HARPY: Human-AI Collaboration for Database Error Resolution via LLMs

Maximilian Plazotta[1], Meike Klettke[1]

[1]Data Engineering Group, Faculty of Informatics and Data Science, University of Regensburg

Motivation

- Real-world relational data often contains subtle logical, format, or range errors
- Conventional statistical / rule-based cleaners miss many of these anomalies
- Large Language Models (LLMs) detect them at scale but are prone to hallucinations
- The Human-in-the-loop (HITL) approach is essential to validate LLM responses

The HARPY idea

We combine the *efficiency* of LLM automation with *expert validation* via human-in-the-loop oversight — a reliable, modular and cost-efficient database error remediation pipeline.

Demonstration Scenarios

Scenario 1 – Scientific data

Data never leaves the host machine. Applied to a marine-biology dataset (100 rows, 5 cols): typos, NULL, and logical errors were detected locally.

Scenario 2 – Enterprise data

High-volume: cost, scalability and trust. Claude Haiku 4.5 for precision and speed — all errors found with correct suggestions.

Scenario 3 – Bring your own data

Anyone can run HARPY on private datasets with sufficient local RAM or an API connection to the chosen LLM vendor.

The HARPY System: High-Level Architecture

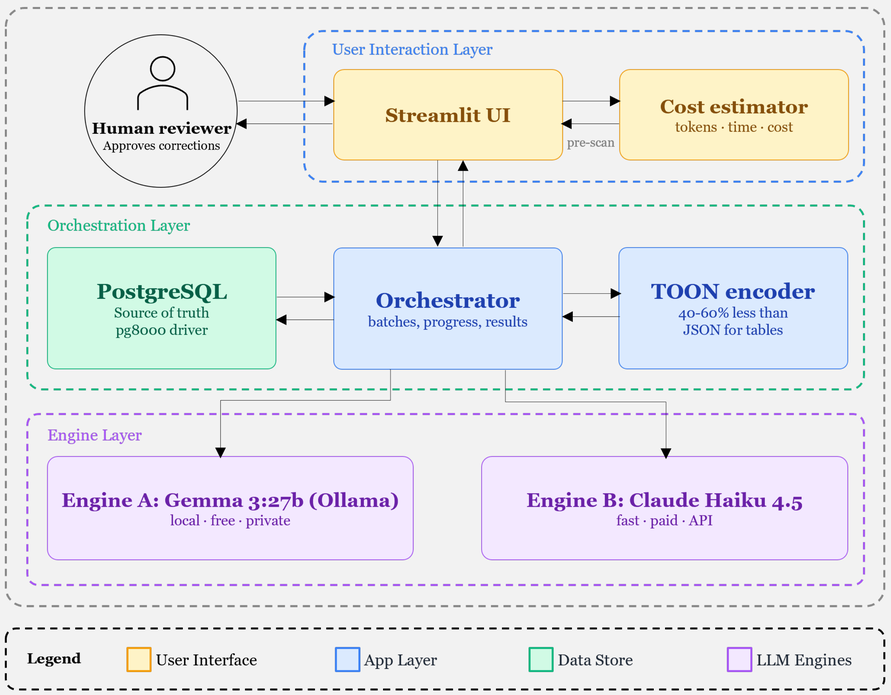


Figure 1: The HARPY system: user-facing interaction layer, orchestration layer, and modular engine layer.

- **Interaction layer:** Streamlit UI offering pre-scan cost estimation and an interactive error review queue for expert validation.
- **Orchestration layer:** segments the PostgreSQL table into batches, converts them into TOON format, dispatches the prompt, and collects structured error reports.
- **Engine layer:** modular and pluggable – any open- or closed-source LLM can be swapped in within minutes (Gemma 3, Claude Haiku 4.5).

