XnY: Data Assetification for Open AI Economy

XnY Core Developers and Codatta Labs

November 23, 2024

Contents

1 Abstract

This paper introduces XnY, a blockchain network designed to address data challenges in the AI economy by transforming data into assets. XnY classifies data into three main types: X-Data (user and entity information), Y-Data (intelligent task responses), and Frontier Data (domain-specific knowledge). By combining these data types, the platform enhances AI capabilities in sectors like healthcare and finance.

Currently, data marketplaces are inefficient, with data creators underpaid and quality data accessible mainly to large corporations. XnY tackles these issues by turning data into liquid assets and establishing clear ownership and usage rights, thereby improving data liquidity. Using blockchain technology, it ensures data sovereignty with robust privacy, security, and fair value distribution.

Key features of XnY include data governance and rights management, data lifecycle transparency, data quality assurance, secure data storage, and on-chain ownership verification. By incorporating data attribution and valuation methods such as influence functions, in-context probing, and Shapley values, XnY enables the creation of data marketplaces and exchanges. These platforms facilitate efficient data trading and value discovery, resulting in a complete ecosystem where data can be traded seamlessly and transparently.

By leveraging these capabilities, XnY builds an ecosystem that empowers data creators, AI developers, and researchers to collaborate directly. This fosters innovation while sharing both risks and rewards, establishing a fair and equitable data economy. The platform also provides the necessary infrastructure to advance toward Artificial General Intelligence (AGI) and unlock the full potential of AI applications.

2 Introduction

XnY is a blockchain network dedicated to the open AI economy, addressing AI's pressing data issues. The blockchain's mission is to assetify data. Data serves as the foundational building block for AI and millions of data-driven applications. Unlike traditional raw materials, data indirectly generates value by enabling AI models and empowering applications [15, 17]. In the near term, data is crucial for building vertical AIs, which address the pressing issues of insufficient grounding that foundational models experience. In the long run, as we progress toward AGI (Artificial General Intelligence), data will gain even more significance due to the end-to-end training strategy [4, 12].

However, current AI models, especially large language models (LLMs), mainly rely on open internet data that fails to adequately recognize or compensate the individuals who created it [3, 4]. Domain AI solutions also suffer due to their reliance on traditional outsourcing models, where intermediaries capture significant premiums [13, 14]. This leads to data creators being under-rewarded for their contributions, resulting in prohibitively high data access costs, which are affordable only to big tech companies or elite teams with substantial funding. This massive inefficiency in the data marketplace impedes the development of AGI and limits the potential of AI-native applications.

Today, AI faces several critical challenges related to data. Large language models rely on publicly available internet data, providing no compensation to original data creators. Data

for domain-specific AI remains prohibitively costly, primarily accessible to major corporations due to inefficiencies in the data marketplace. Additionally, data creators often receive inadequate compensation, leading to a monopolized and unequal ecosystem.

These issues limit access to quality data, making domain-specific AI expensive and restricting the development of AI-native applications, particularly for smaller players. XnY aims to change this by enabling equitable and efficient access to data.

3 Data Types and Composition

3.1 Data Types: X-Data, Y-Data and Frontier Data

The XnY ecosystem categorizes data into three primary types, each serving a different purpose within AI development and data-driven services. These data types can also be composed together to advance AI capabilities:

- **X-Data: User and Entity-Centric Information** X-Data encompasses a diverse range of user-specific and entity-related data, such as de-identified profiles, healthcare records, fitness logs, video content, encrypted Web2 data, design products, and artwork. X-Data has a two-fold value: it can be used independently to empower applications like cross-dApp recommendations, advertising systems, and other data-driven services, while also fueling the arts-reference economy and serving as a resource for Generative AI. Additionally, X-Data serves as the foundational base to derive Y-Data, contributing to the creation of structured intelligence, and helps construct Frontier Data by providing essential context and information for specialized fields.

- **Y-Data: Intelligent Task Responses** Y-Data consists of responses or outputs generated through cognitive tasks. It involves solving complex problems, reasoning through scenarios, or providing expert annotations. Y-Data is typically produced through user interaction that requires a higher level of intelligence, such as solving math problems, completing surveys, or offering professional insights. This type of data contributes structured outputs necessary for building task-specific AI solutions.

