



City Research Online

City St George's, University of London

Citation: Lukmanjaya, W., Halmich, C., Butler, T., Cook, D. & Karystianis, G. (2026). Computational text analysis on unstructured police data: a scoping review. *Crime Science*, 15(1), 6. doi: 10.1186/s40163-026-00272-2

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/37598/>

Link to published version: <https://doi.org/10.1186/s40163-026-00272-2>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

SYSTEMATIC REVIEW

Open Access



Computational text analysis on unstructured police data: a scoping review

Wilson Lukmanjaya^{1*} , Christina Halmich², Tony Butler¹, Darren Cook³ and George Karystianis¹

Abstract

Introduction Police reports made following attendance at various events (e.g., crashes, domestic violence, theft) often contain rich contextual details including indicators of mental health issues or abuse types, and persons/entities involved and their relationships, which are not typically captured in structured administrative data, interviews or official statistics. However, the sheer volume of information along with strict data access protocols render manual analysis impractical. Computational text analysis methods offer a feasible and effective approach to automatically process this underutilized data source.

Aim This article is an overview of studies using computational text analysis (e.g., text mining, natural language processing (NLP)), on unstructured police data, serving as a guide for researchers interested in employing similar methodologies.

Methods This scoping review was conducted in accordance with the PRISMA-SCR guidelines, following the two screening processes (title/abstract and full text screening) and the development of a pre-defined protocol. A search was conducted across seven electronic databases (ProQuest, IEEE Xplore, Scopus, PubMed, Web of Science, Criminal Justice Abstracts, Google Scholar) covering the past 20 years.

Results A total of 5426 records were identified. After removing duplicate entries and screening titles/abstracts and full-text publications, 61 studies met the inclusion criteria. Included studies were published between 2004 and 2024, with most from the United States, Australia and the Netherlands. Most studies used opensource tools: Bidirectional Encoder Representations from Transformers (BERT), natural language tool kit (NLTK), scikit-learn, or General Architecture for Text Engineering (GATE) to analyze unstructured police data. Our review indicates applications of computational text analysis on unstructured police data have moderate to high performance. Common limitations included variable data quality, with reliability depending on the level of detail provided by the police report's author, and failure to report ethical implications or methodological limitations.

Conclusions Computational text analysis can extract key information from unstructured police data. However, future research should clearly report ethics approvals and implications, and methodological limitations. Establishing a structured data-sharing framework between law enforcement and researchers is also crucial to facilitate access and support high quality, impactful research in this field.

*Correspondence:
Wilson Lukmanjaya
w.lukmanjaya@unsw.edu.au

Full list of author information is available at the end of the article

Keywords Automated text analysis, Text mining, Natural language processing, Machine learning, Police data, Unstructured data

Introduction

Police data are widely utilized for many purposes, ranging from criminal investigations and predictive policing to public safety planning and policy development. Datasets typically consist of structured administrative records (e.g., date, time, crime classifications, sex, postcode) and unstructured (i.e., textual) narratives that describe contextual and chronological details of various event types (e.g., crashes, domestic violence) (Karystianis et al., 2021). Police forces routinely utilize structured data to produce national crime statistics and inform decision-making (BOCSAR, 2025; Federal Bureau of Investigation, 2025). However, the New South Wales Bureau of Crime Statistics and Research (BOCSAR) acknowledges that such statistics may not accurately reflect the true incidence of crime within communities, due to reporting bias, changes in law enforcement and reporting policies, and transient visitor populations (BOCSAR, 2025). Still, reliance on structured data remains the most efficient approach for prevalence analysis, as it eliminates the need for additional data collection or complex data transformation (Mishra & Misra, 2017). While structured data support the systematic reporting of crime statistics that drive research, policy evaluation, and evidence-based decision-making, unstructured narratives remain largely untapped, despite offering rich contextual details, due to sheer volume, strict access protocols, data privacy, and lack of technological expertise (Birks et al., 2020; Karystianis et al., 2018; Sarzaeim et al., 2023).

Developing structured police data usually involves specific data processing procedures or coding that is often not independently verifiable due to strict access (Goldstein et al., 2023; Rahwan et al., 2019). Overreliance on this structured data could diminish context and reduce transparency in reporting the broader narrative, particularly when legal frameworks evolve or when nuanced, topic-specific inquiries arise (Lukmanjaya et al., 2025). Reports from government institutions that rely on these structured data have been shown to limit the representativeness of the reported statistics. For instance, Karystianis et al. (2024a) found that coercive control is often underreported in domestic violence events. Cases of underreporting and underrepresentation are not exclusive to domestic violence, or to the area of justice overall. Octoman et al. (2023) concluded that child abuse and neglect was also underreported after analyzing child protection narratives in South Australia. Whereas Golder et al. (2025) in their scoping review on automated analysis on unstructured health data concluded that identifying adverse events and previously unknown safety signals

would not be possible by using structured data alone. Solely using structured data can result in blind spots, where important patterns or emerging issues can be overlooked, contributing to the underreporting of crime along with a distorted evidence base that informs legal and policy decisions (Joh, 2018; Moses & Chan, 2016). These limitations underscore the emerging opportunity of using computational text analysis, a set of techniques designed to systematically and automatically extract information from unstructured data, offering a powerful means to uncover patterns, themes, and signals that structured data may otherwise miss (Shah et al., 2024).

