LEEC for Judicial Fairness: A Legal Element Extraction Dataset with Extensive Extra-Legal Labels

Zongyue Xue[1,2,#], **Huanghai Liu**[1,#], **Yiran Hu**[1,#,*], **Yuliang Qian**[1], **Yajing Wang**[1], **Kangle Kong**[1], **Chenlu Wang**[1], **Yun Liu**[1] and **Weixing Shen**[1,*]

[1]School of Law, Tsinghua University

[2]Yale Law School

{xuezy21, liuhh23, huyr21, qianyl23,wangyj20, kkl22}@mails.tsinghua.edu.cn,

{wangchenlu, liuyun89, wxshen}@mail.tsinghua.edu.cn

[*]Corresponding author [#]Equal contribution

Abstract

An extensive label system is pivotal to facilitating judicial fairness and social justice. Prior empirical research and our interview with legal professionals underscore the importance of extra-legal factors in criminal trials. To help identify sentencing biases and facilitate downstream applications, we introduce the **L**egal **E**lement **E**xtra**C**tion (LEEC) dataset comprising 15,919 judicial documents and 160 labels. This dataset was constructed through two main steps: **First**, designing the label system by legal experts based on prior empirical research which identified critical factors driving and processes generating sentencing outcomes in criminal cases; **Second**, employing legal knowledge to annotate judicial documents according to the label system and annotation guideline. LEEC represents the most extensive and domain-specific legal element extraction dataset for the Chinese legal system. Our experiments reveal that despite certain capabilities, both Document Event Extraction (DEE) models and Large Language Models (LLMs) face significant restrictions in legal element extraction tasks. Finally, our empirical analysis based on LEEC provides **evidence for judicial unfairness** in Chinese criminal sentencing and confirms the applicability of LEEC for future empirical research and other downstream applications. LEEC and related resources are available on https://github.com/THUlawtech/LEEC.

1 Introduction

Extracting key elements and their relations from judicial documents is valuable to facilitate further research and promote judicial fairness. Previous research has discerned that the practical application of the law is influenced significantly not only by legal factors but also by extra-legal ones. For instance, previous studies in Western jurisdictions have indicated that disparities in gender and ethnicity have significant impacts on criminal sentencing [Ulmer, 2012]. The divergence between the "law in books" and "law in action" can precipitate instances of judicial unfairness, thereby undermining social justice [Pound, 1910]. Empirical research on a large volume of precedents can facilitate researchers in validating these assumptions. However, previous legal datasets have almost exclusively incorporated legal factors. Consequently, researchers who endeavor to empirically investigate the influence of extra-legal factors are impeded by the lack of comprehensive datasets. With the help of the extensive label system constructed on the legal knowledge graph by our team of legal experts, the Legal Element ExtraCtion (LEEC) dataset aims to provide comprehensive element mentions, trigger words, and values manually annotated from large-scale judicial documents with high quality. This could facilitate the automatic extraction of elements, benefiting numerous LegalAI applications, such as Legal Judgement Prediction and Similar Case Retrieval, as well as the replication and innovation of empirical legal research.

Inspired by the success of general-domain element extraction [Guo *et al.*, 2020; Hogenboom *et al.*, 2011; Liao and Grishman, 2010], previous studies [Feng *et al.*, 2022; Shen *et al.*, 2020; Sierra and others, 2018] attempted to construct an element extraction system in the legal domain, leveraging both hand-crafted features and neural networks. For instance, LeCaRD [Ma *et al.*, 2021], the first Legal Case Retrieval Dataset in China, contains 107 query cases and 10,700 candidate cases selected from over 43,000 Chinese criminal judgments. LEVEN is a large-scale Chinese Legal event detection dataset [Yao *et al.*, 2022], with 8,116 legal documents and 150,977 human-annotated event mentions in 108 event types. At present, the existing datasets in China also include CAIL [Xiao *et al.*, 2019], Criminal [Ma *et al.*, 2021], LERD [Yao *et al.*, 2023], CJO[1], and PKULAW[2], etc. However, there are several main challenges in the existing work:

(1) Incomprehensive Label System. Existing label systems [Li *et al.*, 2023a; Li *et al.*, 2023b; Li *et al.*, 2023c; Richards *et al.*, 2016; Liu *et al.*, 2023; Li *et al.*, 2023d] of prior studies mainly lay emphasis on a limited scope of charge-oriented elements, which is far from enough. Existing studies predominantly focus on the legally prescribed fac-

[1]Please see the official website of *China Judgement Online*.

[2]Please see the official website of *PKULAW*.

tors in sentencing, overlooking extra-legal elements. However, a wealth of empirical research suggests that these elements, such as the defendant's and victim's age, gender, race/ethnicity, etc., may significantly influence trial and sentencing outcomes [Chen *et al.*, 2023; Doerner and Demuth, 2010; Richards *et al.*, 2016; Tran *et al.*, 2019; Ulmer, 2012]. The absence of these factors in the label system may compromise empirical studies on judicial fairness and the performance of downstream tasks.

(2) Lack of Domain Focus. An overwhelming majority of existing datasets [Nguyen *et al.*, 2016; Veyseh *et al.*, 2022] for element extraction mainly focus on the element or event extraction in the general domain. However, such datasets may not be well suited to applications in the legal domain. For example, Recidivist (*Leifan* in Chinese) and Previous Criminal Record (*Qianke* in Chinese) are closely connected yet distinct legal concepts in Chinese criminal law, which could be difficult to distinguish without adequate legal knowledge. Furthermore, judicial documents may depict different interpretations and perspectives on the same legal elements from various court participants, such as whether the defendant voluntarily surrendered, confessed, or pled guilty. This can cause confusion without legal knowledge. Therefore, existing datasets from general domains are hardly applicable for comprehensive analysis and tasks based on legal texts owing to their lack of adequate understanding of legal knowledge and contexts.

To provide a solid foundation for legal element extraction, LEEC makes the following improvements:

(1) Extensive Label System. Our team of legal experts not only expanded the coverage of legally prescribed factors that may significantly impact Chinese criminal trials, but also actively conducted interviews with legal professionals and utilized a comprehensive collection of empirical studies in Chinese contexts published in Chinese core legal journals and internationally. In this way, we were able to construct an extended and comprehensive label system in the legal domain. As a result, we systematically compiled both legal and extra-legal key labels that may have a substantial impact on Chinese judicial practice.

(2) Large Scale. LEEC is annotated based on the publicly available cases of both LEVEN and LeCaRD, with a total of 15,919 cases. Therefore, the high coverage of cases could help alleviate the problem of the limited number of cases in few-shot charges, leading to an increased ability to meet needs in real-world court settings. Besides, the annotation of LEEC could be combined with the previous annotation from LEVEN and LeCaRD, providing more comprehensive information to facilitate the analysis of judicial documents.

(3) Broader Application. The knowledge graph we developed for our annotation system encapsulates significant relationships among various elements. For instance, since it is common for a single Chinese judicial document to involve multiple defendants, crimes, and victims, our team of legal experts has linked defendant and victim characteristics to their respective individuals and affiliated crime characteristics to the corresponding offenses. This integration of crucial interrelations among labels enhances performance in various downstream applications, including the prediction of a specific defendant's crime and sentencing, and also expands LEEC's applicability in real-world court settings and future empirical research.

To validate the quality and applicability of LEEC, we implement various DEE models and LLMs and evaluate them on LEEC, which shows that these models exhibit insufficient accuracy in extracting elements from the LEEC dataset. Moreover, our preliminary empirical analysis based on LEEC uncovers several defendant demographic characteristics that significantly impact sentencing decisions, highlighting potential judicial unfairness in Chinese criminal trials. It is shown that this dataset has the potential to facilitate future empirical research and other downstream applications, enabling the identification and resolution of issues within the criminal justice system and contributing to social justice.

2 Interview

To dig deeper into the value of a labeled legal dataset and how to build it, we interviewed 11 Chinese legal researchers, 3 officers in two legal aid agencies in Jiangsu Province and Shanghai, and one lawyer, who occupies a leadership position in the Shanghai Bar Association.

Among the 11 legal researchers, none of them has ever utilized any legal resource research. This confirmed the severe limitations of these works as discussed in Introduction. However, they all speculated that at least some extra-legal factors, including demographic characteristics, may influence criminal sentencing. For example, one researcher witnessed a judge commenting on a female offender:"How could this happen to a little girl? She must have suffered a lot." Therefore, that researcher speculated that female offender may be treated more leniently sometimes as they may be perceived as vulnerable. Naturally, over 80% of them (9 in 11) welcome more high-quality empirical analyses to discover the potential sentencing disparities and contribute to judicial fairness.

Seven out of the nine researchers asked about the current situation of Chinese empirical legal studies believe that the proportion of these studies remains low. Several barriers were identified: 1) As Chinese legal education does not typically involve methods for empirical legal research, most legal researchers lack basic knowledge regarding such research; 2) All researchers identified significant barriers in processing and extracting labels from large judicial text data. For example, one researcher mentioned that a scholar may need to spend two to three months to extract labels for such analysis. Meanwhile, building algorithms like regular expression matching require considerable time and often fail to achieve high accuracy because of the complexity of judicial documents. Finally, all researchers confirm the value of a labeled judicial dataset with high quality based on criminal judicial documents, which could facilitate empirical legal research and discover valuable patterns regarding "law in action".

Three officers from legal aid agencies provide unique and valuable insights, highlighting potential judicial unfairness faced by socio-economically disadvantaged defendants. They unanimously affirm that the financial compensation provided to attorneys representing legal aid defendants is notably low. Concurrently, both agencies typically allocate legal aid cases

to attorneys who have actively expressed a willingness to handle such cases. This indicated that legal aid cases may be disproportionately handled by inexperienced attorneys who struggle to secure their own clients. One officer, when questioned about such concern, said: "We have regulations that prohibit treating legal aid cases as a means for training or gaining experience for attorneys as every case is important. However, we cannot control attorneys' minds." The interview with these officers underpins the necessity of conducting more empirical analyses regarding judicial unfairness.

The lawyer we interviewed claimed to have handled over 2,000 cases throughout her legal career. She confirmed that as she became a more sophisticated lawyer, she gradually ceased handling legal aid cases. By the time of the interview, she had neither read nor heard of anyone in the judicial practice citing or mentioning any empirical legal research or legal resource studies. This clearly highlights the dearth and limitations of these studies in China, which led to its lack of visibility in judicial practice. Based on her extensive experience, she believes that judges are influenced by a wide range of extra-legal factors, including the defendant's gender, ethnicity, age, employment status, etc. Judges often consider these factors comprehensively and may even emotionally resonate with the defendant in some cases. She welcomes insightful legal research regarding this issue.

Based on our interview, we confirm the value of a meticulously annotated legal dataset with a more extensive label system that covers both legal and extra-legal factors. Such a dataset could significantly facilitate empirical legal research and uncover the potential influence of extra-legal factors in Chinese criminal trials, thereby enhancing judicial fairness and social justice in China.

3 Data Analysis

This section summarizes the preprocessing of LEEC and discusses its detailed distribution in important dimensions.

3.1 Corpus and Preprocessing

Our selection of cases draws from the publicly available LEVEN [Yao *et al.*, 2022] and LeCaRD [Ma *et al.*, 2021] datasets. After deduplication, the number of total unique cases is reduced to 15,919 cases from a total of 17,352 cases. All documents within this collection represent criminal cases filed between 2001 and 2021. Among the complete dataset, we preserve 689 judicial documents as a non-published dataset for future evaluation, and publish the annotated data from 15,230 judicial documents. We compare LEEC with two types of datasets: (1) General-domain ED datasets. Compared with ACE2005 [Grishman *et al.*, 2005] and MAVEN [Wang *et al.*, 2020], LEEC's label system is specifically designed for legal texts. Moreover, the judicial documents of LEEC were annotated by law school students with a deep understanding of legal knowledge and concepts. (2) Legal-domain datasets. Compared with LEVEN and LeCaRD, our dataset offers a comprehensive expansion of the label system with a finer level of granularity, encapsulating both legal and extra-legal labels. Furthermore, LEEC encompasses independent annotations of distinct entities such as victims, defendants, and crimes. This methodological approach considerably enhances the dataset's precision and applicability, thereby contributing significantly to downstream applications and empirical research.

