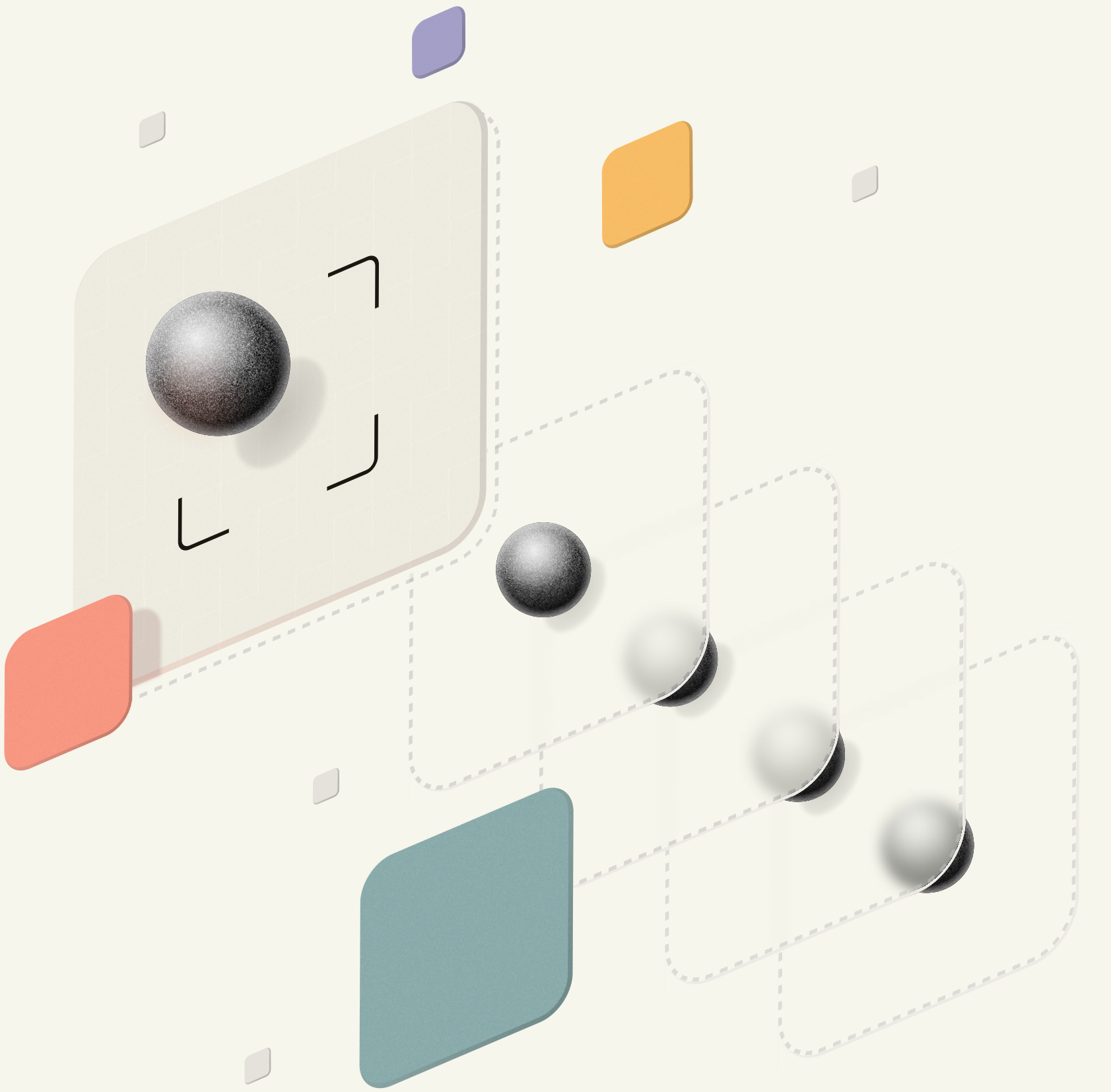
ChainLabel

Decentralized AI Data Labeling for Smart Contracts



Introduction to ChainLabel

ChainLabel is a DAO that runs a data labeling platform that focuses on Web3-specific data labeling, starting with smart contract code annotation. The goal of ChainLabel is to be the leading Web3 data annotation platform, empowering the future of AI agents that can transact in a Web3 and blockchain world. By utilizing labeled smart contract code, developers can build AI agents that can autonomously code, leading to prompt-generated smart contracts, automated scam detection, trading bots built with a simple explanation, and much more.