- **Frontier Data: Domain-Specific Knowledge** Frontier Data is the most advanced data type in the XnY ecosystem, combining X-Data and Y-Data enriched with domain-specific expertise. Frontier Data captures nuanced, highly specialized knowledge essential for developing domain-specific AI solutions. It requires contributions from trained professionals, such as doctors, lawyers, or engineers, making it indispensable for applications that demand high accuracy and deep domain expertise.

3.2 Composability of X-Data and Y-Data

Intelligently pairing X-Data with Y-Data is crucial for advancing AI and building task-specific solutions that range from understanding user behavior to generating expert insights. For example, combining X-Data with Y-Data helps:

- **Recommendation Systems**: Enhance personalization by understanding users and their preferences.
- **Finance**: Construct detailed risk profiles for lending and investment advice [17].

Used in conjunction, X-Data can involve temporal and multidimensional integration, merging data across different time points or attributes to improve comprehensiveness and accuracy. Meanwhile, Y-Data, being labeled data, can be updated over time to reflect evolving target outcomes. Together, X-Data and Y-Data enable AI models to adapt to dynamic changes and provide more accurate predictions in various verticals, such as:

- **Healthcare**: Integrating patients' temporal electronic medical records, health monitoring data, and medication histories to better monitor health changes and treatment effects, improving medical service quality [2].
- **Finance**: Combining data from various DeFi protocols, including transactions, lending, and asset records, to compile a comprehensive user financial profile. This enables more accurate, personalized investment advice and risk management capabilities [17].

4 Assetification of Data

4.1 Data as an Economic Asset

More people increasingly view data as a valuable asset because of its numerous use cases and the growing demand across different verticals [17, 12]. Its ability to generate cash flow makes it a true economic asset. To maximize this value, data must also be treated as a liquid asset, allowing it to change hands frictionlessly through ownership transfers and efficient markets [13, 6]. By allowing market dynamics to efficiently determine data's monetary value, data can achieve greater liquidity and usability.

At the core of the data marketplace's current issues lies the lack of proper rewards for data creators [14]. Data creators often need to be more-rewarded for their contributions, especially for highly specialized Frontier Data. By making data more easily assetified and enhancing market efficiency, the XnY ecosystem aims to ensure proper pricing and a fair, sustainable reward system. This incentivizes experts to provide high-quality, domain-specific data, ultimately enriching the entire data marketplace.

4.2 Transforming Data into Liquidity Assets

XnY tackles the above issues through data assetification. By improving data liquidity, XnY aims to enhance market efficiency, allowing data to be exchanged more seamlessly. This focus on liquidity will help incentivize data creators and drive fair value determination, which can ultimately build a more efficient and equitable data economy. This process transforms raw data into structured, tradable assets, creating value within the data ecosystem. XnY defines data assets with the following characteristics:

- **Verifiable**: Data sources, quality, and integrity must be verifiable, a feature ensured through blockchain's immutability.

- **Interoperable**: Data must be accessible, easily shared, and applicable across various platforms and scenarios.
- **Privacy-Preserving**: Data privacy must be protected, ensuring only authorized parties can access it while preventing unauthorized leaks or misuse.

Assetification allows data to be recognized as an economic asset, with standards for valuation, liquidity, and traceability using blockchain. This recognition ensures that data creators are compensated fairly and data demanders can use data more effectively.

5 Data Sovereignty as the Foundation for Data Liquidity

The ability to transform data into a liquid asset requires clear definitions of ownership and usage rights, as well as robust frameworks for maintaining privacy, security, and value distribution. This section provides the necessary background to understand why these features are essential for enabling data liquidity and how they contribute to a more efficient data marketplace.

5.1 Data Sovereignty: Ownership, Access Rights, and Legal Frameworks

We define data creators as the rightful holders of data ownership and data demanders as those granted data usage rights. Beyond the data creators themselves, many data demanders are willing to act as consumers of the data, generating greater value. This interaction allows data creators to benefit from the consumption of their data, fostering the circulation of data assets and enabling their preservation and appreciation of value.

