



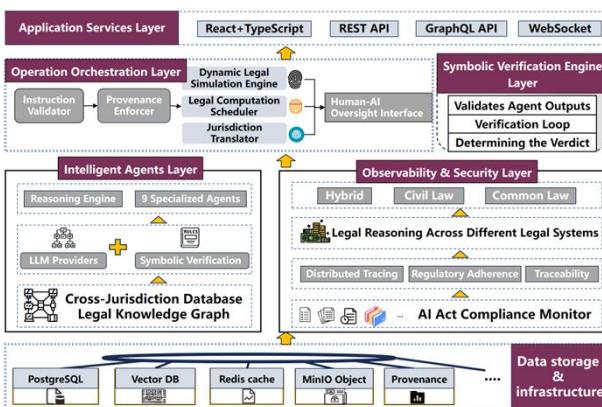
# Legal-ISA: A Modular Framework for Systematic Legal AI Evaluation

Daisong Gong  
Nankai University

## Introduction

Current legal AI research is constrained by **horizontal fragmentation** (narrow domain specialization lacking holistic reasoning) and **vertical fragmentation** (jurisdictional confinement preventing cross-border application), as evidenced by recent benchmarks like LEXam and MSLR. To address this, we introduce **Legal-ISA**, an integration layer inspired by computer architecture. By defining standardized operation interfaces for retrieval and reasoning, Legal-ISA ensures component substitutability and enables systematic, unified evaluation across diverse legal domains and jurisdictions.

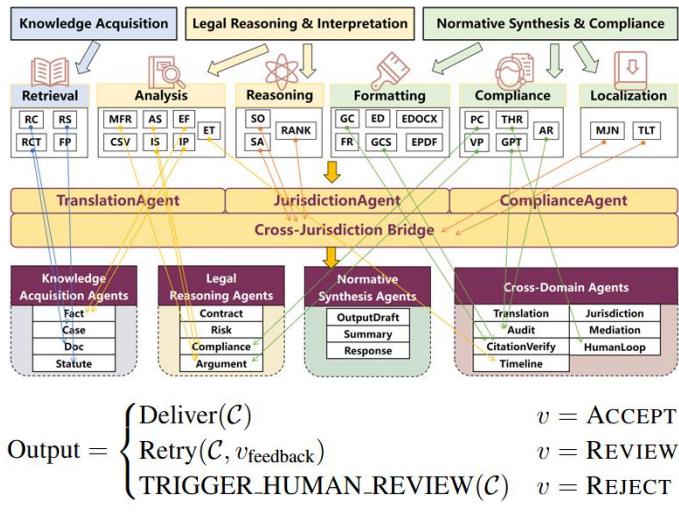
## Architecture



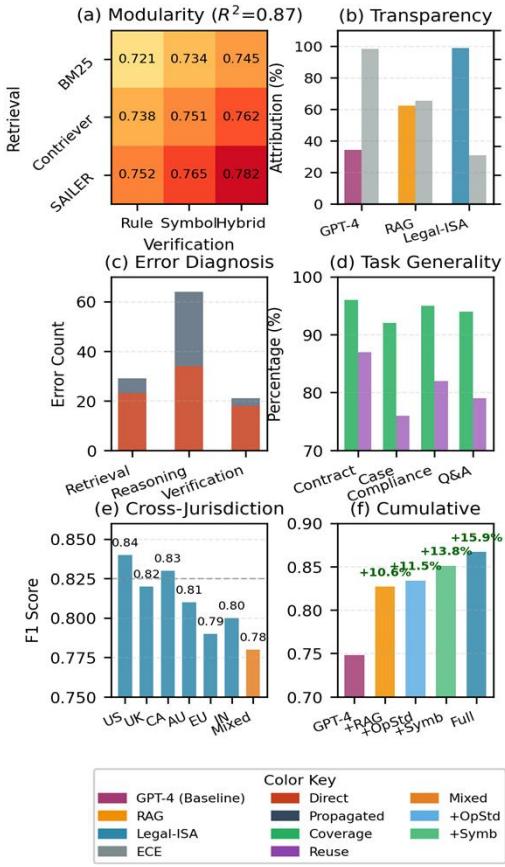
$$\text{conf}_{\text{final}} = \text{conf}_{\text{LLM}} \times \left( 1 - \beta \times \frac{\#\text{unattrb}}{\#\text{total}} \right)$$

- Architecture:** Six-layer framework integrates reasoning techniques via standardized interfaces and mandatory provenance protocols.
- Orchestration:** Orchestration layer manages validation, scheduling, cross-jurisdictional translation, and human oversight escalation.
- Verification:** Agents execute via a mandatory neuro-symbolic loop correcting outputs against formal legal constraints.
- Assurance:** Assurance layer guarantees observability and auditability backed by polyglot persistence and immutable ledgers.

## Mapping



## Experiment



## Conclusion

- Legal-ISA delivers high transparency and calibrated confidence, enabling principled human oversight and reasoning across diverse legal tasks.
- The framework supports cross-jurisdictional reasoning through human-in-the-loop knowledge engineering, opening collaborative research avenues.
- Our current work explores a legal-agnostic intermediate representation to unify cross-jurisdictional disparities and advance a global legal reasoning.
- We welcome discussions with fellow researchers—feel free to reach us at [2312325@mail.nankai.edu.cn](mailto:2312325@mail.nankai.edu.cn)!

**System Reliability:** Validates high modular attribution and superior transparency with minimized calibration error compared to baselines.

**Diagnostic Precision:** Localizes primary errors to reasoning bottlenecks (89% propagation) while maintaining 94.3% functional coverage across diverse legal tasks.

**Global Impact:** Demonstrates robust cross-jurisdictional performance and achieves a cumulative +15.9% improvement over standard models.

$$R^2 = 0.87$$

$$\rho = 0.94$$

QLAR	CAIL2018-Article
LegalBench	MultiLegalPile