We recognize the pivotal role of data in advancing AI technologies like GPTs and LLMs. These AI models built a fantastic foundation using a combination of unstructured and structured data. That being said, for AI models built to operate in a Web3 world, blockchain specific structured (also known as labeled) data is missing. ChainLabel solves this gap, as the first and only platform to specialize on annotating the world of Web3 for use by AI, leading to unlimited possibilities at the intersection of AI and blockchain. Notably, the platform operates through community governance, utilizing the $LABEL token to empower labelers with a greater voice in decision-making processes.

Whether you're a data labeler looking to contribute on ChainLabel.ai or an AI developer eager to leverage the platform's labeled data, ChainLabel invites you to join in shaping the future of AI and blockchain technology.

Importance of Labeled Data

Modern AI models leverage both structured and unstructured data to enhance their predictions and analyses. Unstructured data is just raw data– imagine an AI trained on every article ever written on the internet. Structured data takes those articles and labels them with a specific framework, known as an Ontology. Training models with structured OR unstructured data is effective, but combining them both leads to superior results.

In the world of blockchain, everything is publicly readable, so there is a near-unlimited amount of unstructured data. However, structured, labeled data is sorely missing. That's where ChainLabel comes in: a labeling platform focused on Web3 data. Through the provision of labeled smart contract data, ChainLabel empowers our community to train sophisticated models capable of unlocking the synergies between blockchain and AI. Envision a future where AI systems autonomously draft smart contracts, detect fraudulent activities, and undertake a myriad of other advanced tasks.

ChainLabel

Why Structured Data Benefits LLMs

Enhanced Learning Effectiveness —

Structured data follows a clear and systematic format, facilitating AI algorithms in efficient learning. Whether through labeled datasets, tabular structures, or tagged text elements, this organization aids AI systems in swiftly identifying patterns and correlations with precision.

Improved Precision and Operational Efficiency —

Structured data empowers LLMs to grasp context and interrelationships within the dataset more effectively. This structured methodology minimizes data ambiguities, resulting in more precise analyses and outputs. For example, when sentences are labeled with their respective parts of speech in a dataset, LLMs gain a deeper understanding of grammatical intricacies and nuances.

Streamlined Feature Extraction —

Structured data streamlines the feature extraction process, a crucial step in training machine learning models. With explicitly defined features, such as columns in a database, structured data enables algorithms to readily identify and leverage these features for making predictions or producing responses.

Enhanced Scalability and Management —

As datasets expand, handling and processing unstructured data can pose challenges in terms of complexity and resource utilization. Structured data addresses these concerns by providing a consistent format that is more manageable and scalable.

Reduced Error Rates —

Structured data adheres to predefined formats and expectations for each data component, minimizing errors in data processing and analysis. This consistency mitigates misunderstandings and inaccuracies that can arise from ambiguous or unstructured data.

Training and Validation Support —

Structured data plays a vital role in supervised learning, a prevalent training approach for LLMs. It enables distinct separation between input features and target outputs, streamlining the training process as the model learns to associate inputs with accurate outputs.

Our Governance and Incentive Structure

The *$LABEL* token serves as our DAO's governance token, where our community controls the future of the project.

The *$LABEL* token is also offered as an incentive and reward for our labelers. If you log onto ChainLabel and accurately label data, you earn *$LABEL* tokens.Higher accuracy leads to more *$LABEL* rewards, incentivizing high-quality labeling. Inaccurately labeled data does not earn *$LABEL* tokens.

$LABEL tokens have an additional purpose: AI developers can use *$LABEL* tokens to compensate the DAO for access to the labeled smart contract code.