Clearly defining data sovereignty is a prerequisite for effectively extracting and utilizing the data's value. At its core, this involves precisely delineating data ownership and data usage rights, two pivotal aspects of data value creation and utilization. These elements are essential because the realization of data value and its associated benefits arise from the interaction between data creators and data demanders.

- **Data Ownership**: Refers to those who hold the rights to a dataset, granting them authority over actions such as viewing, transferring, copying, modifying, or deleting the data.
- **Data Access Rights**: These refer to who is authorized to use the data. The scope of these rights can vary depending on the permission granted, ranging from basic rights like viewing and copying to more advanced rights such as modification or deletion.

Ownership inherently includes full management rights over data, while access rights are limited to specific permissions granted by the owner. This distinction allows data to be used without transferring ownership, ensuring creators maintain control while allowing specific, authorized usage.

5.2 Legal Frameworks for Data Sovereignty

Globally, data sovereignty has been defined and regulated through various legal frameworks established by major economies, including the following:

- **European Union - General Data Protection Regulation (GDPR)**: As one of the most influential personal data protection laws worldwide, the GDPR sets principles for safeguarding personal data and strengthens individuals' control over their information [16]. It grants rights such as data access, correction, deletion, transfer, and transparency in processing. The GDPR emphasizes the need for explicit consent from data subjects (individuals), mandates security measures during storage and transfer, and ensures privacy rights in data handling processes.
- **United States - California Consumer Privacy Act (CCPA)**: The CCPA targets consumer data protection in California and provides consumers with rights to know about, access, delete, and opt out of the sale of their personal data. The law also emphasizes consumer privacy control by requiring companies to disclose how they collect, use, and share data [5].
- **China - Personal Information Protection Law (PIPL)**: Effective in 2021, the PIPL adopts principles similar to the GDPR, explicitly defining requirements for the collection, use, and storage of personal information. It grants users the right to request deletion or transfer of their data and requires companies to obtain user consent when processing personal information, ensuring security and transparency in data handling [7].

Given the unique nature of data as a new form of asset, data sovereignty cannot rely solely on conceptual or legal definitions. Unlike traditional assets such as real estate, gold, or stocks, data sovereignty is more elusive, making it difficult to efficiently track and identify ownership or ensure its validity and integrity through legal measures alone. This necessitates technical solutions to define and safeguard the exercise of data sovereignty and enable the extraction of associated benefits [19, 10].

5.3 Technical Considerations for Data Sovereignty

At the technical level, Data Ownership and Data Access Rights address issues beyond legal frameworks, such as data control, privacy protection, security, and benefit distribution. These include, but are not limited to:

- **Data Ownership**:
  - ◇ **Data Circulation Control**: Data creators can decide whether to upload data on-chain, participate in data transactions, or share it. Ownership is verified through the issuance of XDAT as Proof of Ownership.
  - ◇ **Privacy and Storage Management**: Data creators can store and manage their data in decentralized networks, applying encryption to ensure security and confidentiality.

- ◇ **Data Revenue Rights**: Blockchain technology enables data creators to authorize, share, or trade their data within decentralized markets, earning rewards or revenue in return.

- **Data Access Rights**:
  - ◇ **Access Authorization**: Data creators can authorize data access via smart contracts, setting conditions like time limits, access frequency, and usage scope. The usage process is traceable and immutable, ensuring transparent data flow.
  - ◇ **Usage Control Management**: Smart contracts and access control mechanisms allow data creators to revoke or modify data usage permissions, maintaining complete control over data usage at all times.
  - ◇ **Compliance and Transparency**: Blockchain technology clarifies the obligations of data creators, ensuring compliance, accountability, and responsible usage under predefined conditions to safeguard security and privacy.

Understanding data sovereignty and its components is essential to transforming data into a liquidity asset. By outlining ownership and usage rights, and leveraging both legal and technical frameworks, we lay the groundwork for enabling efficient data exchange, fair valuation, and ultimately building a more liquid and equitable data economy. In the following sections, we will discuss the specific capabilities XnY will deliver and how they align with these requirements.

