tcs Research

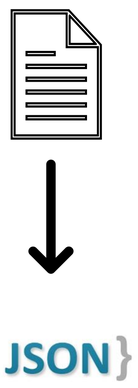
Inventing for impact

LexChronos: An Agentic Framework for Structured Event Timeline Extraction in Indian Jurisprudence

Anka Chandrahas Tummepalli, Preethu Rose Anish

Motivation

- Legal Reasoning needs structured case understanding
- Event extraction – connects facts, actors and time
- Structured timelines – key for legalAI, yet underexplored

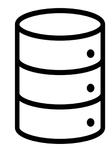


Challenges

- Judgments are dense and complex
- Need entity tracking, temporal links
- No public event-level annotated datasets for judgments

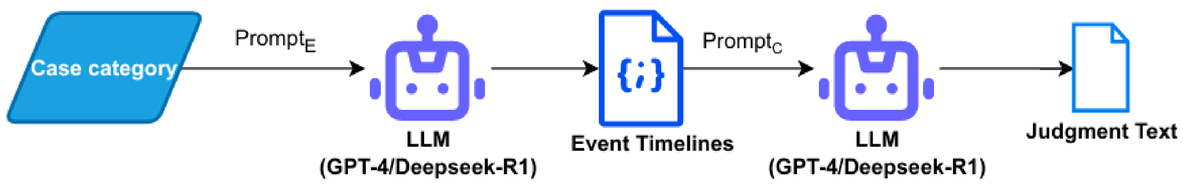
Contributions

- Dual Agent framework for structured timeline extraction
- Synthetic dataset of 2000 annotated Supreme Court judgments



Dataset Creation

- Case Category Selection
  - Randomly select from 25 case categories(Criminal, Civil, Cyber Law etc.)[1]
- LLMs used – DeepSeek-R1, GPT-4
- Event Timeline Generation
  - Creates structured timelines with LexChronos Event Schema (Timestamp, Event, Judge(s), Precedent(s))
  - Validated against 8 core judgment components[2]
- Judgement Text Generation
  - Generates synthetic Supreme Court-style text from timelines
- Final dataset - 2000 samples across 25 case categories



[1]https: //www.sci.gov.in/case-category/
[2]https://indiankanoon.org

Methodology

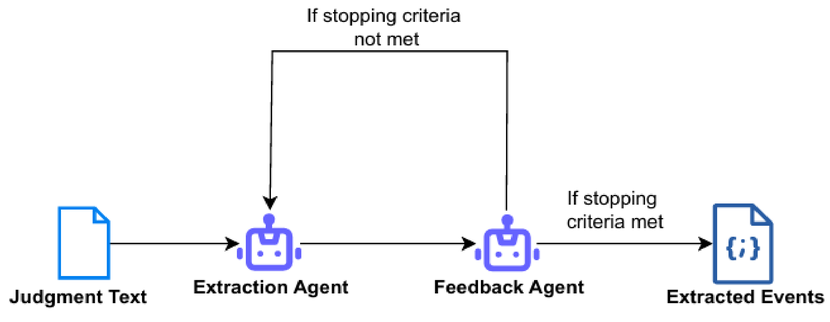
- Dual-Agent Architecture

Extraction Agent
builds timelines
↓
Feedback Agent
critiques
↓
Refinement loop
Until stopping criteria met

- Stopping Criteria
  - Patience limit – $\forall j \in \{i-2, i-1, i\} : S_j \leq S_{best}$
  - Tolerance threshold – $S_{i-2} = S_{i-1} = S_i$

  S_i is the confidence score of i^{th} iteration

- Downstream Task – Judgment Summarization
  - Compare two approaches
    - ✓ Unstructured: Judgment text ⟶ summary
    - ✓ Structured: Event timeline ⟶ summary
- Evaluation: GPT-4 based pairwise comparison using 8 legal quality criteria



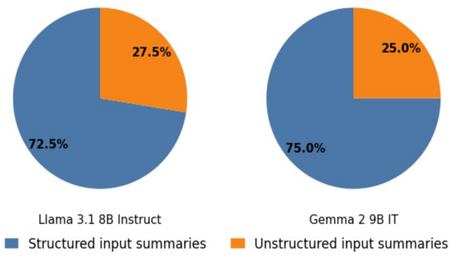
Evaluation

- LLMs used for Dual-Agent Architecture:
  - Instruct-tuned Llama 3.2 3B Instruct (Extraction Agent)
  - Gemma 2 2B IT (Feedback Agent)[1]
- Evaluated using BERT-based Precision, Recall, F1 score

Extraction Agent's performance without Feedback (%)

Performance of Dual Agent architecture (%)

- LLMs used for summarization:
  - Llama 3.1 8B Instruct[2]
  - Gemma 2 9B IT[3]



% of cases GPT-4 preferred for both models

[1]https://huggingface.co/google/gemma-2-2b-it
[2]https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct
[3]https://huggingface.co/google/gemma-2-9b-it

Conclusion

- LexChronos: Dual-Agent framework achieves 87.5% BERT-based F1 score, improves summarization
- Future Work:
  - Real-world expansion: Human annotated datasets; cover all case categories and multilingual judgments
  - Advanced Applications: Extend to Precedent Mapping, Argument Generation and Judgment Prediction