

A Neuro-Symbolic AI-based Framework for Holistic SOP Consistency Management in Regulated Industries

Malte Scheuvens¹, Lisa Katzenschlager¹, Annika Loos¹ and Felix Demmelbauer-Ebner¹

¹cdot GmbH, Stubenring 24/13, 1010 Vienna, Austria

Abstract

Managing Standard Operating Procedures (SOPs) in regulated industries such as healthcare and pharmaceuticals presents a dual challenge: growing regulatory complexity driven by frameworks like GxP, NIS2, and ISO standards requires increasingly detailed documentation, while demographic change threatens the loss of tacit process knowledge that has never been formally captured. Existing Document Management Systems (DMS) manage only versions and metadata but cannot detect semantic contradictions across documents. We present an early-stage framework for holistic SOP lifecycle management integrating rule-based consistency checking, neuro-symbolic AI, large language models and emerging world models. The Knowledge Intelligence Layer addresses immediate industry pain points through hybrid contradiction detection, pairing an LLM-based semantic layer for candidate identification with a formal reasoning layer based on First-Order Logic and Answer Set Programming (ASP)-based Defeasible Deontic Logic for deterministic verification. The Knowledge Acquisition Layer extends the framework toward the capture and formalization of implicit expert knowledge through egocentric video observation and targeted clarification mechanisms. We demonstrate the framework through a use case with a regulated healthcare provider, where the combination of rule-based consistency checking together with LLM-based semantic analysis surfaced contradictions across interdependent SOPs that semantic analysis alone could not deterministically verify, suggesting that declarative rules combined with neural approaches provide a principled and auditable foundation for SOP management in settings where correctness is legally mandated.

Keywords

neuro-symbolic AI, SOP management, consistency checking, defeasible deontic logic, world models, compliance

1. Introduction

Standard operating procedures (SOPs) are required by law in regulated industries and are audited on a regular basis. Failure to comply with them has immediate legal and financial consequences, as illustrated by the temporary production halt at Euroapi's Brindisi site in 2024 [1]. This work was motivated by several interrelated challenges: regulatory pressure from GMP, NIS2 Art. 20, the EU AI Act, and ISO norms continuously expand the documentation burden [2]. Hidden SOP contradictions are latent audit risks, with review and harmonization consuming 20–50% of quality management capacity and violation costs exceeding proactive investment by 2.7 times [3, 4]. Furthermore, demographic change threatens irreversible loss of tacit process knowledge as approximately 7 million workers are forecast to leave the German labor market by 2035 [5].

Existing Document Management Systems such as MasterControl, Veeva, and DocuWare manage versions and metadata but cannot detect whether the content of one SOP contradicts another [6, 7, 8]. Recent AI initiatives by these vendors, such as MasterControl's SOP Analyzer and Document Summarizer [9], rely on LLM-based generation and single-document analysis without cross-document reasoning. Pure LLM approaches do not close the consistency gap: non-deterministic outputs and hallucinations are disqualifying in GxP environments where compliance decisions must be verifiable and auditable, and where a conflict proof must be reproducible and explainable rather than probabilistic.

We present an early-stage industrial framework combining rule-based consistency checking, neuro-symbolic AI, and large language models to address key stages of the SOP lifecycle. Beyond consistency validation, the framework supports guided SOP authoring, ensuring newly created documents are

RuleML+RR'26: Companion Proceedings of the 10th International Joint Conference on Rules and Reasoning, August 24–26, 2026, Vilnius, Lithuania

✉ malte.scheuvens@cdot.ai (M. Scheuvens); lisa.katzenschlager@cdot.ai (L. Katzenschlager); annika.loos@cdot.ai (A. Loos); felix.demmelbauer-ebner@cdot.ai (F. Demmelbauer-Ebner)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

consistent with the existing corpus from the point of creation. We demonstrate its applicability through a pilot with a regulated healthcare provider, where the combination of rule-based consistency checking together with LLM-based semantic analysis surfaced contradictions across interdependent SOPs that semantic analysis alone could not deterministically verify.

2. Related Work

Existing solutions for SOP management fall into two categories. Legacy platforms such as MasterControl and Veeva focus on audit-safe archival and signature workflows, while newer tools such as Scribe and Trainual prioritize content creation speed [6, 7, 10, 11]. All share the same fundamental limitation: they verify whether a document exists and is approved, not whether its content is consistent with other documents. DeepHow represents the closest prior work, segmenting video captures into steps via LLMs, but captures only the phenomenology of a process without providing cross-document validation [12].