3.2 Data Distribution

Our dataset reveals the presence of multiple defendants in 33% of cases, multiple victims in 19% of cases, 49% of cases that contain at least one victim, and multiple crimes in 41% of cases. This underscores the necessity and effectiveness of introducing a sophisticated, domain-specific label system to handle such complexity. The distribution of cases is displayed in Table 1.[3] We measure the data quality by Kappa, with a value of 0.71. Specifically, 3,990 documents were annotated with mentions of elements, trigger words, and values for each element, while the remaining 11,240 documents underwent extended annotation without element mentions and trigger words. Detailed annotation process of the dataset can be found in Appendix B, and the Annotation Guideline is available on our *GitHub repository*.

Case Characteristics	**Case Distribution**					
Defendant number	1	2	3	4	5	>5
	10162	2167	1122	600	378	801
Victim number	0	1	2	3	4	>4
(Missing Value=1297)	6483	4510	919	547	320	1154
Crime number	1	2	3	4	5	>5
	8936	2836	1258	705	428	1067

Table 1: Case distribution on defendants, victims, and crimes.

4 Label System

This section encompasses the compilation of the extensive labels and the crucial relationships among them.

4.1 Label Compilation

Our team of legal experts, led by law professors, incorporated a wide range of legal and extra-legal elements to build a comprehensive knowledge graph covering key elements within the Chinese legal domain. First, our team of legal experts compiled the crucial legal circumstances and factors stipulated by Chinese criminal law and judicial interpretations[4], such as whether the defendant confessed, pled guilty, voluntarily surrendered, conducted a justifiable defense, etc. Furthermore, It has widely been revealed that extra-legal factors may significantly impact judicial practice. Therefore, we utilized elements and theories developed and validated by empirical legal research to comprehensively capture the important factors in Chinese criminal trials.

[3] A minor fraction of the annotated victim number contains missing values. This is attributable to instances where the precise number of victims cannot be ascertained due to the insufficient information provided within the judicial document. For a detailed explanation concerning the occurrence of missing values within the annotation, please see the Annotation section in Appendix B.

[4] Judicial interpretations of the Supreme People's Court have binding effects for courts of lower levels in China.

Our team of legal experts systematically compiled 178 quantitative legal studies from 2018 to 2022, published across 22 journals in Chinese, and listed in the China Legal Science Citation Index (CLSCI). The CLSCI is curated by the Law Institute of China Law Society (LICLS), which provides a list of core legal journals in China.[5] As published judicial documents are among the most commonly used data for empirical analyses on sentencing factors, the labels, theories, and results of these studies serve as valuable sources of potentially salient factors influencing judicial decisions. In addition, we drew upon a wide range of empirical legal studies published in SSCI journals, particularly those in Chinese contexts. Our team meticulously collected the core theories and labels used in these studies and incorporated them into our legal system.

For instance, the Group Threat Theory suggests that when majority groups feel threatened by minority populations, criminal justice systems may treat racial or ethnic minorities adversely [Ulmer and Johnson, 2004]. This theory has been validated and developed in the Chinese context by prior empirical research, which found that minorities perceived as "problem minorities" that might disrupt public order may face discrimination in Chinese criminal cases [Hou and Truex, 2022]. Therefore, we included the ethnic status of offenders in our knowledge graph. Moreover, the Focal Concerns Theory highlights four crucial factors influencing sentencing decisions: the defendant's culpability, redeemability, the risk posed to the community, and pragmatic considerations such as the court's workload [Ulmer *et al.*, 2023]. Research in Chinese contexts has shown that, in line with the Focal Concern Theory, the defendant's being a rural-to-urban migrant – measured by his or her registered permanent residence (*Hukou*) – significantly impacts sentencing outcomes [Jiang and Kuang, 2018]. As a result, we also included the *Hukou* information of defendants as an important element in our knowledge graph. Following this scheme, we effectively constructed an extended, multi-level knowledge graph to cover 160 important elements – both legal and extra-legal – in Chinese criminal sentencing. The elements in the knowledge graph are divided into four main categories: defendant characteristics, victim characteristics, case characteristics, and crime characteristics. [6]

4.2 Relation Construction

We integrated the relations among elements into the knowledge graph as they have a significant influence on judicial decisions. This is particularly important in documents involving multiple crimes, offenders, victims, or defenders. The circumstances and characteristics of one defendant or victim may differ from those of the others. Therefore, all characteristics pertaining to a victim or defendant are linked directly to each individual. Besides, as a defendant may have up to two defenders in Chinese criminal trials, the defender characteristics are connected to each individual defender of a specific defendant. Furthermore, as each defendant could be sentenced for multiple crimes in Chinese judicial documents, all characteristics of a crime are connected to the specific crime committed by a particular defendant.[7] The elements within the knowledge graph are depicted in Appendix D. For details regarding the annotation process based on this label system, please see Appendix B.

5 Experiments

In this section, we conduct experiments on representative Document-level Event Extraction (DEE) models, as well as the general and legal Large Language Models (LLMs). Based on the experiment results, we discuss the challenges we identified in legal element extraction.

5.1 DEE Models

Experiment Settings

For traditional DEE models, we select 21 labels about defendants' characteristics and sentencing in LEEC label system, which are shown in Table A1. We use the LEEC dataset with triggers as the experimental data. For each label, the annotation content or trigger word corresponding to the original text is regarded as the entity.

The dataset is split into the train, dev, and test sets at a ratio of 8:1:1. We use the same vocabulary as [Zheng *et al.*, 2019] and randomly initialize all the embeddings where dh=768 and dl=32. We employ the Adam optimizer with the learning rate 5e-4 and the batch size is 16. All models are trained for 100 epochs and the checkpoints with the best F1 scores on the dev set are selected for evaluation on the test set.

Baselines and Metrics

Baselines. We introduce the following models as baselines: 1) DCFEE [Yang *et al.*, 2018] is the first model that introduced Distance Supervision (DS) into the DEE task. there are two variants included: DCFEE-O only extracts one event record from one document while DCFEE-M tries to extract multiple possible event records. 2) Doc2EDAG [Zheng *et al.*, 2019] is an end-to-end DEE model that constructs event records in an auto-regressive way by generating entity-based Directed Acyclic Graphs (DAGs). 3) GreedyDec is a baseline proposed in Doc2EDAG [Zheng *et al.*, 2019] which fills one event table greedily. 4) PTPCG [Zhu *et al.*, 2021] is a lightweight model for end-to-end DEE task based on pruned complete graphs with pseudo triggers.

Metrics. We follow the same evaluation setting in the previous studies [Zheng *et al.*, 2019; Zhu *et al.*, 2021; Peng *et al.*, 2023]. For each prediction record, we select a

[5]For details, please visit *the official website of LICLS*.

[6]Some elements, such as court name, judge name, case title, and year of judgment, can be easily and accurately extracted from Chinese criminal verdicts using keyword identification or regular expression matching. For useful references, please visit *this GitHub program*. These elements do not necessarily require manual annotation, and thus, are not included in the knowledge graph of this study.

[7]Specifically, it is noteworthy that in Chinese criminal cases where an individual defendant committed multiple crimes, the court typically adjudicates a sentence for each individual crime, followed by an overall aggregated sentence. This final sentence, which is usually subject to a certain degree of the judge's discretion, may not necessarily align with the sum of the individual sentences. Consequently, in our knowledge graph, we deliberately included both the sentencing elements, linked to each distinct crime of a specific defendant, and the final, aggregated sentence, linked to each defendant.

golden record by matching records with the same defendant name and the most shared arguments, and calculate the F1 score by comparing the parameters between them.

Results

Table 2 shows the experimental results. we have the following observations: 1) Some baselines cannot converge well on LEEC, such as Doc2EDAG and similar structured Greedy-Dec. One reason is that legal documents are longer and the arguments are more dispersed, which is not conducive to Doc2EDAG's sequential path extension method for reasoning. 2) The DEE models can only extract the words in the document, for example, for the sentence of imprisonment, it can only extract "fixed-term imprisonment", indirectly deriving yes or no. This two-step approach does not sufficiently meet the requirements in real-world applications. 3) We selected 21 relatively simple elements for the DEE task. However, the LEEC label system includes labels that are either sparse[8] or in need of complex judgment and high-level reasoning capabilities[9]. Developing DEE models to extract these labels may present a greater challenge.

Model	Precision	Recall	F1 score
DCFEE-O	62.98	**83.13**	71.67
DCFEE-M	59.56	82.54	69.19
Greedy-Dec	80.27	56.00	65.97
Doc2EDAG	42.73	70.01	53.07
PTPCG	**86.82**	77.99	**82.17**

Table 2: Overall performance of DEE models

5.2 Large Language Models

Experiment Settings

We selected some advanced general LLMs and legal LLMs for our experiments.

(1) General LLMs. I. GPT-3.5 [OpenAI, 2023]: An advanced LLM by OpenAI that excels in understanding and generating text. II. ChatGLM3 [Zeng *et al.*, 2022]: A bilingual open-source LLM for the general domain. We used GLM3-6B-32K as our baseline for its larger context length. III. LLaMA3 [Meta, 2024]: Meta's SOTA open-source LLM. For Chinese documents, we used Llama3-Chinese-8B-Instruct [Community, 2024].

(2) Legal LLMs. I. Lawyer-LLaMA [Huang *et al.*, 2023]: A Chinese legal LLM based on LLaMA [Touvron *et al.*, 2023]. The model without a retrieval module was used in this experiment. II. Tongyi Farui [Aliyun, 2024]: A legal LLM launched by Aliyun, capable of performing various legal tasks such as answering legal questions, assisting in case analysis, and generating legal documents.

[8]Some elements do not occur frequently in Chinese criminal trials, such as the Not_guilty element or the Excluding_evidence_decision element. Extracting such labels may be more difficult for DEE models as they rely heavily on the quantity of effective training data.

[9]For example, one such label is the JOINT_CRIME element. The corresponding part of the Annotation Guideline is available in Appendix B.2.

LLM models make up for some inherent limitations of DEE. Compared to traditional DEE models that can only extract triggers of labels, LLMs possess more advanced reasoning abilities. Consequently, the outputs of LLMs could be closer to human annotation results.[10] Therefore, we provide prompts to LLMs asking them to produce final annotation results in the format of LEEC json data. The dataset used is the same as in the DEE task. For evaluation, we compare the accuracy of the predicted and annotated results of humans.

Results

Table 3 shows the experiment statistics and the accuracy of LLMs. Table 4 provides an example of how LLMs extract judgment documents. Although compared with DEE models, LLMs make up for some of the inherent shortcomings as discussed before, there are still many challenges when using LLMs for legal element extraction as shown in Tables 3 and 4:

Model	Max length	Truncation	Unformatted output	Accuracy
GPT-3.5	16385	12.71%	0.13%	0.7070
GLM3-32k	32768	1.48%	4.49%	0.4920
LLaMA3	8192	20.70%	7.17%	0.6392
Lawyer-LLaMA	2048	80.20%	*	*
Tongyi Farui	12000	12.41%	0.08%	0.6456

Table 3: Overall performance of LLMs. For each model, Max length is the maximum context length, Truncation is the proportion of truncated data. Unformatted output is the proportion of the output not in the given format. Since most responses in Lawyer-LLaMA outputs don't follow the correct format, its result isn't included.