6 Core Features of XnY for Achieving Data Asset Liquidity

To transform data into a truly liquid asset, XnY must deliver a comprehensive set of capabilities that address critical challenges around data ownership, traceability, quality, and security. These capabilities ensure the effective assetification of data while enabling data creators to retain control and earn value from their contributions. This section discusses the core capabilities that XnY delivers, which are essential for realizing data asset liquidity.

6.1 Data Governance and Rights Management

Data governance is at the core of transforming data into a tradable asset [11]. It involves establishing clear ownership and access rights, ensuring data creators can maintain control while allowing authorized use under agreed terms. The XnY platform provides robust features to manage data asset ownership, access permissions, and maintain authorization.

- **Data Ownership and Rights Management**: Each data asset is represented as an XDAT, with explicitly defined ownership and access rights. This ensures data assets are clearly identifiable and validated, supporting transparent rights management.

- **Dynamic Role Authorization**: XnY provides a flexible system for dynamically authorizing and transferring permissions between Data Creators and Data Consumers. This flexibility is crucial for enabling efficient data exchange while maintaining full control.

- **Smart Contract-Based Revenue Sharing**: XnY uses a smart contract-based revenue sharing model to ensure fair compensation for Data Creators and Data Consumers. This mechanism promotes shared value realization from the usage of data assets.

6.2 Data Lineage and Lifecycle Transparency

Lifecycle transparency is a critical feature for ensuring the transparency of data usage throughout its lifecycle. XnY uses blockchain to track data from its origin, through its transformation processes, to its final use [10].

- **Data Lineage Management**: XnY records every step of data collection, annotation, validation, and publication as XDAT. This lineage ensures data credibility and provides a full history of modifications, which is vital for building trust in the data.

- **Transaction Transparency**: Every data transaction, including participants, time, and price, is logged on-chain, providing an immutable record that guarantees transparency in the trading of data assets.

- **Usage Monitoring**: Data usage is also recorded, including the frequency and nature of use, which helps ensure that consumption adheres to the conditions set by Data Creators.

6.3 Data Quality Assurance

Data quality is central to transforming data into an economic asset with measurable value [18]. XnY employs several tools and processes to ensure that data meets the required standards for various applications.

- **Foundational Quality Checks**: Foundational checks such as format standardization and completeness ensure that data meets basic quality requirements.

- **Domain-Specific Workflow Customization**: Custom workflows offer flexibility and help balance cost with acceptable quality standards. Greater participation by experienced data creators can further enhance data quality, though it may increase costs.

- **Reputation-Driven Quality Scoring**: A scoring system rates Data Creators based on multiple considerations, such as data quality, behavioral norms, usage feedback, and other relevant factors. These scores are linked to data revenue-sharing, which incentivizes the maintenance of high data quality.

6.4 Data Storage and Security

XnY ensures that data assets are stored, managed, and transmitted securely to support data asset liquidity. Effective storage and privacy measures are crucial for both enabling data exchange and building trust in the data economy.

- **Data Fingerprinting for Integrity**: Each data asset is assigned a unique fingerprint, which is a compressed representation for identification and integrity verification. This fingerprint guarantees the uniqueness of each data asset.
- **Data Availability Layer**: XnY ensures that data is stored on-chain with mechanisms for real-time retrieval. This infrastructure guarantees data is readily available for transactions, which is vital for maintaining data asset liquidity.
- **Hybrid Storage Architecture**: To balance scalability and accessibility, XnY adopts a hybrid storage approach that combines decentralized storage (e.g., IPFS) with traditional cloud storage. This approach ensures both off-chain data scalability and the completeness of data assets.
- **Advanced Encryption Protocols**: Advanced encryption technologies, such as OpenPGP, are employed to secure data throughout its lifecycle—during transmission, storage, and processing [19]. Future plans include integrating Fully Homomorphic Encryption (FHE) to further enhance data security [8].

6.5 Data Ownership and Verification

To enable data asset liquidity, XnY uses blockchain technology to represent data as tradeable units, allowing for secure and transparent transactions.

- **Data Representation Framework**: The platform provides a flexible framework for representing data assets, facilitating the allocation of ownership and access rights. Data can be categorized and represented in varying levels of granularity, such as slicing X-Data into multiple features (attributes, columns) to enable individual features to be tradable and composable, making it easy to trade a selection of records of all sizes.
- **On-Chain Ownership Verification**: Ownership, access rights, and contributor information are all recorded on-chain, ensuring that data ownership is publicly verifiable, which enhances trust and traceability.

