Judging by the Rules: Compliance-Aligned Framework for Modern Slavery Statement Monitoring

Wenhao Xu[*1,2], **Akshatha Arodi**[*1], **Jian-Yun Nie**[2], **Arsène Fansi Tchango**[1]

[1]Mila - Quebec AI Institute
[2]Université de Montréal
{akshatha.arodi-nagaraja, arsene.fansi.tchango}@mila.quebec,
{wenhao.xu, jian-yun.nie}@umontreal.ca

Abstract

Modern slavery affects millions of people worldwide, and regulatory frameworks such as Modern Slavery Acts now require companies to publish detailed disclosures. However, these statements are often vague and inconsistent, making manual review time-consuming and difficult to scale. While NLP offers a promising path forward, high-stakes compliance tasks require more than accurate classification: they demand transparent, rule-aligned outputs that legal experts can verify. Existing applications of large language models (LLMs) often reduce complex regulatory assessments to binary decisions, lacking the necessary structure for robust legal scrutiny. We argue that compliance verification is fundamentally a rule-matching problem: it requires evaluating whether textual statements adhere to well-defined regulatory rules. To this end, we propose a novel framework that harnesses AI for rule-level compliance verification while preserving expert oversight. At its core is the Compliance Alignment Judge (CA-Judge), which evaluates model-generated justifications based on their fidelity to statutory requirements. Using this feedback, we train the Compliance Alignment LLM (CALLM), a model that produces rule-consistent, human-verifiable outputs. CALLM improves predictive performance and generates outputs that are both transparent and legally grounded, offering a more verifiable and actionable solution for real-world compliance analysis.

Code — https://github.com/mila-ai4h/aims-alignment
Extended version — https://arxiv.org/pdf/2511.07803

1 Introduction

Modern slavery continues to affect over 50 million people worldwide (Walk Free 2022). To address this urgent issue, several countries have enacted Modern Slavery Acts (MSAs), requiring companies to disclose how they assess and address slavery risks within their supply chains. While these laws have led to the publication of thousands of corporate statements each year (UK Government 2025; Australian Government 2025; Public Safety Canada 2025), the disclosures vary widely in clarity, structure, and substance. The current reliance on fully manual review hinders large-scale enforcement, leaving substantial gaps that allow harmful practices to persist (Chambers and Vastardis 2020). With over 80,000 modern slavery statements published globally and limited capacity for expert review, enforcement remains a major challenge (Bora et al. 2025a). This gap presents a unique opportunity for AI-driven solutions that are engineered for demonstrable social benefit when guided by strong domain expertise and rigorous traceability.

Natural Language Processing (NLP) offers a promising path toward scalable compliance monitoring. However, the interdisciplinary nature of this task, spanning legal interpretation, policy enforcement, and societal accountability demands capabilities beyond those of standard classification models (Chen, Rinderle-Ma, and Wen 2025; Pistilli et al. 2023; Rinderle-Ma, Winter, and Benzin 2023). Modern slavery statements often contain vague, unstructured and promotional language that blends compliance content with corporate messaging, complicating the extraction and verification of criteria (Bora et al. 2025b), where conventional NLP approaches frequently fall short (Ariai and Demartini 2024).

Effective compliance verification must prioritize not only accuracy but also traceability. For example, under the Australian Modern Slavery Act (Australian Government 2018), the *Approval* criterion demands explicit attribution to the principal governing body or CEO. Generic or vague language fails to meet this standard. Systems that rely on superficial cues risk costly errors: false negatives may unfairly penalize compliant firms, while false positives may allow violations to go undetected. In such high-stakes domains, outputs must provide transparent, rule-aligned justifications that enable human auditors to trace model outputs back to recognizable legal standards. Additionally, standard evaluation metrics such as precision and recall offer only aggregate insights and fail to capture whether individual regulatory criteria have been correctly justified (Lipton 2018; Doshi-Velez and Kim 2017). This limitation is especially critical in domains like compliance verification, where decisions often hinge on nuanced, criterion-specific judgments. As a result, in current practice, human reviewers remain essential for verifying whether disclosures meet statutory obligations (Chaleshtori et al. 2024; Zwickel et al. 2023). Consequently, prior work has recommended human-in-the-loop systems (Bora et al. 2025b,a).

[*]These authors contributed equally.

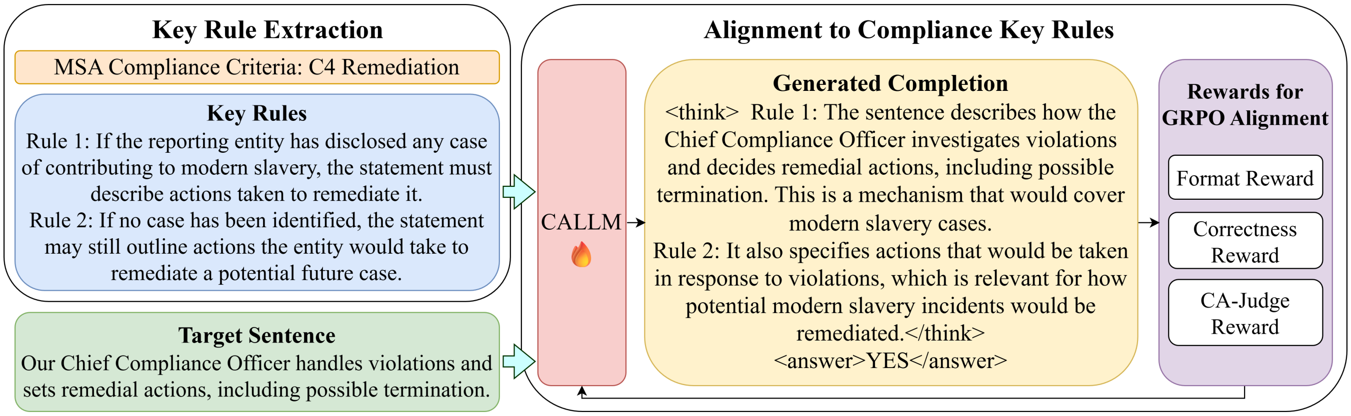

Figure 1: Overview of our framework, which consists of two steps: (1) *Key Rule Extraction* derives natural language rubrics for a compliance criterion. (2) *Alignment* trains CALLM to generate rule-aligned outputs using feedback from CA-Judge. The figure shows an example for the *C4 Remediation* criterion under the Australian Modern Slavery Act, including the corresponding key rules, a target sentence to be classified as compliant (YES) or non-compliant (NO), and CALLM's rule-aligned generation.