Computational text analysis encompasses methods such as text mining, natural language processing (NLP), and machine learning, all of which automate the extraction of information from narrative data without requiring intensive manual effort. These techniques have been widely applied across various fields such as law, healthcare, computer science, and business (Joorabchi et al., 2016; Netzer et al., 2012; Ramon-Gonen et al., 2023; Vidler et al., 2023). Within unstructured police data, researchers have employed approaches including NLP (Parker, 2025), machine learning (e.g., Bidirectional long-short term memory (Bi-LSTM)) (Attota et al., 2024), and topic modelling (Kuang et al., 2017) with varying degrees of success. These methods can efficiently extract information, generate labels for supervised machine learning, and process large volumes of narrative reports, thereby reducing time and human error. More recently, advances in large language models (LLMs) have enabled pipelines capable of detecting information from text (Liu et al., 2025) and predicting events such as crimes or vehicle crashes (Dakalbab et al., 2022; Jaradat et al., 2025). Despite these developments, there has been little systematic evidence on how and why these methods have been applied to unstructured police data. Providing such an overview would present researchers with a clearer direction and a deeper understanding of the opportunities and limitations to inform appropriate use of methods on police data. The aim of this paper is to conduct a scoping review to identify studies that have applied computational text analysis techniques in police unstructured data and inspect their findings, performance and limitations.

This article is structured as follows: firstly, it outlines the aim and objectives of the research, secondly it describes the methodology of the scoping review (i.e., the search strategy, inclusion criteria, data extraction process, and included studies), and thirdly, it presents the results, grouped by the extracted information. Lastly, a

discussion on the findings, limitations, future implications and concluding remarks is presented.

Research aim and objectives

This scoping review aims to map and synthesize existing research within the past 20 years from six bibliographic databases and Google Scholar on the application of computational text analysis techniques (NLP, machine learning, LLMs) for analyzing unstructured police data, with a focus on four objectives:

1. To identify and summarize current methodologies used to analyze unstructured police data.
2. To examine challenges and limitations associated with the collection, processing, and interpretation of unstructured police data, including issues related to data quality, bias, and privacy.
3. To assess the impact of unstructured police data analysis on policy decisions, law enforcement outcomes, and public safety practices.
4. To highlight knowledge gaps and recommend future research directions for improving the use and integration of unstructured data in policing.

Methods

A scoping review of existing academic literature was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses—Scoping Reviews (PRISMA-SCR) guidelines (Tricco et al., 2018). To ease the inter-annotator agreement, Covidence (Covidence systematic review software, 2025) was used to streamline the screening, data extraction, and management processes. No review protocol was prepared for this scoping review.

Eligibility criteria

Studies were included in the review if they met all the following criteria:

- (A) Peer-reviewed academic publication.
- (B) Use of computational text analysis techniques.
- (C) Utilized unstructured police-produced data (e.g., narrative reports, event descriptions).
- (D) Published within the last 20 years at the time of the search (1 January 2004 to 18 November 2024).
- (E) Published in English or have been translated in English.

Information sources

Seven databases, ProQuest, IEEE Xplore, Scopus, PubMed, Web of Science, Criminal Justice Abstracts, and Google Scholar, were searched on 18 November 2024 for all records containing terms relating to computational text analysis *and* police *and* unstructured data

for the Title and Abstract fields, dating back to 2004 (see “Search” section). An iterative process of trial and error along with a librarian consultation was conducted to refine the search query. To validate the effectiveness of the search strategy, relevant articles known to the authors prior to the search were compared with the results to ensure they were included. Inclusion of articles known to the authors ensures that the search strategy encapsulates relevant articles, minimizing search bias.