Interactive UI

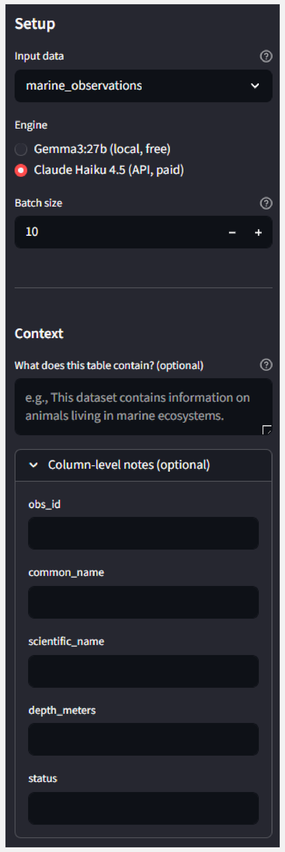


Figure 2: Configuration sidebar.

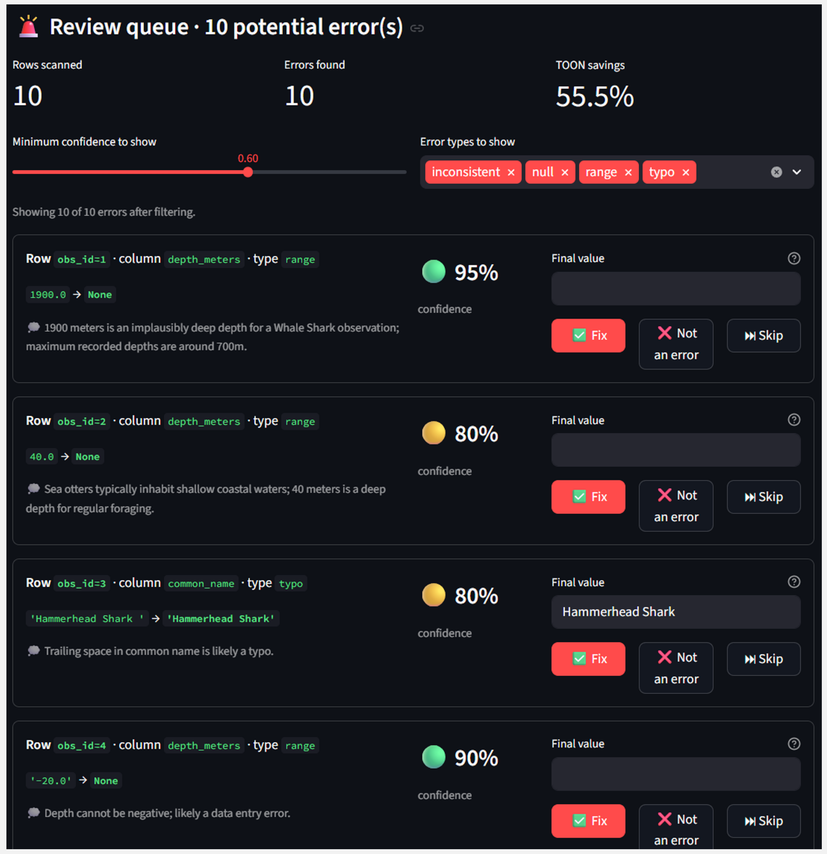


Figure 3: Error review queue with per-anomaly remediation.

Conclusion & Future Work

Conclusion. HARPY is a well-suited example of human-AI collaboration:

- *HITL* mitigates the hallucination risks inherent to AI-only systems.
- *TOON encoding* achieves significant cost reduction.
- The *modular architecture* enables seamless transitions between privacy-sensitive and cost-focused use cases.

Future Work.

- Extend detection to learned, past errors.
- Support *multi-table* cross-references error detection.
- Benchmark across additional domains and larger scale-out datasets.

Contact & Paper

Email: maximilian.plazotta@informatik.uni-regensburg.de
meike.klettke@informatik.uni-regensburg.de

Conference: ADBIS 2026: Demo Paper

Universität Regensburg