In this work, we propose a compliance-aligned framework and train the Compliance Alignment LLM (CALLM) for AI-assisted verification of modern slavery disclosures. Our approach is designed to produce rule-consistent outputs for human audit. At the center of our framework is a Compliance Alignment Judge (CA-Judge), a rule-aware evaluator that assesses whether the output of a model satisfies the statutory requirements of a given compliance criterion. Unlike general-purpose LLM-as-judge models (Zheng et al. 2023b; OpenAI 2023; Korbak, Muennighoff et al. 2023), CA-Judge is grounded in domain-specific regulatory logic. Its structured evaluations are intended to mirror how humans assess compliance: through attention to rule coverage, specificity, and clarity. CALLM uses the feedback of CA-Judge as a reward signal during training, encouraging it to generate outputs that are explicitly aligned with relevant criteria rules (Figure 1). Our goal is to verify whether this improves both task performance and the auditability of generated outputs. By generating rule-aligned assessments, CALLM aims to support faster, more transparent compliance checks.

Recent work has shown that chain-of-thought (CoT) reasoning does not reliably reflect model decision-making and should not be treated as an interpretability method (Barez et al. 2025). Instead of using model outputs to explain final decisions, we generate rule-aligned outputs intended for human verification and not as explanations to be trusted on their own. In our framework, the compliance verifier must review each rule-level rationale and make the final determination. Rather than self-justification, our goal is to generate rule-grounded outputs for expert audit at the criterion level. This human-in-the-loop, rule-aligned design supports oversight, addresses enforcement bottlenecks, and improves real-world deployability. Moreover, CALLM uses relatively small models (3B), promoting adoption and reproducibility.

We evaluate CALLM using both quantitative and qualitative metrics. It outperforms the baselines in both compliance classification and the rule-adherence of generated rationales. A human preference study confirms that outputs are better aligned with statutory rules, making them easier to verify. These results support our hypothesis: aligning model outputs with domain-specific rubrics improves both performance and usability. Further cross-jurisdictional analysis demonstrates that the framework generalizes effectively across jurisdictions.

Our core contribution lies in the training framework that leverages a compliance-specific CA-Judge to produce rule-grounded outputs. The novelty of our approach is in integrating regulatory rubrics (key rules derived from compliance statutes) directly into the training process, enabling more structured generation. While we focus on modern slavery disclosures, the proposed methodology is applicable to other regulatory domains where decisions depend on rule adherence. We release code and implementation guidelines to facilitate adoption and extension across domains. We hope this work contributes to broader efforts in using AI for social good and inspires the community to engage more deeply with underexplored modern slavery compliance challenges.

2 Related Work

Compliance Verification Recent work has explored the use of NLP models for compliance classification, particularly in the regulatory domain (Sun, Luo, and Li 2025; Lorè et al. 2023). Bora et al. (2025b) introduced a large-scale corpus of modern slavery statements from Australia and small evaluation sets for UK and Canada (Bora et al. 2025a). However, the models in these works typically focus on surface-level features and lack explicit reasoning in line with legal rules, limiting their utility in high-stakes applications. Similarly, approaches developed for the COLIEE competition (Rabelo et al. 2022) are not suitable, as they emphasize cross-document retrieval or entailment. In contrast, compliance setting is normative: it requires verification of statutory criteria with traceable, rule-grounded justifications, redefining the decision target and evaluation goal.

KEY RULES for C4 Remediation:
Rule 1: If one or more cases have been declared by the reporting entity where it caused or contributed to modern slavery, the statement should describe the actions taken to remediate these cases.
Rule 2: If no modern slavery case has been identified by the reporting entity, it may still describe actions used to remediate hypothetical cases if one should occur in the future.
Relevant examples: Actions proposed by the reporting entity to remediate modern slavery cases. Corrective actions and sanctions to remediate modern slavery cases include, for example: conducting inquiries and investigations involving stakeholders.
Irrelevant examples: Actions proposed to mitigate the risks of modern slavery instead of remediating existing cases. E.g., "We understand the importance of workers knowing their rights and addressing violations when necessary."

EVALUATION DIMENSIONS:
1. *Accuracy*: Identifies all relevant compliance gaps based on the key rules. Applies legal concepts exactly as defined—no misinterpretation or omission. No partial credit—any incorrect rule application makes the reasoning inaccurate.
2. *Clarity*: Reasoning is logically structured with clear, step-by-step justification. Avoids vague terms, ambiguity, or unsupported claims. Final answer must clearly follow from the reasoning.
3. *Fidelity to Key Rules*: All relevant key rules must be explicitly mentioned and addressed. Paraphrasing is allowed only if legal meaning is preserved. Irrelevant or external standards are penalized.
4. *Consistency*: No internal contradictions; reasoning and conclusion must align.
5. *Evidence Use*: Cites or paraphrases relevant rule clauses accurately. No new rules shall be introduced.
6. *Cognitive Behaviors (Verification & Reflection)*: Demonstrates explicit self-checking, cross-referencing, or reflection.

Figure 2: Evaluation dimensions used by the CA-Judge. The key rules for the *C4 Remediation* criterion under the Australian Modern Slavery Act are also shown, along with examples of relevant and irrelevant sentence types.

