Technical Report

DeFi Protocol Contracts and ABI Processing Pipeline

This report documents the data collection and processing pipeline for DeFi protocol contracts, their ABIs, and decoded functions/events.

1. Introduction

The goal of this project is to build a structured dataset linking DeFi protocols to their deployed smart contracts, verified ABIs, and decoded interfaces (functions and events). The workflow is executed in three phases: dataset gathering, ABI fetching, and ABI decoding. Supporting JSON files from DeFiLlama provide protocol-level metadata.

2. Stepwise Workflow

Step 1: Dataset Gathering

Input file: *protocols.csv* (12,623 rows). Each row represents a protocol–network entry. However, not all entries include contract addresses. Extracted subset: *defillama_extracted_contracts.csv* (3,545 rows). Columns: protocol, slug, contract_address. This subset represents the actual contracts retrievable from DeFiLlama JSON/adapters.

Step 2: ABI Fetching

Input: 3,545 contract addresses from Step 1. Process: Queried blockchain explorer APIs (Etherscan, BscScan, SnowTrace, etc.) using (network, contract_address) pairs. Returned ABI if verified. Output: *contracts_with_abis.csv* (3,545 rows). Columns: protocol, slug, contract_address, abi, network. Results: - Verified ABIs obtained: 944 (26.63%) - Contracts without ABIs: 2601 (73.37%) - Network missing for 978 rows (~27.59%)

Step 3: ABI Decoding

Input: contracts_with_abis.csv Process: Parsed ABI JSON into discrete and signatures. Output: *contracts_events_and_functions.csv* (24773 rows). Coverage: - Unique contracts decoded: 780 - Functions: 20,419 - Events: 4,354 All decoded rows map back to contracts in Step 2. One ABI-present address did not yield decoded rows.

Supporting Metadata: llama_protocols.json

File contains protocol-level metadata (6,445 entries). Includes fields: name, slug, chains, categories, TVL, audits, etc. All Step-2 slugs were found in this file (no missing matches).

Supporting JSON: all_protocols_contracts_with_abis.json

Aggregated JSON combining protocols and ABIs (3,545 entries). Matches Step-2 one-to-one at the address level (excluding a few null address placeholders).

3. Funnel Summary

Stage	Count
Protocol-network entries (protocols.csv)	12623
Contracts extracted (Step 1)	3545
Contracts processed for ABIs (Step 2)	3545
Verified ABIs obtained	944
Decoded contracts (Step 3)	780
Decoded functions+events (rows)	24773

4. Key Observations

- The project did not attempt ABI fetching for all 12,623 entries. Instead, it correctly filtered to 3,545 contracts with extractable addresses. - ABI coverage remains low (26.6%). The majority of contracts are either unverified or inaccessible via explorers. - Network field propagation is incomplete; ~28% missing in Step 2, ~98% missing in Step 3. Needs backfilling for better analysis. - Contract_name field is empty in Step 3. If required, can be enriched from explorers. - Duplicate addresses exist across Step-2 rows (327 addresses repeat, max 9×). Deduplication by unique address is recommended for ABI stats.