

Get paid for the data that trains AI.

A Solana-powered, decentralised marketplace for AI training data where contributors own the upside.

§ 00 — ABSTRACT

Modern foundation models are trained on data labeled, ranked, and curated by humans who capture almost none of the resulting value. Meta paid **\$14.3B** for one such pipeline in 2024. The workers who produced that data received industry-standard piece-rates and zero residual ownership.

\$DATA is a protocol for a public, on-chain training-data marketplace. Contributors submit labeled samples, validators verify them with ZK-attested proofs, and buyers (labs, startups, sovereign AI programs) settle in USDC on Solana. Eighty percent of revenue is routed to contributors and validators by protocol rule — not by policy.

§ 01 — THE PROBLEM

Value capture is broken. The AI training-data market is projected at \$23.0B in 2025, growing 38% YoY. Less than 3% of that flows to the humans whose labor produces the underlying data.

Provenance is opaque. Buyers cannot prove where data came from, who labeled it, or whether it was sourced ethically. Datasets are black boxes.

Settlement is slow. Net-30 to net-90 payouts are standard. Contributors in emerging markets bear FX risk and bank friction.

§ 02 — ARCHITECTURE

\$DATA is a set of Solana programs plus an off-chain validator network. Three actors: **contributors** submit samples, **validators** stake \$DATA and verify, **buyers** escrow USDC and consume datasets.

Submission. A contributor uploads a sample with metadata (modality, schema, license). The CLI computes a content hash and posts a commitment on-chain.

Validation. Three independent validators score the sample. A Groth16 proof attests to validator identity, stake, and scoring honesty without revealing the sample itself. Median latency: 180–300ms.

Settlement. When a buyer purchases a dataset, the program splits payment atomically: 80% to contributors (weighted by accepted-sample count), 12% to validators, 8% to the protocol treasury. Finality in ~400ms.

§ 03 — TOKENOMICS

Fixed supply: **1,000,000,000 \$DATA**. No inflation. No team unlock cliffs that dump on the market.

ALLOCATION	%	PURPOSE
Contributors & rewards	55%	Streamed to contributors over 4 years
Liquidity & market-making	15%	DEX seeding, CEX listings
Treasury	15%	Governance-controlled grants
Team & advisors	10%	4-year linear vest, 12-month cliff
Public sale (pump.fun)	5%	Fair launch, no allocation gate

§ 04 — GOVERNANCE

\$DATA holders vote on validator slashing parameters, fee splits, treasury grants, and protocol upgrades. Voting is quadratic with a 7-day timelock on execution. No multisig backdoor.

§ 05 — SECURITY

Programs audited by **OtterSec**. ZK circuits reviewed by independent cryptographers. Validator stake is slashable for collusion or downtime. Bug bounty: up to \$500K via Immunefi.

§ 06 — ROADMAP

Q3 2026 — Mainnet launch, image/text modalities.

Q4 2026 — Audio + video modalities, validator delegation.

Q1 2027 — Cross-chain settlement (Base, Solana, Sui).

Q2 2027 — On-chain dataset versioning + lineage proofs.

§ 07 — DISCLAIMER

This document is for informational purposes only and does not constitute an offer to sell or a solicitation to buy any security. \$DATA is a utility token. Participation is restricted in certain jurisdictions. Read the full terms at data.xyz/legal before contributing or purchasing.

