

# An Architectural Framework for the Construction of a Crime Narrative Corpus from Judicial Records

Giovanni Acampora<sup>1</sup>, Bilal Ahmed<sup>1,\*</sup> and Autilia Vitiello<sup>1</sup>

<sup>1</sup>Department of Physics “Ettore Pancini”, University of Naples Federico II, via Cintia 21, 80126 Naples, Italy

## Abstract

Forensic investigation relies on reconstructing how criminal events unfold in time, space, and human interaction. This task is inherently complex due to the involvement of multiple sources of evidence; however, it could benefit from information reported in judicial judgments of previous criminal cases, as such documents often contain detailed reconstructions of events that can support investigative reasoning and hypothesis generation. Unfortunately, judicial judgments embed the narrative of criminal events within legally dense and procedurally driven text, which hinders their straightforward reuse for forensic reasoning purposes. In this context, Artificial Intelligence (AI) techniques could facilitate the generation of clear and structured crime narratives from judicial records, making them suitable for use by law enforcement officers and, in the future, by AI-assisted investigative systems. Based on these considerations, the goal of this paper is to present an architectural proposal in which judicial records can be responsibly repurposed into a forensic crime narrative corpus through a human-centered pipeline that prioritizes transparency, traceability, and selective expert oversight. In detail, this framework combines lexicon-guided recall, transformer-based functional classification, uncertainty-driven active learning, and lightweight event and timeline structuring. By generating a publicly available corpus that curates crime narratives from judicial records, this framework establishes a solid foundation for future research on AI-assisted investigative reasoning and mixed-reality crime-scene support systems.

## Keywords

Forensic Natural Language Processing, Judicial Records, Criminal Event Extraction, Crime Narrative Corpus

## 1. Introduction

Crime Scene Investigation (CSI) is a cognitively demanding process in which investigators synthesize fragmented observations, witness accounts, and physical evidence to form coherent hypotheses about past events [1]. Effective investigation depends not only on isolated facts, but also on reasoning about sequences of actions, temporal progression, spatial relations, and interactions among actors. Due to the complexity of this task for law enforcement officers, Artificial Intelligence (AI) techniques are emerging as good methods to support forensic investigations [2, 3, 4, 5, 6]. However, current AI deployments in the criminal justice ecosystem focus primarily on downstream tasks such as forensic laboratory analysis, crime pattern analysis, sentencing support, or prediction of legal outcomes [7, 8]. These systems rarely assist investigators during the early reasoning stages of a case, when decisions about evidence collection and hypothesis generation have the greatest operational consequences. A central obstacle to progress is the scarcity of accessible training data that captures investigative narratives while respecting ethical, legal, and privacy constraints [9].

Based on this consideration, publicly available judicial judgments could be a valuable and under-used source of structured narrative information for forensic investigation and AI-assisted forensic reasoning. Indeed, many criminal judgments contain factual accounts that describe the conduct of the offense, victim-offender interactions, and investigative observations of previous criminal events that could support forensic reasoning and hypothesis generation. However, such narratives are typi-

---

*1st Workshop on supporting CRIME reSolution Through Artificial INtelligence (CRISTAIN), held in conjunction with CHITALY 2025, Salerno, Italy, October 6–10*

\*Corresponding author.

✉ giovanni.acampora@unina.it (G. Acampora); bilal\_ahmed\_raja@hotmail.com (B. Ahmed); autilia.vitiello@unina.it (A. Vitiello)

🌐 <https://www.docenti.unina.it/giovanni.acampora> (G. Acampora); <https://www.docenti.unina.it/autilia.vitiello> (A. Vitiello)

🆔 0000-0003-4082-5616 (G. Acampora); 0000-0001-5562-9226 (A. Vitiello)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

cally interwoven with legal reasoning, precedent discussion, and procedural commentary. Extracting investigator-relevant narratives from these documents therefore requires careful linguistic separation and systematic validation, based on insights from rhetorical role labeling and legal discourse analysis [10, 11, 12].

