TrialGuardianAI

An Integrated Insight-Driven Data-Flow Model for Real-Time Clinical Trial Oversight

Bridging the gap between siloed clinical data and scientific decision-making with Agentic & Generative AI.

Project Name: TrialGuardianAI

Subtitle: An Integrated Insight-Driven Data-Flow Model for Real-Time Clinical Trial Oversight

Tagline: Bridging the gap between siloed clinical data and scientific decision-making with Agentic & Generative AI.

Created By: Yash Dahiya

Event: Novartis Hackathon 2026

The Problem Statement – Fragmented & Siloed Data

The Pain Points:

- **Heterogeneous Sources: 23+ studies, each with 9+ disjointed files (EDC, Lab, Monitoring).**
- **Operational Bottlenecks: Manual review of ~58,000 rows leads to delayed identification of "Red" sites.**
- **Fragmented Communication: High cycle times for CRAs and CTTs to resolve queries.**

Current State: **High operational risk and inconsistent data quality due to lack of a "unified spine."**

My Solution – TrialGuardianAI

The Vision: A single source of truth that harmonizes multi-study data into a single master dataset using Automated Data Ingestion.

Core Value Proposition:

- **Ingest & Harmonize: Automated concatenation of fragmented study files into a "Master Analytical Spine" using the concept of Primary and Foreign Keys.**
- **Derived Metrics: Calculating percentages for all key parameters, including missing visits, missing pages, open queries, non-conformant data, verification status.**
- **Advanced Analytics: Deterministic business logic for Patient Status (Clean/Not Clean), Site Readiness and Subject DQI.**
- **Agentic AI: Rule-driven agents that proactively flag risks.**
- **Generative AI: Context-aware performance summaries for immediate action.**

The Integrated Data-Flow (Technical Architecture)

The "Spine" Table: Using the CPID file as the join-key to merge "inactivated," "pages," and "query" logs.

The 5-Phase Pipeline:

- **Phases 1: Feature engineering (missing visits, missing pages, uncoded terms, safety_queries, problematic_crf, protocol_deviations, crf_not_signed, days_outstanding, problematic_reviews, missing_crf (days), no_data_on_form, open_issues.**
- **Phase 2-3: Deriving the Patient_Clean_Status, Blocking_Reason and DQI_Subject_Score. Aggregating Subject Level dataset to Site Level KPIs, and assigns each site a Red/Amber/Green Site_Risk_Status based on DQI and safety queries and then computing CRA-level performance metrics (sites handled, average site DQI, high-risk patients, average queries per patient).**
- **Phase 4: Performing Interim Analysis at Site, Country and Region level then creating a column named Analysis_Readiness.**
- **Phase 5: The AI Layer (Deterministic Logic + Agentic flagging + GenAI summaries).**

Defining the Quality Engine (Metrics & Logic)

1. Data Quality Index (DQI)

A weighted scoring algorithm (100 - penalties) evaluating missingness, query aging, and safety flags.

- **Base Score: 100 points.**
- **Deductions: Missing Visits (-5.0), Protocol Deviations (-4.0), Open Queries (-1.5), and Unverified/Unsigned Pages (-1.0).**
- **Safety Override: Any Safety Query triggers an immediate 20-point penalty and Hard Caps the total score at 60, ensuring medical risks are instantly visible therefore no subject with a Safety Query can never score above 60, regardless of other data completeness.**

2. Patient Clean Status

The status is determined using a Zero-Tolerance Policy.

- "Clean": Only if every single metric (Derived in phase 1) is 0. This means the patient has no missing data, no open queries, and is fully verified/signed.
- "Not Clean": If even one metric is greater than 0 (e.g., a single open query), the patient is immediately disqualified from "Clean" status.

3. Blocking Reasons

For every subject marked as "Not Clean," the system dynamically generates a report to tell the CRA exactly why.

- The Logic: The script identifies which specific columns in the set are greater than 0.
- Concatenation: It extracts the names of those failing columns and joins them into a comma-separated string (e.g. "Missing_Pages, Open Queries").

4. Analysis Readiness

A site is only flagged as 'Ready' if it simultaneously passes three distinct quality hurdles, failure in any single dimension results in a 'Not Ready' status, mandating further CRA intervention:

- **Clean Patient Rate (> 90%): At least 90% of the patients at the site must have a "Clean" status (zero missing visits, zero open queries, and zero signature/verification gaps). This ensures that the vast majority of the site's data is complete and verified, preventing significant data shifts during final analysis.**
- **Total Safety Queries (== 0): The site must have zero outstanding queries related to patient safety (Adverse Events/Serious Adverse Events). A site cannot be considered ready if there are unresolved questions regarding the safety profile of the trial participants.**
- **Average Site DQI (> 80): The mean Data Quality Index (DQI) across all subjects at the site must be at least 80/100. Even if the patients are "clean," the DQI check ensures that the overall quality of the data entry and management at the site is high and not just "barely passing."**

5. Site Risk Strategy (Traffic Light System)

🟢 Green (On Track)

High DQI, negligible backlog.

🟡 Amber (Monitor)

Dropping DQI or mounting queries; requires targeted CRA review.

🔴 Red (Critical)

Safety queries present or extremely low clean patient rate. Requires immediate site intervention.