(1) Truncated Input and Unformatted Output. One notable discovery is that the limited context window of current LLMs leads to significant problems when analyzing the often lengthy judicial documents. Moreover, despite the instructions provided, it was found in all LLMs evaluated that in some cases the output could not be generated in the expected format. These issues include repeated outputs, incomplete responses, irrelevant answers, and other similar circumstances. Lawyer-LLaMA is the one that deviates the most from the correct format. This may be due to the small Chinese corpus, the limited context window, and a degraded general understanding ability after fine-tuning in the legal domain.

(2) Incomplete Defendant Coverage. LLMs may overlook certain defendants in element extraction. Both GPT and GLM only extracted one of the three defendants. Legal LLMs have an advantage here. While Lawyer-LLaMA did not provide specific extraction details, it did note the presence of two other defendants. This is important for legal research as the absence of defendants can significantly impact the nature and analysis of the case.

[10]For example, LLMs can directly judge whether the sentencing of "Imprisonment" is delivered by the court with 0/1, and convert the sentencing length into numeric months, rather than only extracting the trigger as traditional DEE models do.

Prompt	**Please extract the defendants' information from the following judgment documents. If a label doesn't exist or can't be extracted, fill in the empty string "". For 0/1 labels such as "Control", if the judgment is yes, then fill in 1, otherwise fill in 0. For labels like [ControlTime (months)], you just need to fill in the number. Output example (in the case of one defendant, if there are multiple defendants, please select Defendant 1, Defendant 2...) :** "Defendant 1" : "Name" : "", "Gender" : "", "Birth" : "", "Place" : "", "Control" : "", "ControlTime (months)" : "", "Detention" : "", "DetentionTime (month)" : "", "Imprisonment" : "", "ImprisonmentTime (month)" : "", "PoliticalRights" : "", "PoliticalRightsTime (month/life)" : "", "Fine" : "", "FineNum (yuan)" : "", "PartofProperty" : "", "PartofPropertyNum (yuan)" : "", "AllProperty" : "", "AllPropertyNum (yuan)" : "", "EcoCompensation" : "", "EcoCompensationNum (yuan)" : ""
Input	"Judgment documents: ....Defendant Xie, male, born on September 27, 1992, Han nationality, farmer, civilian, junior high school education, registered in Gaoyang, Hebei Province ... Xie collided with Zhang 1, who was standing on the road, and then collided with a small van of Hebei Fxxxxx parked on the road, causing damage to both vehicles and the death of Zhang 1. After the accident, Ge called Fu to drive to the scene of the accident and pull Xie away to help him escape. Xie takes full responsibility for this accident, and Zhang 1 and Zhang 2 have no responsibility. **After the accident, the relatives of the defendant Xie took the initiative to compensate the economic loss of the victim Zhang 1 family of 460,000 yuan, and obtained the understanding of the victim's family**.... The public prosecution organ provided the corresponding evidence that the **defendant Xie**'s behavior violated the provisions of Article 133 of the Criminal Law of the People's Republic of China, and he should be investigated for the crime of causing traffic accidents. The actions of the defendants, **Ge and Fu**, violated the provisions of Article 310 of the Criminal Law of the People's Republic of China and should be investigated for criminal responsibility for harboring. The defendant Xie fled and surrendered himself, so it is recommended that he be sentenced to between three and five years in prison. The defendant Ge, Fu surrendered the plot, please the court on the defendant Ge, Fu's judgment according to law. The defendants Xie, Ge and Fu have no objection to the criminal facts, charges and sentencing recommendations charged by the public prosecution. ... By the provisions of Article 133, Article 301, Article 67, paragraph 1, and Article 72, paragraph 1 of the Criminal Law of the People's Republic of China, the verdict is as follows: 1. **The defendant Xie, who committed the crime of causing a traffic accident, is sentenced to fixed-term imprisonment of three years,** with a suspension of four years...."

LEEC	DEE Model	GPT-3.5	GLM3-32k	Lawyer-LLaMA

{**"Defendant 1"** : {
"Name" : "Xie",
"Gender" : "Male",
"Birth" : "1992/9/27",
"Nation" : "Han",
"Place" : "Gaoyang, Hebei province",
"Control" : "0",
"ControlTime (months)" : "",
"Detention" : "0",
"DetentionTime (month)" : "",
"Imprisonment" : "1",
"ImprisonmentTime (month)" : "36",
"PoliticalRights" : "0",
"PoliticalRightsTime (month/life)" : "",
"Fine" : "0",
"FineNum (yuan)" : "",
"PartofProperty" : "0",
"PartofPropertyNum (yuan)" : "",
"AllProperty" : "0",
"AllPropertyNum (yuan)" : "",
"EcoCompensation" : "0",
"EcoCompensationNum (yuan)" : "",
}
"Defendant 2" : {
"Name" : "Ge",
...}
"Defendant 3" : {
"Name" : "Fu",
...}}
	{**"Defendant 1"** : {
	"Name" : "Xie",
	"Gender" : "Male",
	"Birth" : "September 27, 1992",
	"Nation" : "Han",
	"Place" : "Gaoyang, Hebei Province",
	"Control" : "",
	"ControlTime (months)" : "",
	"Detention" : "",
	"DetentionTime (month)" : "",
	"I..." : "fixed-term imprisonment",
	"I...Time (month)" : "three years",
	"PoliticalRights" : "",
	"PoliticalRightsTime (month/life)" : "",
	"Fine" : "",
	"FineNum (yuan)" : "",
	"PartofProperty" : "",
	"PartofPropertyNum (yuan)" : "",
	"AllProperty" : "",
	"AllPropertyNum (yuan)" : "",
	"EcoCompensation" : "",
	"EcoCompensationNum (yuan)" : "",
	}
	"Defendant 2" : {
	"Name" : "Ge",
	...}
	"Defendant 3" : {
	"Name" : "Fu",
	...}}
		"Defendant 1" : {
		"Name" : "Xie",
		"Gender" : "Male",
		"Birth" : "1992/9/27",
		"Nation" : "Han",
		"Place" : "Gaoyang, Hebei province",
		"Control" : "0",
		"ControlTime (months)" : "",
		"Detention" : "",
		"DetentionTime (month)" : "",
		"Imprisonment" : "1",
		"ImprisonmentTime (month)" : "36",
		"PoliticalRights" : "",
		"PoliticalRightsTime (month/life)" : "",
		"Fine" : "",
		"FineNum (yuan)" : "",
		"PartofProperty" : "",
		"PartofPropertyNum (yuan)" : "",
		"AllProperty" : "",
		"AllPropertyNum (yuan)" : "",
		"EcoCompensation" : "1",
		"EcoCom...Num (yuan)" : "460000"
		}}
			"Defendant 1" : {
			"Name" : "Xie",
			"Gender" : "Male",
			"Birth" : "1992/9/27",
			"Nation" : "Han",
			"Place" : "Gaoyang, Hebei province",
			"Control" : "0",
			"ControlTime (months)" : "",
			"Detention" : "0",
			"DetentionTime (month)" : "",
			"Imprisonment" : "0",
			"ImprisonmentTime (month)" : "",
			"PoliticalRights" : "0",
			"PoliticalRightsTime (month/life)" : "",
			"Fine" : "0",
			"FineNum (yuan)" : "",
			"PartofProperty" : "0",
			"PartofPropertyNum (yuan)" : "",
			"AllProperty" : "0",
			"AllPropertyNum (yuan)" : "",
			"EcoCompensation" : "0",
			"EcoCompensationNum (yuan)" : ""
			}
				According to your description, I have extracted the information of the **defendant Xie** for you. The defendant Xie, male, Han nationality, farmer, civilian, junior high school education, registered in Gaoyang County, Hebei Province, now lives in Gaoyang County. ... Defendant Xie's behavior violated the provisions of Article 133 of the Criminal Law, and he should be investigated for criminal responsibility for the crime of causing a traffic accident. Notably, defendant Xie surrendered to the plot, **compensated the victim's economic losses, and requested the court for a lighter punishment.**In addition, you mentioned that the **defendant Ge, Fu** also surrendered, please the court on the defendant Ge, Fu's judgment by the law. Finally, it is recommended that you provide more facts and evidence to better analyze the situation.

Table 4: An example of extraction. For LLMs, the input message is *[prompt, input]*, and then it is truncated according to the context length of each model. *LEEC* is the manual annotation in LEEC dataset. *DEE Model* is the golden record based on LEEC for traditional DEE models, *GPT-3.5, GLM3-32K and Lawyer-LLaMA* represents the output of LLMs. Due to page limitations, We illustrate only representative outputs, omitting some unimportant information and the outputs of LLaMA3 and Tongyi Farui.

(3) Erroneous Legal Reasoning. LLMs show the ability to reason according to annotation requirements, but this ability does not yet meet application requirements. LLMs excel at extracting basic labels such as "Name" and "Birth", but there are still challenges in extracting labels that require the understanding and knowledge of legal concepts. For instance, GPT correctly identified Imprisonment labels but mistakenly categorized the voluntary compensation by Xie's family as "FineNum". GLM identified *0* for all sentencing labels. Lawyer-LLaMA observed that Defendant Xie had compensated the victim's family for economic losses, but it failed to distinguish it from the judge's ruling.

Overall, current LLMs cannot achieve the accuracy required for legal empirical research without introducing substantial bias and errors.[11] The LEEC dataset annotated by legal experts can serve as a benchmark to identify existing challenges, evaluate extraction accuracy, and function as the training data for LLMs so that they can be more effectively applied to empirical research and other downstream applications in the future.

[11]It is shown in Table 3 that the highest accuracy of LLMs is a little over 70%, which is far from satisfactory for legal research.

6 Empirical Analysis

Utilizing our LEEC dataset, we undertake an empirical legal analysis with a threefold objective: firstly, to verify the suitability and applicability of the LEEC dataset for empirical analysis; secondly, to determine whether it yields patterns that are reasonable or coherent with related findings from prior empirical studies, thereby attesting to the quality and robustness of this dataset; and thirdly, to provide preliminary evidence regarding the judicial (in)equality within Chinese criminal trials.

Stratification and inequality in criminal sentencing have garnered considerable attention from scholars across the social sciences [Ulmer, 2012]. To investigate such issues in Chinese contexts, we aim to explore the sentencing impact of defendant demographic characteristics, including gender, ethnicity, and age, based on labels in the "Demographic Characteristics" section (refer to Figure A1) within the LEEC dataset. The dependent variable is the length of limited imprisonment. In line with the predominant approach in em-

pirical legal research for investigating causal effects in sentencing [Peng and Cheng, 2022; Liu *et al.*, 2021; Ulmer, 2012], we employ the Ordinary Least Squares (OLS) regression model as our methodological tool. For details regarding our regression model, please refer to Appendix C.

Figure 1 presents the forest plot displaying the estimation coefficients of the defendant demographic variables and their respective 95% confidence intervals. Several interesting results emerge. Firstly, we found that female defendants are likely to be sentenced more leniently in our dataset. This finding is consistent with a series of studies in Western jurisdictions [Embry and Lyons Jr, 2012; Fernando Rodriguez *et al.*, 2006]. These studies provide evidence supporting the chivalry hypothesis, suggesting that due to gender patriarchy, women may be perceived as vulnerable, less blameworthy, and in need of extra protection in criminal sentencing. Secondly, we discovered that as defendants age, their sentences may become more lenient, aligning with prior studies in the U.S. [Ryon *et al.*, 2017; Steffensmeier *et al.*, 1995]. Thirdly, unemployed defendants tend to receive harsher sentences, illustrating a concern for Chinese courts to maintain social stability [Trevaskes *et al.*, 2014]. Additionally, those who are less educated are likely to be sentenced more leniently, possibly because the courts may perceive these defendants as disadvantaged and less blameworthy for their wrongdoings.

