KDASO: Knowledge-Data Dual-Driven Automated Skill Optimization for Liability Adjudication Tasks

Zi-Jian Cheng*, Weiming Wu*, Jie Meng, Zhen Peng, Shan Huang, Qun Li, Guobing Wu, Lan-Zhe Guo

TL; DR KDASO enhances the liability adjudication capability of large language models through knowledge- and data-driven self-evolution of skills.

Problem & Analysis

In skill-based liability adjudication, LLMs face **three intrinsic challenges**:

- Static skill libraries fail to adapt to dynamically evolving real-world scenarios, creating **a skill coverage gap** under distribution shift.
- Skill descriptions suffer from **a semantic execution gap**, where LLM reasoning trajectories may deviate from intended expert logic
- As the skill library grows, the context budget causes critical **skill omission** and attentional interference.

■**Observation 1**: **Static and lagging skills fail to adapt to dynamic-evolving scenarios.**

Real-world scenarios continually introduce novel patterns (e.g., new fraud behaviors). Existing expert-curated skills suffer from temporal lag, creating a coverage gap that prevents models from recognizing emerging cases.

→ Need continuous skill induction from real-time data.

■**Observation 2**: **Logic and description quality of skills determine skill executability by LLMs.**

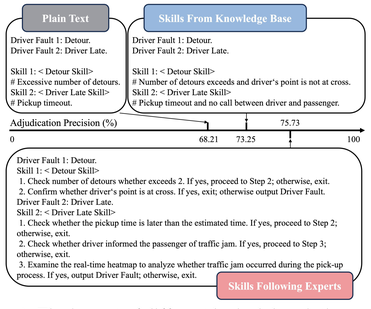
When skills are vague or miss submerged expert logic, LLMs rely on literal pattern matching and deviate from intended reasoning. Higher-quality, execution-aligned descriptions significantly improve adjudication performance.

→ Need automated refinement to bridge the semantic gap.

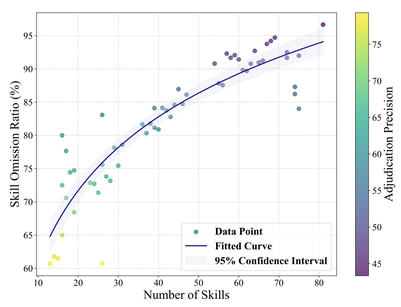
■**Observation 3**: **Skill omission ratio correlates positively with input skill count.**

Due to the Transformer's self-attention selectivity, feeding too many skills exceeds context budgets and dilutes attention, causing the LLM to ignore critical skills and leading to severe performance decay.

→ Need precise routing to select minimal sufficient skills.



▲ The impact of different logical description quality of skills on model performance.



▲ Number of skills as input v.s. skill omission ratio with adjudication precision evaluated on LLMs.

To address these challenges, we propose a **knowledge-data dual-driven framework (KDASO)** featuring **Neighborhood-Conditioned Skill Induction**, **Execution-Aligned Skill Refinement**, and **Context-Budgeted Skill Routing** to achieve dynamic evolution, precise execution, and efficient scalability.

Method

KDASO is a knowledge-data dual-driven automated skill optimization framework with three synergistic modules:

