International Journal of Computer Science Engineering Techniques – Volume 8 Issue 5, October - 2024

Automating Regulatory Compliance Review in Financial Institutions Using Natural Language Processing

Viswatej Seela
Independent Researcher
viswatej1998@gmail.com

Abstract

Financial institutions face a growing volume of regulatory documents, guidance notes, and policy updates that must be monitored and translated into internal actions. Manual review is labor-intensive, error-prone, and difficult to scale. This study examines whether natural language processing can reduce the effort involved in document intake, obligation extraction, and change-impact analysis. The workflow combines named entity recognition, sentence-level obligation classification, and semantic similarity matching to connect new regulations with existing internal policies. On a corpus of 340 regulatory documents from U.S. and EU financial regulators, the approach achieves 87.4% precision and 82.1% recall in obligation extraction and correctly maps 79% of regulatory changes to affected policy sections. The results suggest that NLP can shorten the first stage of compliance review without removing the need for human legal and compliance judgment.

Keywords: regulatory compliance, natural language processing, obligation extraction, financial regulation, RegTech

1 Introduction

The regulatory landscape for financial services has expanded significantly since the 2008 financial crisis. Institutions operating across jurisdictions must track updates from dozens of regulatory bodies, including the Securities and Exchange Commission (SEC), the European Banking Authority (EBA), the Basel Committee, and national regulators. These bodies publish hundreds of pages of guidance, rules, and enforcement actions every year (Thomson Reuters, 2024). Compliance teams in large banks often spend substantial time reviewing these releases, identifying relevant obligations, and assessing their impact on internal policies and procedures.

This manual process creates three immediate risks. First, delays in identifying regulatory changes can lead to non-compliance, which can result in financial penalties and reputational damage. In 2023 alone, global financial regulators imposed more than $6 billion in fines for compliance failures (Fenergo, 2024). Second, legal text is often interpreted inconsistently, so different analysts may extract different obligations from the same document. Third, mapping new requirements to the internal policies they affect depends heavily on institutional memory and specialized expertise.

Recent advances in NLP, particularly transformer-based language models, provide practical tools for compliance workflows. Rather than replacing human judgment, these tools can speed up the identification and triage of regulatory content, allowing compliance officers to focus on interpretation and decision-making.

A three-stage NLP pipeline is evaluated on a multi-source corpus of financial regulatory documents. The purpose is not to automate compliance decisions, but to reduce manual effort in the first stage of review so that specialists can focus on interpretation, escalation, and policy decisions.

2 Related Work

The application of NLP to legal and regulatory text has expanded substantially. Chalkidis et al. (2020) introduced Legal-BERT, a domain-specific pre-trained model that outperforms general-purpose BERT on legal text classification tasks. Hendrycks et al. (2021) released the Contract Understanding Atticus Dataset (CUAD) for contract review, demonstrating the feasibility of automated clause extraction.

Specifically in the regulatory domain, Bholat et al. (2023) surveyed the adoption of RegTech solutions by UK financial institutions, finding that NLP-based tools were the most commonly adopted technology category. Zhang et al. (2024) proposed a transformer-based model for extracting normative statements (obligations, permissions, prohibitions) from regulatory text, achieving strong results on EU financial regulations. Sleimi et al. (2018) developed earlier rule-based approaches to regulatory requirement extraction.

This study builds on that foundation but focuses on end-to-end pipeline integration, including the downstream task of mapping extracted obligations to internal policy documents, a step that matters operationally but is less studied in the literature.

3 Methodology

3.1 Document Corpus

We assembled a corpus of 340 regulatory documents published between 2021 and September 2024 from the following sources: the SEC (enforcement actions, staff guidance), CFPB (circulars, advisory opinions), EBA (guidelines, technical standards), and OCC (bulletins, interpretive letters). Documents were converted from PDF to structured text using Apache Tika, with manual quality checks on a 10% sample to verify extraction fidelity.

For the internal policy mapping task, we obtained a set of 85 anonymized compliance policy documents from a mid-sized financial institution, covering areas including anti-money laundering (AML), know-your-customer (KYC), fair lending, data privacy, and market conduct.