Overall, our empirical analysis reveals multiple extra-legal factors that may contribute to sentencing disparities and judicial unfairness in Chinese criminal trials. However, it should be noted that our investigation is exploratory and preliminary. Deeply investigating the impact of each of these defendant demographic characteristics may require conducting individual studies or even a series of studies in the social sciences, with detailed theoretical construction, robustness tests, further analyses, etc., far beyond the scope of this paper. Nevertheless, our results provide reasonable and interesting findings that validate the applicability and quality of the LEEC dataset, while offering insightful direction and evidence for future researchers' deeper and broader investigations.

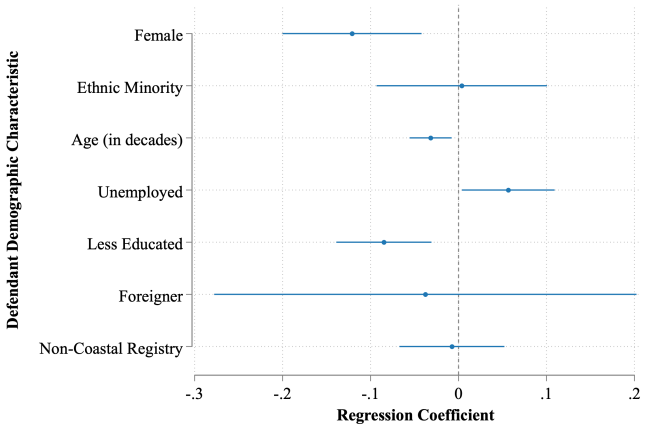

Figure 1: The impact of defendant demographic characteristics on the length of fixed-term imprisonment.

7 Discussion and Conclusion

In this study, we introduce LEEC, a unique dataset designed for legal elements extraction in Chinese legal system. LEEC stands out as its label system is enriched with both legal and extra-legal labels, integrating crucial legal knowledge drawn from Chinese law, empirical legal studies, our interview, and legal experts' understanding of Chinese legal contexts and practices. Each of the 15,919 cases in the dataset is annotated by law school students. Experimental results underline the challenges for traditional models and LLMs in element extraction and the biases in Chinese sentencing, signifying areas of focus for future research.

This study has several limitations that we hope will be addressed by future research: 1) Potential Selection Bias: About 75% of all judicial verdicts in China have been disclosed for cases not processed through mediation in recent years [Tang and Liu, 2019]. Consequently, cases in LEEC may not fully represent those in actual courts due to potential selection bias. Future studies should consider this when utilizing LEEC; 2) Context-Specificity: The data, label system, and annotation methods are inherently rooted in Chinese contexts. However, although laws vary across jurisdictions, the underlying logic, concepts, and biases of the criminal system have much in common in modern societies. Therefore, we believe that the majority of the labels and methods of data curation could serve as useful references for future legal resource research in other jurisdictions; 3) Bias of LLMs: We identified several types of biases in Chinese real criminal sentencing using LLEC, which are typically embedded in social structure, stratification, and ideologies. Whether the biases we identified in LLMs are randomly distributed or similar to those in the real world requires further examination to ensure that AI could better serve judicial fairness and social good.

Ethical Statement

To support downstream applications, the knowledge graph and annotation in this study exhibit high granularity. It is of paramount importance that users of LEEC exercise due caution. We strongly oppose the use of LEEC for any purposes that could lead to discrimination or violations of the rule of law. The personal information included in the published judicial documents was collected and processed in strict compliance with Chinese law. Any future utilization of LEEC must also adhere to applicable laws and commit to responsible, ethical handling of the data.

Acknowledgments

We extend our deep gratitude to Huaiyu Hu, Senyu Li, Bifan Zhao, Jin LI, Shuchen Tang, Yige Fan, Zhiwei Zhang, Caixuan Huang, Dikun Zhu, Kun Liu, and other anonymous interviewees for their insightful opinions and invaluable help. We also wish to express our sincere appreciation to Zexia Yang, Huihan Li, and Zhijie He for their dedicated assistance. This work is supported by the National Key Research and Development Program of China (No.2022YFC3301504).

References

[Aliyun, 2024] Aliyun, 2024. https://tongyi.aliyun.com/farui/chat.

[Chen *et al.*, 2023] Andong Chen, Feng Yao, Xinyan Zhao, Yating Zhang, Changlong Sun, Yun Liu, and Weixing Shen. Equals: A real-world dataset for legal question answering via reading chinese laws. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pages 71–80, 2023.

[Community, 2024] Llama Chinese Community. Llama3-chinese-8b-instruct, 2024. https://github.com/LlamaFamily/Llama-Chinese.

[Doerner and Demuth, 2010] Jill K Doerner and Stephen Demuth. The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in us federal courts. *Justice Quarterly*, 27(1):1–27, 2010.

[Embry and Lyons Jr, 2012] Randa Embry and Phillip M Lyons Jr. Sex-based sentencing: Sentencing discrepancies between male and female sex offenders. *Feminist Criminology*, 7(2):146–162, 2012.

[Feng *et al.*, 2022] Yi Feng, Chuanyi Li, and Vincent Ng. Legal judgment prediction via event extraction with constraints. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 648–664, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[Fernando Rodriguez *et al.*, 2006] S Fernando Rodriguez, Theodore R Curry, and Gang Lee. Gender differences in criminal sentencing: Do effects vary across violent, property, and drug offenses? *Social Science Quarterly*, 87(2):318–339, 2006.

[Grishman *et al.*, 2005] Ralph Grishman, David Westbrook, and Adam Meyers. Nyu's english ace 2005 system description. *ACE*, 5:2, 2005.

[Guo *et al.*, 2020] Kaihao Guo, Tianpei Jiang, and Haipeng Zhang. Knowledge graph enhanced event extraction in financial documents. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1322–1329. IEEE, 2020.

[Hogenboom *et al.*, 2011] Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska De Jong. An overview of event extraction from text. *DeRiVE@ ISWC*, pages 48–57, 2011.

[Hou and Truex, 2022] Yue Hou and Rory Truex. Ethnic discrimination in criminal sentencing in china. *The Journal of Politics*, 84(4):2294–2299, 2022.

[Huang *et al.*, 2023] Quzhe Huang, Mingxu Tao, Zhenwei An, Chen Zhang, Cong Jiang, Zhibin Chen, Zirui Wu, and Yansong Feng. Lawyer llama technical report. *arXiv preprint arXiv:2305.15062*, 2023.

[Jiang and Kuang, 2018] Jize Jiang and Kai Kuang. Hukou status and sentencing in the wake of internal migration: The penalty effect of being rural-to-urban migrants in china. *Law & Policy*, 40(2):196–215, 2018.

[Li *et al.*, 2023a] Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Yueyue Wu, Yiqun Liu, Chong Chen, and Qi Tian. Sailer: Structure-aware pre-trained language model for legal case retrieval, 2023.

[Li *et al.*, 2023b] Haitao Li, Weihang Su, Changyue Wang, Yueyue Wu, Qingyao Ai, and Yiqun Liu. Thuir@coliee 2023: Incorporating structural knowledge into pre-trained language models for legal case retrieval, 2023.

[Li *et al.*, 2023c] Haitao Li, Changyue Wang, Weihang Su, Yueyue Wu, Qingyao Ai, and Yiqun Liu. Thuir@coliee 2023: More parameters and legal knowledge for legal case entailment, 2023.

[Li *et al.*, 2023d] Qingquan Li, Yiran Hu, Feng Yao, Chaojun Xiao, Zhiyuan Liu, Maosong Sun, and Weixing Shen. Muser: A multi-view similar case retrieval dataset. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 5336–5340, 2023.

[Liao and Grishman, 2010] Shasha Liao and Ralph Grishman. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 789–797, 2010.

[Liu *et al.*, 2021] Lin Liu, Christy A Visher, and Daniel J O'Connell. Strain during reentry: A test of general strain theory using a sample of adult former prisoners. *The Prison Journal*, 101(4):420–442, 2021.

[Liu *et al.*, 2023] Bulou Liu, Yiran Hu, Yueyue Wu, Yiqun Liu, Fan Zhang, Chenliang Li, Min Zhang, Shaoping Ma, and Weixing Shen. Investigating conversational agent action in legal case retrieval. In *European Conference on Information Retrieval*, pages 622–635. Springer, 2023.

[Ma *et al.*, 2021] Yixiao Ma, Yunqiu Shao, Yueyue Wu, Yiqun Liu, Ruizhe Zhang, Min Zhang, and Shaoping Ma. Lecard: A legal case retrieval dataset for chinese law system. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2342–2348, New York, NY, USA, 2021. Association for Computing Machinery.

[Meta, 2024] Meta, 2024. https://llama.meta.com/docs/model-cards-and-prompt-formats/meta-llama-3.

[Nguyen *et al.*, 2016] Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besançon. A dataset for open event extraction in english. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1939–1943, 2016.

[OpenAI, 2023] OpenAI, 2023. https://openai.com.

[Peng and Cheng, 2022] Yali Peng and Jinhua Cheng. Ethnic disparity in chinese theft sentencing. *China Review*, 22(3):47–71, 2022.

[Peng *et al.*, 2023] Hao Peng, Xiaozhi Wang, Feng Yao, Kaisheng Zeng, Lei Hou, Juanzi Li, Zhiyuan Liu, and Weixing Shen. The devil is in the details: On the pitfalls of

event extraction evaluation. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9206–9227, Toronto, Canada, July 2023. Association for Computational Linguistics.

[Pound, 1910] Roscoe Pound. Law in books and law in action. *American Law Review*, 44:12, 1910.

[Richards *et al.*, 2016] Tara N Richards, Wesley G Jennings, M Dwayne Smith, Christine S Sellers, Sondra J Fogel, and Beth Bjerregaard. Explaining the "female victim effect" in capital punishment: An examination of victim sex–specific models of juror sentence decision-making. *Crime & Delinquency*, 62(7):875–898, 2016.

[Ryon *et al.*, 2017] Stephanie Bontrager Ryon, Ted Chiricos, Sonja E Siennick, Kelle Barrick, and William Bales. Sentencing in light of collateral consequences: Does age matter? *Journal of Criminal Justice*, 53:1–11, 2017.

[Shen *et al.*, 2020] Shirong Shen, Guilin Qi, Zhen Li, Sheng Bi, and Lusheng Wang. Hierarchical Chinese legal event extraction via pedal attention mechanism. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 100–113, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[Sierra and others, 2018] G Sierra et al. Event extraction from legal documents in spanish. In *1st Workshop on Language Resources and Technologies for the Legal Knowledge Graph*, page 36, 2018.

[Steffensmeier *et al.*, 1995] Darrell Steffensmeier, John Kramer, and Jeffery Ulmer. Age differences in sentencing. *Justice Quarterly*, 12(3):583–602, 1995.

[Tang and Liu, 2019] Yingmao Tang and John Zhuang Liu. Mass publicity of chinese court decisions. *China Review*, 19(2):15–40, 2019.

[Touvron *et al.*, 2023] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[Tran *et al.*, 2019] Vu Tran, Minh Le Nguyen, and Ken Satoh. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, pages 275–282, 2019.

[Trevaskes *et al.*, 2014] Susan Trevaskes, Elisa Nesossi, Flora Sapio, and Sarah Biddulph. *The politics of law and stability in China*. Edward Elgar Publishing, 2014.

[Ulmer and Johnson, 2004] Jeffery T Ulmer and Brian Johnson. Sentencing in context: A multilevel analysis. *Criminology*, 42(1):137–178, 2004.

[Ulmer *et al.*, 2023] Jeffery T Ulmer, Eric Silver, and Lily S Hanrath. Back to basics: A critical examination of the focal concerns framework from the perspective of judges. *Justice Quarterly*, 40(6):813–836, 2023.

[Ulmer, 2012] Jeffery T Ulmer. Recent developments and new directions in sentencing research. *Justice Quarterly*, 29(1):1–40, 2012.

