Lean-based Compliance Training Pipeline

September 7, 2025

1 Datasets

- **Rule → Predicate**:
  - Input: Regulation text clause.
  - Output: Lean predicate (e.g., `def rule_6_1_b (i : Issuer) : Bool := ...`).
- **Filing → Issuer Instance**:
  - Input: Prospectus snippet (Red Herring).
  - Output: Structured Issuer JSON (operating profits, net worth, etc.).
- **Issuer → Compliance Report**:
  - Input: Issuer JSON.
  - Output: Lean report (pass/fail, reasons, remedies).

2 Training Objectives

Short-term (POC to MVP)

- **Supervised Fine-tuning (SFT)** on:
  - Regulation text → Lean predicate.
  - Filing snippet → Issuer JSON.
- **Rule-based rewards** (from Lean verifier and validators):
  - $r_{compile} = 1$ if Lean code compiles, else 0.
  - $r_{agree} = 1$ if Lean report matches gold label, else 0.
  - $r_{schema} = 1$ if Issuer JSON passes schema checks, else 0.
- Use these rewards in training:
  - *RS-SFT*: sample K candidates, keep top-q% by reward, fine-tune on them.
  - *DPO with synthetic prefs*: construct (y^+, y^-) pairs from rewards and train preferences.

Long-term (Scaling and Alignment)

- Incorporate **auditor preferences** for explanations/remedies.
- Train a **reward model** (RM) on ranked outputs.
- Apply **PPO/DPO** with a mixed reward:

$$r(y) = w_c r_{compile} + w_a r_{agree} + w_s r_{schema} + w_h r_{human}$$

- Expand across multiple jurisdictions and regulation domains.

3 Training Loop

Short-term

1. SFT warm-start on gold data.
2. For each input x: sample K candidates $y_1..y_K$.
3. Score with reward function $r(y)$.
4. RS-SFT: keep top-q% candidates for further fine-tuning.
5. Or: DPO with pairs (y^+, y^-) where $r(y^+) > r(y^-)$.

Long-term

1. Continue RS-SFT/DPO for stability.
2. Add PPO with reward signals (Lean + human RM).
3. Maintain KL regularization to prevent drift from the base SFT model.

4 Evaluation

Short-term

- **Compile Rate**: fraction of Lean predicates that compile.
- **Verifier Agreement**:
- **Field F1**: precision/recall for Issuer JSON extraction.
- **Runtime Efficiency**: average verifier runtime.

Long-term

- **Explanation Quality**: auditor ratings.
- **Generalization**: performance on unseen clauses and issuers.
- **Cross-domain Transfer**: robustness across new regulations.
- **User Satisfaction**: qualitative auditor feedback.

5 Deliverables and Milestones

Short-term (0–3 months)

- M1: Curated dataset (rules $\rightarrow$ Lean, filings $\rightarrow$ JSON).
- M2: SFT baseline with $\geq$70% compile rate.
- M3: RS-SFT/DPO pipeline with Lean reward integration; end-to-end demo.

Long-term (3–12 months)

- M4: RLHF with auditor preferences; explanations/remedies tuned.
- M5: Multi-jurisdiction expansion (SEBI, SEC, EU).
- M6: Productization: API, UI, batch tools, evidence linking.
- M7: Research publication or prototype demo.