Paper TT14

Human in the Middle: Accelerating with AI, Anchored in Quality

Melanie Hullings, Formation Bio, New York, NY, USA
Christopher H. White, Formation Bio, New York, NY, USA

ABSTRACT

The integration of artificial intelligence (AI) in regulated environments presents unique challenges, particularly regarding data integrity, traceability, and accountability. In industries governed by strict compliance standards, such as life sciences, finance, and healthcare, leveraging AI requires careful consideration of how decisions are made, validated, and documented. This talk explores the concept of "Human in the Middle" as a practical framework to bridge the gap between AI capabilities and regulatory requirements. By maintaining human oversight in critical decision-making processes, organizations can harness the efficiency and power of AI while preserving compliance and ensuring transparency. The session will examine real-world applications of this model, highlighting how “Human in the Middle” approaches can enable trustworthy AI deployment without compromising on regulatory expectations.

Attendees will gain insights into balancing innovation with oversight, and how to design AI systems that are not only powerful, but also aligned with commitments to data integrity, traceable decision-making, and risk-based quality assurance.

INTRODUCTION

Artificial intelligence is transforming regulated industries by accelerating data processing, insight generation, and decision-making. In clinical research and pharmaceutical development, AI enables automation of data ingestion, quality control, and analysis at unprecedented scale. However, the fundamental expectations of regulators such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) remain rooted in the principles of data integrity - ALCOA+: Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, Available, and Traceable (ICH E6(R3) §5.1; FDA Data Integrity Guidance, 2018).

The rapid acceleration of AI in GxP environments introduces both opportunity and risk. Innovation without an equivalent investment in quality systems risks inefficiency, rework, or non-compliance. Quality creates the stable foundation from which innovation can scale, ensuring reliability, consistency, and trust. In this context, “Human in the Middle” (HITM) operationalizes this balance, anchoring AI-assisted innovation in principles of oversight and integrity.

CONTEXT AND DRIVERS

Regulatory Backdrop

Recent regulatory initiatives underscore the importance of human oversight in AI-enabled systems:

- FDA Computer Software Assurance (CSA) Guidance (Sept 2025) promotes risk-based validation and emphasizes human review of automated outputs.
- EMA Reflection Paper on AI (EMA/705725/2020) explicitly requires explainability, accountability, and human oversight across the medicinal product lifecycle.

Both agencies advocate for governance models where humans remain responsible for validation and decision-making even when AI assists in data generation or analysis.

Regulatory Explainability Gap

While both the FDA and EMA have issued principles for AI governance, explicit methods for ensuring explainability and traceability remain under-defined. This regulatory gap underscores the necessity of frameworks like HITM that explicitly embed explainability and accountability checkpoints across the AI lifecycle.

Industry Trends

Automation, digital health technologies (DHTs), and real-world data (RWD) integration are accelerating. However, these advances have introduced new dependencies on AI models that can be opaque or biased. The emergence of generative AI adds further complexity — while these systems can summarize, predict, and generate code or specifications, they must be subject to structured human checkpoints to ensure compliance and reliability.

Table 1: Collaborative, risk-based methods are driving innovation

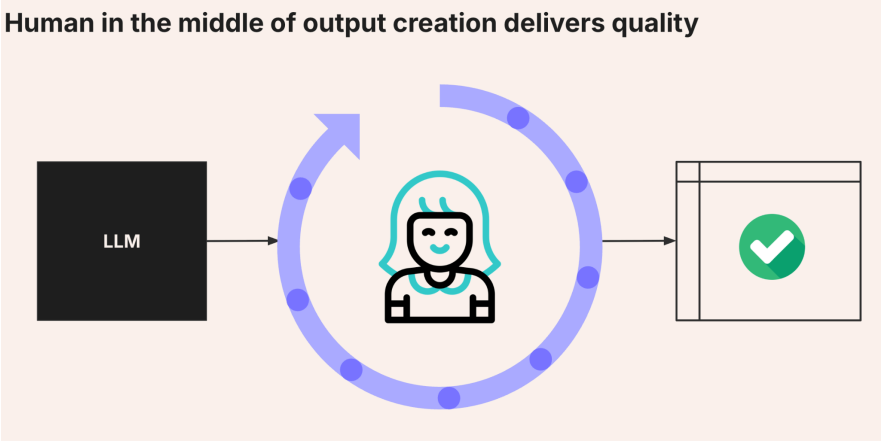
Category	Description	Key Drivers
Collaborative Ecosystem Evolution	Cross-functional alignment of data, quality, and IT teams; growing use of open-source, shared frameworks.	Open innovation, governance, standardization
Risk-Based Quality & Automation	Risk-based validation models combining automation with SME oversight; AI-assisted QC and validation	Efficiency, reproducibility, risk-based assurance
Digital Transformation Across R&D	End-to-end digitalization integrating capture, analytics, and reporting; focus on real-time, interoperable data	Automation, interoperability, cloud ecosystems
Shift Toward Trustworthy AI	Move from experimental to operational AI with focus on explainability and auditability	Ethical AI, transparency, human accountability