On the reasoning side, defeasible deontic logic and ASP-based implementations provide the formal foundations for our consistency checking layer [13, 14, 15]. Nonaka and Takeuchi’s SECI model [16] frames our module design along the full knowledge lifecycle from tacit observation to explicit formalization. Recent advances in JEPA-based World Models [17, 18] and egocentric procedural AI assistants [19] motivate the Knowledge Acquisition Layer, while hybrid RAG architectures [20] underpin the *Explore* module. To our knowledge, no existing system integrates these strands into a unified framework for SOP lifecycle management in regulated industries.

3. The Rule- and Neuro-Symbolic AI-based Framework

3.1. Architecture Overview

The framework consists of five modules organized across two functional layers, all connected through a shared ontology infrastructure that encodes domain concepts, roles, process relationships, and regulatory requirements (see Figure 1).

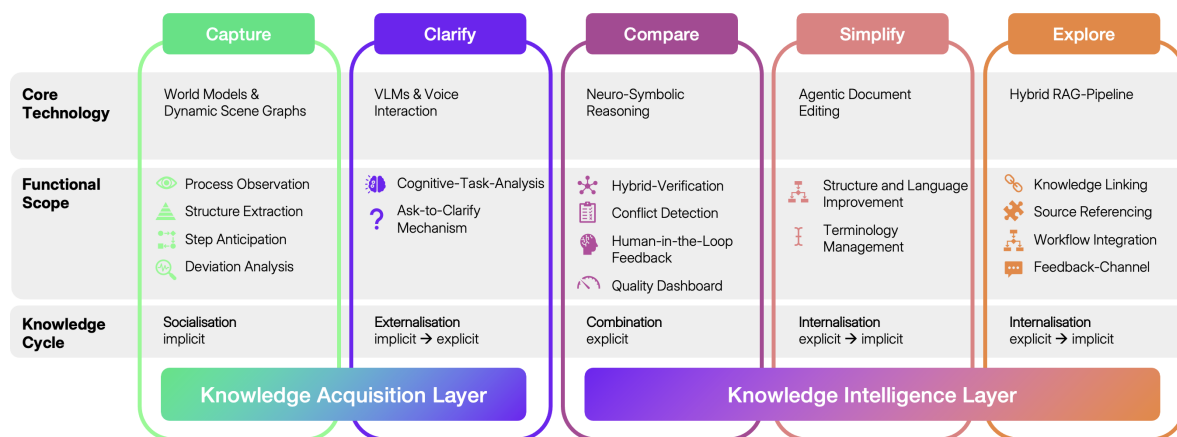


Figure 1: Overview of the SOP Management Framework and its Five Modules

3.2. Knowledge Intelligence Layer

3.2.1. Compare: Rule-based Consistency Checking

Compare identifies semantic contradictions and redundancies across the SOP corpus through a hybrid two-layer approach. A semantic LLM layer detects content overlaps and conflict candidates, including inconsistent process sequences, conflicting responsibilities, and contradictory parameter thresholds.

A formal reasoning layer then verifies conflict candidates deterministically, pairing an ASP-based Defeasible Deontic Logic solver built on clingo [13, 14, 15] with a First-Order Logic component. The DDL solver reasons non-monotonically under the Closed World Assumption (negation as failure), evaluating normative rules with exceptions and defaults, precisely the structure of regulatory documents. FOL on the other hand handles structural constraints: reasoning over the domain ontology and enforcing integrity constraints. Identified conflicts are submitted to domain experts via a Human-in-the-Loop mechanism, with each decision feeding back as a training signal for a reranker that continuously reduces false positives without suppressing true conflicts.

3.2.2. Simplify: Language & Structure Optimization

Simplify optimizes the linguistic and structural quality of SOPs and provides the foundation for agentic document editing. It draws on specialized language models for readability assessment [21, 22], enforces GxP-specific structure requirements such as scope tables, responsibility assignments, change histories, and supports domain-specific thesauri for controlled vocabulary [23]. All changes are tracked via an automatic audit trail. User queries and feedback from *Explore* feed directly into *Simplify*'s improvement suggestions, closing the loop between knowledge access and document quality.