Reasoning Models Recent advances in reasoning, such as Chain-of-Thought prompting (Wei et al. 2022), self-consistency decoding (Wang et al. 2022), and instruction-tuned LLMs (OpenAI 2023) improve accuracy by explaining intermediate steps. However, they often lack alignment with domain-specific rules, which is crucial in compliance contexts. Recently, it is shown that these explanations frequently fail to reflect the model's actual internal reasoning process, creating illusions of interpretability, problematic in high-stakes applications (Barez et al. 2025).

Reinforcement Learning for Alignment Reinforcement Learning with Human Feedback (RLHF) aligns large language models to human preferences (Ouyang et al. 2022). Reward signals based on final answer correctness are often too sparse, which limits training efficiency (Liu et al. 2024). Group Relative Policy Optimization (GRPO, Shao et al. 2024b) compares candidate outputs within groups, removing the need for an explicit reward model.

LLM-as-a-Judge The LLM-as-a-Judge framework leverages language models to evaluate outputs via preference-aligned feedback rather than rigid metrics (Zheng et al. 2023a; Gu et al. 2024). While effective in open-ended reasoning tasks (Saha et al. 2025), these approaches typically operate in general domains without structured rules.

Our Contribution We unify reasoning-based generation, alignment, and judge feedback in a compliance setting, where explicit rules guide both training and evaluation, and the CA-Judge provides fine-grained supervision to ensure outputs are coherent and aligned with regulatory criteria.

3 Dataset and Task

We use the AIMS.au dataset (Bora et al. 2025b), which contains annotated sentences from 5,731 modern slavery statements submitted by Australian companies in relation to the Australian Modern Slavery Act. The task is framed as sentence-level binary classification across multiple reporting criteria. These criteria differ in complexity. For example, *Signature* is a simple criterion that checks if the document is signed, while *C4 Mitigation* requires assessing whether a company has described concrete steps to mitigate modern slavery risks. We focus on 7 complex and 2 simple criteria in our experiments, selected based on the availability of expert-defined key rules for each. Each criterion is governed by a distinct set of key rules, defined by domain experts in (Bora et al. 2025b) based on the Australian Modern Slavery Act. Examples are shown in Figure 2. Models are trained on the dataset's training split and evaluated on the test split. To address class imbalance, we apply random downsampling to the training set.

Each training instance includes a target sentence, its surrounding context, and the key rules for a specific criterion. These are formatted into a structured prompt. The model is trained to generate a predicted label (`Yes`/`No`) along with a justification grounded in the provided rules, encouraging rule-aligned outputs. Further dataset details and a full prompt template are provided in the extended version.

4 Compliance Alignment LLM

To align model reasoning with domain-specific rule requirements, we propose a rule-aligned training framework, shown in Figure 1. This framework consists of two main stages: key rule extraction and alignment to these rules.

4.1 Key Rule Extraction

We translate compliance criteria into structured, rule-based rubrics that we define as key-rules. The key rules can be written by experts or extracted using LLMs and then re-

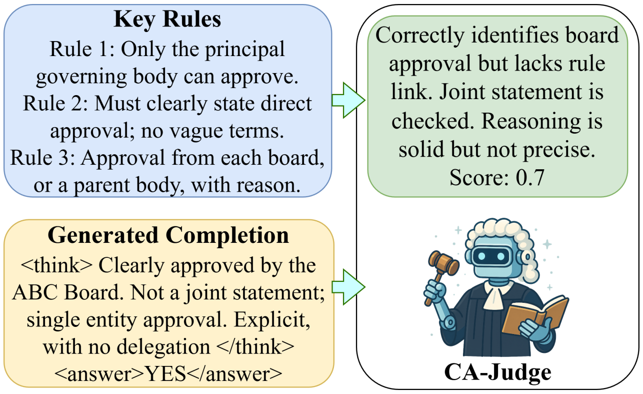

Figure 3: The Compliance Alignment Judge evaluates generated completion from a model against predefined key rules for compliance and generates a decision score with justification. The score reflects the degree of rule compliance and quality, enabling fine-grained, rule-aligned evaluation.

viewed. These rules distill regulatory knowledge into deterministic rubrics framed in natural language. Each rule clearly specifies what constitutes a valid versus invalid response. These serve as both training targets and evaluation anchors. Key rules of *C4 Remediation* criterion under the Australian MSA is provided in Figure 2.

4.2 Alignment to Compliance Key Rules

We train a model to align its outputs to the key-rules using feedback from the CA-Judge. To achieve this, we apply *Group Relative Policy Optimization* (GRPO) (Shao et al. 2024b), a reward-based fine-tuning method that uses scalar scores, selected for its adaptability and efficiency in policy optimization, particularly in low-resource settings.

4.3 Compliance Alignment Judge (CA-Judge)

We introduce the CA-Judge as an evaluation method for compliance verification. CA-Judge employs a large language model to assess the alignment between a model's outputs and a set of predefined rules. Given the model's rationale, predicted label, and the associated key rules for a specific criterion, CA-Judge returns a scalar score that captures overall alignment across six evaluation dimensions (see Figure 2 for their definitions). The six dimensions target auditability (*accuracy*, *fidelity*), explainability (*clarity*, *evidence use*), and reliability checks (*consistency*, *verification*). It also produces a justification for the score. This approach provides rule-grounded feedback and closely mirrors how humans assess compliance, making it well suited for regulatory domains (see Figure 3).

4.4 Reward Design and Training Pipeline

We design a composite reward function to guide the fine-tuning of our CALLM model (base model and training details are in Section 5). The reward integrates surface-level formatting checks, prediction correctness, and alignment with key rules. During training, each instance includes input text, key compliance rules, and a gold label.