Search

For Google Scholar, a simplified approach was adopted due to limited search string capability, as follows: “police data text analysis”. The complete search strategy for the remaining six bibliographic databases is as follows:

(“artificial intelligence” OR “intelligent system” OR “cognitive computing” OR “artificial cognition” OR “intelligent automation” OR “machine learning” OR “statistical learning” OR “algorithmic modeling” OR “adaptive learning” OR “supervised learning” OR “unsupervised learning” OR “unlabeled learning” OR “clustering” OR “cluster detection” OR “pattern discovery” OR “data driven learning” OR “automatic detection” OR “automated identification” OR “machine based detection” OR “machine-based detection” OR “machine detection” OR “autonomous detection” OR “automatic recognition” OR “machine based recognition” OR “machine-based recognition” OR “machine recognition” OR “computer recognition” OR “computer-based recognition” OR “computer based recognition” OR “pattern recognition” OR “autonomous recognition” OR “recognition method” OR “labeled learning” OR “guided learning” OR “predictive modeling” OR “predictive modelling” OR “task-driven learning” OR “task driven learning” OR “text mining” OR “text analytic” OR “textual data mining” OR “document mining” OR “information discovery” OR “natural language processing” OR “computational linguistic” OR “language technology” OR “text understanding” OR “textual analysis” OR “document analysis” OR “language analysis” OR “deep learning” OR “neural network modeling” OR “neural network modelling” OR “deep neural network” OR “deep structured learning” OR “knowledge mining” OR “data discovery” OR “knowledge extraction” OR “insights discovery” OR “data preprocessing” OR “algorithmic technique” OR “algorithmic method” OR “text analysis” OR “data retrieval” OR “data extraction” OR “entity extraction” OR “automated reading” OR “machine reading” OR “autonomous reading” OR “machine comprehension” OR “word vectorization” OR “semantic embedding” OR “language representations” OR “information retrieval” OR “information extraction” OR “machine reading” OR “word embedding” OR “feature extraction” OR “knowledge discovery” OR “data engineering” OR “knowledge engineering” OR “exploratory data analysis” OR “quantitative content

analysis" OR "automatic content analysis" OR "computational method" OR "big data" OR "analysis" OR "decision tree" OR "random forest" OR "support vector machine" OR "TF/IDF" OR "TF-IDF" OR "regular expression" OR "regex" OR "rule-based" OR "rule based" OR "classification model" OR "prediction model" OR "descriptive statistic") AND ("police" OR "policing" OR "constabulary" OR "law enforcement" OR "legal enforcement" OR "public security") AND ("unstructured" OR "free-text narrative" OR "free text narrative" OR "narrative report" OR "non-standardized" OR "non standardized" OR "non-standardised" OR "non standardised" OR "non codified" OR "non-codified" OR "raw text" OR "unformatted text" OR "open narrative" OR "descriptive text" OR "descriptive report" OR "case summary" OR "case summaries" OR "case report" OR "report narrative" OR "qualitative report" OR "incident description" OR "incident report" OR "textual report" OR "textual data" OR "field note" OR "evidence report" OR "forensic statement" OR "police report" OR "police narrative" OR "police record" OR "police data" OR "police document" OR "investigative text" OR "interrogation text" OR "police-reported" OR "police reported"))).

Selection of sources of evidence

Literature identified through database searches was imported into EndNote Version 20.6 (Clarivate Analytics, 2023) to remove duplicates and Covidence for data management. After duplicates were removed, two reviewers (WL, GK) performed title and abstract screening, and full-text screening according to the inclusion criteria. Inter-Annotator Agreement was calculated with a Cohen's Kappa of 0.49, indicating moderate agreement (Landis & Koch, 1977). Disagreements were resolved following a discussion between the two reviewers and a third expert (CH) on machine learning with a majority vote ruling for points of disagreement.

Data charting process

The final dataset was inspected and confirmed by the two reviewers (WL, GK) to create a template containing information on:

Article type; Title; Year; First-author affiliation; Country of study; Topic of study; Methods; Open source; Tools used; Study aim; Language; Type of Police data; Dataset size; Mention of ethics; Evaluation methods used; Best performing tool; Performance of best performing tool; Method limitations.

Data items (Table 1)

Critical appraisal of individual sources of evidence

No critical appraisal was conducted for this study, consistent with the objectives of scoping reviews to map the

extent and nature of evidence rather than assessing its quality.

Synthesis of results

Article type was included to ensure the exclusion of non-peer-reviewed publications. Country of study was recorded and further categorized as high-income, upper middle-income, lower middle-income and low to middle income countries based on the World Bank's classification for statistical purposes (Metreau, Young & Eapen, 2024). First-author affiliations were also recorded to identify relevant institutions involved in the studies.

The topic of study was categorized into three broad domains (crime, crash, administrative). Sources of police data were classified by jurisdictional level (e.g., local, state, national). If a study analyzed multiple datasets and police data was one of them, we included only the police dataset in this review. For instance, Barros et al. (2023) employed both a dataset of 1000 federal police documents and a large dataset of over 110,000 WikiHow articles. For our purposes, only the police dataset was reported.

