
phuse
connect·share·advance
Single Day Events
phuse.global


# Open-Source AI augmented workflow for Automated TLF Generation and Quality Control

Anshu Aman, Sunny R Choudhary
ZS Associates

phuse.global

# Agenda



- Understanding TLF Bottleneck
- A Multi-Agent Architecture for TLF generation
- Automated Quality Control, Traceability & Compliance
- Q&A: Open discussion.

phuse

connect·share·advance

# TLF Generation

Understanding the operational challenges and limitations in standard clinical reporting

# TLF creation process

connect·share·advance

The manual generation of Tables, Listings, and Figures (TLFs) comprises a multi-stage workflow encompassing data extraction, manipulation, and formatting. This methodology is inherently resource-intensive and prone to human error.

## Upstream data

- Approved SAP and upstream metadata
- A structured Data Analysis Plan
- Additional skills files (if required) with CDISC glossary and guardrails

## Coding

- Demands redundant, line-by-line manual programming against ADaM datasets.
- Programmers does tedious reverse-engineer manual mockups into working code

## Input

- Structured ADaM dataset (.parquet/.sas7bdat)
- Study inputs and requirements are tracked using spreadsheets and manual specification files.

## ARD Generation

- Quality control is bottlenecked by resource-intensive double-programming.
- Requires significant manual time for data validation, cleaning, and cross-checking

## TLF Designing

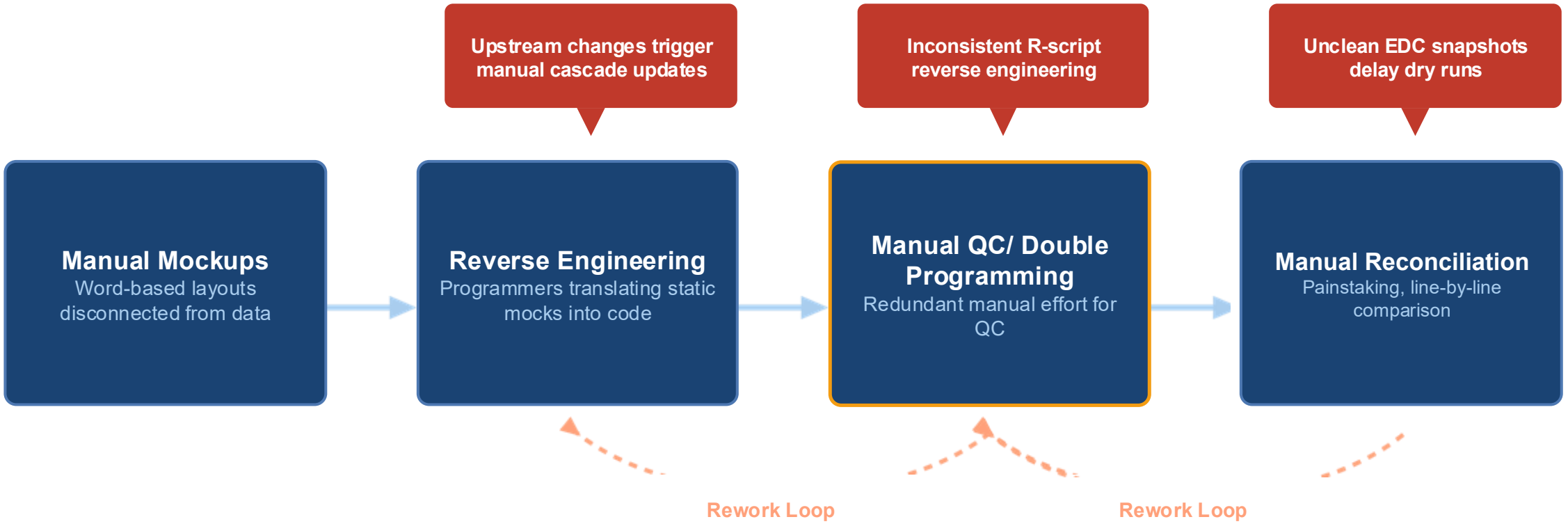
- Manual mockup creation to define the intended outputs on Word
- Creates static targets that remain completely separate from actual production

## TLF Output

- Struggles to meet the demand for faster reporting workflows caused by escalating trial complexity.

