TrustNLP, 2026

KoLegalQA: A Korean Legal QA Dataset for Trustworthy and Explanation-Grounded Legal AI

Yongtae Lee, Surin Lee, Sumin Kim, S M Wahidur Rahman, Heung-No Lee

Gwangju Institute of Science and Technology (GIST)

Motivation

Current Limitations

- Few Korean legal QA benchmarks
- Classification-focused resources
- Limited expert-verified supervision
- Weak statutory grounding
- Limited support for citation-based legal QA
- Few real-world consultation datasets

KoLegalQA

- 19,266 real-world legal consultations
- Expert-authored or verified responses
- Statutory references + clause-level summaries
- Benchmark for trustworthy legal QA

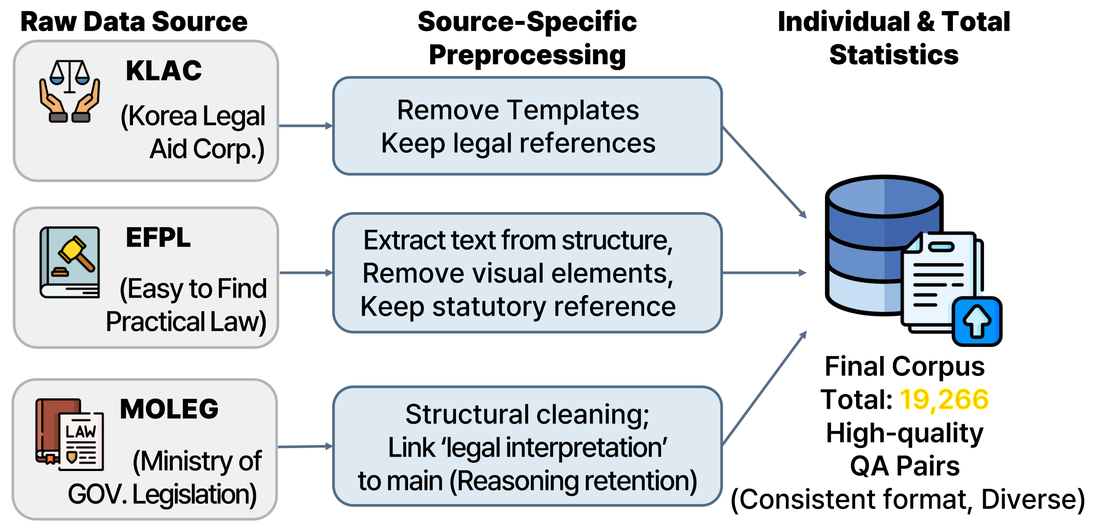
Dataset Construction

Data Sources

- From government-operated platforms in South Korea
  - Korean Legal Aid Corporation
  - Easy-Find Practical Law
  - Ministry of Government Legislation

Processing Steps

- QA extraction
- Cleaning & normalization
- Statutory reference preservation
- Clause-level summary attachment



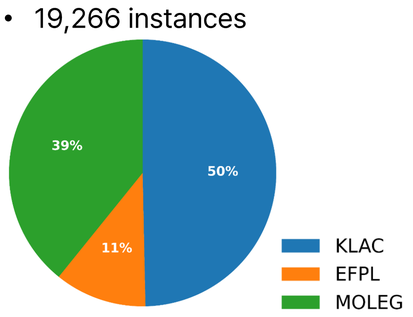
Statistics

Characteristics

- Real legal consultations
- Korean language only
- Statutory provisions
- Clause-level summaries
- Multiple legal domains

Final Dataset

- 19,266 instances



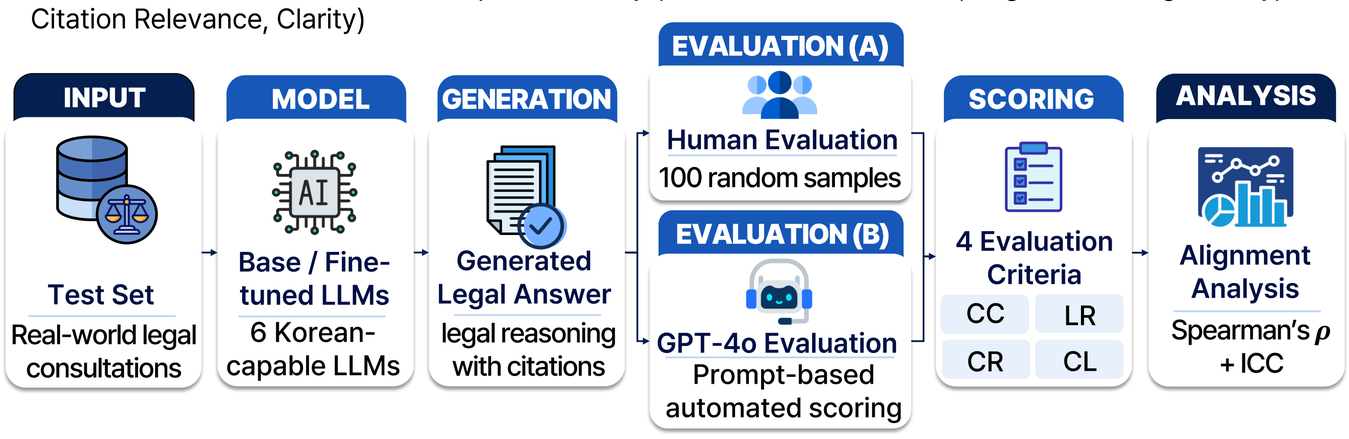
Annotation Density

Subset	Instances	Statutory Annotation Density	
		Mean	Median
KLAC	9,618	0.48	0
EFPL	2,156	0.86	0
MOLEG	7,492	11.71	10

