Paper ML11

Reimagining CSR: From TFLs to Dynamic Text with Generative AI

Kala Shivali, Sycamore Informatics, North Carolina, United States
Sandeep Sudhakar, Sycamore Informatics, Bangalore, India
Madan Kumar Pernati, Sycamore Informatics, Hyderabad, India
Sai Sharat Paila, Sycamore Informatics, Boston, United States

Abstract

CSRs are one of the most critical—and labor-intensive—deliverables in the regulatory submission lifecycle. Traditionally developed through manual synthesis of protocols, outputs, and medical narratives, CSR generation is ripe for disruption. This paper introduces a breakthrough framework that leverages generative AI and NLP to automate the creation of high-quality, submission-ready CSRs. By combining structured study metadata, statistical outputs, and regulatory templates, the solution dynamically generates narrative content aligned to ICH E3 guidance. It ensures consistency, traceability, and quality—while dramatically reducing time and manual effort.

The paper will demonstrate how AI models can ingest and interpret data tables, annotate key insights, and generate drafts for medical writer review, all within a GxP environment. You will gain a clear understanding of how AI augments—not replaces—the medical writing process, enabling teams to shift focus from repetitive tasks to strategic insight and quality assurance. As regulators move toward digital-first submissions, this approach positions organizations to deliver faster CSRs.

Keywords: Generative AI, Clinical Study Reports, CSR Automation, Natural Language Processing, ICH E3 Compliance, Regulatory Writing, Large Language Models, Quality Control, Human-in-the-Loop, Digital-First Submissions

1. Introduction

1.1 Role of CSRs in Regulatory Submissions

The Clinical Study Report is the authoritative account of a clinical investigation, providing integrated efficacy, safety, and methodological detail for regulatory review. CSRs are central to:

- **Regulatory Review:** The CSR undergoes thorough regulatory evaluation
- **Vital Decisions:** Risk-benefit evaluation for pivotal trials
- **Labeling Decisions:** Drug labeling and approval determinations

Yet the end-to-end process—curating analysis results, assembling narratives, enforcing style and consistency, and managing review cycles—is time-consuming and error-prone. The traditional process involves analysis, narrative development, and review management, often leading to late-stage findings, re-work, and delay.

1.2 Challenges in Traditional CSR Life Cycle

Current workflows require manual extraction and interpretation of tables, figures, and listings (TLFs), alignment with protocol/SAP, and harmonization of cross-references across sections.

Typical Pain Points:

Pain Point	Description
Duplicative Effort	Redundant manual work across multiple document sections
Inconsistencies	TLF data misaligned with narrative text
High Review Burden	Extensive manual QC and reconciliation cycles
Re-work & Delay	Late-stage findings cause timeline slippage

1.3 Why Now? Drivers for Change

Three trends make CSR automation timely—together, these enable a shift from **manual drafting** to **augmented authoring**:

- **LLM Maturity:** Domain-tuned NLP for technical writing
- **Data Standards:** ADaM/Define-XML adoption
- **Digital-First:** Agency momentum for structured content

This represents a convergence of technology, standards, and regulatory direction.

1.4 Scope and Contributions

Proposed AI Framework:

- **AI-assisted CSR authoring framework** — Leveraging LLMs for structured regulatory document generation
- **Grounded in compliance, quality, and human oversight** — Regulatory-ready with built-in validation controls
- **Technical architecture and workflows detailed** — End-to-end process from data to publication
- **Includes validation controls and implementation playbook** — Practical guidance for organizational adoption

2. Technical Foundations

2.1 Inputs to Automation

Structured Data and Metadata	Study Documentation	Templates and Guidance
ADaM Datasets, TLFs	*Protocol / Amendments*	*ICH E3 Structure*
Standardized analysis datasets and tables, figures, listings that form the quantitative foundation	Study design, objectives, endpoints, and statistical analysis plan (SAP) specifications	Section outlines aligned to international regulatory guideline structure

CSR automation requires three input streams. Structured data and metadata—ADaM datasets, TLFs, Define-XML, and controlled terminology—form the quantitative foundation. CDISC standardization enables machine-readable provenance, allowing every numerical claim to trace back to its source.

Study documentation provides context: protocols establish objectives and endpoints, the SAP governs result interpretation, and medical narratives supply clinical detail that transforms numbers into meaningful statements.

Templates and guidance documents enforce consistency. Style guides standardize terminology, ICH E3-aligned outlines ensure regulatory compliance, and publisher requirements dictate technical specifications. Together, these streams create a closed system where every generated statement can be validated.