HUMAN IN THE MIDDLE FRAMEWORK

Definition and Principles

The Human in the Middle (HITM) framework positions humans as active overseers of AI systems at predefined control points. Unlike "Human in the Loop," which focuses on iterative model training and feedback cycles, HITM emphasizes oversight during AI deployment and operational use, especially in GxP contexts. In practice, HITM converts AI outputs into modular components that humans can review efficiently. Outputs are sequenced such that high-risk or high-impact elements are reviewed first (e.g., endpoint definitions before derived variables). At each step, subject matter experts assess both the output and the AI's rationale, providing iterative feedback that the model incorporates in subsequent steps. This method ensures that each AI-assisted deliverable is not only accurate but explainable and auditable, producing a trusted first draft rather than an unchecked result.

Figure 1. Embedding Human in the Middle Across AI Development and Use



Core elements of the HITM model include:

1. Defined Oversight Points: Structured checkpoints where human experts validate AI-generated outputs.
2. Traceable Audit Trails: Complete documentation of decisions made, data reviewed, and approvals recorded.
3. Risk-Based Oversight: Review intensity scaled to the criticality of the AI decision or output.
4. Explainability: Ensuring every AI-driven action can be interpreted by domain experts and regulators.

Human in the Middle (HITM) is essential not only during AI deployment but also throughout the development lifecycle. During development, human experts define model objectives, curate and label training data, and interpret model behavior to ensure that outputs align with clinical, ethical, and regulatory expectations. This early-stage oversight reduces risks of bias, data leakage, and misalignment between model performance metrics and real-world validity. In the use phase, HITM establishes deliberate checkpoints where subject matter experts validate intermediate and final AI outputs before they are accepted into regulated workflows. This dual-layer approach ensures continuous traceability—from data curation to decision execution—creating an end-to-end feedback loop that preserves explainability, accountability, and compliance. Rather than viewing AI as an autonomous decision-maker, HITM repositions humans as active co-pilots who enable both speed and trustworthiness in innovation.

Table 2: Humans need to be in the middle of both AI development and use

Aspect	AI Model + Tool Development	AI Model + Tool Use
When	Development phase	Deployment phase
Process	Agile, iterative	Waterfall checkpoints
Human's Role	Trainer/teacher	SME gatekeeper/inspector
Goal	Improve model	Assure quality & compliance
Regulatory Lens	Design controls	GxP oversight, ALCOA+

APPLICATIONS AND CASE EXAMPLES

Clinical Data Management & Analytics

AI systems like Apollo AI Spec Builder at Formation Bio have demonstrated that when human experts are embedded as quality checkpoints, AI can safely accelerate specification generation for electronic data capture (EDC) systems. In this workflow:

1. LLMs generate draft fields and form structures.
2. Subject matter experts (SMEs) review key outputs such as Schedule of Assessments alignment, data type consistency, and coding accuracy.
3. Human feedback is incorporated iteratively, resulting in validated, audit-ready EDC specifications.

The outcome: AI acts as a trusted co-pilot, reducing manual programming effort while maintaining traceability and GxP compliance.

