on the user's device
- TEE Verification: Secure computation in the Satya Network's trusted execution environments

All verification methods follow a standardized attestation schema that provides a way to verify data quality without revealing the underlying data.

Each DLP implements its own unique Proof of Contribution function tailored to the specific type of data it handles, as different forms of data have inherently different measures of quality and importance. For instance, a DLP focused on financial data might prioritize factors like transaction accuracy, completeness of records, and consistency of reporting in its scoring mechanism. In contrast, a social media-focused DLP might weigh factors such as user engagement levels, account longevity, and content interaction metrics more heavily. For health data, a DLP might emphasize data freshness, frequency of measurements, and device accuracy ratings. These customized scoring functions allow DLPs to accurately assess and reward the most important contributions within their specific domains - a Twitter-focused DLP might reward users who consistently generate high-engagement content and maintain active presence, while a genetic data DLP might prioritize completeness of genetic panels and verification of testing sources. This flexibility in defining contribution importance enables each DLP to optimize for the unique characteristics that make its particular type of data important for AI training and other applications.

Data validation occurs through a network of Trusted Execution Environments (TEEs) called the Satya Network. These nodes provide verifiable attestations about data quality while preserving privacy of the underlying data. Each DLP defines its own validation criteria, enabling a market-driven approach to data quality assessment.

The validation process ensures:

- Data authenticity - proving the data is genuine and unaltered
- Ownership verification - confirming the contributor owns the data
- Quality assessment - measuring data completeness and utility, for example, by leveraging model influence functions.
- Uniqueness checks - preventing duplicate contributions.
- DLP-specific criteria - validating format and content requirements

This architecture enables DLPs to maintain high data quality standards while preserving individual privacy. Once the data is validated, the data score is written onchain, and users receive DLP token rewards based on their contribution's quality score, creating economic incentives for high-quality data submission. See the Economics section for more details on DLP tokens.

3.3 Data Structure and Cross-Platform Datasets

A key advantage of the DLP architecture is its ability to coordinate data across platforms and applications. Each DLP defines standardized metadata schemas and data structures for their domain, ensuring consistency and usability of the aggregated dataset. For example, a user's email data, browsing history, and social media activity can be combined into a structured, normalized training dataset for AI models, or healthcare records along with sleep and other wearable data can be aggregated across DLPs for research while maintaining privacy.

The DLP enforces structural consistency through its validation mechanisms, ensuring data quality and proper formatting across sources. This creates powerful network effects where applications can access unified user profiles across platforms, and researchers can train AI models on previously impossible combinations of data.

The coordination layer ensures that as datasets grow in size and importance, the benefits flow back to data contributors through their DLP tokens. This creates a sustainable ecosystem where users are incentivized to contribute data across platforms.

4. Security Model

Vana implements a dual-layer security model addressing both traditional blockchain security and data privacy considerations. The security architecture ensures that while data access rights can be tokenized and traded freely, the underlying personal data remains private and sovereign.

4.1 Blockchain Security

The network's security foundation builds on established blockchain principles while adapting them for data-specific requirements. At the consensus layer, the network operates under a Proof-of-Stake mechanism requiring validators to stake a minimum of 35,000 VANA tokens. This creates an economic security model where malicious behavior results in stake slashing, aligning validator incentives with network security. The smart contract layer enforces access permissions and governs data rights transitions.

4.2 Data Privacy Architecture

The data privacy layer extends beyond traditional blockchain security to ensure user sovereignty. All personal data remains encrypted with keys controlled by users, never leaving personal servers or secure enclaves in unencrypted form. Multi-layer encryption enables selective data sharing while preserving privacy of the broader dataset.

Computation on private data occurs exclusively within Trusted Execution Environments (TEEs), providing hardware-level isolation from host systems. Remote attestation verifies the integrity of TEE code and execution, while secure channels protect data in transit. This enables complex operations like arbitrary python code for proof-of-contribution and model training without exposing raw data.

Access control is enforced through a combination of smart contracts and cryptographic mechanisms. All data access within secure enclaves is logged for auditability. Data Liquidity Pools maintain privacy boundaries through tokenized access rights that are cryptographically bound to specific approved uses.

4.3 Trust Model

The security architecture relies on several core trust assumptions. The system requires trust in TEE hardware security guarantees for individual data, the underlying cryptographic primitives, and an honest majority (>2/3) of validator stake.

The threat model protects against malicious validators up to 1/3 of total stake, compromised applications, network-level attacks, and unauthorized attempts to access private data. However, the system cannot protect against fundamental breaks in TEE security, physical attacks on user devices, or social engineering of users who might be coerced into giving up their data by a malicious actor.

This comprehensive security model enables the key innovation of Vana: maintaining individual data sovereignty while enabling collective creation of AI models through tokenized data rights. By solving both the blockchain security and data privacy challenges, the protocol creates the foundation for a new kind of data economy.

5. Economic Model

Vana employs a single-token model with a native protocol token (VANA) that secures the network and facilitates transactions. It also provides the infrastructure for data-specific tokens issued by Data Liquidity Pools (DLPs) that represent rights to specific datasets. This dual-layer system enables efficient data operations while maintaining flexibility for specialized data markets.