In order to bridge this gap, the goal of this work is to propose a human-centered path toward a forensic crime narrative corpus derived from judicial records, together with a hybrid Human-in-the-Loop (HITL) modeling workflow that can isolate, validate, and structure narrative content with high precision and feasibility. The core claim is not that the process should be fully automated, but that it should be designed around selective human oversight to ensure reliability, accountability, and practical utility [13, 9]. In detail, we outline a hybrid HITL framework that combines lexicon-guided recall to prioritize narrative-bearing text, transformer-based functional classification that builds on advances in legal NLP [10, 14], uncertainty-driven active learning to reduce annotation burden [15, 16], and lightweight event and timeline assembly to provide incident-level representations [17, 18]. The proposed design aims to maximize narrative coverage while controlling annotation effort and maintaining auditability, consistent with best practices in corpus construction and HITL systems [15, 16, 13]. Unlike conventional legal Natural Language Processing (NLP) work, which often targets judgment outcomes or doctrinal structure [7], the proposed corpus and workflow are geared toward modeling crime narratives to support forensic reasoning and investigative decision support.

The intended outcome is a dataset that supports functional annotations at the sentence-level that distinguish narrative events, contextual descriptions, and procedural or legal discourse. In addition, it is expected to include event templates that capture actors, actions, targets, locations, and temporal ordering, drawing on principles from event extraction and narrative schema learning [19, 20]. Each case is also expected to support a canonical narrative reconstruction with explicit links to the source text in order to preserve provenance and auditability [21, 13]. The resulting data set is intended to enable a range of downstream research tasks that are currently limited by the lack of narrative-centric forensic data. These include event-sequence modeling of criminal behavior, comparative analysis of offense patterns across cases, temporal reasoning over investigative actions, and knowledge grounding for interactive decision-support systems. In particular, such structured crime narratives can serve as training and evaluation material for AI agents designed to assist law enforcement personnel during crime-scene investigations, including mixed-reality and situated AI systems that provide contextual guidance, procedural prompts, and hypothesis support during evidence collection. Within this scope, the planned release of the dataset is accompanied by a reproducible baseline sentence-level functional classification model and annotation protocols, intended to establish a shared reference point for future forensic NLP and HITL investigative AI research.