3.2.3. Explore: Knowledge Retrieval

Explore implements a hybrid RAG pipeline for transparent, low-threshold access to process knowledge, referencing the exact source location for every answer. The retrieval strategy combines relational metadata, a vector database for semantic retrieval, and a knowledge graph encoding document dependencies and organizational context [24, 20, 25]. Frequently asked questions feed into *Simplify* as SOP simplification candidates, while answer feedback is aggregated and forwarded to document owners. An MCP interface enables integration with existing Q&A systems.

3.3. Knowledge Acquisition Layer

3.3.1. Capture: Physical Process Observation

Capture transforms egocentric video from smart glasses or helmet cameras into dynamic scene graphs using JEPA-based World Models [17, 18], where nodes represent objects, edges represent actions, and temporal dependencies encode causal relationships. Where current VLM-based approaches show limitations in real-time processing [19], JEPA-based models infer goals, anticipate action sequences, and model state transitions at 1/50th the planning duration of comparable architectures [18]. Scene graphs are validated against the shared ontology for ontological consistency, temporal ordering, and completeness. Deviations are quantified as node- and edge-level uncertainty, with high uncertainty automatically triggering the *Clarify* module. Crucially, the same normative rules used by *Compare* to check written SOPs also verify physically observed processes via the shared ontology.

3.3.2. Clarify: Implicit Knowledge Extraction

Clarify operationalizes the Ask-to-Clarify mechanism to maximize information gain from expert interaction. Drawing on Cognitive Task Analysis [26] and proactive interaction research [27, 28], it uses uncertainty signals from *Capture*, including logical contradictions, undocumented steps, high action-anticipation entropy, and missing ontology entries as learning triggers. Information-Gain optimization [28] ensures only highest-yield questions are posed, minimizing workflow interruption. Responses are classified and fed back into the scene graph and document corpus via *Compare* and *Simplify*.

4. Use Case: SOP Management at a Regulated Healthcare Provider

4.1. Context and Starting Situation

The pilot partner is a large Austrian provider of care and social services operating across multiple facilities. Its regulatory environment, covering care standards, hygiene protocols, medication administration, and data protection requirements, closely parallels pharmaceutical GxP in its demands on documentation correctness and auditability. Several hundred SOPs are maintained manually across facilities with no systematic cross-document consistency checking. For the pilot, a smaller cross section of 50 documents was analyzed. Initial assessment revealed contradictions accumulated from parallel SOP development across care areas, broken references to superseded guidelines, and inconsistent role designations across documents. Acute demographic pressure compounds the problem: severe nursing staff shortages drive long onboarding times, while almost no implicit care knowledge has been formally captured in written SOPs.

4.2. Framework Application

The Knowledge Intelligence Layer was applied to the provider's SOP corpus. In the following, we focus on the rule-based consistency checking pipeline of the *Compare* module. Its ontology was bootstrapped from domain-relevant source documents including DNQP standards [29]. The hybrid consistency checking pipeline was then run across the subset of the corpus, with the formal reasoning layer verifying conflict candidates identified by the semantic LLM layer.

A selection of applied rule categories is presented below:

- **Metadata completeness:** All required metadata fields (e.g. approved by, created on, version number) must be completed before a document is considered valid.
- **Responsibility assignment:** No two SOPs in the same domain may assign conflicting RACI (Responsible, Accountable, Consulted, Informed) responsibilities for the same action.
- **Cross-reference validity:** If a reference to a document is broken or outdated, it must be updated to a valid and current source.
- **Document structure conformance:** If a document type defines a required table of contents and section structure, all SOPs of that type must adhere to the defined structure, including required section titles and classified section contents.
- **Terminology consistency:** Terms defined in the master glossary must be used identically across all referencing documents.
- **Regulatory cascade:** If an SOP cites a regulation, all SOPs in the same process chain must be consistent with that regulation's constraints.
- **Temporal ordering:** Steps with safety-relevant preconditions must precede dependent steps in any procedure referencing both.

4.3. Results and Limitations

Across the applied rule categories, multiple conflicts and structural deficiencies were identified and subsequently validated with quality management staff via the Human-in-the-Loop interface, which presented formal proofs of contradiction alongside resolution options. Validated decisions were logged for reranker training and ontology enrichment. Initial validation demonstrates reduced manual review effort and improved document landscape transparency, providing a structured foundation for audit preparation. The pilot also establishes the domain foundation for a future *Capture* and *Clarify* deployment, where abundant undocumented tacit care knowledge represents a high-value acquisition target. Limitations are acknowledged: the pilot is in early stages and quantitative metrics including false-positive rate, precision, and review hour reduction are being collected but are not yet reportable. *Capture/Clarify* integration remain as future work and come with their own challenges, like privacy concerns, user acceptance, ergonomics, the current experimental nature of world models and the difficulties of grounding their outputs in a normative layer, among others.