Meaning for CRA: This logic provides a prioritized roadmap that flags high-risk sites for immediate intervention while validating the data quality of healthy sites. It transforms raw metrics into a clear action plan, ensuring they focus their time on resolving safety queries and critical data gaps at "Red" and "Amber" locations.

Rule-Driven Agentic AI

- **Proactive Signal Detection: Automatically monitors every site for specific "Risk Signals" such as SAFETY_ESCALATION (any safety query), LOW_DQI (Avg DQI < 70%), and QUERY_BACKLOG (>20 queries).**
- **Automated Action Playbook: Maps detected signals to high-priority clinical tasks, ensuring CRAs receive a pre-defined resolution plan (e.g., 'Immediate CRA outreach and site action plan required') rather than just raw data.**
- **Dynamic Dashboard Tagging: Generates "Traffic Light" status indicators (🔴, 🟠, 🟢) to allow executive teams to manage by exception and identify bottlenecks in real-time.**
- **Efficiency Impact: Bridges the gap between data identification and action, ensuring critical safety and quality risks are escalated immediately without waiting for manual human review.**

Generative AI (Gemini)

- **Cognitive Data Synthesis: Utilizes Gemini-2.5-Flash to transform technical site metrics and agentic signals into clear, natural language performance narratives.**
- **Stakeholder Personas: Employs specialized prompt engineering to create bulleted "on-the-move" summaries for CRAs and formal, audit-ready technical reports for Clinical Study Managers (CSM's).**
- **Priority Batching: Optimizes oversight by batch-processing sites based on severity—focusing AI resources on "Red" and "Amber" sites first to summarize the most critical trial risks.**
- **Human-Centric Reporting: Consolidates all insights into the Full_CRA_Site_Performance_Reports.txt, providing a searchable "single source of truth" for site-specific qualitative analysis.**

The Interactive Dashboard (Demo)

The TrialGuardianAI Dashboard is a multi-tier intelligence hub that converts raw clinical data into a real-time oversight tool. It allows users to filter by global trends, drill down into specific countries, navigate to Subject, Site, Country, Region level and show AI-generated performance summaries for individual sites under the section named "AI Insights".

1. Executive Visibility (Region & Country Level)

- Global Health Mapping: Monitors AMERICA, ASIA, and EMEA regions using executive scorecards that track Average DQI and Database Lock Readiness (%).
- Performance Scatter Plots: Visualizes the relationship between data quality and readiness at the country level. This allows leaders to identify geographic bottlenecks (e.g., countries with high site counts but low DQI).

- Trend Classification: Countries and regions are automatically tagged as "Excellent," "Stable," or "At Risk" based on cumulative performance scores.

2. Operational Deep-Dive (Site Level)

- Dynamic Risk Filters: Users can filter sites by Risk Status (🔴 Red, 🟠 Amber, 🟢 Green) and Analysis Readiness to focus resources on critical interventions.
- AI Insight Summary: Integrates Gemini-1.5-Flash directly into the interface. By selecting a site, the dashboard retrieves a natural-language narrative that interprets complex metrics into a brief performance story.
- Actionable Playbook: Displays "Recommended Actions" directly on-screen, such as "Immediate CRA outreach" or "Query aging review," closing the gap between data and action.

3. Patient-Level Integrity (Subject Level)

The "Gatekeeper" View: Detailed visibility into individual patient metrics (Missing Visits, Protocol Deviations, etc.).

Blocking Reason Identification: Transparently lists why a patient is marked as "Not Clean," providing CRAs with a specific task list for data remediation.

Issue Distribution: Interactive bar charts visualize the volume of queries vs. deviations per subject to detect site-wide behavioral patterns.

Impact & Scalability

Operational Acceleration

Automated the harmonization of 23 fragmented studies and ~58k rows of data, reducing data preparation and cross-referencing time from weeks to minutes.

Intelligent Risk Mitigation

The Agentic AI layer proactively detects safety escalations and quality bottlenecks, allowing Clinical Trial Teams to intervene weeks before a scheduled database lock.

Global Scalability

Built on a modular "Analytical Spine," the system can scale to hundreds of global trials by simply ingesting new study directories into the standardized pipeline.

Conclusion & Next Steps

1

The Winning Formula

TrialGuardianAI proves that combining Deterministic Logic (DQI/Readiness) with Agentic and Generative AI provides a complete solution for both data accuracy and human-led decision-making.

2

Predictive Forecasting

Future iterations will move from "detection" to "prediction," using historical query aging patterns to forecast which sites will likely fail readiness checks in the next 30 days.

3

Real-Time Integration

Next steps include transitioning from batch processing to direct API hooks with Live EDC systems (e.g., Rave/Veeva) for instantaneous global trial oversight.

Call-to-Action

Explore Today

1. Live Dashboard: https://trial-oversight-ai.streamlit.app/
    - Click through all 4 tabs: Subject → Site → Country → Region
    - See real data from 23 studies, 57,974 records, 15 sites
    - Test filters, drill-down, export functions
2. Full Documentation: https://github.com/YASHDAHIYA449/TrialOversightAI
    - Code (Python notebooks + Streamlit app)
    - Architecture diagrams
    - DQI formula & validation
    - Deployment instructions
    - API reference