2.2 Model Capabilities and Guardrails

LLM/NLP for CSRs:

CAPABILITIES	GUARDRAILS	POLICY CONSTRAINTS
Table Understanding Structured-to-Text conversion **Consistency Checks** Templated Drafting **Style Enforcement** Regulatory tone compliance	**Strict Source-Linking** Every claim traced to data **Deterministic Prompting** Reproducible outputs	**Block Unsupported Claims** No data = no statement **Block Hallucinations** Prevent fabricated data

LLMs bring multiple relevant capabilities: table understanding (parsing TLF structures and extracting quantitative relationships), structured-to-text transformation, consistency checks (rule-based and learned), templated drafting with contextual retrieval, and style enforcement for regulatory tone.

Guardrails constrain these capabilities. Strict source-linking requires every claim to reference its originating TLF—no assertion without provenance. Deterministic prompting ensures reproducible outputs. Policy constraints block unsupported claims categorically: no data means no statement. Hallucination triggers immediate rejection.

2.3 Mapping to ICH E3 Guidelines

Automation yields the greatest benefit where language is formulaic yet data-rich:

Section 11: Efficacy Results	Section 12: Safety Results	Sections 9-10: Methods
• Disposition summaries • Protocol deviations • Endpoint analysis summaries	• TEAEs and serious AEs • Deaths and discontinuations • Lab/vital sign shifts	• Study population • Analysis sets definitions • Statistical methods

Automation yields maximum return where language is formulaic yet data-rich. Section 11 (Efficacy) covers disposition summaries, protocol deviations, and endpoint analyses—all following predictable structures requiring precise TLF alignment. Section 12 (Safety) presents TEAE overviews, serious AEs, deaths, discontinuations, and lab shifts in standardized formats tied to MedDRA hierarchy.

Sections 9-10 (Methods) contain boilerplate varying only by study-specific parameters: population definitions, analysis sets, and statistical methodology.

The boundary matters: benefit-risk discussions requiring interpretive judgment remain human-authored. AI assists with phrasing and cross-reference verification while humans retain accountability.

Human-Authored Sections: Free-text sections for clinical relevance remain human-authored

AI Assistance Available: Phrasing options and cross-reference verification

3. Framework Architecture

3.1 Components

The framework consists of seven integrated layers:

- **Data Layer (D):** Ingests ADaM, TLFs, and metadata; validates with schema checks; aligns controlled terminology
- **Retrieval Layer (R):** Indexes study documents and metadata; provides section-specific context to the model (RAG)
- **Orchestration and Templates (O):** Encodes ICH E3-aligned section skeletons; parameterizes boilerplate; orchestrates drafting steps
- **Generation Layer (G):** Domain-tuned LLMs produce drafts with inline citations to sources (e.g., TLF IDs, Define-XML value-level metadata)
- **Quality and Compliance (QC):** Auto-checks for numeric concordance, terminology usage, and cross-section consistency; logs full lineage
- **Human-in-the-Loop (H):** Medical writer and statistician review, redline, and sign-off
- **Security and Audit (S):** Role-based access, immutable audit trails, electronic signatures, and release workflows

3.2 End-to-End Workflow

The workflow progresses through six stages:

01	02	03	04	05	06
Data Readiness	**Automated QC**	**Context Assembly**	**HITL Review**	**Draft Generation**	**Finalization**

1. **Data readiness:** Load final TLFs, ADaM, and Define-XML. Run integrity checks.
2. **Context assembly:** Select CSR sections to draft; retrieve relevant protocol/SAP excerpts and TLF IDs.
3. **Draft generation:** Produce structured narratives per section, including parameterized statements and data-linked assertions.

4. **Automated QC:** Validate numbers (e.g., percentages rounding), cross-check totals against TLFs, flag unreferenced claims.
5. **HITL review:** Writers edit for scientific nuance; statisticians verify alignment to SAP and derivations.
6. **Finalization & publishing:** Apply style rules; freeze with eSignatures; package for submission.

3.3 Template-Driven Drafting

Templates enforce structure without constraining scientific content:

TEMPLATE PATTERN:

"A total of {N_treated} participants received {drug} ({dose_regimen}). The overall incidence of TEAEs was {teae_pct}%."

EXAMPLE OUTPUT:

"A total of **120** participants received **Drug A (100 mg once daily)**. The overall incidence of TEAEs was **23.5%**."

TEMPLATE PATTERN:

"The most frequently reported TEAEs ({threshold}%) were {pt_list}. Serious AEs occurred in {sae_pct}% of participants."

EXAMPLE OUTPUT:

"The most frequently reported TEAEs (**5%**) were **headache (8.0%), nausea (6.7%), and fatigue (5.3%)**."

Key Benefits: Consistent structure • Dynamic data population • Traceable to source TLFs

Parameter values are sourced directly from TLFs; footnotes track dataset/table versions. Writers can raise the reporting threshold or re-order lists without re-running analyses.

