Transforming regulatory intelligence with AI driven methodologies in HEOR

Devyani Biswal[1], Yuntzu Yin[1], Purushartha Singh[1], Luk Arbuckle[1]

1. IQVIA Applied AI Science

Disclaimer: At the time of work, DB,YY, PS, and LA were employed at IQVIA

MSR7

Background

Health economics and outcomes research (HEOR) depends on consistent, transparent interpretation of evolving regulatory and methodological standards. Fragmentation across frameworks like the FDA's Quality System Regulation (QSR), the European Medicines Agency (EMA) GxP, and International Council for Harmonisation (ICH) guidelines creates barriers to alignment in clinical trial design, risk management, and reimbursement strategy. Manual analysis is time-consuming, error-prone, and limits scalability as new frameworks emerge.

Objectives

Regulatory intelligence involves the collection, analysis, and application of regulatory information to support decision-making and ensure adherence to evolving standards. As health regulations evolve, and privacy, data protection, and AI regulations grow in complexity, managing these actionable insights requires efficient and scalable solutions. This study introduces a methodology leveraging large language models (LLMs) and AI Agents to automate the mapping of diverse regulatory and operational standards to structured frameworks, from quality assurance and patient safety to privacy and AI risk management. The objective is to streamline controls mapping and requirements management while maintaining accuracy, traceability, and scalability.

This conceptual framework also lays the groundwork for structuring regulatory content in a way that is adaptable to domains such as HEOR. By abstracting guidance into functional categories and control matrices, the approach supports future alignment with evaluation criteria used in clinical development, market access, and post-market surveillance.

Methods

Our methodology uses a pipeline that ingests global regulatory documents and applies large language models (LLMs) to extract, structure, and align key requirements into a unified control matrix. Overlap and score matrices are then generated to identify convergence across frameworks, filtered by relevance and compliance risk.

The pipeline consists of five components:

1. ***Collection & Aggregation*** of source documents (regulations, standards, guidance, and precedent).
2. ***Prompting & Mapping*** using LLMs to identify requirements, followed by expert human-in-the-loop validation.
3. ***Overlap Matrix Construction*** from a control matrix to identify where regulations align or diverge.
4. ***Score Matrix Filtering*** using domain-specific thresholds for risk, relevance, and applicability.
5. ***Strategy Development*** for aligning findings with HEOR use cases (e.g., trial design optimization, GCP adherence, or CHEERS-based reporting).

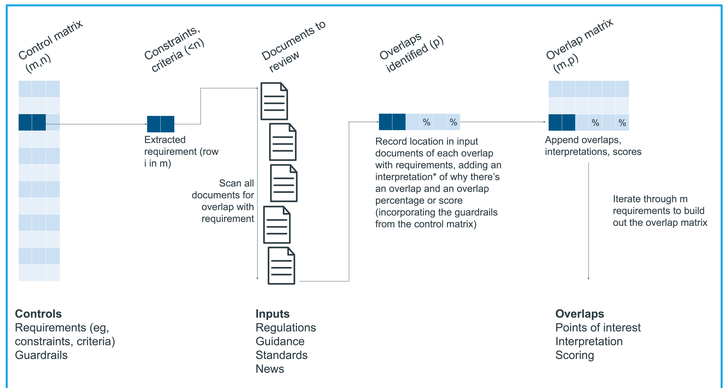

FIGURE: Building the overlap matrix (starting from a control matrix)

The diagram illustrates how the overlap matrix is constructed to be scored with domain-specific thresholds and expert-in-the-loop adjustments.

Results

To demonstrate feasibility, the proposed AI pipeline was validated using complex regulatory frameworks from NIST[1] and ISO[2], using regional privacy laws, regulatory guidance documents, and functional requirements. These use cases confirmed the framework's ability to:

› Extract and structure diverse regulatory actions into interpretable matrices.

› Identify overlaps and discrepancies across frameworks through transparent scoring.

› Support expert-driven validation through explainability and traceable mappings.