Algorithm 1: Rule-Aligned Training

Input: Dataset $\mathcal{D}$ with (context, key rules, label), LLM f_θ, CA-Judge, weights λ_i
policy model $\pi_\theta \leftarrow \pi_{\text{init}}$
while not converged **do**
    Sample batch $\{(x_i, r_i, y_i)\}_{i=1}^{B} \sim \mathcal{D}$
    for all x_i in batch **do**
        Generate completions $\{c_i^{(1)}, \dots, c_i^{(K)}\}$ using f_θ
        for all $c_i^{(k)}$ **do**
            Get: $r_{\text{format}}, r_{\text{xml}}, r_{\text{correct}}, r_{\text{judge}}$
            $r_{\text{total}} = \sum_{j=1}^{4} \lambda_j \cdot r_j$
        end for
        Rank completions by r_{total} and compute GRPO loss
    end for
    Update model parameters θ; $\pi_{\text{old}} \leftarrow \pi_\theta$
end while
Return: Fine-tuned model f_θ

The model generates multiple completions, each containing structured reasoning and a final prediction. These outputs are scored along three dimensions: surface-level fidelity, correctness, and rule-alignment. A total reward is computed as a weighted sum of these components, and GRPO is used to rank completions and update model parameters. The training regime is shown in Algorithm 1.

Surface-Level Fidelity To encourage structurally correct outputs, we apply two rewards (Shao et al. 2024b):

1. Format Match. A binary reward indicating whether the output matches the expected format:

$$R_{\text{format}}(i) = \mathbf{1}\{\,\text{match}(c_i)\,\} \tag{1}$$

where c_i is the model's completion.

2. XML Tag Count. Rewards the presence of required tags while penalizing excess length:

$$\begin{aligned} R_{\text{xml}}(i) = \min[\,1,\ \max[\,0,\ & \\ \xi_r \sum_{t \in \mathcal{T}} \mathbf{1}\{t \in c_i\} - \xi_p\,\Delta_i\,]\,] & \end{aligned} \tag{2}$$

where $\mathcal{T}$ is the set of required tags, Δ_i is excess character count. $\xi_r = \frac{1}{|\#\mathcal{T}|}$ is the reward for required tags. ξ_p is the penalty for excess text.

Correctness. We assign a binary reward based on whether the model's predicted label matches the gold label:

$$R_{\text{corr}}(i) = \mathbf{1}\{\,\hat{y}_i = y_i\,\} \tag{3}$$

Rule-alignment. The Compliance Alignment Judge provides a scalar score that evaluates the alignment of the model's reasoning and final prediction with the key-rules:

$$R_{\text{judge}}(i) = \text{CAJudge}(\text{rules}, \text{reasoning}, \hat{y}) \in [0, 1] \tag{4}$$

Total Reward. We define the total reward as a weighted sum of the above components:

$$\begin{aligned} R_{\text{total}}(i) = \lambda_1 \cdot R_{\text{format}}(i) + \lambda_2 \cdot R_{\text{xml}}(i) \\ + \lambda_3 \cdot R_{\text{corr}}(i) + \lambda_4 \cdot R_{\text{judge}}(i) \end{aligned} \tag{5}$$

5 Base Models and Experimental Setup

We evaluate the proposed framework against a range of models as baselines, spanning Zero-Shot, Few-Shot and Fine-tuned settings. We include GPT-4o in zero/few-shot configurations using Chain-of-Thought prompts. In few-shot settings, we use 3 examples that were randomly selected from a diverse list of real-world cases to maximize coverage across typical compliance scenarios. Additionally, we evaluate DeepSeek-R1 and its distilled variant, DeepSeek-R1-Distill-Qwen-7B (DeepSeek-AI 2025), in zero-shot settings. We also include Pre-Trained and Fine-tuned variants of the CALLM base model (ablations). We do a full evaluation one criterion at a time for all 9 criteria.

CA-Judge: We use JudgeLRM-7B (Chen et al. 2025), a judgment-oriented LLM trained to be good at judging tasks, as our Compliance Alignment Judge, responsible for scoring the alignment between model-generated justifications and the compliance key rules (see Section 4.3). We selected JudgeLRM-7B, as it has fast inference speed and is trained in judgment tasks and has strong performance in judging-based evaluation benchmarks, consistently outperforming general-purpose models on such tasks. An example of the scoring rubric and full prompts are in the extended version.

CALLM: We use Qwen2.5-3B-Instruct (Yang et al. 2025) as the base model, chosen for its strong performance in instruction tuning and generation efficiency on compliance-style prompts. The model is fine-tuned using our framework as described in Section 4. In our experiments, we use equal weights for all rewards: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$. We set all λ values to 1 as fixed training hyperparameters, following the prior work (Shao et al. 2024a; DeepSeek-AI et al. 2025; Dao and Vu 2025). We set ξ_r to 0.25 as there are 4 required tags, to make the rewards fall in [0,1] and ξ_p to 0.001.

Evaluation Metrics: We report F1 score as the primary metric for classification performance. To assess the rule adherence of model justifications, we additionally assess the model outputs using CA-Judge score, which reflects the degree to which model outputs align with compliance key rules. We opted against reference-based metrics as they correlate poorly with human judgments, and require high-quality gold reasoning references, which are difficult to obtain in compliance settings without substantial expert annotation. We leverage CA-Judge that evaluates across multiple axes such as accuracy, clarity, and correctness, which is seen in recent works (Liu et al. 2023; He, Zhang, and Roth 2024; Li et al. 2024). We also conduct human evaluation.