[Veyseh *et al.*, 2022] Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Huu Nguyen. Mee: A novel multilingual event extraction dataset. *arXiv preprint arXiv:2211.05955*, 2022.

[Wang *et al.*, 2020] Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. MAVEN: A massive general domain event detection dataset. In *Proceedings of EMNLP*, pages 1652–1671, 2020.

[Xiao *et al.*, 2019] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Tianyang Zhang, Xianpei Han, Zhen Hu, Heng Wang, and Jianfeng Xu. CAIL2019-SCM: A dataset of similar case matching in legal domain. *CoRR*, abs/1911.08962, 2019.

[Yang *et al.*, 2018] Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. Dcfee: A document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55, 2018.

[Yao *et al.*, 2022] Feng Yao, Chaojun Xiao, Xiaozhi Wang, Zhiyuan Liu, Lei Hou, Cunchao Tu, Juanzi Li, Yun Liu, Weixing Shen, and Maosong Sun. LEVEN: A large-scale chinese legal event detection dataset. In *Findings of ACL*, pages 183–201, 2022.

[Yao *et al.*, 2023] Feng Yao, Jingyuan Zhang, Yating Zhang, Xiaozhong Liu, Changlong Sun, Yun Liu, and Weixing Shen. Unsupervised legal evidence retrieval via contrastive learning with approximate aggregated positive. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4783–4791, 2023.

[Zeng *et al.*, 2022] Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*, 2022.

[Zheng *et al.*, 2019] Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. Doc2edag: An end-to-end document-level framework for chinese financial event extraction. *arXiv preprint arXiv:1904.07535*, 2019.

[Zhu *et al.*, 2021] Tong Zhu, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Min Zhang. Efficient document-level event extraction via pseudo-trigger-aware pruned complete graph. *arXiv preprint arXiv:2112.06013*, 2021.

A Labels for the DEE Task

In this section, we present the labels of the sub-dataset in the Defendant Event Extraction (DEE) task, as well as and their corresponding label names in LEEC, as shown in Table A1. The table is organized into two main categories: Demographic Factors and Aggregated Sentencing. The Demographic Factors include basic information about the defendant such as name, gender, birth date, nationality, and birthplace. The Aggregated Sentencing category encompasses various types of sentences and penalties, including probation, detention, imprisonment, political rights deprivation, fines, property confiscation, and economic compensation, along with their respective durations and amounts. This structured labeling schema is designed to systematically capture and analyze the diverse aspects of legal cases in the DEE task.

Defendant	Role type	Corresponding label
Demographic Factors	Name	Defendant_name
	Gender	Defendant_gender
	Birth	Defendant_birth
	Nation	Defendant_nationality
	Place	Defendant_birthplace
Aggregated Sentencing	Control	Probation_aggregated
	ControlTime	Probation_term_aggregated
	Detention	Limited_incarceration_aggregated
	DetentionTime	Limited_incarceration_term_aggregated
	Imprisonment	Fixed_imprison_aggregated
	ImprisonmentTime	Fixed_imprison_term_aggregated
	PoliticalRights	Political_deprivation_aggregated
	PoliticalRightsTime	Political_deprivation_term_aggregated
	Fine	Fine_aggregated
	FineNum	Fine_amount_aggregated
	PartofProperty	Partial_property_confiscation_aggregated
	PartofPropertyNum	Partial_confiscated_amount_aggregated
	AllProperty	Total_property_confiscation_aggregated
	AllPropertyNum	Total_confiscated_amount_aggregated
	EcoCompensation	Loss_compensation_aggregated
	EcoCompensationNum	Compensation_amount_aggregated

Table A1: Roles in the defendant event table and their corresponding labels in the LEEC label system

B Annotation

This section shows the translated samples of our Annotation Guideline for LEEC. For full texts (in Chinese) of this guideline, please refer to our *GitHub repository*.

B.1 Annotation Method

The annotation of LEEC requires the annotators to find and determine the element mentions, trigger words, values of each elements from the documents. Specifically, the annotation is conducted manually by a team of graduates and undergraduates majoring in law, trained and led by professors in law. All of them are interviewed before joining the team to ensure their ability to comprehend Chinese legal concepts and knowledge, and practiced for several hours before formal annotating. We have compiled a comprehensive annotation guideline of over 150 pages in Chinese to assist our annotators. It provides an in-depth understanding of each element and their respective annotation methods. The guideline includes definitions, potential values for each element, common locations within the judicial documents where the element frequently appears, detailed rules for annotation, and real-world examples of document annotation for the elements, etc. The examples of the annotation guideline are provided in Appendix B.2. We did not require annotators to memorize the entire guideline before starting due to its excessive length. Instead, they received a briefing on the guideline's structure, annotation methods, and areas needing special attention. Additionally, they participated in a practice annotation session. During actual annotation, they primarily relied on their legal knowledge and skills from prior training to make decisions for each label. When they encountered uncertainties, they would consult the relevant sections of the guidelines, which served as a reference tool. As they gained more experience, their dependence on the guideline generally decreased.

During actual annotation, we adopted a two-stage process. In the first stage, we performed a fine annotation of 3,990 documents randomly selected from the public datasets of LeCaRD and LEVEN. We annotated the element mentions, trigger words, and the values of each element, which are the basis for evaluating several baselines of the document-level event extraction task to validate the quality and applicability of the annotations for extraction tasks. The second stage is the extended annotation of the values of elements on the remaining public datasets of LeCaRD and LEVEN, which covers 11,240 documents, to assist the verification of the results of element extraction tasks. Besides, as the label system of LEEC is largely originated from prior empirical studies, this extended dataset could also contribute to the replication and exploration of empirical research. In both stages, we performed double annotations on a portion of the work of each annotator to check the consistency and quality of the annotation.

During the annotation process, we observed that a minor fraction of judicial documents contained an unusually high number of defendants or victims, in some cases reaching into the hundreds. These documents were predominantly associated with corporate crimes committed for financial gains. In response to this situation, we implemented a upper limit, treating cases with more than seven defendants or victims as if they had exactly seven, and only annotated the first seven defendants and victims.

It is also noteworthy that a number of elements may not always be explicitly mentioned in judicial documents. For instance, the offender gender, while frequently disclosed, is not always explicitly stated, as discretion is commonly exercised in such circumstances. Similarly, whether an offender has received forgiveness from victims or their close relatives is sometimes clearly affirmed or negated, yet such information may be not mentioned in many judicial documents. As a general rule, we classify these non-mention instances as missing values. However, an exception exists when our team of legal experts determines that, in almost all cases in Chinese judicial practice, a certain element (of binary nature) is specified in the judicial document when it holds a value of 1. Thus, its absence from a judicial document suggests that the element holds a value of 0. An example is the COUNTERCLAIM

element, which denotes whether the defendant has lodged a counterclaim against the private prosecutor or the victim in the incidental civil action portion of the criminal case. In practice, this element is typically documented if and only if a counterclaim has indeed been initiated. Hence, should the element be absent from the judicial document, the annotator would assign it a value of 0. The same principle applies to elements concerning the sentencing type of offenders. If the offender does not receive a specific type of punishment, the judicial document would not mention the corresponding element in the sentencing section. Therefore, such non-mention instances are also assigned a value of 0 during annotation. Tables A2 to A13 contain special notes for elements wherein non-mention does not equate to missing value. We measure the data quality by Kappa, with a value of 0.71. This value demonstrates that the manual annotation of LEEC is conducted with high quality, contributing to the development of legal element extraction and the analysis of legal cases.

B.2 Examples from the Annotation Guideline

I. Joint_crime

1. Label Meaning

A JOINT_CRIME element refers to whether the court determined that a crime was committed by two or more persons or units in collaboration.

2. Potential Value of Element

1 = Yes; 0 = No

3. Annotation Rules

(1) This element generally appears in the fact description section (typically after the statement of "The court finds") or the reasoning section (typically after the statement of "The court considers").

(2) This element cannot be determined solely based on the number of defendants that appear in the judicial document. Having multiple defendants does not necessarily constitute a joint crime; having only one defendant does not necessarily mean there is no joint crime.

(3) If the judicial document explicitly states that "the behavior is a joint crime", then it is a joint crime. If there are other defendants mentioned in the same criminal act, but handled separately, it also counts as a joint crime.

(4) If terms like "in league with others" or "conspiring with others" appear, it is generally considered a joint crime. Any ambiguities or uncertainties should be reported to us.

(5) If the document does not explicitly state that the crime is a joint crime, but contains phrases such as the defendant "is an accomplice", "is a principal offender", or "is a coerced accomplice", it is considered a joint crime.

(6) If the joint crime status is inconsistent across different charges, each crime should be annotated separately.

(7) If it can be clearly determined from the judicial document that there is only one perpetrator, it can be determined that it is not a joint crime.

(8) If a crime contains multiple criminal acts, but only some of the criminal facts are jointly committed, the rest are not, it may still be considered a joint crime.

(9) In corporate crimes, the entity and its directly responsible personnel may be both considered guilty by the court, but this does not constitute a joint crime.

4. Annotation Example [1]

Original Document Text

Case number: (2011) Yong Zhen Xing Chu Zi No.49

From November to December 1994, the defendant FANG Xingdu and FANG Jinqi (already sentenced) **conspired in advance**, using the convenience of FANG Xingdu's position as a guard at the original Ningbo Heqiao Chemical Co., Ltd. (now Ningbo Xinqiao Chemical Co., Ltd.), responsible for the receipt of styrene raw materials, when FANG Jinqi drove Ningbo Chemical Hazardous Goods Transport Company's chemical tank truck to transport styrene raw materials from Zhenhai Port Area to Ningbo Heqiao Chemical Co., Ltd... The court believes that the defendant FANG Xingdu **conspired with others**...

Annotation Result

1

Reason for Annotation

"Conspired in advance" indicates that the two defendants conspired with each other in advance; "conspired with others" is also a typical expression of joint crime. Therefore, even if the judgment document does not explicitly state that the defendant committed a "joint crime", it can also be marked 1.

5. Related Labels: Yes. Principal_offender; Accomplice; Coerced_accomplice.

II. Forgiveness

1. Label Meaning

A FORGIVENESS element refers to whether the court determined that the defendant had obtained forgiveness from victims or their close relatives.

2. Potential Value of Element

1 = Yes; 0 = No

3. Annotation Rules

(1) This element generally appears in the fact description section (typically after the statement of "The court finds") or the reasoning section (typically after the statement of "The court considers").

(2) For this element to be assigned a value of 1, the document usually contains expressions such as "obtained the forgiveness of the victim's relatives" or "obtained the forgiveness of the victim". The forgiveness here includes the victims and their close relatives.

(3) The element should be annotated for each defendant who appears in the judicial document, respectively.

(4) If there are multiple defendants in the case, and one of them obtains forgiveness, it does not mean that the victim also forgives other defendants.

(5) If there are multiple victims in the case, and only some of the victims forgave the defendant, this still constitutes the circumstance of obtaining forgiveness, and thus, the element should be assigned a value of 1.

4. Annotation Example

Original Document Text

Case Number: (2017) Yu 1381 Criminal First Instance 426

The defendants, WANG Congwen and HUANG Jinlian,

[1] In the Annotation Guideline, there are three annotation examples for this label. Here shows one of them.

confessed their crimes truthfully, and thus, could receive lighter punishments. They compensated the economic losses of the victim's close relatives, and **obtained the forgiveness of the victim's close relatives**. Therefore, they can be punished lightly at discretion. The defendant HUANG Jinlian has shown remorse and has no danger of reoffending. A suspended sentence has no significant adverse impact on the community where she lives...

Annotation Result

1

Reason for Annotation

The document clearly states that the defendants obtained the forgiveness of the victim's relatives.