■ Neighborhood-Conditioned Skill Induction

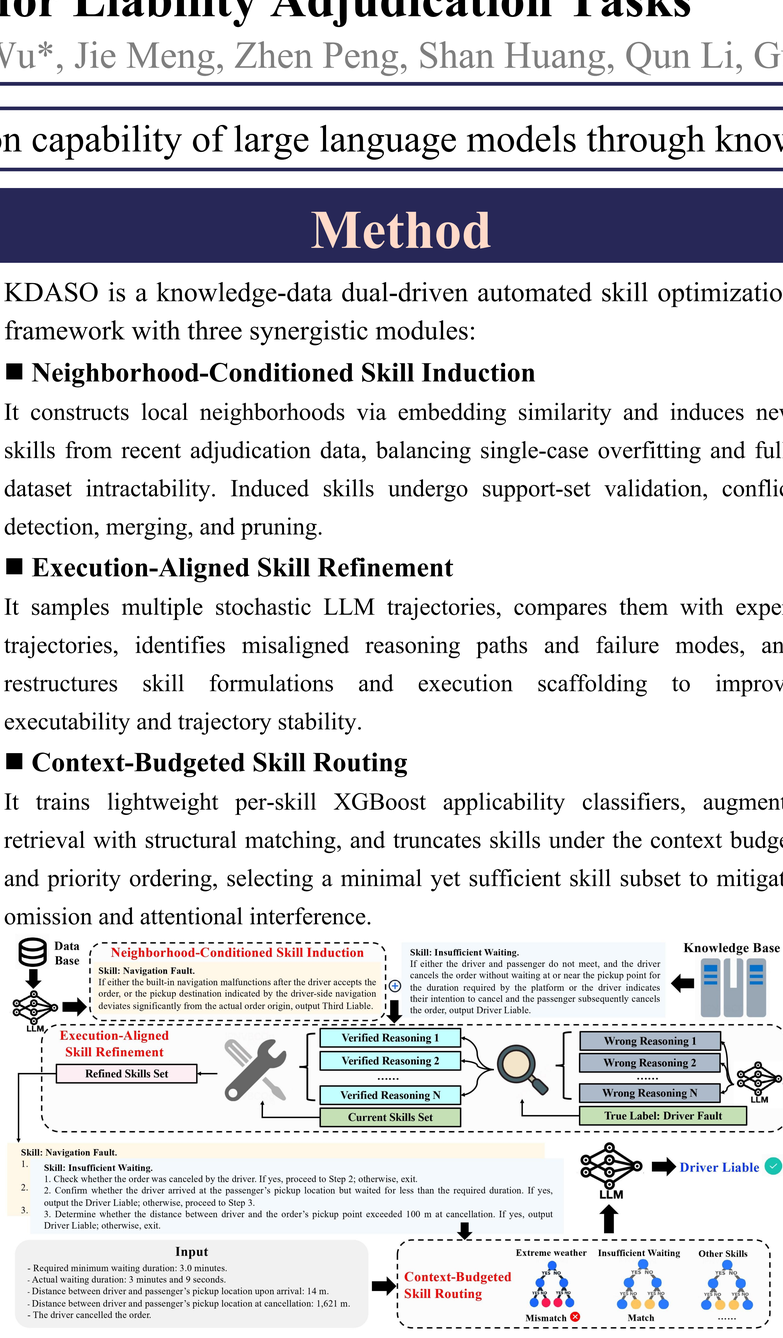
It constructs local neighborhoods via embedding similarity and induces new skills from recent adjudication data, balancing single-case overfitting and full-dataset intractability. Induced skills undergo support-set validation, conflict detection, merging, and pruning.

■ Execution-Aligned Skill Refinement

It samples multiple stochastic LLM trajectories, compares them with expert trajectories, identifies misaligned reasoning paths and failure modes, and restructures skill formulations and execution scaffolding to improve executability and trajectory stability.

■ Context-Budgeted Skill Routing

It trains lightweight per-skill XGBoost applicability classifiers, augments retrieval with structural matching, and truncates skills under the context budget and priority ordering, selecting a minimal yet sufficient skill subset to mitigate omission and attentional interference.



▲ The overall illustration of KDASO framework.

Experiment Setup

We conduct extensive evaluations across three real-world scenario families: **ride-hailing adjudication** on DiDi datasets, **legal liability determination** on LawBench and LegalBench, and **logical reasoning judgment** on ReClor and AR-LSAT. The performance is rigorously compared against four categories of baselines: *proprietary LLMs, liable-adjudicated general-purpose algorithms, domain-specific models, and open-source general LLMs*.

Results

- KDASO significantly outperforms existing baselines across ride-hailing, legal, and logical reasoning tasks, demonstrating strong cross-domain generalization and parameter efficiency.

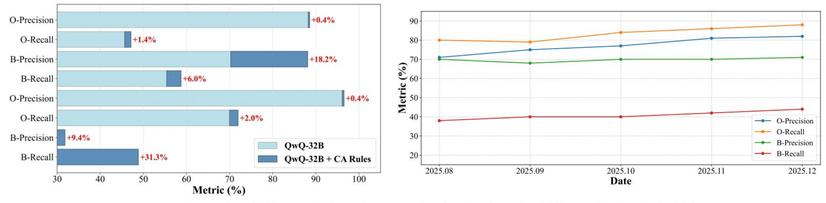
Model†	CA				CR				FD			
	O-Precision	O-Recall	B-Precision	B-Recall	O-Precision	O-Recall	B-Precision	B-Recall	O-Precision	O-Recall	B-Precision	B-Recall
NL2FOL	68.21±0.54%	62.13±1.87%	41.84±0.32%	23.56±0.45%	62.46±0.73%	50.66±0.68%	73.96±0.91%	30.31±0.29%	77.17±0.64%	60.85±0.77%	9.72±0.19%	43.75±0.58%
NeoBERT	57.98±1.62%	11.02±0.21%	52.94±0.48%	36.21±0.56%	63.95±0.71%	24.13±0.33%	56.30±0.59%	35.66±0.47%	71.13±0.83%	9.61±0.18%	**14.43±0.26%**	17.50±0.51%
ModernBERT	63.44±0.75%	18.85±0.28%	47.31±0.48%	46.57±0.69%	59.02±0.66%	16.73±0.04%	50.95±0.52%	24.24±0.35%	70.30±0.79%	9.89±0.20%	13.86±0.23%	17.50±1.31%
SIM-RAG	71.31±0.88%	59.16±0.74%	57.78±0.65%	29.89±0.41%	58.43±0.92%	15.48±0.42%	51.24±0.45%	23.75±0.34%	74.27±0.85%	37.46±0.51%	12.36±0.21%	23.75±0.34%
NodeRAG	72.24±2.89%	61.70%±0.76%	56.82%±0.63%	28.74%±0.39%	61.66%±0.90%	31.21%±0.53%	50.02%±1.33%	24.63%±0.36%	74.91%±0.86%	45.23%±1.89%	14.22%±0.25%	25.49%±0.37%
QwQ-32B	72.49%±0.89%	62.30%±0.97%	57.78%±0.65%	29.88%±0.41%	38.78%±0.52%	85.32%±0.94%	79.53%±0.93%	**83.76%±0.96%**	71.49%±0.86%	63.65%±0.84%	10.03%±0.20%	43.75%±0.58%
KDASO$_{Qwen3-14B}$	69.56%±0.83%	**92.01%±0.98%**	**69.79%±0.84%**	38.50%±0.53%	74.41%±0.85%	79.53%±0.93%	**83.76%±0.96%**	71.49%±0.86%	77.99%±0.89%	63.65%±0.84%	10.03%±0.20%	43.75%±0.58%
KDASO$_{QwQ-32B}$	**73.42%±0.90%**	87.38%±0.97%	69.17%±0.89%	**47.70%±0.60%**	**80.39%±0.94%**	79.09%±0.92%	81.18%±0.95%	**75.67%±0.87%**	**79.07%±0.91%**	**66.29%±0.81%**	10.11%±0.20%	**47.50%±0.60%**