To transform data into liquid assets, XnY delivers essential capabilities that span data ownership management, lifecycle transparency, quality assurance, secure storage, and tokenization. By providing these features, XnY builds the infrastructure necessary to meet the challenges of the modern data economy, ensuring data can flow efficiently, generate fair rewards, and ultimately achieve full liquidity. The following sections will explore how these capabilities are put into practice within real-world applications and how they are integrated into the broader XnY ecosystem.

7 Building an Open AI Economy

Data is the foundation of artificial intelligence (AI) systems, but the challenge is the absence of robust scientific methods to quantify its value, especially in dynamic contexts that consider factors like scarcity, demand, and other vital variables [9, 1]. Central to building a fair AI economy are **data attribution** and **data valuation**—mechanisms that assess the relationship between data and AI models. **Data Attribution** measures the impact of individual data points on model outputs, thereby enhancing transparency and accountability [9]. **Data Valuation** quantifies the utility of data to ensure data creators are fairly rewarded [1]. Together, these mechanisms form the core of a sustainable incentive system that aligns rewards with the actual value contributed by data, fostering a more equitable AI ecosystem.

7.1 Methods for Data Attribution and Valuation

Data attribution and valuation methods that fall into three main categories: **influence functions**, **in-context probing**, and **Shapley** values.

- **Influence functions** measure how changes in training data affect model outputs by calculating gradients and inverse Hessians, giving insight into data impact.
- **In-context Probing (ICP)** estimates data utility by prompting a pre-trained model with specific inputs, providing a lightweight alternative without needing gradient calculations.
- **Shapley values** allocate value among data points based on their marginal contribution to the model's performance, ensuring fairness by assessing each point's role in enhancing the AI model.

7.2 Challenges in Quantifying Data Value

While **influence functions** and **in-context probing** effectively **rank** data samples based on their influence on model performance, assigning actual economic value remains challenging.

Ranking vs. Pricing: Ranking methods provide influence assessments but lack the mechanism to assign actual economic value. Determining fair pricing requires understanding market dynamics such as demand, scarcity, and utility. An **auction-based market** is key to solving the above challenge and discovering the economic value of data points by leveraging insights from attribution and valuation modules [6]. This composite approach facilitates fair pricing and incentivizes participation, which is applicable to both access-based payments and royalty models.

7.3 XnY's Role in Enabling Data Attribution and Valuation

To build a sustainable AI economy, we must consider **data valuation** and **attribution** when determining rewards for contributions. Data valuation and attribution are practical