5. Related Labels: None

C Technical Details for Regression Models

In our empirical analysis, we transformed certain labels to enhance interpretability and address the focal concerns of this analysis. For example, in our preliminary exploration, we do not intend to compare any single ethnic minority (of which there are at least 55 groups in China) with the Han people. Rather, we aim to compare ethnic minorities as a whole with the Han people, similar to the main analyses in previous studies.[2] Therefore, we generated a binary variable in which 1 represents the defendant is ethnic minority, and 0 means Han. Likewise, we reclassified the defendant's nationality to discern whether being a foreigner (1=Yes; 0=No) incurs a significant disparity in sentencing outcomes. Similarly, we generated a binary indicator for unemployment status (1=Yes; 0=No), concentrating on the potential influence of unemployment on sentencing. We further crafted a binary variable to assess whether lacking a junior high school degree (1=Yes; 0=No)-the mandated level of compulsory education in China since 1986—bears any consequence on judicial decisions. Research suggests that China's household registration system engenders unequal provisions of public services and may foster severe discrimination and stigmatization related to an individual's geographic origin.[3] This effect may be particularly pronounced for those hailing from less affluent, underdeveloped, typically non-coastal areas.[4] Therefore, to explore potential disparities in the Chinese criminal justice system, we introduced a binary variable to identify defendants with household registrations from non-coastal provinces (1=Yes; 0=No). It should be noted that the defendant's place of birth and current residence are strongly associated with their household registration. To avoid potential collinearity, we excluded these factors from the current analysis. Moreover, as defendant age is included, we did not include whether the court explicitly determined whether the defendant's age is a legally mitigating factor (age <18 or age ≥ 75) to avoid interfere with the interpretation of the impact of the defendant age variable. However, the rich information of these labels provides future researchers with a robust framework and rich information to develop variables tailored to the specific requirements and topics of their respective research projects. As defendant demographic variables contain missing values, we employed multiple imputation to fill these values based on the control variables within the regression model following a series of prior empirical legal studies.[5] Multiple imputation produces multiple set of imputed data and enables researchers to conduct regression on each dataset, and finally combined the regression results to produce more robust estimation.[6] Adopting the suggestion of researchers[7], we imputed the dataset 20 times to achieve optimal estimations.

The dependent variable of the analysis is the duration (measured in months) of fixed-term imprisonment in sentencing, which is most commonly utilized type of primary punishment within our dataset and also one of the harshest judicial punishment. This introduces impact in three aspects: 1) Given the continuous nature of our dependent variable, we apply the Ordinary Least Squares (OLS) regression model with robust standard errors to our analysis; 2) As only natural person could be sentenced to fixed-term imprisonment, the variable of whether the defendant is a unit and the corresponding data are omitted; 3) For verdicts that involve multiple defendants, we selected the defendants that received the harshest sentence for analysis. We then introduce the set of control variables outlined within the "sentencing factors" section of the label system (refer to Figure A1) to align with the legally mandated considerations for the collegiate bench during sentencing, including the defendant's plea of guilt, voluntary surrender, confession, and acquisition of victims' forgiveness, among others. We also introduced dummy variables of crime type as defined by chapters within the special provisions of Chinese criminal law, categorized as crimes of encroaching on property, crimes of endangering public security, etc. As we selected the defendants that receive the harshest sentence, we omitted the variables regarding whether the

[2]Please see: [Hou and Truex, 2022] Yue Hou and Rory Truex. Ethnic discrimination in criminal sentencing in China. *The Journal of Politics*, 84(4):2294–2299, 2022; [Peng and Cheng, 2022] Yali Peng and Jinhua Cheng. Ethnic disparity in Chinese theft sentencing. *China Review*, 22(3):47–71, 2022.

[3][Song, 2016] Yang Song. Hukou-based labour market discrimination and ownership structure in urban china. *Urban Studies*, 53(8):1657–1673, 2016.)

[4][Gu et al., 2022] Hengyu Gu, Yuhao Lin, and Tiyan Shen. Do you feel accepted? perceived acceptance and its spatially varying determinants of migrant workers among Chinese cities. *Cities*, 125:103626, 2022.

[5][Liu et al., 2021] Lin Liu, Christy A Visher, and Daniel J O'Connell. Strain during reentry: A test of general strain theory using a sample of adult former prisoners. *The Prison Journal*, 101(4):420–442, 2021.

[6]Please see: [Little and Rubin, 2019] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019; [Van Buuren, 2007] Stef Van Buuren. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, 16(3):219–242, 2007.

[7]Please see: [Graham et al., 2007] John W Graham, Allison E Olchowski, and Tamika D Gilreath. How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention science*, 8:206–213, 2007; [Petrie and Coverdill, 2010] Michelle A Petrie and James E Coverdill. Who lives and dies on death row? race, ethnicity, and post-sentence outcomes in texas. *Social problems*, 57(4):630–652, 2010.

defendant is the principal offender, accomplice, or coerced accomplice. Although in Chinese legal theory, one defendant could both be the abettor and receive the the harshest sentence, we identify no such circumstance in our dataset. Therefore, whether the defendant is the abettor is omitted in the regression. The LEEC dataset enables us to control a broader array of variables than most of the empirical legal research in social sciences, and thus, permits us to better reduce bias in estimation results.

D Element Schema and Description

To facilitate the understanding of our label system and each element within it, thereby promoting future application and research, we provide the multi-level label system in Figure A1. Additionally, Tables A2 to A13 provide detailed descriptions of each element, including the element name, explanation, and value type. In addition to the 160 labels within our domain-specific label system, the LEEC dataset includes the full text of the corresponding judicial documents as an extra label to facilitate data processing and downstream applications. For more detailed explanations and annotation methods for each label, please refer to the Annotation Guideline available on our *GitHub repository*.

— — — — — — — —

See Next Page

— — — — — — — —

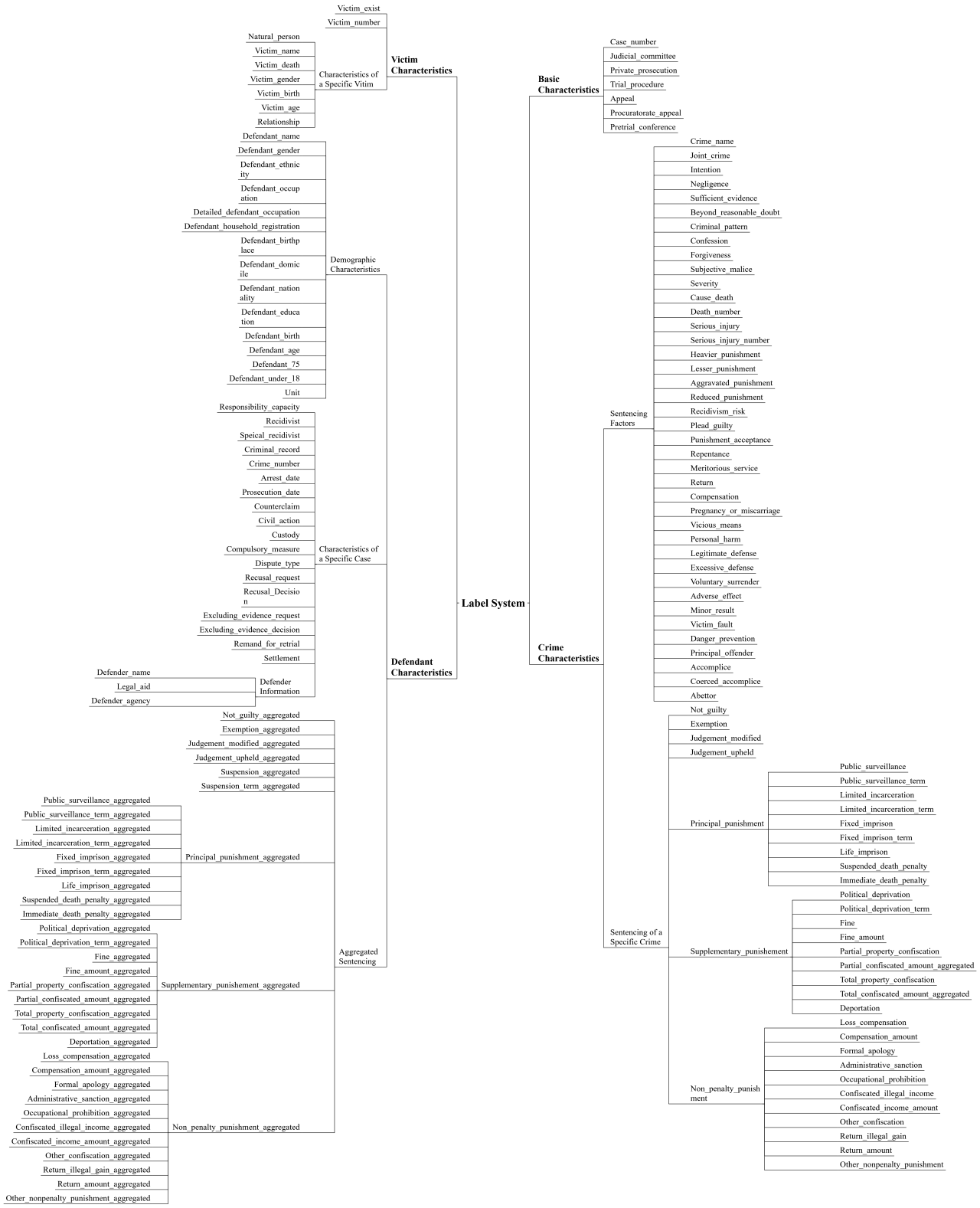

Figure A1: The Detailed Element Schema of LEEC.

Element Name	Explanation	Value Type
Basic Characteristics		
Case_number	A CASE_NUMBER element refers to the ID assigned by the court that uniquely identifies each judicial documents.	Extracted from the Specific Content of Judicial Documents
Judicial_committee	A JUDICIAL_COMMITTEE element refers to whether the court submitted the case to the judicial committee for discussion.	1 = Yes, 0 = No
Private_prosecution	A PRIVATE_PROSECUTION element refers to whether the victim, the victim's legal representative, or close relative institute an action directly in a people's court in a case of private prosecution.	1 = Yes, 0 = No, Non-mention is treated as 0
Trial_procedure	A TRIAL_PROCEDURE element refers to the procedure applied to the trial of the case.	Summary, Formal, Fast-Track Sentencing, Transfer from Summary to Formal Procedure, Transfer from Fast-Track to Formal Procedure
Appeal	An APPEAL element refers to whether the defendant appealed the case.	1 = Yes, 0 = No
Procuratorate_appeal	A PROCURATORATE_APPEAL element refers to whether the people's procuratorate files an appeal to the people's court at the higher level.	1 = Yes, 0 = No
Pretrial_conference	A PRETRIAL_CONFERENCE element refers to whether the court determined that a pretrial conference for a case should be held.	1 = Yes, 0 = No
Victim Characteristics (I)		
Victim_exist	A VICTIM_EXIST element refers to there was a victim whose legal rights and interests have been violated by a criminal act in a criminal case according to the court.	1 = Yes, 0 = No
Victim_number	A VICTIM_NUMBER element refers to the number of victims according to the information provided by the court. This element is annotated only when elements VICTIM_EXIST is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents
Natural_person	A NATURAL_PERSON element refers to whether the victim is a natural person in the biological sense. This element is annotated only when elements VICTIM_EXIST is assigned a value of 1.	1 = Yes, 0 = No
Victim_name	A VICTIM_NAME element refers to the name of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_death	A VICTIM_DEATH element refers to whether a specific victim was dead. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	1 = Yes, 0 = No
Victim_gender	A VICTIM_GENDER element refers to the gender of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_birth	A VICTIM_BIRTH element refers to the date of birth of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_age	A VICTIM_AGE element refers to the age of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table A2: List of detailed element information (I).