3.2 Pipeline Architecture

The compliance review pipeline consists of three stages:

Stage 1: Entity Recognition. A fine-tuned NER model identifies regulatory entities in each document, including regulatory bodies, legal references (statute citations, rule numbers), affected entity types (banks, broker-dealers, fintechs), effective dates, and monetary thresholds. We fine-tuned a RoBERTa-base model on 2,400 manually annotated sentences.

Stage 2: Obligation Extraction. A sentence-level classifier categorizes each sentence as an obligation (“shall,” “must”), a permission (“may”), a prohibition (“shall not”), or non-normative (background, definitions). This model was trained on 8,200 labeled sentences using a fine-tuned

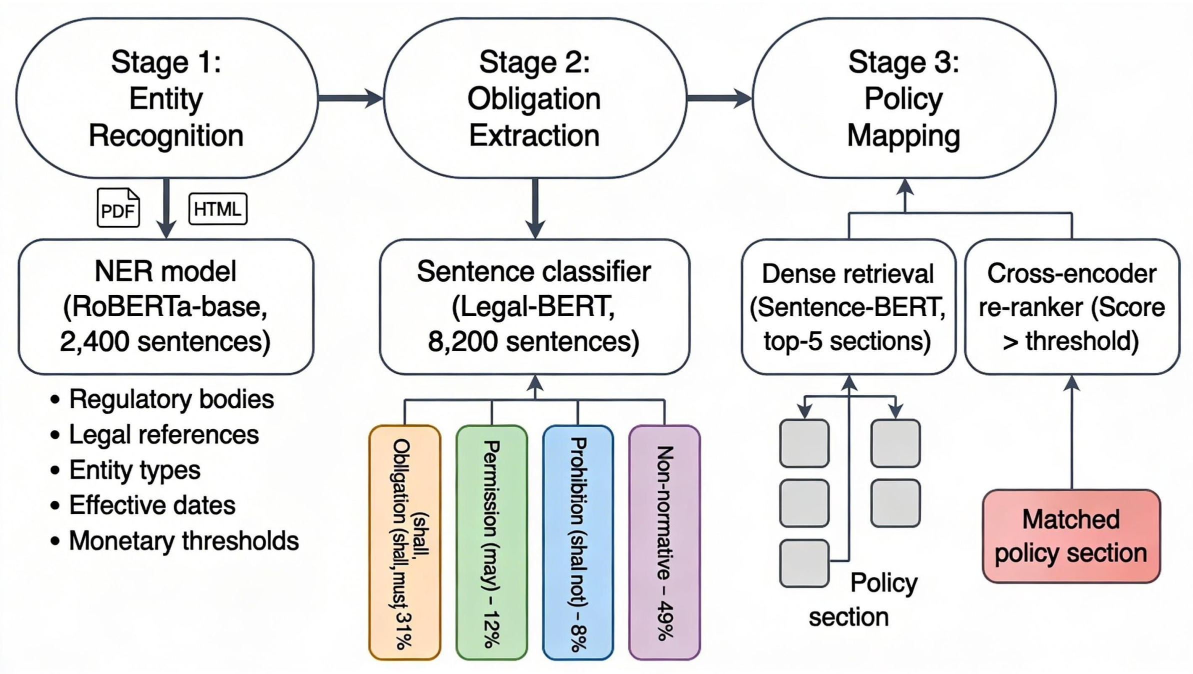

Figure 1: Workflow for document ingestion, obligation extraction, and policy mapping.

Legal-BERT model (Chalkidis et al., 2020), with label distributions of 31% obligation, 12% permission, 8% prohibition, and 49% non-normative.

Let denote a sentence and let $\in \{1, 2, 3, 4\}$ denote its class label. The encoder produces a contextual representation $h = (\;)$, and the classifier estimates class probabilities through a softmax layer:

$$(\;=\;|\;) = \frac{\exp(\;^{\top}h + \;)}{\sum_{=1}^{4} \exp(\;^{\top}h + \;)}.$$

To account for class imbalance across normative and non-normative sentences, the model is trained with weighted cross-entropy loss:

$$\mathcal{L}_{\mathrm{cls}} = -\frac{1}{}\sum_{=1}\sum_{=1}^{4} \mathbf{1}[\;=\;] \log (\;=\;|\;),$$

where is the class weight for category . This formulation keeps the extraction task explicit and makes model behavior easier to compare across alternative sentence classification approaches.