# The Bottleneck






**Takeaway:** Legacy workflows remain brittle and inefficient, creating significant risks for data integrity and timelines.

# The Bottleneck



Relying on manual, human-driven processes introduces massive time sinks into the clinical reporting lifecycle.

Based on the metrics in the graph, we are currently losing hundreds of hours to repetitive tasks:

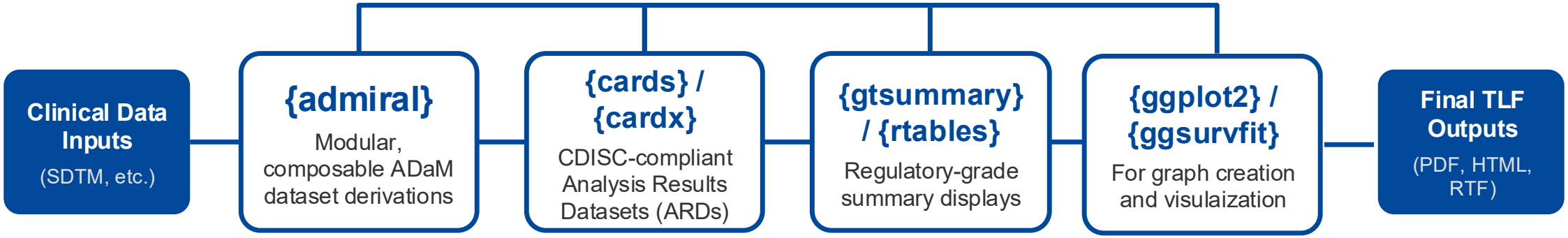
- **Standard Mockup Creation:** Teams requires approximately 80 hours manually drafting and reverse-engineering standard outputs.
- **First-Line Programming:** The initial development phase consumes roughly 440 hours of manual, line-by-line coding effort.
- **Quality Control:** The validation phase introduces a massive operational burden by relying on redundant, black-box double-programming.


Graph presented here is for representation purpose only.

# The Foundation: Standardizing on Pharmaverse

**Before AI can automate, the pipeline must be standardized.**

Open-source, GxP-compliant R packages replace proprietary legacy systems.




**Impact:**

Recent studies demonstrate a **15–20% reduction** in TLF development time simply by transitioning to pharmaverse workflows.

**Regulatory Alignment:**

Delivering consistent, GxP-compliant clinical reporting workflows designed for end-to-end auditability and traceability.