Element Name	Explanation	Value Type
Victim Characteristics (II)		
Relationship	A RELATIONSHIP element refers to the relationship between the defendant and the victim based on the content of judicial documents.	Relationship between Non-natural Persons, Marital Relationship, Close Relatives, Other Relatives, Acquainted, Unknown
Defendant Characteristics (I)		
Defendant_name	A DEFENDANT_NAME element refers to the name of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_gender	A DEFENDANT_GENDER element refers to the gender of the defendant.	1 = Male; 0 = Female
Defendant_ethnicity	A DEFENDANT_ETHNICITY element refers to the ethnicity of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_occupation	A DEFENDANT_OCCUPATION element refers to the occupation of the defendant categorized into four types.	Extracted from the Specific Content of Judicial Documents
Detailed_defendant_occupation	A DETAILED_DEFENDANT_OCCUPATION element refers to the detailed occupation of the defendant as explicitly stated in the judicial document.	Extracted from the Specific Content of Judicial Documents
Defendant_household_registration	A DEFENDANT_HOUSEHOLD_REGISTRATION element refers to the place of registered permanent residence of the defendant, also known as *Hukou* in Chinese.	Extracted from the Specific Content of Judicial Documents
Defendant_birthplace	A DEFENDANT_BIRTHPLACE element refers to the place of birth of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_domicile	A DEFENDANT_DOMICILE element refers to the address of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_nationality	A DEFENDANT_NATIONALITY element refers to the nationality of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_education	A DEFENDANT_EDUCATION element refers to to the level of formal education or academic degree attained by the defendant.	Illiteracy, Under Secondary School, Secondary School, Primary School, Regular Senior High School, Secondary Vocational School, Tertiary Vocational School, Bachelor's Degree, Master's Degree, Doctor's Degree
Defendant_birth	A DEFENDANT_BIRTH element refers to the date of birth of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_age	A DEFENDANT_AGE element refers to the age of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_75	A DEFENDANT_75 element refers to whether the court determined that the defendant is above the age of 75, which is a mitigating circumstance in Chinese criminal law, and thereby may be explicitly stated in the judicial document.	1 = Yes; 0 = No
Defendant_under_18	A DEFENDANT_UNDER_18 element refers to whether the court determined that the defendant is under the age of 18, which is a mitigating circumstance in Chinese criminal law, and thereby may be explicitly stated in the judicial document.	1 = Yes; 0 = No
Unit	A UNIT element refers to whether the defendant is a company, enterprise, institution, organization, or group.	1 = Yes; 0 = No
Responsibility_capacity	A RESPONSIBILITY_CAPACITY element refers to the court determined that the level of the defendant's ability to take responsibility for crimes.	Full Criminal Responsibility Capacity Criminal Responsibility Incapacity Relatively Criminal Responsibility Incapacity Partial Criminal Responsibility Capacity
Recidivist	A RECIDIVIST element refers to whether the court determined that the defendant was a recidivist.	1 = Yes; 0 = No

Table A3: List of detailed element information (II).

Element Name	Explanation	Value Type
Defendant Characteristics (II)		
Special_recidivist	A SPECIAL_RECIDIVIST element refers to whether the court determined that the defendant was a special recidivist as stipulated in the Chinese criminal law.	1 = Yes; 0 = No
Criminal_record	A CRIMINAL_RECORD element refers to whether the court determined that the defendant had a previous criminal record.	1 = Yes; 0 = No
Crime_number	A CRIME_NUMBER element refers to the total number of the crime name of a specific defendant.	Numeric Value Calculated Based on the Content of Judicial Documents
Arrest_date	An ARREST_DATE element refers to the date of the execution of arrest by criminal justice authorities.	Extracted from the Specific Content of Judicial Documents
Prosecution_date	A PROSECUTION_DATE element refers to the date of the initiation of a public prosecution by the procuratorate.	Extracted from the Specific Content of Judicial Documents
Counterclaim	A COUNTERCLAIM element refers to whether the defendant files a counterclaim against the private prosecutor in a private criminal prosecution case or the victim in the incidental civil action part of the incidental civil case.	1 = Yes; 0 = No; Non-mention is treated as 0
Civil_action	A CIVIL_ACTION element refers to whether the court determined that the incidental civil action is instituted.	1 = Yes; 0 = No; Non-mention is treated as 0
Custody	A CUSTODY element refers to holding officials suspected of serious job-related violations or crimes and other related individuals in custody by the oversight authority.	1 = Yes; 0 = No; Non-mention is treated as 0
Compulsory_measure	A COMPULSORY_MEASURE element refers to the methods to restrict a certain degree of personal freedom for criminal suspects and defendants by criminal justice authorities.	Forced Appearance/Granted Bail/Residential Confinement/Detention/Arrest
Dispute_type	A DISPUTE_TYPE element refers to the type of dispute involved in the case, including disputes with clearly identified victims, disputes between neighbors, family disputes, and disputes without clearly identified victims.	Family Disputes / Disputes among Neighbors / Other Disputes with Victims / Disputes without Victims
Recusal_request	A RECUSAL_REQUEST element refers to whether the parties requested for recusal.	1 = Yes; 0 = No
Recusal_decision	A RECUSAL_DECISION element refers to whether the court approved of a recusal request.	1 = Yes; 0 = No
Excluding_evidence_request	An EXCLUDING_EVIDENCE_APPLICATION element refers to whether parties and their defenders or litigation representatives requested for excluding illegal evidence.	1 = Yes; 0 = No
Excluding_evidence_decision	An EXCLUDING_EVIDENCE_DECISION element refers to whether the court approved of a request for excluding illegal evidence.	1 = Yes; 0 = No
Remand_for_retrial	A REMAND_FOR_RETRIAL element refers to whether the court determined that the case shall be remanded for retrial.	1 = Yes; 0 = No
Settlement	A SETTLEMENT element refers to whether both parties reach a settlement.	1 = Yes; 0 = No
Defender_name	A DEFENDER_NAME element refers to the name of the defender.	Extracted from the Specific Content of Judicial Documents
Legal_aid	A LEGAL_AID element refers to whether the defender was designated by the legal aid agency.	1 = Yes; 0 = No
Defender_agency	A DEFENDER_AGENCY element refers to the agency of the defender.	Extracted from the Specific Content of Judicial Documents

Table A4: List of detailed element information (III).

Element Name	Explanation	Value Type
Defendant Characteristics (III)		
Not_guilty_aggregated	A NOT_GUILTY_AGGREGATED element refers to whether the court determined that the defendant was not guilty.	1 = Yes; 0 = No; Non-mention is treated as 0
Exemption_aggregated	An EXEMPTION_AGGREGATED element refers to whether the defendant is exempted from criminal punishment as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_modified_aggregated	A JUDGEMENT_MODIFIED_AGGREGATED element refers to whether the court determined that full or partial revision of sentence is made in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_upheld_aggregated	A JUDGEMENT_UPHELD_AGGREGATED element refers to whether the court decided to uphold the previous judgment in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspension_aggregated	A SUSPENSION_AGGREGATED element refers to whether the court determined a suspension of sentence in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspension_term_aggregated	A SUSPENSION_TERM_AGGREGATED element refers to the probation period for suspension as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element SUSPENSION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Probation_aggregated	A PROBATION_AGGREGATED element refers to whether the defendant was subject to probation as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Probation_term_aggregated	A PROBATION_TERM_AGGREGATED element refers to the term of probation as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element PROBATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Limited_incarceration_aggregated	A LIMITED_INCARNATION_AGGREGATED element refers to whether the defendant was subject to limited incarceration as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Limited_incarceration_term_aggregated	A LIMITED_INCARCERATION_TERM_AGGREGATED element refers to the term of limited incarceration as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element LIMITED_INCARNATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fixed_imprison_aggregated	A FIXED_IMPRISON_AGGREGATED element refers to whether the defendant was subject to fixed incarceration as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Fixed_imprison_term_aggregated	A FIXED_IMPRISON_TERM_AGGREGATED element refers to the length of fixed-term imprisonment as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element FIXED_IMPRISON_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table A5: List of detailed element information (IV).

Element Name	Explanation	Value Type
Defendant Characteristics (IV)		
Life_imprison_aggregated	A LIFE_IMPRISON_AGGREGATED element refers to whether the defendant was subject to life imprisonment as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspended_death_penalty_aggregated	A SUSPENDED_DEATH_PENALTY_AGGREGATED element refers to whether the defendant was subject to a suspended death penalty as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Immediate_death_penalty_aggregated	An IMMEDIATE_DEATH_PENALTY_AGGREGATED element refers to whether the defendant was subject to an immediate death penalty as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation_aggregated	A POLITICAL_DEPRIVATION_AGGREGATED element refers to whether the defendant was subject to the deprivation of political rights as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation_term_aggregated	A POLITICAL_DEPRIVATION_TERM_AGGREGATED element refers to the term of deprivation of political rights as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element POLITICAL_DEPRIVATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fine_aggregated	A FINE_AGGREGATED element refers to whether the defendant was fined as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Fine_amount_aggregated	A FINE_AMOUNT_AGGREGATED element refers to the amount of the fine as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element FINE_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Partial_property_confiscation_aggregated	A PARTIAL_PROPERTY_CONFISCATION_AGGREGATED element refers to whether the defendant was subject to confiscation of a part of his or her property as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Partial_confiscated_amount_aggregated	A PARTIAL_CONFISCATED_AMOUNT_AGGREGATED element refers to the amount of partially confiscated property as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element PARTIAL_PROPERTY_CONFISCATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Total_property_confiscation_aggregated	A TOTAL_PROPERTY_CONFISCATION_AGGREGATED element refers to whether the defendant was subject to confiscation of all of his or her property as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Total_confiscated_amount_aggregated	A TOTAL_CONFISCATED_AMOUNT_AGGREGATED element refers to the amount of totally confiscated property as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element TOTAL_PROPERTY_CONFISCATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table A6: List of detailed element information (V).

Element Name	Explanation	Value Type
Defendant Characteristics (V)		
Deportation_aggregated	A DEPORTATION_AGGREGATED element refers to whether the defendant was subject to deportation as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Loss_compensation_ aggregated	A LOSS_COMPENSATION_AGGREGATED element refers to whether the defendant was ordered to make compensation for the economic loss as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Compensation_amount_ aggregated	A COMPENSATION_AMOUNT_AGGREGATED element refers to the amount of making compensation for the economic loss as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element LOSS_COMPENSATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Formal_apology_ aggregated	A FORMAL_APOLOGY_AGGREGATED element refers to whether the defendant was criticized by the court, ordered to make a guarantee of repentance, or make formal apology as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Administrative_sanction_ aggregated	An ADMINISTRATIVE_SANCTION_AGGREGATED element refers to whether the defendant was subjected to administrative sanctions by the relevant department as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Occupational_prohibition_ aggregated	An OCCUPATIONAL_PROHIBITION_ AGGREGATED element refers to whether the defendant was subjected to occupational prohibition as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_illegal_ income_aggregated	A CONFISCATED_ILLEGAL_INCOME_ AGGREGATED element refers to whether the defendant was subjected to the confiscation of illegal income as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_income_ amount_aggregated	A CONFISCATED_INCOME_AMOUNT_ AGGREGATED element refers to the amount of confiscated income as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element CONFISCATED_ILLEGAL_INCOME_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_confiscation_ aggregated	An OTHER_CONFISCATION_AGGREGATED element refers to whether the defendant was subjected to the confiscation of other objects related to the crime as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_illegal_gain_ aggregated	A RETURN_ILLEGAL_GAIN_AGGREGATED element refers to the order for returning illegal gains and compensations as determined by the court in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_amount_aggregated	A RETURN_AMOUNT_AGGREGATED element refers to the amount of returning illegal gains and compensations as ordered by the court in the aggregated sentencing of a defendant. This element is annotated only when element RETURN_ILLEGAL_GAIN_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_nonpenalty_ punishment_aggregated	An OTHER_NONPENALTY_PUNISHMENT_ AGGREGATED element refers to the order of other types of nonpenalty punishment as determined by the court in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0

Table A7: List of detailed element information (VI).