6 Results

Quantitative Analysis Table 1 presents F1 scores across the nine compliance criteria for a range of baseline models and our proposed model, CALLM. Despite having only 3B parameters, CALLM achieves the highest overall macro-F1 score (0.639), outperforming much larger models such as GPT-4o [1] and DeepSeekR1 (671B). To ensure that performance gains are not solely due to fine-tuning, we include a fair baseline, "base model FT", identical in architecture and training data to CALLM but lacking our alignment feedback with CA-Judge. CALLM consistently outperforms this baseline, highlighting the value of our compliance-aligned optimization. CALLM shows particularly strong performance on challenging criteria such as *C2 Supply Chains*, *C3 Risk Description*, and *C4 Mitigation*, where it requires interpreting complex and sometimes subjective compliance rules. These results demonstrate the benefits of our alignment strategy, especially in tasks that demand structured rule-grounded reasoning. Interestingly, while CALLM excels in complex criteria, its performance on simpler ones like *Approval* is slightly lower than GPT-4o. Our error analysis reveals that CALLM is stricter in rule adherence, and is sensitive to data issues—such as broken target sentences that omit key terms like "approved by", which are required by the rule definitions (see the extended version for examples). Overall, CALLM strikes a favorable balance between model size and performance, surpassing all baselines under comparable or even more favorable resource settings.

We evaluate model outputs using CA-Judge to assess whether training with compliance-aligned feedback leads to outputs that better reflect the intended regulatory rules. While final qualitative validation is done with human evaluation, this automated assessment serves as a key intermediate check to validate our training framework on a held-out test set. In our setting, as expected, training with CA-Judge feedback leads to reasoning that better aligns with the compliance rules, specific to the task of regulatory rule alignment. Specifically, we compare CALLM to the second-best model from Table 1, GPT-4o with Few-shot chain-of-thought prompting. Across all nine criteria, CALLM consistently outperforms GPT-4o, achieving a substantially higher overall average score (0.74 vs. 0.53). These results suggest that CALLM generates justifications that are more grounded in the key compliance rules, while GPT-4o often produces generic reasoning with limited alignment to compliance rubrics. CALLM shows particularly strong gains on challenging criteria such as C2 Structure, C3 Risk Description, and C4 Remediation, where precise rule application is both complex and critical. These results highlight the value of our framework guided by compliance-specific evaluations. Detailed results are in the extended version.

Qualitative Analysis To assess the quality of model-generated justifications, we conducted a human study comparing CALLM and GPT-4o. The study involved 270 total responses from five volunteer participants (two with prior knowledge of modern slavery compliance and three without). For each of the nine compliance criteria, we randomly sampled six representative examples (three compliant, three non-compliant), excluding low-quality or overly short sentences. Both models were prompted using the same setup to generate structured rule-aligned outputs and a final answer. Responses were uniformly formatted and randomly assigned as Option A or B to ensure blinding.

Participants were asked to choose which response better satisfied the compliance rules, using the same evaluation dimensions as the CA-Judge: accuracy, clarity, fidelity to

[1] Estimated at 1800B parameters

Group	Criterion	GPT-4o ZS	GPT-4o FS	DeepSeek R1	DeepSeek Distill	Base model PT	Base model FT	CALLM (Ours)
# Params		1800B	1800B	671B	7B	3B	3B	3B
General	Approval	**0.855**	0.843	0.837	0.464	0.349	0.755	0.786
	Signature	0.409	0.636	0.250	0.154	0.422	0.524	**0.692**
C2	Structure	0.619	0.658	**0.678**	0.350	0.310	0.535	0.572
	Operations	0.529	**0.651**	0.537	0.290	0.181	0.596	0.632
	Supply Chains	0.420	0.556	0.399	0.260	0.192	0.550	**0.601**
C3	Risk Description	0.422	0.450	0.495	0.260	0.237	0.564	**0.712**
C4	Mitigation	0.709	0.664	0.700	0.243	0.451	0.714	**0.749**
	Remediation	0.552	**0.601**	0.529	0.159	0.225	0.397	0.570
C5	Effectiveness	**0.518**	0.492	0.504	0.342	0.267	0.394	0.439
Overall	(macro)	0.559	0.617	0.548	0.280	0.293	0.559	**0.639**

Table 1: F1 Scores Across Compliance Criteria for Baseline Models and our CALLM model. Best score in each row is highlighted. ZS = Zero-shot, FS = Few-shot, PT = Pre-trained, FT = Fine-tuned. Base model here indicates the base model of CALLM without CA-Judge supervision. Overall, CALLM outperforms the baselines.

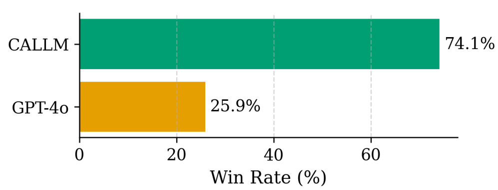

Figure 4: Human-preference comparison between our model and the baseline, showing that CALLM was preferred.

key rules, and consistency. Across 54 paired comparisons, CALLM was preferred in 74.1% of cases, receiving 200 preference votes compared to 70 for GPT-4o, as shown in Figure 4. CALLM was especially favored in criteria requiring nuanced compliance interpretation (e.g., distinguishing corporate structure from operational involvement). To further assess alignment between human and CA-Judge evaluations, we compared preference direction per criterion. In 7 out of 9 criteria (77.8%), human judgments matched the CA-Judge's scoring direction, demonstrating strong agreement between the two. More details are in the extended version.

Ablation Study Table 1 includes ablations of our CALLM model. We compare: (1) the pre-trained base model (**PT**), (2) a model fine-tuned using surface-level rewards for format and label correctness (**FT**), and (3) our full model, CALLM, trained with all the rewards, including rule-aligned feedback from the Compliance Alignment Judge (**CALLM**). CALLM achieves the highest overall performance, with improvements on complex criteria such as *C3 Risk Description*, and *C4 Mitigation*. These results demonstrate that fine-tuning based on format and correctness improves upon the base model, but the greatest gains are achieved by aligning reasoning with key rules from the CA-Judge.

7 Discussion

Case Studies We present qualitative case studies demonstrating that our training framework yields rule-aligned justifications. Figure 5 (top) shows an example. We observe that in several instances, baseline models predicted the correct label but either omitted justification or provided vague rationales unrelated to the key compliance rules. In contrast, CALLM generates explanations that explicitly referenced relevant risk-related criteria (e.g., identifying when descriptions failed to specify industries, regions, or supply chain elements associated with modern slavery). Additional detailed cases are shown in the extended version.