5. Conclusion

We presented an early-stage industrial framework integrating rule-based reasoning, LLMs, and World Models across five modules on a shared domain ontology, covering the SOP lifecycle from physical process observation to cross-document consistency verification. The hybrid two-layer consistency checker in *Compare* addresses the fundamental limitation of pure LLM approaches: regulated environments require deterministic, formally verifiable conflict proofs, and obligations, prohibitions, and exceptions must be reasoned about explicitly rather than approximated. The pilot with a regulated healthcare provider demonstrates the applicability of the framework in an early deployment stage, validating the Knowledge Intelligence Layer and illustrating the rule-based consistency checking pipeline of the *Compare* module through a representative selection of rule categories.

Future work includes further quantitative evaluation in different enterprises and domains, rule base expansion for pharmaceuticals production, MedTech, and NIS2 compliance with full defeasibility hierarchies, and the *Capture/Compare* connection prototype. The shared ontology enables a key architectural continuity property: the same normative rules that verify written SOPs will extend to physically observed processes once *Capture* reaches prototype stage, providing a unified compliance layer across both document and physical process dimensions.

Two open challenges are worth highlighting for the RuleML+RR community. First, ASP is NP-hard in the general case; scaling defeasible consistency checking to inventories exceeding 500 SOPs requires modular decomposition strategies that preserve soundness while bounding solver complexity. Second, ontology quality is a hard dependency of the formal reasoning layer; automated bootstrapping from regulatory source documents reduces this burden but introduces its own coverage and alignment challenges.

Acknowledgments

The authors would like to thank Austria Wirtschaftsservice (aws) for its support through the aws First Incubator Program, and NVIDIA for its support through the NVIDIA Inception Program.

Declaration on Generative AI

For the writing process, the authors used DeepL for grammar and language improvements. All outputs were reviewed and revised by the authors.

References

- [1] EUROAPI, Full-year 2024 guidance suspended due to the temporary pause of API production at the Brindisi site in Italy, 2024. Accessed: Apr. 25, 2026. Available: <https://www.euroapi.com/en/full-year-2024-guidance-suspended-due-to-the-temporary-pause-of-API-production-at-the-Brindisi-site-in-Italy>.
- [2] European Union, Artikel 20, governance – NIS2-Richtlinie EU 2022/2555, 2022. Accessed: Apr. 10, 2026. Available: <https://nis2resources.eu/de/richtlinie-20222555-nis2/artikel-20/>.
- [3] D. M. Schwartz, SMEs spend seven per cent of their working time on administrative processes, KfW SME Panel 2024, KfW, 2025.
- [4] Ponemon Institute LLC, Study: Total Cost of Compliance with Data Protection Regulations, Technical Report, Globalscape, 2017. Accessed: Apr. 27, 2026. Available: <https://www.globalscape.com/resources/whitepapers/data-protection-regulations-study>.
- [5] dpa, Arbeitsmarkt verliert bis 2035 rund sieben millionen menschen, 2026. Accessed: Apr. 25, 2026. Available: <https://www.handelsblatt.com/politik/konjunktur/iab-studie-deutscher-arbeitsmarkt-verliert-bis-2035-rund-sieben-millionen-menschen/28822072.html>.