Element Name	Explanation	Value Type
Crime Characteristics (I)		
Crime_name	A CRIME_NAME element refers to the charges against the defendant.	Extracted from the Specific Content of Judicial Documents
Joint_crime	A JOINT_CRIME element refers to whether the court determined that an intentional crime was committed by two or more persons or units jointly.	1 = Yes; 0 = No
Intention	An INTENTION element refers to the subjective aspect of the defendant in an intentional crime as determined by the court.	Direct Intention / Indirect Intention
Negligence	A NEGLIGENCE element refers to the subjective aspect of the criminal in a negligent crime as determined by the court.	Carelessness/Overconfidence
Sufficient_evidence	A SUFFICIENT_EVIDENCE element refers to whether the court determined that evidence is sufficient.	1 = Yes; 0 = No
Beyond_reasonable_doubt	A BEYOND_REASONABLE_DOUBT element refers to whether the court determined that the evidence is strong enough to rule out any reasonable doubt.	1 = Yes; 0 = No
Criminal_pattern	A CRIMINAL_PATTERN element refers to the pattern of the crime.	Preparation / Discontinuation / Criminal Attempt / Consummation
Confession	A CONFESSION element refers to whether the court determined that the defendant confessed to authorities.	1 = Yes; 0 = No
Forgiveness	A FORGIVENESS element refers to whether the court determined that the defendant had obtained forgiveness from victims or their close relatives.	1 = Yes; 0 = No
Subjective_malice	A SUBJECTIVE_MALICE element referes to the degree of the subjective malice of the defendant as determined by the court.	No or Low Subjective Malice/High Subjective Malice
Severity	A SEVERITY element refers to the level of severity of the crime as determined by the court.	Circumstances Clearly Minor / Circumstances Minor / Circumstances Serious or Execrable / Circumstances Very Serious or Execrable
Cause_death	A CAUSE_DEATH element refers to whether the court determined that whether any death occurred because of the crime.	1 = Yes; 0 = No
Death_number	A DEATH_NUMBER element refers to the total number of victims who died because of the crime according to the court. This element is annotated only when element VICTIM_DEATH is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents
Serious_injury	A SERIOUS_INJURY element refers to whether the court determined that serious injury of any victim is caused by the crime.	1 = Yes; 0 = No
Serious_injury_number	A SERIOUS_INJURY_NUMBER element refers to the total number of victims who were seriously injured by the crime according to the court. This element is annotated only when element SERIOUS_INJURY is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents

Table A8: List of detailed element information (VII).

Element Name	Explanation	Value Type
Crime Characteristics (II)		
Heavier_punishment	A HEAVIER_PUNISHMENT element refers to the number of circumstances leading to heavier punishment within the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Lesser_punishment	A LESSER_PUNISHMENT element refers to the number of circumstances leading to lesser punishment within the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Aggravated_punishment	An AGGRAVATED_PUNISHMENT element refers to the number of circumstances leading to aggravated punishment above the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Mitigated_punishment	A MITIGATED_PUNISHMENT element refers to the number of circumstances leading to reduced punishment below the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Reoffending_danger	A REOFFENDING_DANGER element refers to whether the court determined that the defendant would likely to commit any crime again.	1 = Yes; 0 = No
Plead_guilty	A PLEAD_GUILTY element refers to whether the court determined that the defendant pled guilty.	1 = Yes; 0 = No
Punishment_acceptance	A PUNISHMENT_ACCEPTANCE element refers to whether the court determined that the defendant accepted punishment.	1 = Yes; 0 = No
Repentance	A REPENTANCE element refers to whether the court determined that the defendant had showed genuine repentance.	1 = Yes; 0 = No
Meritorious_service	A MERITORIOUS_SERVICE element refers to whether the court determined that the defendant performed meritorious service.	1 = Yes; 0 = No
Return	A RETURN element refers to whether the court determined that the defendant actively returned the property that the defendant acquired illegally or equivalent amount of money.	1 = Yes; 0 = No
Compensation	A COMPENSATION element refers to whether the court determined that the defendant actively made compensation to victims actively.	1 = Yes; 0 = No
Pregnancy_or_miscarriage	A PREGNANCY_OR_MISCARRIAGE element refers to whether the court determined that the defendant was pregnant or had a miscarriage during prosecution or trial.	1 = Yes; 0 = No
Vicious_means	A VICIOUS_MEANS element refers to whether the court determined that the crime was conducted using vicious means.	1 = Yes; 0 = No
Personal_harm	A PERSONAL_HARM element refers to whether the court determined that the defendant posed a high risk of or caused personal harm.	1 = Yes; 0 = No
Legitimate_defense	A LEGITIMATE_DEFENSE element refers to whether whether the court determined that the act of the defendant was legitimate defense.	1 = Yes; 0 = No

Table A9: List of detailed element information (VIII).

Element Name	Explanation	Value Type
Crime Characteristics (III)		
Excessive_defense	An EXCESSIVE_DEFENSE element refers to whether the court determined that defendant's defense noticeably exceeded the necessary limits.	1 = Yes; 0 = No
Voluntary_Surrender	A VOLUNTARY_SURRENDER element refers to whether the court determined that the defendant voluntarily surrendered to the police and gave a true account of one's crime after committing it.	1 = Yes; 0 = No
Adverse_effect	An ADVERSE_EFFECT element refers to whether the court determined that the social effects caused by the defendant were significantly adverse.	1 = Yes; 0 = No
Minor_result	A MINOR_RESULT element refers to whether the court determined that the circumstances of the alleged conduct are obviously minor, causing no serious harm.	1 = Yes; 0 = No
Victim_fault	A VICTIM_FAULT element refers to whether the court determined that victim was in fault for the crime.	1 = Yes; 0 = No
Danger_prevention	A DANGER_PREVENTION element refers to whether the court determined that the defendant's act was a legitimate prevention of urgent danger.	1 = Yes; 0 = No
Principal_offender	A PRINCIPAL_OFFENDER element refers to whether the court determined that the defendant organized and leads a criminal group in conducting criminal activities or played a principal role in a joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Accomplice	An ACCOMPLICE element refers to whether the court determined that the defendant played a secondary or supplementary role in a joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Coerced_accomplice	A COERCED_ACCOMPLICE element refers to whether the court determined that the defendant was coerced to participate in a crime shall. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Abettor	An ABETTOR element refer to whether the court determined that the defendant was the one who instigated others to commit a crime and should be punished according to the role he played in the joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Not_guilty	A NOT_GUILTY element refers to whether the court determined that the defendant was not guilty of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0
Exemption	An EXEMPTION element refers to whether the defendant is exempted from criminal punishment of a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0

Table A10: List of detailed element information (IX).

Element Name	Explanation	Value Type
Crime Characteristics (IV)		
Judgement_modified	A_JUDGEMENT_MODIFIED element refers to whether the court determined that full revision and partial reversal with partial revision of sentence should be made of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_upheld	A JUDGEMENT_UPHELD element refers to whether the court decided to uphold the previous judgment of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0
Public_surveillance	A PUBLIC_SURVEILLANCE element refers to whether the defendant was subject to public surveillance for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Public_surveillance_term	A PUBLIC_SURVEILLANCE_TERM element refers to the term of public surveillance of a specific crime as determined by the court. This element is annotated only when element PROBATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Limited_incarceration	A LIMITED_INCARNATION element refers to whether the defendant was subject to limited incarceration for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Limited_incarceration_term	A LIMITED_INCARCERATION_TERM element refers to the term of limited incarceration as determined by the court for a specific crime of a defendant. This element is annotated only when element LIMITED_INCARNATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fixed_imprison	A FIXED_IMPRISON element refers to whether the defendant was subject to fixed incarceration for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Fixed_imprison_term	A FIXED_IMPRISON_TERM element refers to the length of fixed-term imprisonment as determined by the court for a specific crime of a defendant. This element is annotated only when element FIXED_IMPRISON is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Life_imprison	A LIFE_IMPRISON element refers to whether the defendant was subject to life imprisonment for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspended_death_penalty	A SUSPENDED_DEATH_PENALTY element refers to whether the defendant was subject to a suspended death penalty for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Immediate_death_penalty	An IMMEDIATE_DEATH_PENALTY element refers to whether the defendant was subject to an immediate death penalty for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation	A POLITICAL_DEPRIVATION element refers to whether the defendant was subject to the deprivation of political rights for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation_term	A POLITICAL_DEPRIVATION_TERM element refers to the term of deprivation of political rights for a specific crime of a defendant as determined by the court. This element is annotated only when element POLITICAL_DEPRIVATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table A11: List of detailed element information (X).

Element Name	Explanation	Value Type
Crime Characteristics (V)		
Fine	A FINE element refers to whether the defendant was fined for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Fine_amount	A FINE_AMOUNT element refers to the amount of the fine of a specific crime as determined by the court. This element is annotated only when element FINE is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Partial_property_confiscation	A PARTIAL_PROPERTY_CONFISCATION element refers to whether the defendant was subject to confiscation of a part of his or her property for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Partial_confiscated_amount	A PARTIAL_CONFISCATED_AMOUNT element refers to the amount of partially confiscated property as determined by the court for a specific crime of a defendant. This element is annotated only when element PARTIAL_PROPERTY_CONFISCATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Total_property_confiscation	A TOTAL_PROPERTY_CONFISCATION element refers to whether the defendant was subject to confiscation of all of his or her property for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Total_confiscated_amount	A TOTAL_CONFISCATED_AMOUNT element refers to the amount of totally confiscated property as determined by the court for a specific crime of a defendant. This element is annotated only when element TOTAL_PROPERTY_CONFISCATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Deportation	A DEPORTATION element refers to whether the defendant was subject to deportation for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Loss_compensation	A LOSS_COMPENSATION element refers to whether the defendant was ordered to make compensation for the economic loss for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Compensation_amount	A COMPENSATION_AMOUNT element refers to the amount of making compensation for the economic loss as determined by the court for a specific crime of a defendant. This element is annotated only when element LOSS_COMPENSATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Formal_apology	A FORMAL_APOLOGY element refers to whether the defendant was criticized by the court, ordered to make a guarantee of repentance, or make formal apology for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Administrative_sanction	An ADMINISTRATIVE_SANCTION element refers to whether the defendant was subjected to administrative sanctions by the relevant department for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Occupational_prohibition	An OCCUPATIONAL_PROHIBITION element refers to whether the defendant was subjected to occupational prohibition for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_illegal_income	A CONFISCATED_ILLEGAL_INCOME element refers to whether the defendant was subjected to the confiscation of illegal income for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0

Table A12: List of detailed element information (XI).

Element Name	Explanation	Value Type
Crime Characteristics (VI)		
Confiscated_income_amount	A CONFISCATED_INCOME_AMOUNT element refers to the amount of confiscated income for a specific crime of the defendant as determined by the court. This element is annotated only when element CONFISCATED_ILLEGAL_INCOME is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_confiscation	An OTHER_CONFISCATION element refers to whether the defendant was subjected to the confiscation of other objects related to a specific crime of the defendant as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_illegal_gain	A RETURN_ILLEGAL_GAIN element refers to the order of returning illegal gains and compensations for a specific crime of the defendant as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_amount	A RETURN_AMOUNT element refers to the amount of returning illegal gains and compensations as ordered by the court for a specific crime of the defendant. This element is annotated only when element RETURN_ILLEGAL_GAIN is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_nonpenalty_punishment	An OTHER_NONPENALTY_PUNISHMENT element refers to the order of returning illegal gains and compensations for a specific crime of the defendant as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0

Table A13: List of detailed element information (XII).