The results below used the NIST Privacy Framework to illustrate the method's capacity to scale across domains, suggesting strong potential for adaptation to HEOR-relevant frameworks such as ICH E6, CHEERS, and the European Network for Health Technology Assessment (EUnetHTA) .

Switching from category-level to subcategory-level alignment notably increased precision across all domains processing)

Using initialization and structured glossaries helped refine outputs and boost scores in harder-to-capture domains.

Processing by individual chapters, rather than full documents, led to the highest overall performance (Recall = 0.95). processing)

Recall improved from 0.15 (GPT-4-32k) to 0.95 (GPT-4o with subcategory granularity and chapter-level processing)

Recall							
	Model	GPT4-32k	GPT-4o	GPT-4o	GPT-4o	GPT-4o	GPT-4o
	Granularity	Category	Category	Subcategory	Category Subcategory	Category Subcategory	Category Subcategory
	Initialization	Yes	Yes	Yes	Yes	Yes	Yes
	Glossary	No	No	No	Yes	Yes	Yes
	Document Name	Yes	Yes	Yes	Yes	Yes	Yes
	Strategy	Full Document	Full Document	Full Document	Full Document	Full Document	Indv. Chapters
Identity-P	ID.IM-P	0.00	0.38	0.87	0.81	0.88	0.94
	ID.BE-P	0.50	0.50	0.87	0.50	1.00	1.00
	ID.RA-P	0.10	0.30	0.58	0.90	1.00	1.00
	ID.DE-P	0.00	0.50	0.33	0.50	0.50	1.00
Govern-P	GV.PO-P	0.00	0.22	0.63	1.00	1.00	1.00
	GV.RM-P	0.00	1.00	1.00	1.00	1.00	1.00
	GV.AT-P	0.33	0.67	0.67	0.67	0.67	1.00
	GV.MT-P	0.13	0.13	0.67	0.56	0.63	0.94
Control-P	CT.PO-P	0.10	0.67	0.50	0.81	0.95	1.00
	CT.DM-P	0.20	0.53	1.00	0.87	1.00	1.00
	CT.DP-P	0.08	0.17	1.00	0.75	0.83	0.92
Communication-P	CM.PO-P	0.17	0.67	0.80	0.94	0.94	0.94
	CM.AW-P	0.20	0.60	0.44	0.67	0.87	1.00
Protect-P	PR.PO-P	0.20	0.20	0.60	0.60	0.80	1.00
	PR.AC-P	0.19	0.31	0.31	0.38	0.38	0.69
	PR.DS-P	0.31	0.23	0.46	0.39	0.46	1.00
	PR.PT-P	0.00	0.00	0.33	0.67	0.67	0.67
Overall		**0.15**	**0.42**	**0.65**	**0.71**	**0.80**	**0.95**

Conclusion

These use cases confirmed the framework's ability to:

› This conceptual framework transforms regulatory intelligence into a scalable system for interpreting global standards using LLMs and human validation.

› Initial validation with NIST and ISO frameworks shows strong feasibility, with clear pathways for extension to HEOR domains such as health technology assessment, clinical trial design, and market access.

› By enabling structured, explainable mappings, the approach supports strategic alignment across regulatory, quality, and evaluation functions.

References

1. Susnjak T, Hwang P, Reyes NH, Barczak AL, McIntosh TR, Ranathunga S. Automating research synthesis with domain-specific large language model fine-tuning. arXiv preprint arXiv:2404.08680. 2024 Apr 8.
2. Tian K, Mitchell E, Zhou A, Sharma A, Rafailov R, Yao H, Finn C, Manning CD. Just ask for calibration: Strategies for eliciting calibrated confidence scores from language models fine-tuned with human feedback. arXiv preprint arXiv:2305.14975. 2023 May 24.
3. Xiong M, Hu Z, Lu X, Li Y, Fu J, He J, Hooi B. Can llms express their uncertainty? an empirical evaluation of confidence elicitation in llms. arXiv preprint arXiv:2306.13063. 2023 Jun 22.

1. National Institute of Standards and Technology (NIST)
2. International Organization for Standardization (ISO)

Presented at ISPOR 2025, Montreal, Canada