$LABEL Token Economics

Category	Tokens	Total %	Unlocked %	Tokens Unlocked TGE	Vesting Schedule
Marketing	80,000,000	8%	10%	8,000,000	6 Month Lock - 18 Month Vest
Team	110,000,000	11%	-	-	6 Month Lock - 36 Month Vest
Development / Operations	100,000,000	10%	3%	3,000,000	36 Month - Daily Vest
Advisors	60,000,000	6%	-	-	6 Month Lock - 24 Month Vest
Liquidity	330,000,000	33%	80%	264,000,000	48 Month - Daily Vest
Ecosystem / Rewards	220,000,000	22%	5%	11,000,000	48 Month - Daily Vest
Treasury / Reserve	100,000,000	10%	5%	5,000,000	6 Month Lock - 24 Month Vest
Total	**1,000,000,000**	**100%**	**29.1%**	**291,000,000**	

Maintaining Data Labeling Quality

In the realm of AI model training, the principle of 'garbage in, garbage out' is well understood — quality inputs yield quality outputs. For structured data, this underscores the critical importance of quality control, as training AI models on subpar data can yield inaccuracies in their results.

To ensure top-tier output quality, ChainLabel employs a meticulous quality assurance framework. Initially, each piece of data is labeled by three individuals. Subsequently, the results from these labelers are cross-checked for consensus. If unanimous agreement is reached among the three labelers, the data is marked as 'finalized' and becomes downloadable. In cases of disagreement among labelers, a larger group is engaged to achieve consensus before finalizing the data. Additionally, if 'finalized' data is found to be inaccurate, it is promptly removed from circulation and reassigned to additional labelers until a consensus is reached.

Uses for ChainLabel

Autonomous coding —

AI developers can leverage our labeled data to train models proficient in crafting and executing smart contracts. Rather than mastering the intricacies of smart contract coding, one can simply input a query into a large language model. A model trained on ChainLabel's data will then autonomously generate the smart contract code. Drawing from extensive smart contract data across various blockchains, ChainLabel's dataset represents top-tier AI training data.

ChainLabel paves the way for financial empowerment —

Blockchain technology offers the promise of financial liberation, yet achieving parity requires the ability to develop trading algorithms that automate strategies. Writing effective smart contracts, the backbone of such algorithms, demands considerable expertise. ChainLabel bridges this gap by offering access to its labeled smart contract data for training AI platforms. These AI systems can then autonomously generate trading smart contracts, empowering you to leverage your insights and outperform other traders

Shielding against scams with ChainLabel's data —

While smart contracts underpin Web3, delving into their code can be daunting and prone to obfuscation. By training AI models on labeled smart contract data, developers can create algorithms that decipher smart contract code, providing insights into transactions before they are finalized. Envision a future where scams are preemptively exposed, safeguarding users from potential harm.

Product Roadmap

Functions Labeled on ChainLabel —

Our initial developers have started labeling data to get the project started. At launch, the project boasts over 3,500 individually labeled smart contract functions, which have all been run through ChainLabel's thoroughly automated quality assurance process.

Launch

- Smart Contract Function Labeling
  - Label individual functions with a pre-defined ontology

Q3 2024

- Governance Platform
  - Snapshot fork allows community to vote or propose new ontologies

- AI-enabled Quality Assurance Process
  - Shift from the majority driven quality assurance (QA) to AI-enabled, leading to more valuable data

Q4 2024

- Label wallets
  - Add a new labeling tool: label wallets, letting the community tag wallets with different attributes for AI training purposes

- Community-built programming-assistance AI agent
  - Using ChainLabel data, build an AI agent capable of assistive coding

- Community-built scam-detection AI agent
  - Using ChainLabel data, build an AI agent capable of identifying malicious smart contracts

Importance of Community

Labeled smart contracts are essential for AI. Cutting-edge artificial intelligence models like GPTs and LLMs depend on data for their development. Providing these models with structured, labeled smart contract data (the most common form of blockchain code) unlocks limitless potential at the intersection of AI and blockchain.

ChainLabel was built by and for our community. We firmly believe that making labeled smart contract data accessible to everyone is crucial for fostering an open and equitable future. It's time to break down siloed data sources and democratize AI and Blockchain. If you're interested in participating in our community governance and governance initiatives launching soon, start accruing *$LABEL* tokens and get engaged.

ChainLabel is still in its early stages, and we aim to expand our community with your assistance. Enhance and augment tags, train models, and explore endless possibilities. Eventually, the community might decide to broaden the platform to include AI data beyond smart contract labels. Eager to support our cause? Visit ChainLabel.ai

AI is fueled by properly labeled data

Let’s label and catalog the world’s information in order to achieve the full potential of artificial intelligence.