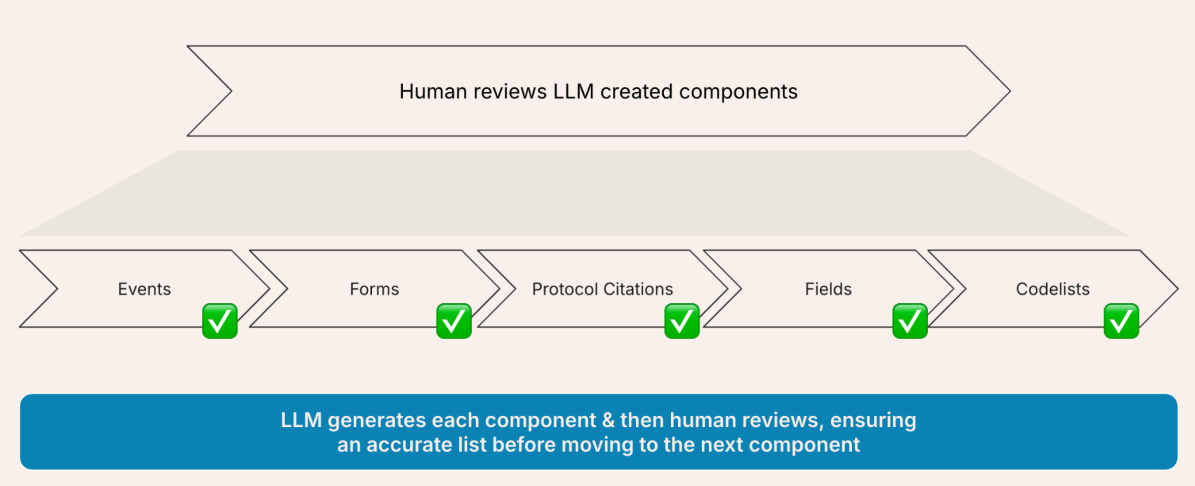
Data Models and Pipelines

HITM principles are equally valuable in extract-transform-load (ETL) pipelines for clinical and real-world data. When AI suggests mappings between datasets (e.g., trial data to EHR-derived data), SMEs review intermediate mappings rather than only the final datasets. LLM mapping suggestions are validated at the concept and feature levels before integration into data models or analysis.

This incremental oversight ensures that:

- Mapping rules remain scientifically and clinically interpretable.
- Downstream analyses are based on validated transformations.
- Data lineage is preserved, supporting reproducibility and auditability.

Figure 2: Data capture form design from protocol



Quality Assurance and Validation

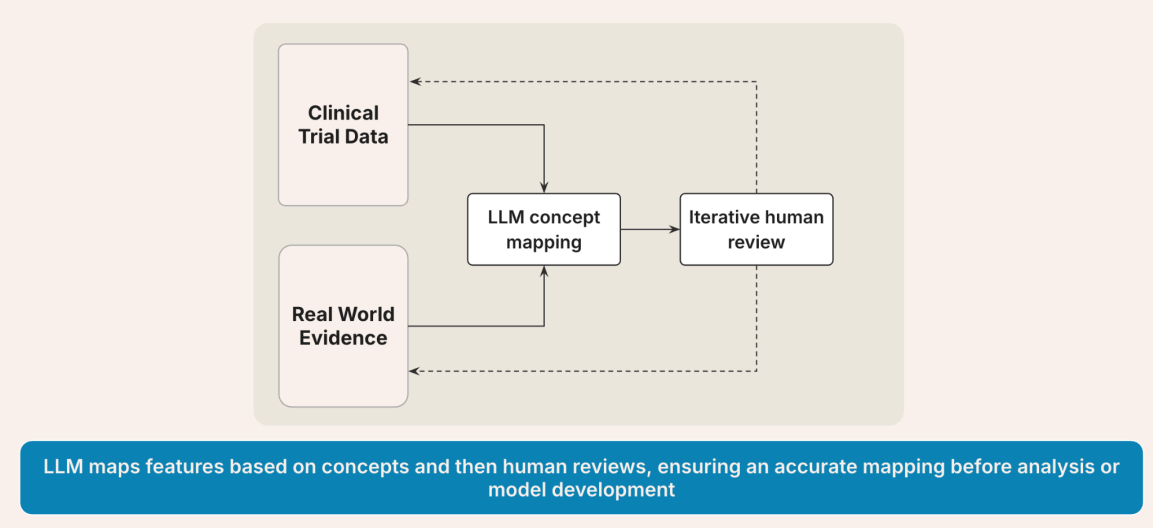
AI can automate parts of validation documentation, such as system capability assessments, data migration plans, and risk assessments, when guided by regulatory frameworks. LLMs trained on regulatory and internal validation templates guide reviewers step-by-step, ensuring completeness and traceability across system capability, validation plan, and migration documentation. Using a HITM approach:

- AI drafts content based on structured templates and prior validation records.
- QA reviewers evaluate, correct, and approve each section, creating a traceable record of human confirmation.