Experimental Setup

Task Question → Legal Answer Generation

- 6 pre-trained Korean models (*e.g., princeton-nlp/Llama-3-8B-Prolong-512k-Base*) from HuggingFace
- Fine-tuning with same training configuration (QLoRA, 4-bit quantization)
- Evaluation: Automatic (G-Eval) & Human evaluation
- 4 Evaluation Criteria for Model's Response Quality (Conclusion Correctness, Legal Reasoning Validity, Citation Relevance, Clarity)



Dataset Example

Administrative Law · Traffic/Driving Life Law Q&A

Question: I left my wallet on the subway during my commute — it has my credit card and ID inside. What should I do?

Answer: If you left an item on the subway

- Contact the terminal station of the line you rode and the station where you got off. If you transferred, contact the terminal stations of every section traveled.
  - Describe the item in detail — time of loss, appearance, contents, distinguishing features.
- Inquire at a nearby police station, or check and report on the National Police lost-and-found portal.
- Check the Seoul Metro site or your regional transit lost-and-found center.

Reporting and reissuing identification

- Resident Registration Card: report and reapply at your district office or via Government24 (www.gov.kr).
- Driver's license: report and reapply at a license exam office or via the Korea Road Traffic Authority (www.safedriving.or.kr).

Credit card

- Call the card company immediately to suspend use, as a lost ID may also enable identity theft.

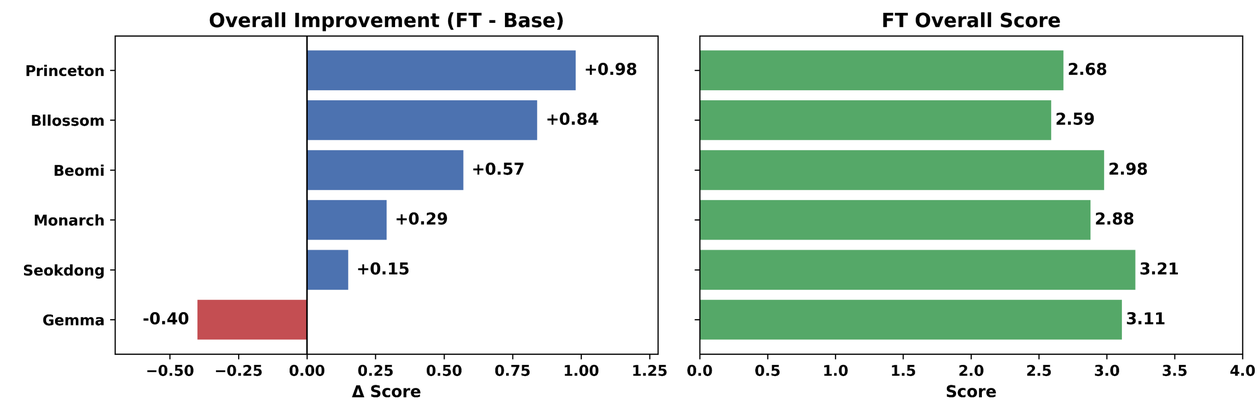
Source · Ministry of Government Legislation — Easy Life Law Information

Annotated cues of expert-grade output

Hierarchical, multi-step procedure	Names official channels + URLs
Specific, actionable instructions	Source-attributed to statute

G-Eval Results

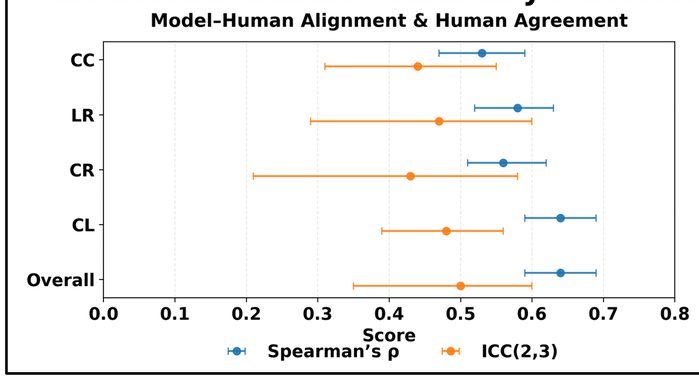
Base vs Fine-tuned Overall Score



- G-Eval shows moderate-to-strong agreement with human judgments (Spearman's ρ = 0.53–0.64).
- These results suggest that automated evaluation provides a reliable proxy for assessing legal QA quality.

Human Evaluation Results

Automated evaluation reasonably reflects human judgement



- Fine-tuned models received higher human ratings.
- Strong agreement between G-Eval and human evaluation.
- Spearman's ρ ranged from 0.53–0.64.
- Highest alignment for Overall quality (ρ = 0.64).
- Moderate inter-rater agreement (ICC = 0.43–0.50).
- Supports G-Eval as a reliable evaluation proxy.

Takeaways

- KoLegalQA establishes a large-scale benchmark with 19K+ **real-world legal consultations**.
- **Clause-level summaries** enable explanation-grounded legal reasoning.
- **Fine-tuning with KoLegalQA** improves legal response quality across most models.
- **G-Eval aligns well with human evaluation**.
- **KoLegalQA offers a practical benchmark** for Korean legal LLM research.

Github

https://github.com/cryptoecc/INFONET-Research

Contacts

lyt98313@gm.gist.ac.kr, heungno@gist.ac.kr

SCAN ME