The VANA token functions as a fundamental index of data utility across the network. As Data Liquidity Pools create and operate their own tokens tied to specific datasets, activity on these pools is naturally reflected in VANA through transaction fees and network usage. This creates an efficient proxy for the aggregate state of data assets in the network without requiring direct ownership of individual dataset tokens.

The native token is used for network security through validator staking, DLP staking which determines emission rewards for different DLPs, transaction fee payments for network operations, and as the primary trading pair on DEXes for all DLP tokens. The token targets a total supply of 120 million VANA, with fees supporting ongoing network operations and development.

5.1 Data as an Asset Class: Data Liquidity Pool Tokens

Each DLP creates its own token with custom economics for their specific data type. This is essential because the importance of different types of personal data (email, messages, photos, etc.) varies widely based on context. Email data might be important for business models but not social ones, photos might be crucial for vision models but useless for others, and message data importance depends on factors like conversation length and topic diversity.

Rather than attempting to centrally price these different data types, DLPs enable specialized teams to focus entirely on solving the valuation and validation challenge for specific data types. Through their proof-of-contribution implementation and tokenomics, DLP operators can design pricing models, set validation criteria, create incentive structures, and develop reward distribution mechanisms.

To help bootstrap the ecosystem, the top 16 DLPs by stake weight receive VANA emissions. This number, chosen to prioritize quality over quantity, is governed by token holders and adjusts with network growth. The rewards are split between the top 16 based on a set of performance metrics governed by the Vana DAO. Future updates may also reward a lottery of DLPs, even if they are not in the top 16.

5.2 Data Token Core Principles

DLP tokenomics typically incorporate three key components that help ensure sustainable growth and alignment between participants. The first focuses on Data Contribution Rewards, where tokens are allocated to ecosystem contributors through a points-based system that verifies all contributions. Contributors who provide high-quality or unique data receive additional incentives through performance multipliers, ensuring the highest standards of data quality within the ecosystem. This is captured through a DLP's proof of contribution implementation.

Economic Alignment forms the second core principle, establishing a robust token utility framework. All data access requires burning both the DLP token and a VANA fee in a coordinated mechanism, creating consistent demand for both assets. DLPs must maintain a minimum stake of 10,000 VANA to operate, ensuring commitment to the broader ecosystem. Liquidity is exclusively paired with VANA, centralizing trading activity, while revenue from data sales is shared between active contributors and token holders.

The third principle focuses on Supply Management, implementing deflationary mechanics through a fixed total supply and consistent burn pressure. We recommend that data access burns both DLP tokens and VANA, reducing supply over time as usage grows. The burn mechanics can extend beyond basic access to include feature activation and cross-pool integration.

Importantly, DLP creators maintain full control over their tokenomics design. While the above framework provides a starting point, we encourage innovative approaches that align rewards with the marginal benefit of data contributions and create meaningful economic flows based on the utility derived from data access. This flexibility enables DLPs to evolve and adapt their token systems to best serve their specific data communities and use cases.

5.3 Transaction Types and Fees

The network charges VANA fees for various operations, each priced according to their computational cost and network impact. These fees help prevent spam while enabling smooth network operations.

Data operations form the foundation of network activity. Users pay fees when contributing new data to DLPs, running proof-of-contribution verifications, granting access rights to data consumers, and managing data access permissions. These operations ensure data quality and controlled access across the network.

Token operations facilitate the exchange of economic rewards within the ecosystem. This includes trading DLP tokens, using data tokens for model training, executing AI model tokens for inference, and converting between different DLP tokens. The standardization of these operations through VANA enables seamless interaction between different data pools and applications.

Market operations support the broader ecosystem development. Users pay fees when creating buy/sell orders for data access, setting up automated data licensing agreements, establishing new DLPs, and modifying DLP parameters. These operations enable sophisticated data markets to emerge organically.

This multi-token system creates natural price discovery mechanisms for data, where markets rather than central authorities determine relative importance. Competition between DLPs drives innovation in data valuation models, while token holders can signal priorities through their allocation choices. As the network grows, increasing transaction volume creates steady demand for VANA tokens, while the fixed supply helps maintain long-term economic sustainability.

6. Applications and Market Infrastructure

Vana enables entirely new categories of applications that were previously impossible due to data sovereignty constraints. While AI model training represents an initial use case, the protocol's architecture supports a new frontier of user-owned data applications, from data coordination, to specialized AI models and health research.

6.1 AI Development Applications

The most immediate application of Vana is enabling privacy-preserving AI development at scale. Imagine a foundation model trained on data from 100 million users – owned not by a corporation, but by its data contributors. Through Data Liquidity Pools, users pool their messages, emails, and photos while maintaining cryptographic control of their data. Rather than training in centralized data centers, the model uses federated learning across secure enclaves, coordinated by DLPs that validate quality and prevent gaming. Contributors receive tokens based on their data's utility to the model, and when companies license the API, that revenue flows back to token holders, further incentivizing high quality data coming to the AI model [3].