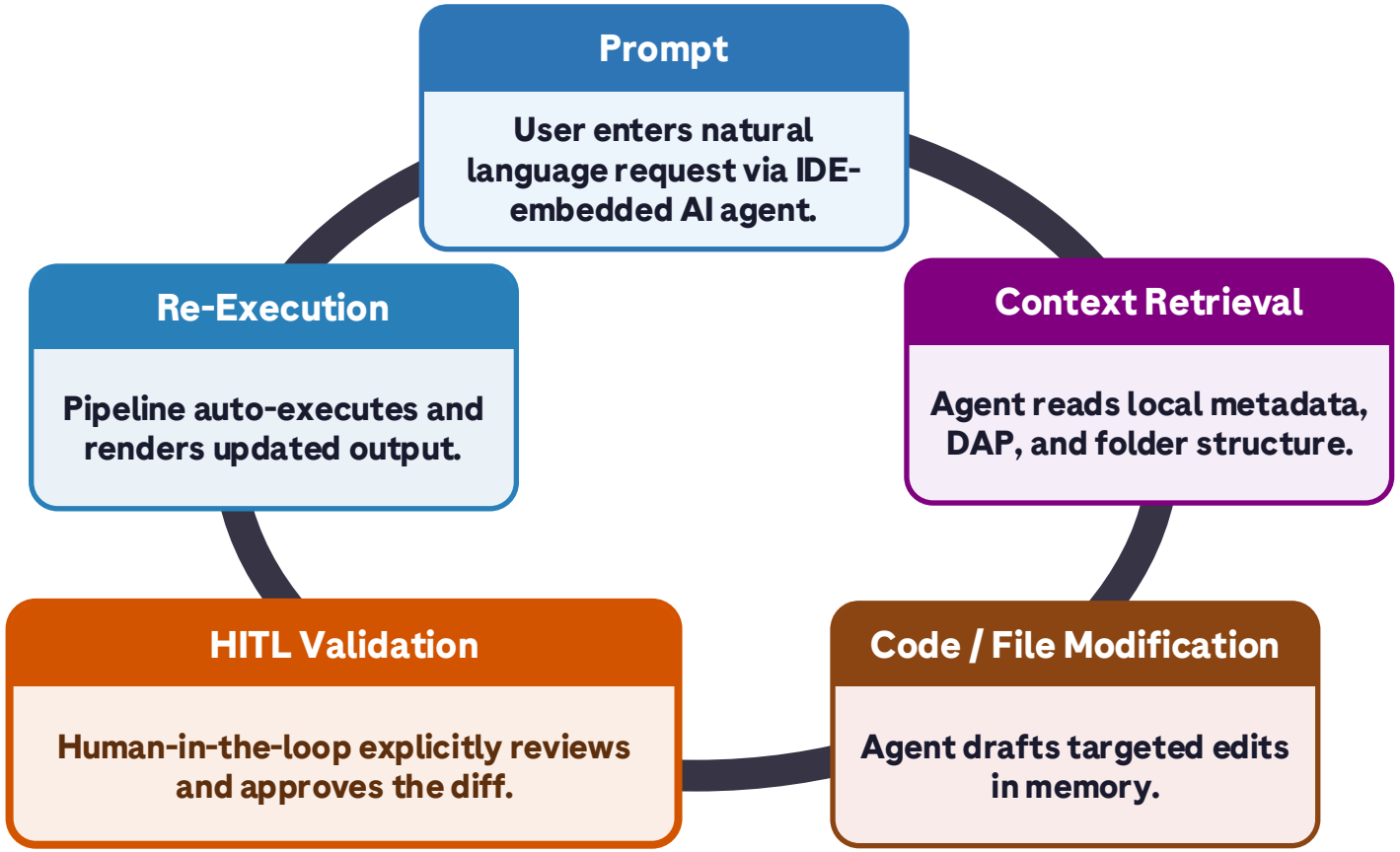
# Multi-Agent Architecture

HITL framework for multi agent based TLF creation process

# Conversational Edit Loop

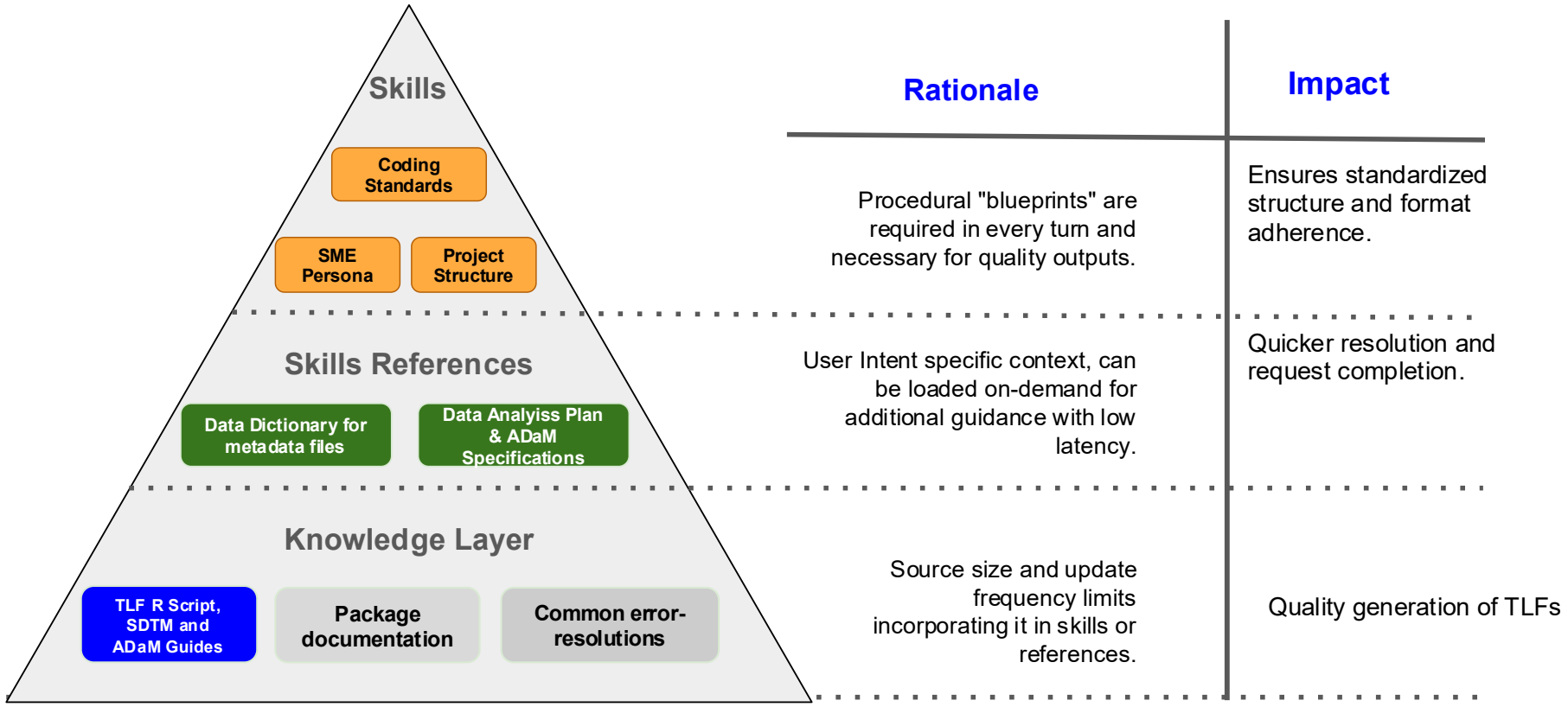


Pharmaverse augmented TLF agent's conversational edit loop enables rapid, iterative refinements while maintaining strict human-in-the-loop validation for every modification.



# Multi-Layer Architecture for TLF context

- To ensure consistent, high-quality TLF generation, we employ a multi-layered architectural approach that integrates specific skills, project references, and a broader knowledge base.
- This structure ensures compliant and accurate output by providing the necessary procedural blueprints required at each stage of the generation process.