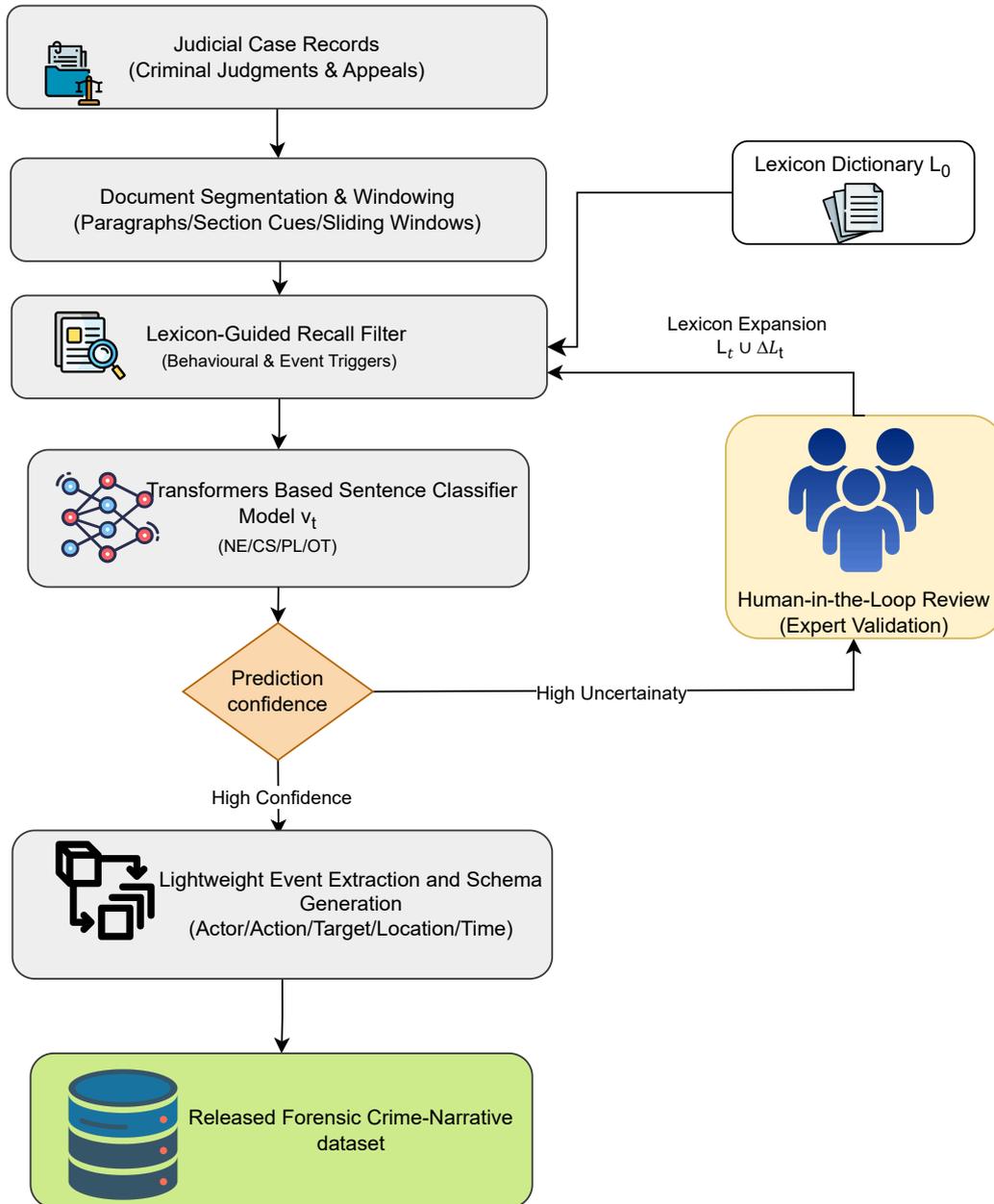
## 2. The Proposed Framework

The proposed architecture follows an incremental and auditable pipeline designed to balance narrative coverage, annotation efficiency, and robustness across heterogeneous judicial writing styles. The core design principle is to maximize the recall of crime narrative content while limiting human effort through selective verification and iterative refinement. Human oversight is incorporated where automated confidence is insufficient, supporting scalability without sacrificing traceability [15, 13]. The general workflow is illustrated in Figure 1.

In the following, all components of our framework are discussed in detail.

### 2.1. Judicial Case Records

The source material consists of publicly available criminal judgments in which the conduct of the offense and the interactions between the victims are described with sufficient detail to support the narrative reconstruction. These documents vary substantially in length, rhetorical structure, and format. Some judgments provide explicit sections such as *Facts* or *The Offending*, whereas others distribute narrative information across multiple paragraphs without clear demarcation. In this work, extraction is restricted



**Figure 1:** Proposed hybrid Human-in-the-Loop pipeline for constructing a forensic crime narrative corpus from judicial records. The workflow is designed to support both dataset creation and iterative model refinement under explicit provenance constraints.

to factual descriptions of the crime incident and its immediate context. Legal principles, sentencing rationales, and procedural arguments are treated as contextual metadata and excluded from narrative modeling. This separation reduces the risk of conflating factual accounts with judicial interpretation and aligns with established practices in the labeling of the rhetorical role of legal documents [10, 12]. At the same time, judicial narratives are post hoc reconstructions and should not be treated as exhaustive accounts of crime-scene dynamics. The goal is therefore not to recreate all investigative detail, but to provide a consistent narrative signal that enables event-oriented analysis, comparative study across cases, and grounded downstream modeling.

## 2.2. Document Segmentation

Each judicial judgment is segmented into paragraphs. Let a document  $d$  be represented as an ordered sequence of paragraphs  $(p_1, p_2, \dots, p_n)$ . To preserve local narrative coherence while limiting contextual drift, overlapping paragraph windows of fixed length  $k$  are constructed:

$$W(d) = \{w_i = (p_i, p_{i+1}, \dots, p_{i+k-1}) \mid 1 \leq i \leq n - k + 1\}.$$

Where available, section headings such as *Facts*, *The Offending*, or *Background* are treated as weak structural signals to prioritize narrative-dense regions. When headings are unreliable or absent, sliding windows preserve coverage throughout the document.

## 2.3. Lexicon-Guided Recall

An initial domain-aware lexicon  $L_0$  is constructed by forensic domain inspection. The lexicon includes behavioral verbs, motion indicators, references to force or instruments, and expressions commonly associated with crime narratives. The lexicon is used as a recall-oriented prior rather than as a strict rule-based filter, which aligns with previous work on weak supervision and lexicon-guided annotation [16].