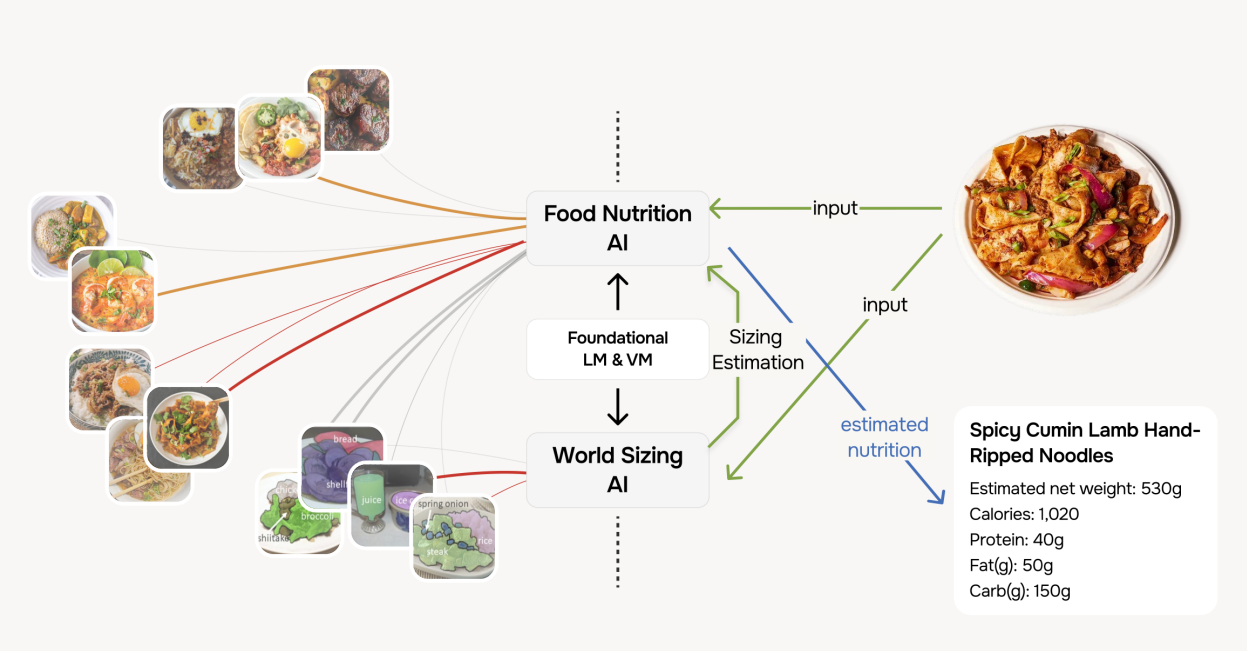

Figure 1: Illustration of Data Attribution and Valuation in AI-Driven Food Nutrition Analysis (Description: Line weight represents the valuation of each data point in terms of its contribution to AI model improvement, while line color indicates attribution during inference, highlighting which data points provide the most information for analyzing the given input.)

mechanisms that allow us to understand the value each data point brings to improving AI systems. Within the same category, the rewards received by data creators should correlate positively with the value their data generates in downstream AI models.

Proper data pricing should consider market dynamics such as scarcity and demand, which can be effectively managed through an **auction-based market** [6]. The auction mechanism helps determine the economic value of data points by integrating insights from attribution and valuation modules, providing a fair and transparent pricing system suitable for both instant access-based payments and long-term royalty models.

To optimize market efficiency, data should be tradable as an asset with minimal friction, allowing easy applications, resale, and further contribution to AI model improvements. XnY supports **data assetification**, transforming data into a tradable and liquid asset that drives economic value.

- **Data Marketplace and Exchange**: XnY creates a data marketplace where data can be traded transparently and fairly. Methods such as attribution, valuation, and auction-based value discovery ensure fairness in transactions, allowing data creators, AI developers, and other stakeholders to discover the actual value of their data.

- **Alignment of Incentives**: By integrating valuation metrics with royalty models, XnY guarantees that data creators receive long-term rewards, proportional to the value generated by their data. This incentivizes sustainable collaboration and encourages