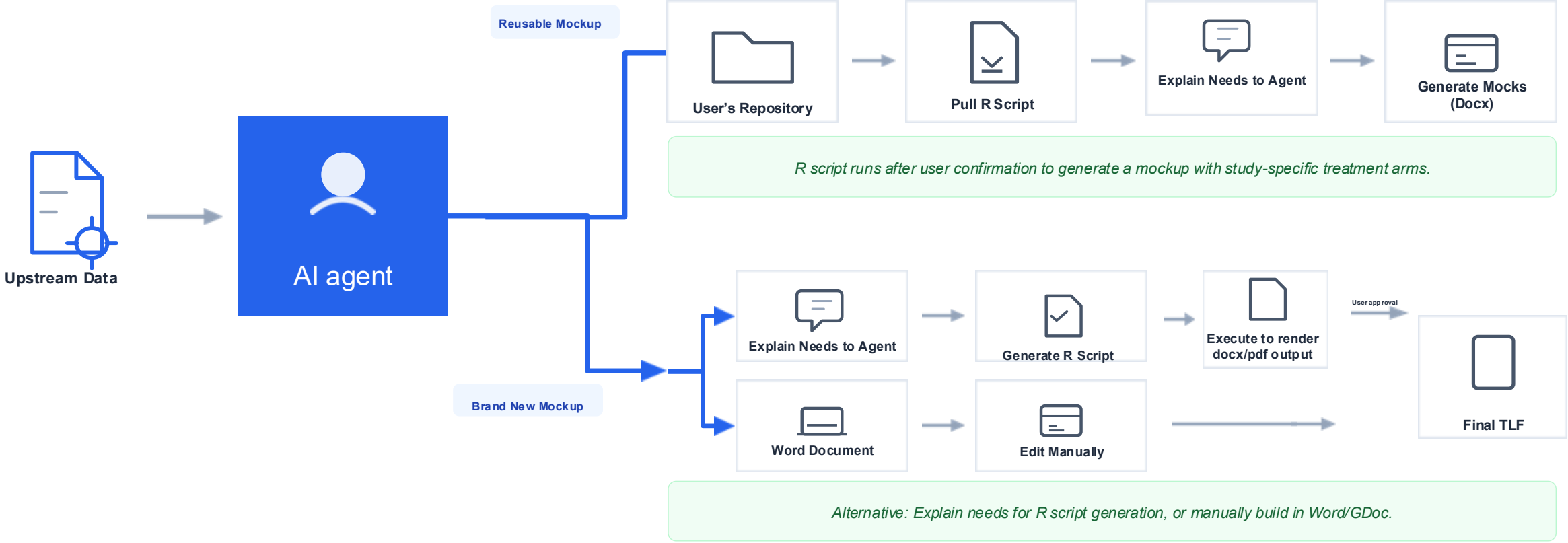


# TLF Generation using AI agents

connect·share·advance

## Pillar 1: Smart context ingestion

This AI-driven workflow streamlines the TLF generation process by automating context ingestion, script execution, and output rendering while maintaining critical human oversight.



# TLF Generation using AI agents

connect·share·advance

## Pillar 2: Human-in-the-Loop Conversational Refinement

The AI does not generate a black-box image; it precisely modifies the underlying **R script**. Explicit human confirmation is required before code is committed.

### User Chat Interface

**User**
"Add BMI to this demographic table using continuous2 styling."

**AI Assistant**
I have analyzed your request. I will inject bmi into the select statement and map the continuous variables to the double statistic layout.

**Execution Pipeline Actions**
- ✓ Inserted bmi into target select statement
- ✓ Expanded continuous2 formatting profile
- ✓ Verified structural integrity against ADaM spec

### </> Interactive R Script Editor

```
library(gtsummary) library(dplyr) # Create demographic table tbl_demographics <- trial %>% select(age, grade, bmi) %>% # BMI added here tbl_summary( type = list(all_continuous() ~ "continuous2"), statistic = list(all_continuous() ~ c("{median} ({p25}, {p75})", "{mean} ({sd})")) )
```