4. Quality, Validation, and Compliance

4.1 Source–Narrative Traceability & Automated Checks

Source–Narrative Linking: Each generated sentence retains pointers to originating sources (TLF ID, ADaM variable, Define-XML VLM record). Lineage metadata is preserved through edits, enabling auditors to reconstruct evidence chains.

Automated Consistency Checks:

NUMERIC CONCORDANCE	TERMINOLOGY ALIGNMENT	CROSS-SECTION COHERENCE
Percentages recomputed from counts; discrepancies flagged beyond tolerance	Preferred terms and MedDRA levels checked; spellings/style harmonized	Population counts and denominators reconciled across Efficacy and Safety

Data Flow: TLF Source → Generated Text → QC Validation → Audit Trail

4.2 GxP Controls & Human Oversight

GxP Controls

- **Change control & audit trail:** Every draft, prompt, and model output is versioned and time-stamped.
- **Access control:** Role-based permissions separate authorship, QC, and approval.
- **Electronic signatures:** Formal sign-off gates for major milestones.

Human Oversight and Model Governance

- **Scope of use:** Clearly defined sections approved for AI drafting vs. human-only
- sections.
- **Model risk management:** Dataset shift monitoring; periodic re-validation; bias and performance testing on representative studies.
- **Operational safeguards:** "No-extrapolation" policy; blocking of speculative language; mandatory citations for quantitative claims.

GxP Controls	Human Oversight & Model Governance
CHANGE CONTROL: Audit trail	**SCOPE OF USE:** Defined boundaries
ACCESS CONTROL: Role-based	**MODEL RISK:** Assessment
E-SIGNATURES: 21 CFR Part 11	**OPERATIONAL:** Safeguards
VALIDATION: IQ/OQ/PQ	**MONITORING:** Continuous

5. Discussion and Future Directions

AI-assisted CSR authoring is a practical near-term application of LLMs in regulatory writing. As sponsors modernize their submission pipelines (e.g., eCTD v4.0; structured authoring), deeper integration becomes possible: linking Define-XML value-level metadata to narrative statements; auto-population of clinical overviews; and bidirectional traceability across ISS/ISE. Over time, reusable content libraries and study-agnostic boilerplate can compress timelines while improving consistency. The North Star remains unchanged: preserve scientific integrity and regulatory trust through transparent, auditable, human-supervised systems.

The Future of CSR

- **AI-assisted CSR authoring** uses LLMs for regulatory writing, enabling faster, more consistent document generation.

- **Modernized pipelines** allow deeper integration of metadata and narratives through standardized data formats.
- **Reusable content libraries** improve consistency and reduce timelines across multiple studies.
- The goal is to **maintain scientific integrity and regulatory trust** while improving efficiency.
- This is achieved through **transparent, auditable, human-supervised systems** with rigorous controls.

KEY INSIGHT: The future of CSR development combines AI efficiency with human expertise for regulatory excellence.

6. Conclusion

AI does not replace medical writers; it amplifies their impact. With rigorous controls, source-linked drafting, and human oversight, organizations can reduce manual burden, shorten cycle time, and improve narrative consistency—without compromising compliance. Early pilots that focus on well-structured sections lay the groundwork for portfolio-level adoption and readiness for digital-first submissions.

- **AI amplifies medical writers' impact** — not replaces them. Human expertise remains central.
- **Rigorous controls, source-linked drafting** and human oversight reduce manual burden.
- **Improves narrative consistency** without compromising compliance through automated checks.
- **Early pilots on well-structured sections** prepare for wider adoption across portfolio.

KEY INSIGHT: From Tables to Text: Enabling faster, more consistent CSR development through human-AI collaboration

Appendices

A. Implementation Playbook

A.1 Getting Started

01 ASSESS	02 PILOT	03 VALIDATE	04 SCALE
Evaluate current CSR processes, identify pain points, and define success metrics	Select low-risk CSR sections for initial AI-assisted drafting trials	Implement QC processes, measure accuracy, gather medical writer feedback	Expand to additional sections and studies, integrate into standard workflows

Key Implementation Activities:

- Define governance structure and roles
- Establish data quality requirements
- Develop template libraries
- Train medical writing teams
- Configure validation checkpoints
- Integrate with existing systems
- Establish monitoring metrics
- Create feedback loops
- Document lessons learned

A.2 Overview with Milestones

ASSESS MILESTONES: Process assessment complete. Pain points documented. Success metrics defined. Stakeholder buy-in secured.

PILOT MILESTONES: Pilot sections selected. Templates configured. First AI-assisted drafts generated. Writer feedback collected.

VALIDATE MILESTONES: QC processes validated. Accuracy metrics met. SOPs updated. Training completed.

SCALE MILESTONES: Full section coverage. Multi-study deployment. Continuous improvement cycle. ROI demonstrated.