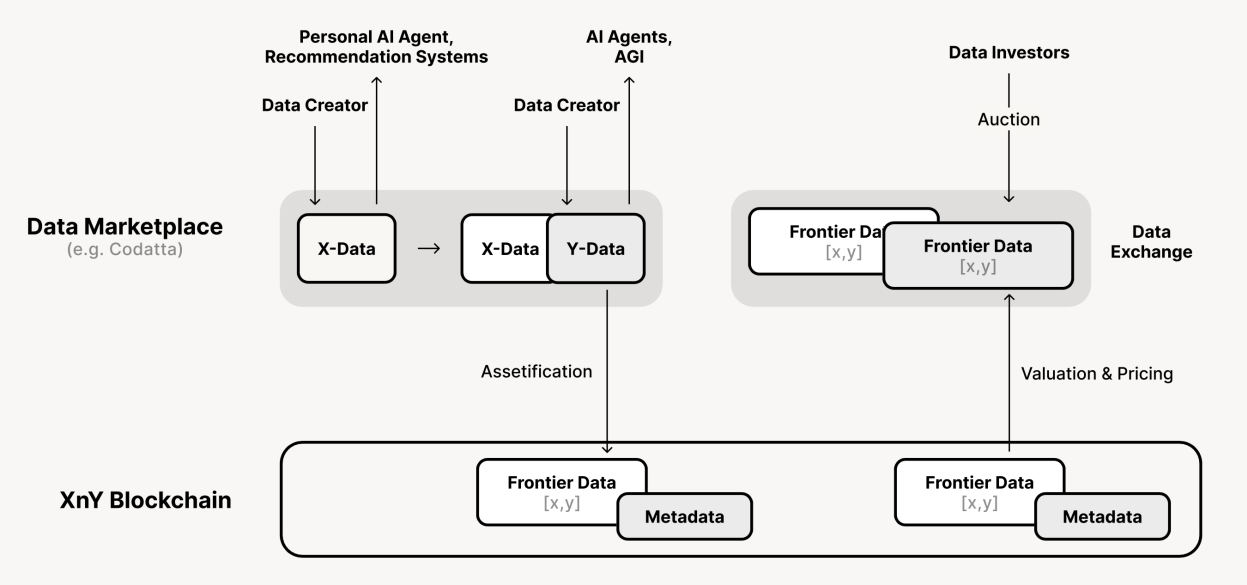

Figure 2: Complete Lifecycle of Data Creation and Assetification in XnY Ecosystem

the continuous provision of high-quality data, essential for AI model improvements.

- **Building Blocks for Data Liquidity**: XnY is focused on assetification and creating tools for tracking data lineage, managing rights, and providing transparent valuation. These capabilities ensure data is treated as a liquid asset, easily tradable, and impactful within the AI economy.

Data attribution, valuation, and auction-based value discovery are central to creating a fair and transparent Open AI Economy. Attribution measures data influence, valuation quantifies worth, and auction-based approaches establish economic value, ensuring fair reward mechanisms for data contributors. Current ranking methods alone are insufficient for assigning economic value, making the XnY platform crucial for a functional and equitable data marketplace.

By uniting data valuation, attribution, and auction-based value discovery, XnY provides the essential infrastructure for data assetification. However, to fully realize business value, we need a marketplace, platform, and exchange to combine these modules effectively and deliver the complete solution. This fosters effective collaboration between AI developers and data creators, enabling fair value distribution and building a sustainable data-driven ecosystem. As the Open AI Economy matures, these capabilities will be essential to recognizing data as a valuable asset, ensuring shared benefits for all contributors.

In today's era of rapid advancements in AGI, the value of data is higher than ever. Yet, the benefits have not been adequately shared with individual Data Creators. The demand for high-quality, domain-specific, reasoning-rich data presents significant challenges, especially for AI developers and researchers who often lack the resources to source such data on a scalable level [13, 14].

XnY provides a foundational building block for data assetification, but this is just one part of the solution. To realize the potential of an Open AI Economy, we will need to build

additional components. These include a marketplace to match business needs, a platform to support AI developers with tools and incentives, and an exchange to facilitate value discovery and improve data liquidity.

XnY aims to create an Open AI Economy where AI developers and researchers can directly collaborate with Data Creators, driving innovation while sharing both risks and rewards. Whether the goal is to build vertical AI products, conduct scientific research, or ultimately contribute to AGI, XnY lays the foundation for these collaborations. By leveraging blockchain-based infrastructure, XnY enables efficient collaboration between Data Creators and AI Developers, laying the foundation but needing additional components to fully address data scarcity and high sourcing costs.

The Open AI Economy empowers Data Creators to become long-term stakeholders in the AI value chain. They contribute their skills and expertise to build high-quality datasets, accepting the risk that some data may not become commercially viable, while still benefiting from the long-term rewards of creating lasting value. This royalty-based model aligns the risks and benefits among all value-creating participants

The platform also establishes a transparent revenue distribution mechanism through multi-role collaboration and data assetification. Key participants benefit in the following ways:

- **Data Suppliers** earn rewards in platform tokens or fiat currency for providing high-quality data, with rewards based on data quality and usage frequency.
- **Data Validators** are rewarded for verifying data accuracy and actively participating in community activities.
- **AI Developers and Researchers** gain direct access to high-quality, domain-specific data that drives innovation and optimizes AI applications.
- **The Platform** earns revenue through transaction fees, increased activity, and the appreciation of governance tokens, enjoying a share of the value created across the ecosystem. These revenues are then shared with Data Creators and other contributors within the community.

The rules and criteria for revenue distribution are initially set by the platform but gradually transition to a community-driven model using token-weighted voting. This governance mechanism allows the platform to adapt to market changes and ecosystem growth over time, fostering an equitable and sustainable Open AI Economy.