UPDATED

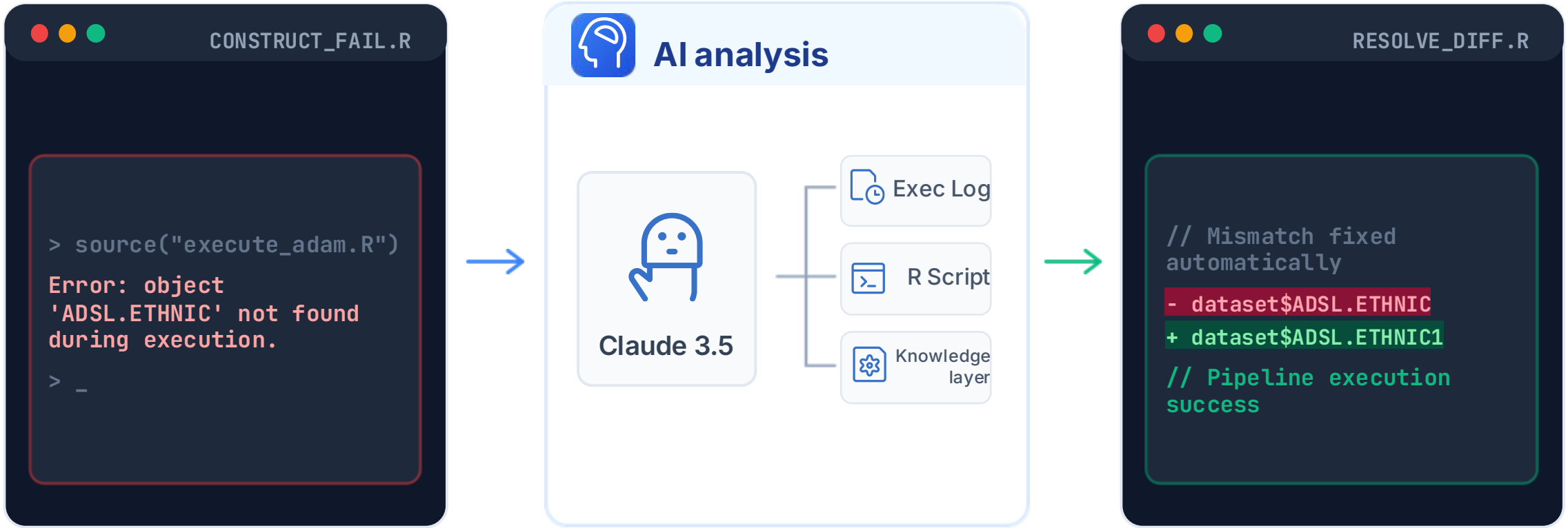
| Characteristic | N = 200 | |
|---|---|---|
| Age, years | 47 (38, 56)<br>47.3 (12.1) | |
| Grade | | |
| I | 68 (34%) | |
| II | 76 (38%) | |
| III | 56 (28%) | |
| BMI, kg/m² | 27.9 (24.8, 31.3)<br>28.1 (4.9) | NEW ROW |

# TLF Generation using AI agents



**Pillar 3: Context aware Debugging**

When executing against actual ADaM datasets, parameter mismatches can cause failures. TLF agent acts as **"The Fixer"** by reading execution logs and required files to propose targeted fixes.



# Multi-Layer Architecture for TLF context

| Agentic Layer | Best For… | Characteristics | Best Match For |
|---|---|---|---|
| Skill Body (In-Context ) | Core Directives & Rules | • Size: < 5K tokens<br>• Volatility: Stable (e.g., core SOPs)<br>• Precision: Procedural (Do X, then Y) | Foundational rules, coding guidelines, folder structures, and interaction protocols. |
| Skill Reference (On-Demand / Local Files) | Exact Specifications | • Size: 5K – 50K tokens<br>• Volatility: Semi-stable (e.g., IG versions)<br>• Precision: Reference (Exact schema text) | Schemas, specifications, and metadata structures that the AI only needs when executing a specific task. |
| MCP / RAG (Dedicated Knowledge Layer) | Deep Exploration & History | • Size: > 50K tokens<br>• Volatility: Frequently updated<br>• Precision: Exploratory (Find examples) | Massive documentation libraries, historical codebases, and expansive catalogues. |

phuse

connect·share·advance

# Automated Quality Control, Traceability & Compliance

Shifting from manual oversight to deterministic, auditable, and GxP-compliant automated workflows.

# The Regulatory Imperative

Regulatory authorities (FDA, EMA) and ICH E6(R3) guidelines require strict traceability, reproducibility, and GxP compliance.

They will not accept **"AI generated codes"** as an audit trail.

### The Standard Solution

Submitted code should be **fully reproducible** without the LLM context.