Model	LA-ITI	LA-CP	LE-OPP	LE-UN	LE-LH
Proprietary LLMs					
Gemini 3 Pro	39.8	62.9	82.9	85.8	83.2
GPT-4o	44.8	52.9	77.0	87.2	76.6
Liable-Adjudicated General-Purpose Algorithms					
NL2FOL	34.0	53.8	69.8	83.5	74.3
NeoBERT	38.9	51.7	71.7	83.0	75.0
ModernBERT	33.4	53.7	68.8	83.7	74.5
SIM-RAG	34.2	53.0	72.8	82.1	74.8
NodeRAG	36.3	55.6	72.6	84.8	74.4
Liable-Adjudicated Domain-Specific Models					
Lawyer-LLAMA	4.4†	31.3	52.4	83.2	75.3
LexiLaw-6B	22.8	40.0	57.2	82.7	74.6
ChatLaw-13B	27.6	27.9	53.8	82.7	73.2
ChatLaw-33B	19.8	12.7	56.1	84.3	71.7
Open-Source General LLMs					
Deepseek-R1	**44.2**	61.0	74.2	84.3	79.0
Qwen3-8B	33.6	53.6	74.2	83.4	73.4
QwQ-32B	39.8	60.9	74.2	85.4	**80.6**
KDASO$_{Qwen3}$	39.6	56.4	73.8	83.4	76.1
KDASO$_{QwQ-32B}$	44.0	**63.5**	**74.4**	**86.8**	80.4

Model	Reclor			AR-LSAT
	Test	Test-E	Test-H	
Proprietary LLMs				
Gemini 3 Pro	96.6	96.1	96.9	94.2
GPT-4o	90.0	89.5	90.3	28.3†
Liable-Adjudicated General-Purpose Algorithms				
NL2FOL	88.1	89.5	87.0	77.0
NeoBERT	91.5	92.4	90.8	80.1
ModernBERT	85.5	88.2	83.4	78.9
SIM-RAG	87.6	88.9	86.6	76.1
NodeRAG	86.7	89.4	84.6	77.2
Liable-Adjudicated Domain-Specific Models				
MERIt	79.3	85.2	74.6	73.0
LReasoner	76.1	87.0	67.5	73.9
IDOL	80.6	87.7	75.0	76.1
AMR-LDA	90.2	91.6	89.1	77.0
Open-Source General LLMs				
Deepseek-R1	**96.0**	**96.1**	**95.9**	**95.2**
Qwen3-8B	87.3	87.9	86.8	77.0
QwQ-32B	92.2	94.1	90.7	83.9
KDASO$_{Qwen3}$	87.7	89.3	86.4	78.3
KDASO$_{QwQ-32B}$	93.7	93.9	93.6	87.8

▲ Comprehensive performance comparison of KDASO across three real-world scenarios.

- Skills transfer seamlessly to concept-similar scenarios, yielding a 8.96% improvement in overall performance without any adaptation. Furthermore, continuous optimization maintains stable and consistent gains, with a substantial portion of the generated skills successfully validated by DiDi experts and integrated into real-world policies.



▲ Transferability and Continuous Optimization Stability of KDASO Skills.

✓ If you are interested in this paper, please contact Zi-Jian Cheng (chengzj@lamda.nju.edu.cn) or visit our project page for more details (https://github.com/LAMDA-NeSy/KDASO) .

Project: **Paper:** **Contact:**