Error Analysis We examine failure cases of CALLM. As shown in Figure 5 (bottom) CALLM incorrectly predicts compliance. Note that the CA-Judge correctly identifies the rule violations and assigns a lower score. In some cases, the model generates verbose or generic reasoning that nominally matches key-rules but lacks specificity. Ambiguous inputs such as implicit approvals or vague governance language remain challenging, suggesting the need for tighter rule definitions or hybrid supervision strategies.

Cross-jurisdiction Generalization To test generalization beyond Australia, we evaluated CALLM on modern slavery statements from the UK and Canada (Bora et al. 2025a). CALLM is trained on the AU train data and evaluated on the test sets of AU, UK and CA. We use the same setup as Table 1. CALLM consistently outperforms larger models and the fine-tuned baseline (macro-F1), yielding robust cross-jurisdictional performance (see Table 2).

Facilitation of follow-up work We release code, prompt templates, and implementation guidelines to support real-world adoption and enable the community to build on our work whether by improving model performance, enhancing evaluation methods, or extending the framework to other

CALLM Success Case for the C4 Remediation criterion

Sentence: ...contributing to society in both developed markets as well as the developing regions of the world where we operate this is enshrined in our Code. ***Our Chief Compliance Officer and our Regional Ethics Counsels administer our Code, investigate and respond to violations and determine remedial and disciplinary action which can include termination of employment.*** We have committed to conducting annual training on the Code for our employees...
CALLM Reasoning: 1. The target sentence clearly states that the Chief Compliance Officer (CCO) and Regional Ethics Counsels (RECs) administrate a Code, investigate and respond to violations, and determine remedial and disciplinary actions, which can include termination of employment. 2. This describes a remediation action for modern slavery cases by specifying what actions can be taken (termination of employment). 3. The sentence describes a process, which includes administering a Code, investigating and responding to violations, and determining remediation and disciplinary measures, aligning with Rule 1 since it refers to existing or potential cases. 4. Even though it mentions the possibility of future cases, the actions outlined are clearly remediation and handling of existing cases (Rule 1 requirements specifically addressing potential future cases as well as existing ones). These actions are directly specified for Rule 2. 5. The actions described are specific and concrete, providing clarity on what remediation measures the entity will take.
CALLM Final Answer: *yes* (Ground truth: yes)
CA-Judge response: The model's reasoning aligns well with the key rules. It correctly identifies that the sentence describes concrete remediation actions, such as termination, satisfying Rule 1. It also notes these actions could apply to future cases, covering Rule 2. The explanation is clear and well-structured but could be improved by more explicitly linking the actions to specific aspects of each rule. The model's final answer of "YES" is correct because the reasoning clearly supports that the described actions meet the criteria for remediation actions as per the key rules. *Score: [[0.9]]*

CALLM Failure Case for the Approval criterion

Sentence: This Statement was approved by the boards of the three reporting entities covered by this Statement... Busways Gosford Pty Ltd. Busways North West Pty Ltd. ***Busways Pacific Pty Ltd.*** Busways South Australia Pty Ltd...
CALLM Reasoning: The key rule is whether the statement shows approval by the principal governing body. The sentence *"This Statement was approved by the boards of each of the reporting entities covered by this statement"* appears to meet this. 1. Approval Authority: Specifies approval by the board of each reporting entity, indicating direct involvement of governing bodies. 2. Direct Approval: No indirect terms like "on behalf of" or "considered by." 3. Verification: "The Board of Busways Pty Ltd" confirms direct, explicit board-level approval. The sentence clearly states board-level approval for each entity, with no mention of delegation or committees.
CALLM Final Answer: *yes* (Ground truth: no)
CA-Judge response: The reasoning identifies board approval but misses ambiguity in "boards" and fails to consider joint statement alternatives. It only partially aligns with the key rules, leading to an incorrect answer. *Accuracy:* Notes board approval but overlooks that "boards" may imply a group rather than a singular principal body. *Clarity:* Logically structured but misses ambiguity in language and does not address required alternatives. *Fidelity to Key Rules:* Skips conditions for joint approval—e.g., higher body or justification for single board. *Consistency:* Internally consistent but incomplete in rule coverage. *Evidence Use:* Cites rules but is vague on "boards" vs. principal body. *Cognitive Behaviors:* Some reflection shown, but alternative scenarios are not explored. *Final Answer:* Incorrect due to incomplete handling of joint approval rules. *Score: [[0.5]]*

Figure 5: Illustrative use cases of CALLM with CA-Judge scoring. *Sentence* includes the target sentence (bold) and context. **Top:** CALLM correctly predicts compliance for C4 Remediation. CA-Judge assigns a high score, reflecting strong alignment with key rules (shown in Figure 2) and well-structured reasoning. **Bottom:** CALLM incorrectly predicts compliance for Approval. CA-Judge detects flaws in the reasoning, correctly lowering the score in line with the rule violations (rules in Figure 3).

Model	Params (B)	AU	UK	CA
GPT-4o ZS CoT	1800	0.559	0.500	0.560
GPT-4o FS CoT	1800	0.617	0.573	0.614
DeepSeek-R1	671	0.548	0.505	0.550
Base Model FT	3	0.559	0.560	0.589
CALLM (ours)	**3**	**0.639**	**0.620**	**0.617**

Table 2: Models are trained on AU and evaluated on AU, UK and CA. CALLM shows generalization across jurisdictions. Here, ZS = Zero-shot, FS = Few-shot, and FT = Fine-tuned.

compliance domains. Additional details, including the experimental setup, hyperparameters, limitations of our work, and ethical considerations are in the extended version.

Future Work We aim to support multi-hop legal reasoning and finer-grained rule signals, further improving framework utility for complex policy and compliance challenges.