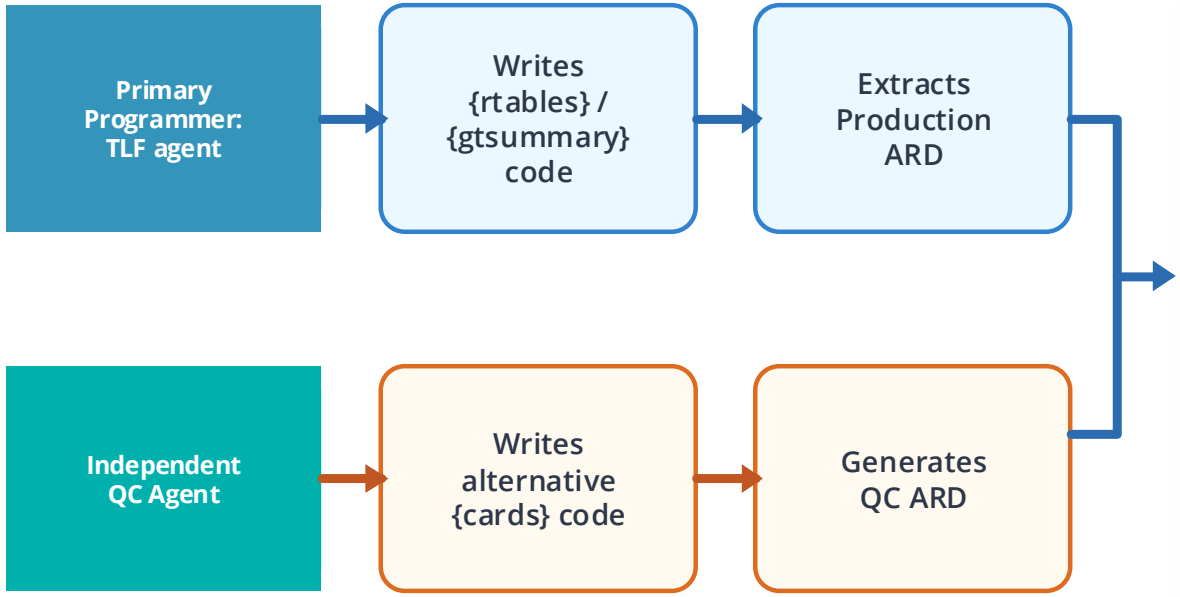
This way, the AI functions as a **co-pilot**, not the runtime environment.

# Redefining Quality Control

connect·share·advance

## Automated ARD Reconciliation

Traditional double-programming requires dual human effort and manual visual comparison of final TLFs.




### Automated Programmatic Diff-Checker

| Characteristic | N = 200 | N = 200 |
|---|---|---|
| Age, years | 47 (38, 56)<br>47.3 (12.1) | 47 (38, 56)<br>47.3 (12.1) |
| Grade | | |
| I | 68 (34%) | 68 (34%) |
| II | 76 (38%) | 76 (38%) |
| III | 56 (28%) | 56 (28%) |
| BMI, kg/m² | 27.9 (24.8, 31.3)<br>28.1 (4.9) | 27.9 (24.8, 31.3)<br>28.2 (4.9) |

Statistical Mismatch

Because ARDs are structured dataframes (not PDFs), the system performs a programmatic, cell-by-cell mathematical comparison. Humans only review flagged statistical mismatches.

# Radical Traceability: Dual-Layer Observability

## Layer 1: Prompt & Agent Tracking

`LangSmith`

- **Records every step:** Captures user interactions, precise agent thought processes, and intermediate generated code snippets in real-time.
- **Captures the 'Why':** Permanently preserves the critical reasoning pathway behind every automation decision and clinical transition.
- **Human-in-the-Loop:** Automatically updates the shared Knowledge Layer based on manual exception handling and error resolutions.