- AI integrates reviewer feedback across downstream documents, reducing redundancy while retaining control.

This not only enhances efficiency but aligns with FDA's CSA principles of critical thinking and proportional validation.

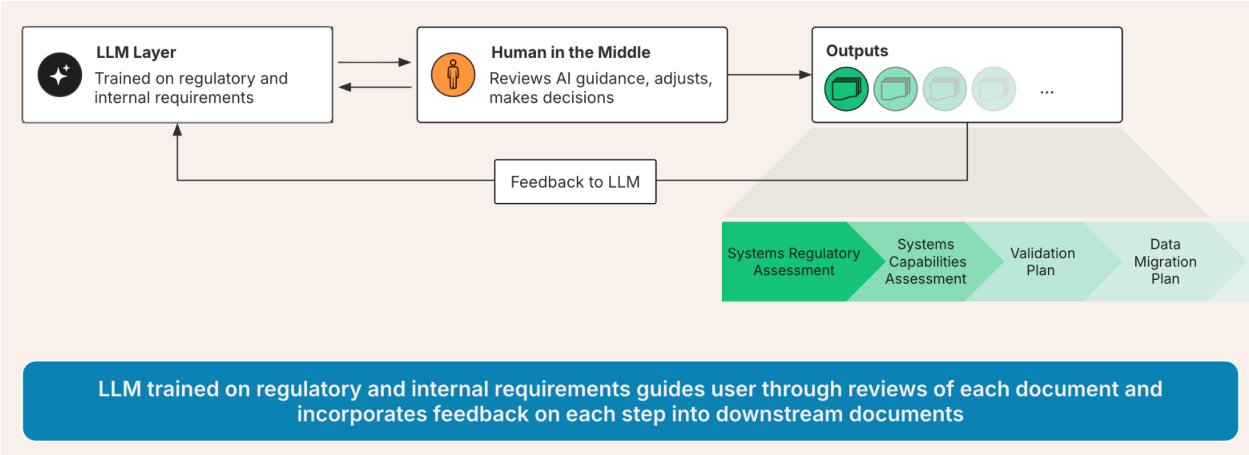
Figure 3: Data models and pipelines



Built-in Quality and Compliance

HITM embeds oversight and accountability directly into AI workflows. By breaking outputs into manageable components, sequencing reviews from high to low risk, and documenting rationale at every step, HITM transforms AI from a "black box" into a transparent co-pilot. This design enables continuous explainability, reinforces trust, and operationalizes the FDA's CSA and EMA's AI reflection principles.

Figure 4: GxP Systems Lifecycle Documentation



RISK, QUALITY, AND TRUST

Risk-Based Oversight

HITM operationalizes risk-based review so human effort can be focused where AI uncertainty or potential impact is highest.

- High-risk areas (e.g., safety data, efficacy endpoints) demand full SME validation.
- Low-risk areas (e.g., performance dashboards, internal KPIs) can rely more heavily on automated review.

Interpretability and Trust

HITM ensures that all AI actions can be explained in a manner that regulators, auditors, and scientists can understand. Interpretability is essential for maintaining public and institutional trust in AI-driven processes.

Failures in explainability have historically led to major data integrity breaches across regulated sectors (e.g., the Ranbaxy case, where lack of traceability undermined confidence in data). HITM provides the structural safeguards to prevent such failures in the AI era.

FUTURE OUTLOOK

HITM represents more than a compliance framework; it is an operating model for trustworthy AI in life sciences. Looking ahead:

- Regulators are likely to issue explicit guidance requiring human oversight in AI-assisted decisions.
- Industry collaborations (PhUSE, CDISC, ISPE) are positioned to harmonize standards for AI explainability and auditability.
- Embedding HITM directly into AI architectures will enable automation that remains transparent, validated, and compliant through "governance by design."

In summary, AI can accelerate drug development only when anchored in quality. HITM balances innovation with compliance, ensuring that humans remain accountable for the integrity of AI-driven outcomes. As AI systems evolve toward multi-agent architectures, integrating HITM directly into orchestration layers will become essential. Future regulatory frameworks may formalize human-in-the-middle checkpoints as part of AI system design controls. Industry collaborations, such as those led by PhUSE, CDISC, and ISPE, will play a critical role in harmonizing methodologies for auditability, explainability, and cross-system governance, enabling "governance by design" in next-generation AI systems.