8 Conclusion

We introduce a novel framework for aligning large language models with compliance requirements in high-stakes domains. Our approach centers on CA-Judge, which evaluates model outputs against key statutory criteria, and trains the Compliance Alignment LLM (CALLM) using a rule-aligned reward signal. We find that CALLM, by generating outputs that explicitly reference relevant rules, outperforms baseline models in both predictive accuracy and human preference. This is intended to enhance human-in-the-loop verification of modern slavery statements by providing rule-based justifications, to enable faster and more reliable review. By supporting rule-aligned reasoning at scale, we aim to reduce manual review burdens, increase accountability, and build trust in real-world deployments. We hope this work encourages broader integration of AI-assisted review in modern slavery compliance and encourages the AI community to join the global fight against modern slavery.

Acknowledgments

We gratefully acknowledge Yuchen Hui for constructive and insightful discussions during the early stages of this project, and Ziyan Wang, Pedro Ferraz, Hager Radi, and Jeremy Pinto for their valuable suggestions throughout the experimental phase. This research was supported by compute resources provided by Mila (mila.quebec).

References

Ariai, F.; and Demartini, G. 2024. Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges. *ArXiv*, abs/2410.21306.

Australian Government. 2018. Modern Slavery Act 2018. https://www.legislation.gov.au/C2018A00153/latest/text. Accessed on 09 February 2025.

Australian Government. 2025. Modern Slavery Register. Accessed on 09 February 2025.

Barez, F.; Wu, T.-Y.; Arcuschin, I.; Lan, M.; Wang, V.; Siegel, N.; Collignon, N.; Neo, C.; Lee, I.; Paren, A.; et al. 2025. Chain-of-Thought Is Not Explainability. *Preprint, alphaXiv*, v2.

Bora, A. E.; Arodi, A.; Zhang, D.; Bannister, J.; Bronzi, M.; Tchango, A. F.; Bashar, M. A.; Nayak, R.; and Mengersen, K. 2025a. AIMSCheck: Leveraging LLMs for AI-Assisted Review of Modern Slavery Statements Across Jurisdictions. *arXiv preprint arXiv:2506.01671*.

Bora, A. E.; St-Charles, P.-L.; Bronzi, M.; Tchango, A. F.; Rousseau, B.; and Mengersen, K. 2025b. AIMS.au: A Dataset for the Analysis of Modern Slavery Countermeasures in Corporate Statements. arXiv:2502.07022.

Chaleshtori, F. H.; Ghosal, A.; Gill, A.; Bambroo, P.; and Marasović, A. 2024. On Evaluating Explanation Utility for Human-AI Decision Making in NLP. *arXiv preprint arXiv:2407.03545*.

Chambers, R.; and Vastardis, A. 2020. Human rights disclosure and due diligence laws: the role of regulatory oversight in ensuring corporate accountability. *Chicago Journal of International Law*, 21: 323.

Chen, N.; Hu, Z.; Zou, Q.; Wu, J.; Wang, Q.; Hooi, B.; and He, B. 2025. JudgeLRM: Large Reasoning Models as a Judge. arXiv:2504.00050.

Chen, Q.; Rinderle-Ma, S.; and Wen, L. 2025. Beyond Yes or No: Predictive Compliance Monitoring Approaches for Quantifying the Magnitude of Compliance Violations. *ArXiv*, abs/2502.01141.

Dao, A.; and Vu, D. B. 2025. AlphaMaze: Enhancing Large Language Models' Spatial Intelligence via GRPO. arXiv:2502.14669.

DeepSeek-AI. 2025. DeepSeek-R1-Distill-Qwen-7B. https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Qwen-7B.

DeepSeek-AI; Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; Zhang, X.; Yu, X.; Wu, Y.; Wu, Z. F.; Gou, Z.; Shao, Z.; Li, Z.; Gao, Z.; Liu, A.; Xue, B.; Wang, B.; Wu, B.; Feng, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; Dai, D.; Chen, D.; Ji, D.; Li, E.; Lin, F.; Dai, F.; Luo, F.; Hao, G.; Chen, G.; Li, G.; Zhang, H.; Bao, H.; Xu, H.; Wang, H.; Ding, H.; Xin, H.; Gao, H.; Qu, H.; Li, H.; Guo, J.; Li, J.; Wang, J.; Chen, J.; Yuan, J.; Qiu, J.; Li, J.; Cai, J. L.; Ni, J.; Liang, J.; Chen, J.; Dong, K.; Hu, K.; Gao, K.; Guan, K.; Huang, K.; Yu, K.; Wang, L.; Zhang, L.; Zhao, L.; Wang, L.; Zhang, L.; Xu, L.; Xia, L.; Zhang, M.; Zhang, M.; Tang, M.; Li, M.; Wang, M.; Li, M.; Tian, N.; Huang, P.; Zhang, P.; Wang, Q.; Chen, Q.; Du, Q.; Ge, R.; Zhang, R.; Pan, R.; Wang, R.; Chen, R. J.; Jin, R. L.; Chen, R.; Lu, S.; Zhou, S.; Chen, S.; Ye, S.; Wang, S.; Yu, S.; Zhou, S.; Pan, S.; Li, S. S.; Zhou, S.; Wu, S.; Ye, S.; Yun, T.; Pei, T.; Sun, T.; Wang, T.; Zeng, W.; Zhao, W.; Liu, W.; Liang, W.; Gao, W.; Yu, W.; Zhang, W.; Xiao, W. L.; An, W.; Liu, X.; Wang, X.; Chen, X.; Nie, X.; Cheng, X.; Liu, X.; Xie, X.; Liu, X.; Yang, X.; Li, X.; Su, X.; Lin, X.; Li, X. Q.; Jin, X.; Shen, X.; Chen, X.; Sun, X.; Wang, X.; Song, X.; Zhou, X.; Wang, X.; Shan, X.; Li, Y. K.; Wang, Y. Q.; Wei, Y. X.; Zhang, Y.; Xu, Y.; Li, Y.; Zhao, Y.; Sun, Y.; Wang, Y.; Yu, Y.; Zhang, Y.; Shi, Y.; Xiong, Y.; He, Y.; Piao, Y.; Wang, Y.; Tan, Y.; Ma, Y.; Liu, Y.; Guo, Y.; Ou, Y.; Wang, Y.; Gong, Y.; Zou, Y.; He, Y.; Xiong, Y.; Luo, Y.; You, Y.; Liu, Y.; Zhou, Y.; Zhu, Y. X.; Xu, Y.; Huang, Y.; Li, Y.; Zheng, Y.; Zhu, Y.; Ma, Y.; Tang, Y.; Zha, Y.; Yan, Y.; Ren, Z. Z.; Ren, Z.; Sha, Z.; Fu, Z.; Xu, Z.; Xie, Z.; Zhang, Z.; Hao, Z.; Ma, Z.; Yan, Z.; Wu, Z.; Gu, Z.; Zhu, Z.; Liu, Z.; Li, Z.; Xie, Z.; Song, Z.; Pan, Z.; Huang, Z.; Xu, Z.; Zhang, Z.; and Zhang, Z. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv:2501.12948.