- [6] MasterControl, SOP management solutions for life science manufacturing, 2026. Accessed: Apr. 12, 2026. Available: <https://www.mastercontrol.com/quality/sop-standard-operating-procedure>.
- [7] Veeva, Veeva QualityDocs | GxP and quality content management, 2026. Accessed: Apr. 12, 2026. Available: <https://www.veeva.com/products/veeva-qualitydocs/>.
- [8] DocuWare, Qualitätsmanagement lösungen, 2026. Accessed: Apr. 12, 2026. Available: <https://start.docuware.com/de/qualitaetsmanagement-loesung>.
- [9] MasterControl, Specialized AI for life science quality manufacturing, 2026. Accessed: May, 2026. Available: <https://www.mastercontrol.com/ai-for-life-sciences-quality-manufacturing/>.
- [10] Scribe, Scribe | create step-by-step guides — fast, 2026. Accessed: Apr. 12, 2026. Available: <https://scribe.com/>.
- [11] Trainual, Manage and automate your SOPs easily with Trainual's system, 2026. Accessed: Apr. 12, 2026. Available: <https://trainual.com/solution/sops>.
- [12] DeepHow, Operational excellence through physical AI, 2026. Accessed: May 27, 2026. Available: <https://deephew.com/>.
- [13] G. Governatori, An ASP implementation of defeasible deontic logic, *KI – Künstliche Intelligenz* 38 (2024) 79–88. doi:10.1007/s13218-024-00854-9.
- [14] G. Governatori, M. Palmirani, Legal explanation in defeasible deontic logic via LegalRuleML, in: *Proceedings of the Twentieth International Conference on Artificial Intelligence and Law*, ACM, Chicago, IL, USA, 2025, pp. 500–501. doi:10.1145/3769126.3769257.
- [15] X. Fang, Z. Li, C. Chen, B. Liao, LLMASPIC: A neuro-symbolic framework for defeasible reasoning, in: I. Lynce, N. Murano, M. Vallati, S. Villata, F. Chesani, M. Milano, A. Omicini, M. Dastani (Eds.), *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2025. doi:10.3233/FAIA250981.
- [16] I. Nonaka, H. Takeuchi, F. Mader, *Die Organisation des Wissens: wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*, Campus, Frankfurt/Main, New York, N.Y., 1997.
- [17] M. Assran, et al., V-JEPA 2: Self-supervised video models enable understanding, prediction and planning, 2025. doi:10.48550/arXiv.2506.09985, arXiv:2506.09985.
- [18] L. Maes, Q. L. Lidec, D. Scieur, Y. LeCun, R. Balestrieri, LeWorldModel: Stable end-to-end joint-embedding predictive architecture from pixels, 2026. doi:10.48550/arXiv.2603.19312, arXiv:2603.19312.
- [19] J. Li, et al., Building egocentric procedural AI assistant: Methods, benchmarks, and challenges, 2026. doi:10.48550/arXiv.2511.13261, arXiv:2511.13261.
- [20] Z. Guo, L. Xia, Y. Yu, T. Ao, C. Huang, LightRAG: Simple and fast retrieval-augmented generation, 2025. doi:10.48550/arXiv.2410.05779, arXiv:2410.05779.
- [21] J. M. Imperial, et al., UniversalCEFR: Enabling open multilingual research on language proficiency assessment, 2025. doi:10.48550/arXiv.2506.01419, arXiv:2506.01419.
- [22] T. Guidroz, et al., LLM-based text simplification and its effect on user comprehension and cognitive load, 2025. doi:10.48550/arXiv.2505.01980, arXiv:2505.01980.
- [23] G. M. D. Nunzio, Terminology-augmented generation (TAG): Foundations, use cases, and evaluation paths, *Frontiers in Artificial Intelligence* 1 (2025).
- [24] H. Han, et al., RAG vs. GraphRAG: A systematic evaluation and key insights, 2025. doi:10.48550/arXiv.2502.11371, arXiv:2502.11371.
- [25] B. J. Gutiérrez, Y. Shu, W. Qi, S. Zhou, Y. Su, From RAG to memory: Non-parametric continual learning for large language models, 2025. doi:10.48550/arXiv.2502.14802, arXiv:2502.14802.
- [26] R. E. Wray, J. R. Kirk, J. E. Laird, Eliciting problem specifications for LLM-Modulo cognitive systems, *Cognitive Systems Research* 94 (2025) 101409. doi:10.1016/j.cogsys.2025.101409.
- [27] X. Lin, et al., Ask-to-clarify: Resolving instruction ambiguity through multi-turn dialogue, 2025. doi:10.48550/arXiv.2509.15061, arXiv:2509.15061.
- [28] S. Yang, et al., Plug-and-play clarifier: A zero-shot multimodal framework for egocentric intent disambiguation, 2025. doi:10.48550/arXiv.2511.08971, arXiv:2511.08971.
- [29] A. Büscher, A. Möller, Deutsches netzwerk für qualitätsentwicklung in der pflege (DNQP): Aktueller stand, *Public Health Forum* 22 (2014) 27–28.