CONCLUSION

The Human in the Middle framework provides a scalable, auditable, and regulatorily aligned model for responsible AI adoption. By embedding human oversight at critical junctures, organizations can achieve the efficiency of automation without compromising on the principles that safeguard patient safety and data quality. The future of clinical data science will depend not only on how fast AI evolves but on how thoughtfully humans remain in control of it.

REFERENCES

1. U.S. Food and Drug Administration (FDA). Data Integrity and Compliance With Drug CGMP: Questions and Answers Guidance for Industry. April 2016. Available at: https://www.fda.gov/media/119267/download
2. U.S. Food and Drug Administration (FDA). Computer Software Assurance for Production and Quality System Software: Guidance for Industry. September 2025. Available at: https://www.fda.gov/media/188844/download
3. European Medicines Agency (EMA). Reflection Paper on the Use of Artificial Intelligence in the Medicinal Product Lifecycle (EMA/705725/2020). July 2023. Available at: https://www.ema.europa.eu/en/documents/reflection-paper/reflection-paper-use-artificial-intelligence-medicinal-product-lifecycle_en.pdf
4. International Council for Harmonisation (ICH). E6(R3) Good Clinical Practice Draft Guideline. 2023. Section 5.1 – Quality Management. Available at: https://www.ich.org/page/efficacy-guidelines
5. U.S. Food and Drug Administration (FDA). Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device (SaMD) Action Plan. January 2021. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
6. Good Automated Manufacturing Practice (GAMP) 5, Second Edition. A Risk-Based Approach to Compliant GxP Computerized Systems. International Society for Pharmaceutical Engineering (ISPE), 2022.
7. PhUSE AI/ML Working Group. White Paper: Responsible AI in Clinical Research – Governance, Transparency, and Risk-Based Validation. PhUSE, 2024.
8. European Commission. Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). 2021/0106 (COD). Available at: https://artificialintelligenceact.eu
9. Murdoch, T. B., & Detsky, A. S. The Inevitable Application of Big Data to Health Care. JAMA, 2013; 309(13): 1351–1352. DOI: 10.1001/jama.2013.393
10. Zhang, Y., et al. Explainable Artificial Intelligence in Regulatory Science: A Review. Frontiers in Artificial Intelligence, 2024; 7: 145872. DOI: 10.3389/frai.2024.145872

ACKNOWLEDGMENTS

Thank you to Laura Ng for her expert slide and figure designs.

RECOMMENDED READING

1. **FDA.** *Framework for Regulatory Use of Real-World Evidence (RWE).* December 2021.
   Outlines principles for trustworthy data sources and oversight of analytical tools, relevant to AI-enabled data integration.

2. **EMA & Heads of Medicines Agencies (HMA).** *Guideline on Computerised Systems and Electronic Data in Clinical Trials.* February 2024.
   Reinforces data integrity expectations under ALCOA+ and provides context for applying Human in the Middle to trial data capture and monitoring.
3. **CDISC.** AI and Automation in Clinical Data Standards Task Force White Paper. 2024.
   Discusses standardization of AI outputs and metadata traceability to support regulatory acceptability.
4. **PhUSE Data Transparency Working Group.** AI Governance and Auditability in Clinical Development. 2023.
   Offers case studies of audit-ready AI workflows and human oversight checkpoints.
5. **Good Machine Learning Practice (GMLP) for Medical Device Development: Guiding Principles.** FDA, Health Canada, and MHRA Joint Publication, October 2021.
   Provides global principles for ensuring reliability, transparency, and human accountability in AI-assisted decision-making.
6. **ISPE.** AI Maturity Model for GxP Environments. Draft Guidance, 2024.
   Introduces a structured approach for scaling AI capability while embedding HITM principles across data, process, and validation layers.

CONTACT INFORMATION

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

Author Name: Melanie Hullings
Company: Formation Bio
Address: 16 E. 34th St. Floor 10 New York, NY 10065
Email: melanie@formation.bio
Website: https://www.formation.bio/

Author Name: Christopher H. White
Company: Formation Bio
Address: 16 E. 34th St. Floor 10 New York, NY 10065
Email: christopher@formation.bio
Website: https://www.formation.bio/