A.3 90-Day (13 Weeks) Pilot

MOBILIZE	CONFIGURE	EXECUTE	EVALUATE
Weeks 0-2	*Weeks 3-6*	*Weeks 7-10*	*Weeks 11-13*
Select 1-2 completed studies with stable TLFs; define success metrics (e.g., time-to-first-draft).	Import templates, load data, and enable retrieval from protocol/SAP; tune prompts for 2-3 target sections.	Generate drafts; run automated QC; conduct writer/statistician reviews.	Compare to historical baselines; finalize SOP addenda; plan scale-out.

A.4 Roles and RACI

Activity	Med Writing	Biostats	Data Mgmt	QA/CSV	Reg Ops
Section selection and scoping	R	C	C	C	I
Template configuration	R	C	C	C	I
Draft generation	R	C	I	I	I
QC and reconciliation	R	R	C	C	I
Final approval and publishing	A	C	I	A	A

R Responsible A Accountable C Consulted I Informed

A.5 Metrics to Track

TIMELINE EFFICIENCY: Time-to-first-draft and time-to-approval by section.
DATA ACCURACY: Concordance rate to TLFs; number of escalations from QC.
QUALITY IMPACT: Defect density (pre- and post-AI) and top error classes.
CONTENT REUSABILITY: Reuse rate of boilerplate across studies/indications.

B. Risk Register

B.1 Summary

RISK CATEGORY	DESCRIPTION	LIKELIHOOD	IMPACT	RISK LEVEL
Data Quality	Incomplete or inconsistent TLF data leads to inaccurate narratives	Medium	High	HIGH
Hallucination	LLM generates unsupported claims or fabricated statistics	Medium	High	HIGH
Regulatory Acceptance	Uncertainty around agency acceptance of AI-assisted documents	Medium	Medium	MEDIUM
Change Management	Writer resistance or skill gaps in adopting new workflows	Medium	Medium	MEDIUM
Model Drift	LLM updates may change output quality or consistency over time	Low	Medium	LOW
IP/Confidentiality	Sensitive trial data exposure through external LLM services	Low	High	LOW

B.2 Risk Mitigations

DATA QUALITY: Implement pre-flight data validation checks. Require complete TLF packages before generation. Automated concordance verification.

HALLUCINATION: Strict source-linking requirement. Block unsupported claims. Deterministic prompting with grounded context. Human review gates.

REGULATORY ACCEPTANCE: Maintain full audit trails. Document AI-human collaboration model. Engage early with regulators. Follow FDA/EMA AI guidance.

CHANGE MANAGEMENT: Phased rollout with champions. Comprehensive training program. Emphasize AI as augmentation not replacement. Feedback loops.

MODEL DRIFT: Version-lock LLM deployments. Regression testing suite. Output quality monitoring. Controlled update cycles with validation.

IP/CONFIDENTIALITY: Private cloud or on-premise deployment. Data anonymization where possible. Enterprise agreements with data protection. No training on customer data.

KEY PRINCIPLE: Proactive risk management with layered controls ensures AI-assisted CSR development meets GxP quality and regulatory expectations

References

1. CSR structure and scope are aligned to ICH E3; sections called out in the paper map to E3's organization.
   https://database.ich.org/sites/default/files/E3_Guideline.pdf
   https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-and-content-clinical-study-reports-step-5_en.pdf
2. EMA's Reflection Paper on AI informed the human-oversight, lifecycle governance, and risk-management language.
   https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf?
3. FDA CDER's AI resources (updated Feb 2025) and the revised discussion paper supported the positioning that AI can aid regulatory decision-making when used with controls.
   https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development
4. **21 CFR Part 11** (and FDA's scope/applicability guidance) guided the audit trail, e-signatures, and access control requirements.
   https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application
5. Data/metadata traceability references to **CDISC ADaM** and **Define-XML** underpinned the source-linked narrative approach.
   https://www.cdisc.org/standards/foundational/adam
6. "Digital-first submissions" were grounded with **eCTD v4.0** acceptance timelines (FDA) to connect structured authoring to near-term operations.

https://www.fda.gov/media/179700/download?attachment
https://globalforum.diaglobal.org/issue/may-2025/ectd-v4-0-a-look-at-2025-and-beyond/

Contact Information

Your comments and questions are valued and encouraged. Contact the authors at:

Authors: Kala Shivali, Sandeep Sudhakar, Madan Kumar Pernati, Sai Sharat Paila

Company: Sycamore Informatics

Location: Bengaluru and Hyderabad, India

Email: kala.shivali@sycamoreinformatics.com, sandeep.sudhakar@sycamoreinformatics.com, madankumar.pernati@sycamoreinformatics.com, saisharat.paila@sycamoreinformatics.com

Website: www.sycamoreinformatics.com