## Layer 2: Execution Logging

`{logrx}`

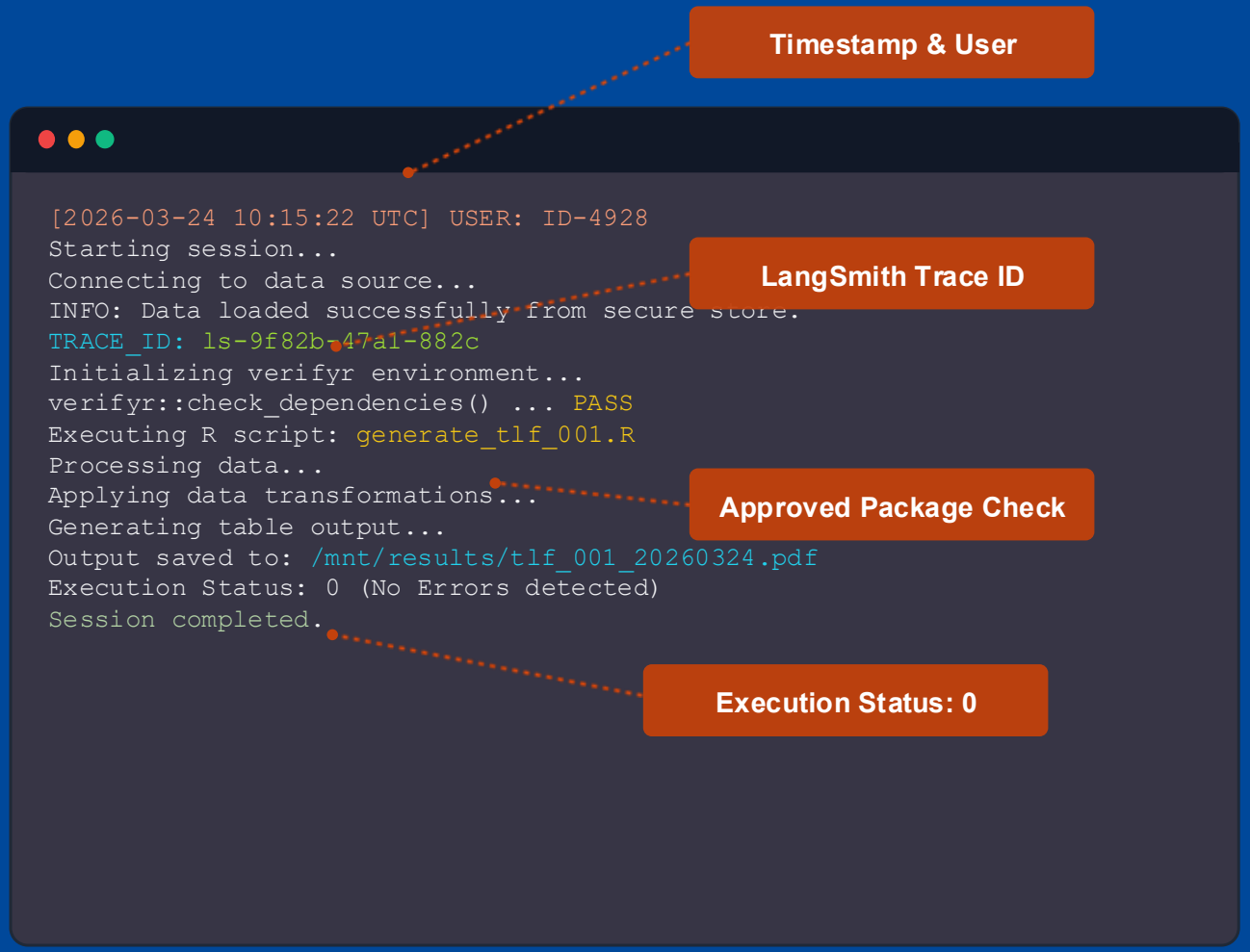
- **SAS-Style Logging:** Generates familiar, robust execution logs designed specifically for pharmaceutical compliance and audit trails.
- **Environment Snapshots:** Captures the complete computational context, exact R package versions, and standard system error logs.

# The Deterministic Audit Trail

connect·share·advance

Every **finalized TLF** can trace its **exact lineage**.

We can perfectly reproduce the output **independent of the AI**, securing GxP compliance and mathematical reproducibility.



# Workflow Diagnostics: Legacy vs. Augmented

| Metric | Legacy Workflow | AI-Augmented Pharmaverse |
|---|---|---|
| **Architecture** | Siloed proprietary macros | Open-source modular<br>R stack ecosystem |
| **Drafting Speed** | 80+ hours manual mockups | Instant synthetic mockups<br>Generated via upstream metadata |
| **Quality Control** | Manual double-programming | Automated ARD checks<br>Deterministic differential comparison |
| **Traceability** | Disconnected Word docs | Reproducible Git history<br>Fully logged execution and prompts |

connect·share·advance


# THANK YOU!


phuse

# The Global Healthcare Data Science Community

**Contact Channels**

office@phuse.global

phuse.global

**Social Media**

/company/phuse

/phusetube

/phuse_global

/phusebook