Stage 3: Policy Mapping. Extracted obligations are matched to relevant internal policy sections using a two-step process: (a) dense retrieval using sentence-BERT embeddings (Reimers and Gurevych, 2019) to identify the top-5 candidate policy sections, followed by (b) a cross-encoder re-ranker to score the relevance of each candidate. A match is accepted if the re-ranker score exceeds a calibrated threshold.

Figure 1 summarizes the workflow used for ingestion, extraction, and policy mapping.

3.3 Evaluation Protocol

Two compliance professionals from the financial services industry independently annotated a held-out set of 60 regulatory documents for (a) normative sentence labels and (b) correct internal policy mappings. Inter-annotator agreement (Cohen's kappa) was 0.81 for sentence classification and 0.74 for policy mapping, reflecting the inherent ambiguity in legal text interpretation.

4 Results

4.1 Obligation Extraction

Table 1 reports the performance of the obligation extraction stage on the held-out test set.

Table 1: Obligation extraction performance by category.

Category	**Precision**	**Recall**	**F1**
Obligation	87.4%	82.1%	84.7
Permission	83.2%	76.8%	79.9
Prohibition	85.1%	79.4%	82.2
Non-normative	91.6%	93.2%	92.4
Macro Average	86.8%	82.9%	84.8

The model performs well on obligations and non-normative text but shows lower recall on permissions and prohibitions, where the wording is often subtle and tied to conditional phrasing. Error analysis shows that 38% of missed obligations involve multi-sentence constructions in which the obligation is separated from its trigger condition.

4.2 Policy Mapping

The policy mapping stage correctly identified the relevant internal policy section in 79% of test cases (top-1 accuracy). This rises to 91% when considering the top-3 retrieved candidates, suggesting the retrieval stage is effective but the re-ranker can be further improved.

Errors in policy mapping frequently involve cross-cutting regulations that affect multiple policy domains simultaneously. For example, a data privacy regulation affecting AML record-keeping obligations may need to be mapped to both the privacy policy and the AML policy, but the model tends to identify only the primary match.

4.3 Time Savings Estimate

In consultation with compliance practitioners, we estimated the time savings from pipeline deployment. Manual review of a typical 30-page regulatory document requires 4–6 hours for obligation identification and 2–3 hours for policy mapping. The automated pipeline reduces the obligation-identification step to approximately 45 minutes of review and the mapping step to approximately 30 minutes, representing a 70–80% reduction in analyst time for the initial triage phase.

Figure 2 compares the manual workflow with the AI-assisted review workflow.