A candidate window  $w$  is retained if it contains at least one lexical trigger:

$$\phi(w; L_t) = \mathbb{1}[\exists \ell \in L_t : \ell \in w].$$

The retained window set at iteration  $t$  is defined as:

$$\mathcal{C}_t(d) = \{w \in W(d) \mid \phi(w; L_t) = 1\}.$$

To reduce selection bias, windows that lack lexical triggers may still be sampled under the active learning regime so that previously unseen narrative expressions can be identified and incorporated.

## 2.4. Functional Classification

Sentences within retained windows are classified according to their functional role in the document, following rhetorical role labeling traditions [10, 11]. Each sentence  $x$  is assigned to one of four categories: Narrative Event (NE), Contextual Setting (CS), Procedural Legal (PL), or Other (OT).

A transformer-based encoder produces a contextual representation  $h = f_\theta(x)$ , which is passed to a linear classification head:

$$P_\theta(y = c \mid x) = \text{softmax}(Wh + b)_c.$$

The model is trained using cross-entropy loss over manually verified samples, consistent with standard fine-tuning approaches in legal NLP [14].

## 2.5. Human Verification and Active Learning

Human review is focused on sentences with high predictive uncertainty, estimated using entropy:

$$U(x) = - \sum_c P_\theta(y = c \mid x) \log P_\theta(y = c \mid x).$$

Sentences exceeding a predefined uncertainty threshold  $\tau$  are routed for expert annotation. Validated narrative expressions are incorporated into the lexicon through incremental updates:

$$L_{t+1} = L_t \cup \Delta L_t.$$

This iterative design follows established HITL and active learning practices that aim to improve annotation efficiency while maintaining quality [15, 16, 13].

## 2.6. Event Extraction and Timelines

Sentences classified as Narrative Event are converted into structured event templates:

$$e_j = (a_j, r_j, o_j, \ell_j, t_j),$$

where  $a_j$  denotes the agent,  $r_j$  the action,  $o_j$  the target,  $\ell_j$  the location reference, and  $t_j$  the temporal index. This representation is based on the principles of semantic role labeling and temporal event extraction [20, 17, 18].

Events are assembled into a case-level timeline:

$$T = (e_1, e_2, \dots, e_m), \quad t_1 \leq t_2 \leq \dots \leq t_m.$$

The representation is intentionally lightweight to ensure robustness across diverse judicial writing styles and to support scalable validation.

## 2.7. Forensic Crime Narrative Dataset

For each judicial case, the proposed framework returns a crime narrative dataset that includes structured metadata, sentence-level functional labels with confidence and verification status, event templates with temporal ordering, and narrative reconstructions linked to the source text. The release is intended to include annotation guidelines, inter-annotator agreement statistics [22], offense-type coverage analysis, and baseline benchmarks.

In addition to the annotated data, a frozen sentence-level functional classification model trained on the final verified corpus is released as a reference artifact. This model is provided to support reproducibility and comparative evaluation, and is not intended to represent an optimized or deployment-oriented system.

## 3. Validation and Limitations

The evaluation is expected to cover functional classification performance, event extraction quality, and annotation efficiency. Classification can be assessed using macro-averaged precision, recall, and F1 score. Event extraction quality can be measured via slot-level accuracy for event arguments, and efficiency can be quantified as a reduction in human annotation effort relative to non-HITL baselines [15, 16, 13].

The resulting corpus is intended to support research on investigative decision support, precedent-based crime pattern comparison, and knowledge grounding for mixed-reality systems that assist investigators in reconstructing event sequences. These applications should be treated as advisory rather than as determinative. The framework is designed to preserve provenance links to enable inspection and to discourage ungrounded inference.

Judicial narratives are post hoc reconstructions and may reflect legal priorities rather than investigative completeness. They may also encode biases embedded in judicial processes. These constraints require careful interpretation and appropriate safeguards in downstream use [9, 23].

## 4. Conclusion

The goal of this paper is to propose a human-centered approach to constructing a forensic crime narrative corpus from judicial records. By combining lexicon-guided recall, transformer-based functional classification, and uncertainty-driven selective human oversight, the proposed framework aims to address data scarcity while supporting transparency and auditability. The design is intended to provide a foundation for future empirical work on forensic narrative modeling and AI-assisted investigative reasoning.