Beyond foundation models, specialized AI services emerge through domain-specific DLPs. These could include sentiment analysis models trained on verified social data, recommendation engines using cross-platform user behavior, or niche language models for specific industries or languages. The key innovation is that these models continuously improve through incentivized data contributions while distributing rewards back to contributors.

6.2 Data Coordination Applications

DLPs function as DataDAOs, enabling collective bargaining for data rights at unprecedented scale. Consider a DAO of 23andMe users collectively negotiating licensing terms with pharmaceutical companies – transforming the dynamics of medical data monetization. Or even negotiating directly with 23andme to change their terms of service, or outright buying 23andme. These DataDAOs standardize data validation, set pricing models, and govern usage rights, creating efficient markets for previously fragmented data assets, while bringing together enough collective power to challenge big tech companies on their existing policies.

For clarity, Data Liquidity Pools refer to smart contracts that instantiate a DataDAO, which in turn refers to the larger ecosystem of data contributors, developers, and researchers that evolve around a particular data ecosystem.

6.3 Personal Data Applications

The protocol enables a new category of personalized AI applications where users simply login with their Vana identity and bring their entire data history. Rather than rebuilding user profiles from scratch or asking for platform-specific permissions, applications can access standardized, verified data through user-controlled permissions. This creates compounding benefits as users accumulate more data and applications become increasingly personalized.

6.4 Financial Applications

The tokenization of data rights enables sophisticated financial applications:

- Data Asset Trading
- Secondary markets for DLP tokens
- Futures contracts on data utility
- Options on model training rights
- Lending against data asset holdings

Examples: Trading venue data futures, AI training right options, even prediction markets on the success of model training

Data Derivatives

- Indices tracking data assets
- Synthetic exposure to data portfolios
- Cross-pool trading pairs
- Yield generation from data assets

Examples: Data sector indices, cross-dataset correlation trading

As more companies just become data companies, we believe that this ecosystem will evolve into a sophisticated market for data as a strategic asset class. The tokenization of data rights will enable more dynamic and liquid mechanisms for making data available to train AI models. Vana will ensure that individuals, rather than companies and platforms, see the rewards of this shift.

6.5 Integration Patterns

Applications can interact with the Vana protocol in three primary ways:

1. Direct Data Access: Applications request specific data directly from users, with cryptographic access control and automatic permission expiration. This enables high-trust personal applications like financial services or health tracking.
2. DLP Integration: Applications access aggregated, validated data pools through standardized APIs, with built-in revenue sharing and governance participation. This suits research platforms and market analysis tools requiring high-quality datasets.
3. Model API Consumption: Applications use pre-trained models from DLPs, paying per call with automatic revenue distribution to data contributors. This enables immediate integration of AI capabilities without direct data access.

7. Conclusion

The core insight of Vana is that AI's unprecedented opportunity should benefit both the users who create the data and the researchers who need it, not just the companies that aggregate it. This is not merely an idealistic vision - users already maintain full legal ownership of their data and can export it from any platform. Just as Bitcoin enabled collective agreement on currency state and Ethereum enabled programmable state transitions, Vana enables programmable data ownership - creating a practical path towards user-owned AI.

The technical challenges ahead are significant. Scaling data from millions of users, implementing efficient distributed model training, and developing the economic primitives around data tokenization all represent substantial technical frontiers. Yet we are already seeing an ecosystem emerge, with applications being built on Vana that were previously impossible. We are only at the beginning - about 1% of the way towards our north star of a user-owned foundation model trained by 100 million users.

The stakes are profound. In an AI-native world, the difference between platform-controlled and user-owned data represents the difference between a centralized monopoly and an open, innovative ecosystem. Data sovereignty is not just about privacy - it is about ensuring the next generation of AI development benefits its contributors rather than merely extracting value from them.

Vana today provides the core primitives - personal servers, data liquidity pools, and programmable ownership rights - that make this future possible. Early signs of adoption, from growing open-source contributions to production applications, suggest the foundations are sound. However, realizing the full vision of user-owned AI will require contributions from many: developers building applications that respect user sovereignty, researchers advancing privacy-preserving machine learning, and users reclaiming their stake in the AI revolution.

The technical foundations are in place. The legal right to data portability exists. The economic incentives align individual and collective interests. Together, this brings true data sovereignty.

8. References

[1] Urbit Developers. Urbit Developer Documentation, 2024: https://developers.urbit.org/reference/arvo/overview.

[2] Solid Project. Solid Protocol Specification, 2024: https://solidproject.org/TR/protocol.

[3] Kazlauskas, A. User-Owned Foundation Models. Anna Kazlauskas' Blog, February 28, 2024: https://anna.kazlausk.as/posts/user-owned-models-update.

[4] Delacroix, S. "Bottom-up data Trusts: disturbing the ‘one size fits all’ approach to data governance" International Data Privacy Law, vol. 9, no. 4, pp. 236–254, 2019: https://academic.oup.com/idpl/article/9/4/236/5579842.