Doshi-Velez, F.; and Kim, B. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Gu, J.; Jiang, X.; Shi, Z.; Tan, H.; Zhai, X.; Xu, C.; Li, W.; Shen, Y.; Ma, S.; Liu, H.; Wang, Y.; and Guo, J. 2024. A Survey on LLM-as-a-Judge. *ArXiv*, abs/2411.15594.

He, H.; Zhang, H.; and Roth, D. 2024. SocREval: Large Language Models with the Socratic Method for Reference-Free Reasoning Evaluation. arXiv:2310.00074.

Korbak, T.; Muennighoff, N.; et al. 2023. Model Evaluation for Long-Form Question Answering: A Case Study on Summarization. *arXiv preprint arXiv:2309.16505*.

Li, H.; Dong, Q.; Chen, J.; Su, H.; Zhou, Y.; Ai, Q.; Ye, Z.; and Liu, Y. 2024. LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods. arXiv:2412.05579.

Lipton, Z. C. 2018. The mythos of model interpretability. *Communications of the ACM*, 61(10): 36–43.

Liu, J.; Wang, C.; Liu, C.; Zeng, L.; Yan, R.; Sun, Y.; Liu, Y.; and Zhou, Y. 2024. Improving Multi-Step Reasoning Abilities of Large Language Models with Direct Advantage Policy Optimization. *ArXiv*, abs/2412.18279.

Liu, Y.; Iter, D.; Xu, Y.; Wang, S.; Xu, R.; and Zhu, C. 2023. G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment. arXiv:2303.16634.

Lorè, F.; Basile, P.; Appice, A.; de Gemmis, M.; Malerba, D.; and Semeraro, G. 2023. An AI framework to support decisions on GDPR compliance. *Journal of Intelligent Information Systems*, 61(2): 541–568.

OpenAI. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744.

Pistilli, G.; Muñoz Ferrandis, C.; Jernite, Y.; and Mitchell, M. 2023. Stronger together: on the articulation of ethical charters, legal tools, and technical documentation in ML. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 343–354.

Public Safety Canada. 2025. Fighting Against Forced Labour and Child Labour in Supply Chains Act. https://www.publicsafety.gc.ca/cnt/rsrcs/lbrr/ctlg/rslts-en.aspx?l=7. Accessed on 09 February 2025.

Rabelo, J.; Goebel, R.; Kim, M.-Y.; Kano, Y.; Yoshioka, M.; and Satoh, K. 2022. Overview and discussion of the competition on legal information extraction/entailment (COLIEE) 2021. *The Review of Socionetwork Strategies*, 16(1): 111–133.

Rinderle-Ma, S.; Winter, K.; and Benzin, J.-V. 2023. Predictive compliance monitoring in process-aware information systems: State of the art, functionalities, research directions. *Information Systems*, 115: 102210.

Saha, S.; Li, X.; Ghazvininejad, M.; Weston, J.; and Wang, T. 2025. Learning to Plan & Reason for Evaluation with Thinking-LLM-as-a-Judge. *ArXiv*, abs/2501.18099.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y. K.; Wu, Y.; and Guo, D. 2024a. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv:2402.03300.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.-M.; Zhang, M.; Li, Y. K.; Wu, Y.; and Guo, D. 2024b. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. *ArXiv*, abs/2402.03300.

Sun, J.; Luo, Z.; and Li, Y. 2025. A Compliance Checking Framework Based on Retrieval Augmented Generation. In *Proceedings of the 31st International Conference on Computational Linguistics*, 2603–2615.

UK Government. 2025. Modern Slavery Statement Registry. https://modern-slavery-statement-registry.service.gov.uk/. Accessed on 009 February 2025.

Walk Free. 2022. Global estimates of modern slavery: Forced labour and forced marriage. Technical Report, International Labour Organization (ILO).

Wang, X.; Wei, J.; Schuurmans, D.; Le, Q.; Chi, E.; Narang, S.; Chowdhery, A.; and Zhou, D. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Yang, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Zhou, C.; Li, C.; Li, C.; and et al., D. L. 2025. Qwen2.5 Technical Report. *arXiv preprint arXiv:2412.15115*.

Zheng, L.; Chiang, W.-L.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Li, D.; Xing, E. P.; Zhang, H.; Gonzalez, J. E.; and Stoica, I. 2023a. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *ArXiv*, abs/2306.05685.

Zheng, S.; Li, X.; Li, A.; et al. 2023b. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In *arXiv preprint arXiv:2306.05685*.

Zwickel, S.; et al. 2023. Compliance Auditing of AI Systems: The Need for Human-in-the-Loop Approaches. *AI & Society*.