## Author Contributions

Bilal Ahmed is the main contributor to this work and led the conceptualization of the study, the design of the proposed methodology, and the writing of the manuscript. Giovanni Acampora and Autilia Vitiello provided supervision, methodological guidance and critical feedback, with particular support on research formulation, machine learning design, and evaluation plan. All authors reviewed the manuscript.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] T. D. Wilson, D. C. Gambino, Crime scene investigation, in: Handbook of Forensic Psychology, 2 ed., Wiley-Blackwell, 2010.
- [2] A. Vitiello, C. Di Nunzio, L. Garofano, M. Saliva, P. Ricci, G. Acampora, Bloodstain pattern analysis as optimisation problem, *Forensic science international* 266 (2016) e79–e85.
- [3] G. Acampora, C. Di Nunzio, L. Garofano, M. Saliva, A. Vitiello, Applying density-based clustering for bloodstain pattern analysis, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2021, pp. 28–33.
- [4] G. Acampora, A. Vitiello, C. Di Nunzio, M. Saliva, L. Garofano, Bloodstain pattern analysis-a new challenge for computational intelligence community, in: International Conference on Fuzzy Computation Theory and Applications, volume 2, SCITEPRESS, 2014, pp. 211–216.
- [5] N. Galante, R. Cotroneo, D. Furci, G. Lodetti, M. B. Casali, Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives, *International journal of legal medicine* 137 (2023) 445–458.
- [6] G. Acampora, A. Vitiello, C. Di Nunzio, M. Saliva, L. Garofano, Towards automatic bloodstain pattern analysis through cognitive robots, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 2447–2452. doi:10.1109/SMC.2015.428.
- [7] D. M. Katz, M. J. Bommarito, J. Blackman, A general approach for predicting the behavior of the supreme court of the united states, *PLOS ONE* 12 (2017).
- [8] I. Chalkidis, I. Androutopoulos, N. Aletras, Neural legal judgment prediction in english, in: Proceedings of EMNLP-IJCNLP, 2019.
- [9] K. Jaidka, N. Khosla, S. Raj, M. Sap, Investigating biases in legal ai systems, in: AAAI/ACM Conference on AI, Ethics, and Society (AIES), 2023.
- [10] S. B. Majumder, D. Das, Rhetorical role labelling for legal judgements using transformer-based models, in: FIRE Workshop on Artificial Intelligence for Legal Assistance (AILA), 2020.
- [11] K. L. Cope, Rhetorical roles of legal documents: A survey, *International Journal of Legal Information* 42 (2016) 198–214.
- [12] A. Gupta, S. Malik, A. Jatowt, Rhetorical role labeling of legal documents using transformers and graph neural networks, in: SemEval-2023 Task 6, 2023.
- [13] L. Weber, P. Minervini, J. Benckendorff, et al., A human-in-the-loop improves annotation error detection, in: Findings of ACL, 2023.
- [14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of NAACL-HLT, 2019.
- [15] U. Hahn, E. Buyko, R. Landefeld, et al., Active learning-based corpus annotation: the PathoJen corpus, *Journal of Biomedical Semantics* 3 (2012).
- [16] S. Clancy, M. Cote, A. Seering, Active Learning with a Human in the Loop, Technical Report, MITRE, 2012.

- [17] O. Kolomiyets, M.-F. Moens, Extracting narrative timelines as temporal dependency structures, in: Proceedings of ACL, 2012.
- [18] W. Yao, B. Haghighi, H. Poon, S. Riedel, Temporal event knowledge acquisition via identifying before/after relations, in: Proceedings of ACL, 2018.
- [19] N. Chambers, D. Jurafsky, Unsupervised learning of narrative event chains, in: Proceedings of EMNLP, 2008.
- [20] Z. Zhang, E. Strubell, E. Hovy, Transfer learning from semantic role labeling to event argument extraction, in: Proceedings of EMNLP, 2022.
- [21] P. Kalamkar, U. Sreejith, B. Nayak, K. Shrivastava, Corpus for automatic structuring of legal documents, in: Proceedings of the Language Resources and Evaluation Conference (LREC), 2022.
- [22] K. Krippendorff, Reliability in content analysis, *Human Communication Research* 30 (2004) 298–310.
- [23] S. Barocas, A. D. Selbst, Big data’s disparate impact, *California Law Review* 104 (2016) 671–732.