

## Leveraging Natural Language Processing for Legal Artificial Intelligence: A Longformer Approach for Taiwanese Legal Cases

**Authors :** Hsin Lee, Hsuan Lee

**Abstract :** Legal artificial intelligence (LegalAI) has been increasing applications within legal systems, propelled by advancements in natural language processing (NLP). Compared with general documents, legal case documents are typically long text sequences with intrinsic logical structures. Most existing language models have difficulty understanding the long-distance dependencies between different structures. Another unique challenge is that while the Judiciary of Taiwan has released legal judgments from various levels of courts over the years, there remains a significant obstacle in the lack of labeled datasets. This deficiency makes it difficult to train models with strong generalization capabilities, as well as accurately evaluate model performance. To date, models in Taiwan have yet to be specifically trained on judgment data. Given these challenges, this research proposes a Longformer-based pre-trained language model explicitly devised for retrieving similar judgments in Taiwanese legal documents. This model is trained on a self-constructed dataset, which this research has independently labeled to measure judgment similarities, thereby addressing a void left by the lack of an existing labeled dataset for Taiwanese judgments. This research adopts strategies such as early stopping and gradient clipping to prevent overfitting and manage gradient explosion, respectively, thereby enhancing the model's performance. The model in this research is evaluated using both the dataset and the Average Entropy of Offense-charged Clustering (AEOC) metric, which utilizes the notion of similar case scenarios within the same type of legal cases. Our experimental results illustrate our model's significant advancements in handling similarity comparisons within extensive legal judgments. By enabling more efficient retrieval and analysis of legal case documents, our model holds the potential to facilitate legal research, aid legal decision-making, and contribute to the further development of LegalAI in Taiwan.

**Keywords :** legal artificial intelligence, computation and language, language model, Taiwanese legal cases

**Conference Title :** ICAILLP 2023 : International Conference on Artificial Intelligence, Law and Legal Practice

**Conference Location :** Tokyo, Japan

**Conference Dates :** December 04-05, 2023