8 Conclusion

The XnY ecosystem is a practical leap towards building a genuine Open AI Economy by enabling data assetification and establishing a robust infrastructure for fair collaboration among data creators, AI developers, and investors. With features like data attribution, valuation, and transparent asset trading, XnY allows contributors to share directly in the value they help generate. This litepaper outlines the foundational concepts, but it's just the first step.

The upcoming whitepaper will dive into the specifics of our engineering approach, from methodologies and incentive models to the nuts and bolts of the XnY blockchain and decentralized governance. We'll cover privacy controls, detailed technical mechanisms, and how we plan to implement a sustainable AI economy.

We are also looking forward to collaboration with the open-source community and decentralized protocols to push this mission forward. The reality is that building a complete suite demands a broad range of infrastructure and tooling—such as on-chain inference attribution and AI model asset management. We are excited to unite collective forces to deliver on this mission. Stay tuned for the next iteration, where we'll map out the full path ahead for XnY.

References

[1] Chirag Agarwal, Cuong Nguyen, and Volker Tresp. "Estimating data influence and its application to active learning". In: *arXiv preprint arXiv:2003.04064* (2020).

[2] Michael Armbrust et al. "A data protection impact assessment method for cloud computing in healthcare". In: *Proceedings of the VLDB Endowment* 2.2 (2019), pp. 1651–1654.

[3] Emily M Bender et al. "On the dangers of stochastic parrots: Can language models be too big?" In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), pp. 610–623.

[4] Rishi Bommasani et al. "On the opportunities and risks of foundation models". In: *arXiv preprint arXiv:2108.07258* (2021).

[5] State of California. *California Consumer Privacy Act of 2018*. `https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375`. Accessed: 2023-10-01. 2018.

[6] Yunan Chen, Wenbo Sun, and Chunyan Miao. "Market mechanisms for fair data trade". In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (2018), pp. 5824–5826.

[7] National People's Congress of China. *Personal Information Protection Law of the People's Republic of China*. `http://www.npc.gov.cn/englishnpc/c23934/202108/06052aa6c8d241b7baf59996ae6b89a3.shtml`. Accessed: 2023-10-01. 2021.

[8] Craig Gentry. "Fully homomorphic encryption using ideal lattices". In: *Proceedings of the 41st annual ACM symposium on Theory of computing*. ACM. 2009, pp. 169–178.

[9] Amirata Ghorbani and James Zou. "Data Shapley: Equitable valuation of data for machine learning". In: *International Conference on Machine Learning* (2019), pp. 2242–2251.

[10] Omar Hasan and K Salah. "Data provenance in blockchain". In: *Encyclopedia of Big Data Technologies* (2019), pp. 1–10.

[11] Meredith Jones, Christine Swiatek, and Elisa Bertino. "A framework for understanding data governance". In: *IEEE Security & Privacy* 18.1 (2020), pp. 15–21.

[12] David Leslie. "Understanding artificial intelligence ethics and safety". In: *The Alan Turing Institute* (2020).

[13] Xiaofeng Li and Kun Yu. "Data markets, data sovereignty and data sharing in the digital economy". In: *China Economic Journal* 14.2 (2021), pp. 168–181.

[14] H Rashidul Rahman et al. "A fair data economy for consumers and enterprises". In: *IEEE Consumer Electronics Magazine* 8.2 (2019), pp. 73–80.

[15] Chen Sun et al. "Revisiting unreasonable effectiveness of data in deep learning era". In: *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 843–852.

[16] European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council*. `https://eur-lex.europa.eu/eli/reg/2016/679/oj`. Accessed: 2023-10-01. 2016.

[17] Hal R Varian. "Artificial intelligence, economics, and industrial organization". In: *NBER Working Paper Series* w24839 (2018).

[18] Yan Wang et al. "A systematic literature review on data quality in blockchain". In: *IEEE Access* 8 (2020), pp. 223232–223248.

[19] Guy Zyskind, Oz Nathan, and Alex Pentland. "Decentralizing privacy: Using blockchain to protect personal data". In: *2015 IEEE Security and Privacy Workshops* (2015), pp. 180–184.