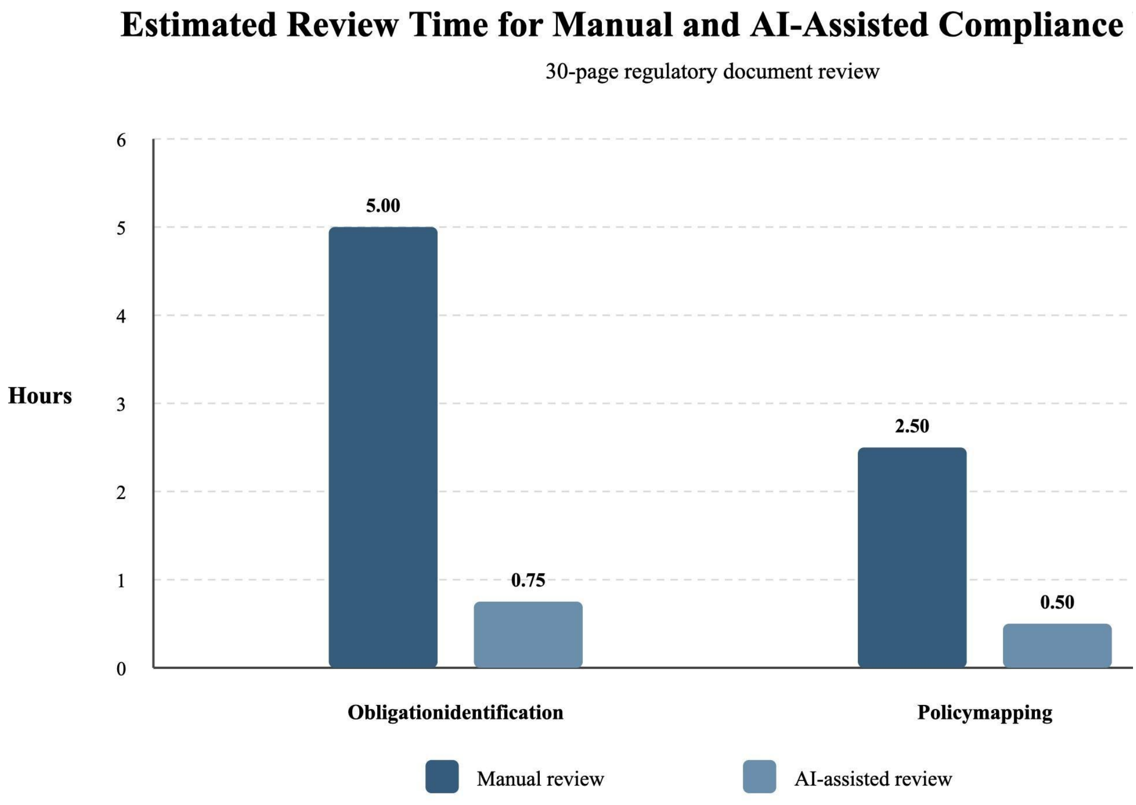

Figure 2. Review time estimate for two core compliance tasks.

Figure 2: Estimated review time for manual and AI-assisted compliance workflows.

5 Discussion

5.1 Practical Deployment Patterns

NLP-based compliance tools are best used as decision-support systems rather than autonomous classifiers. In practice, the pipeline produces a prioritized list of extracted obligations with suggested policy mappings, which compliance officers review, accept, or modify. This human-in-the-loop design aligns with both the accuracy limits of current models and the expectation that compliance decisions require qualified human judgment.

Integration with existing governance, risk, and compliance (GRC) platforms is essential for adoption. The pipeline outputs structured data (entity mentions, obligation text, confidence scores, suggested mappings) that can be ingested by standard GRC systems through REST APIs.

5.2 Limitations

Several limitations warrant discussion. First, our corpus, while diverse, is dominated by English-language U.S. and EU regulations; performance on documents from other jurisdictions and languages has not been evaluated. Second, the pipeline operates at the sentence level, which can miss obligations that span multiple sentences or are embedded in complex referential structures. Third, regulatory language evolves over time, and model performance may degrade as new terminology or phrasing conventions emerge, requiring periodic retraining or fine-tuning.

6 Conclusion

This study showed that a focused NLP pipeline can improve the first stage of regulatory compliance review in financial institutions. The system performs well on obligation extraction and produces useful first-pass policy mapping while leaving final decisions to human reviewers. As regulatory volumes continue to grow, tools of this kind can help compliance teams maintain coverage without increasing manual review effort at the same rate. Future work should extend the pipeline to multi-jurisdictional analysis and test whether larger language models add value without reducing reliability.

References

Bholat, D., Brookes, J., and Cai, C. (2023). The adoption of RegTech and AI in UK financial services. *Bank of England Staff Working Paper*, No. 1035.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., and Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *Findings of EMNLP*, 2898–2904.

Fenergo. (2024). Global financial institution fines report 2023. *Fenergo Research*.

Hendrycks, D., Burns, C., Chen, A., and Ball, S. (2021). CUAD: An expert-annotated NLP dataset for legal contract review. *Proceedings of NeurIPS*, 34, 17606–17616.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of EMNLP-IJCNLP*, 3982–3992.

Sleimi, A., Ceci, M., Pourmasoumi, A., and Susi, A. (2018). Automated extraction of semantic legal metadata using NLP. *Proceedings of RE*, 302–311.

Thomson Reuters. (2024). Cost of compliance report: Navigating regulatory complexity. *Thomson Reuters Regulatory Intelligence*.

Zhang, L., Wang, H., and Chen, Y. (2024). Transformer-based normative statement extraction from financial regulatory text. *Artificial Intelligence and Law*, 32(1), 45–68.