The study methods used were reported to confirm that each study employed a computational text analysis technique. The best performing model was identified from the reported results or author's conclusions. Performance evaluation was gathered from the reported results (if applicable) or summarized as macro-average when deemed appropriate or when reported as such. In cases where studies presented varying results across different models or did not report clear evaluation results, performance was noted as a range when variation occurred across models, or as a macro-average, (average of reported performance metrics) when variation occurred within the same model. Performance levels were categorized as follows: 0–0.5 (Low performance), 0.51–0.8 (Moderate), and 0.81–1.0 (High), respective to the used metrics (e.g., precision, recall, F1-Score, area under the curve).

Due to the diversity in datasets and their size, it was not appropriate to summarize these data quantitatively. Therefore, we did not conduct any analysis for dataset size. Study aims were often closely tied to the nature of the data and the problem context, making aggregation impractical. Evaluation methods were briefly summarized, focusing on the most reported approaches.

Some reports were part of the same dataset and were treated as part of one study (Neubauer et al., 2023). Additionally, conceptual approaches and algorithms such as decision trees, artificial neural networks, and formal concept analysis were not classified as tools for the purposes of this analysis since they are dependent on specific research aims, model parameters, and design. As the aim of this scoping review was to provide an overview

Table 1 Description of information extracted from the final dataset of studies

Variable	Definition	Examples
Article type	Type of the article	Peer-reviewed article
Title	Title of the article	Measuring cybercrime in calls for police service
Year	Year of publication	2024
First-author affiliation	Institution affiliated with the first-author	University College London
Country of study	The country origin of the police data	South Korea, Australia
Topic of study	The topic being studied	Cybercrime, pedestrian crash
Methods	Methods reported by the authors that are relevant to computational text analysis	Machine learning, text classification, naive bayes
Open source	Open-source status of the tools utilized within the study	Yes/No
Tools used	Specific tools mentioned	R, Python, GATE
Study aim	Study aim	Identification, classification
Language	The language of the data	English, German
Type of police data	Descriptions of police data types	Korean calls for Police Service, New South Wales Police Force
Dataset size	Dataset size	15,000,000 calls, 492,393 narratives
Mention of ethics	Whether authors mentioned obtaining ethics or discuss ethical issues	Yes/No/Unspecified
Evaluation methods used (if eligible)	Evaluation methods used for authors' proposed methodologies	Precision, recall, f1-score, accuracy, area under curve
Best performing tool (if eligible)	Reported best performing algorithm/tool respective to authors' proposed methodology and evaluation methods	BERT, novel proposed algorithms/pipelines, XGBoost
Performance of best performing tool (if eligible)	Reported performances of the best performing algorithm/tool respective to authors' proposed methodology and evaluation methods	Accuracy: 93.45% Precision: 96.94% Recall: 88.35% F1-Score: 92.44%
Method limitations	Reported limitations of authors' methods	Data quality challenges, single dataset, overlapping definitions

of existing research and the current state of progress on computational text analysis using unstructured police data, these methods were excluded. They were only included if they were mentioned as the best performing model by the authors.

Results

Selection of sources of evidence

The search was done in two stages, one for the academic databases and the other for Google Scholar. The academic database search yielded 5126 results, of which 1,868 (approx. 36%) were duplicates and 664 were not peer-reviewed and thus, excluded. As Google Scholar provided an estimation of more than two million results, we decided to extract the first 300 as recommended in Haddaway et al. (2015), of which 50 were duplicates. The final set included 2783 abstracts. The abstract screening process resulted in 111 full text studies. In cases of uncertainty, three authors (WL, GK, CH) collectively examined and discussed the publication in question. 50 studies were excluded for the following reasons: use of structured data only (n=18), no use of police data (n=9), not peer-reviewed (n=8), full text not found/not open access (n=7), did not use computational text analysis (n=5), and poor English/not English (n=3). A total of 61 full text studies were included in this review (Fig. 1).

A number of papers (n=26) reported using the same data sources. These were the Karystianis et al. studies on New South Wales Police Force (Karystianis et al., 2018, 2019, 2020), Poelmans et al. studies on Amsterdam-Amstelland Police (Poelmans et al., 2009, 2010, 2011a, 2011b, 2012), Das et al. studies on accident vehicle crashes (Das et al., ; Olliae et al., 2024; Weng et al., 2023), Carnaz et al., (2021a, 2021b) studies on Portuguese criminal-related documents, and Goldstein et al. (2023) and Zhou et al. (2023) on the US National Violent Death Reporting System database.

Time period of studies

The included studies were all published within the last 20 years from 2004 and 2024. A large proportion of studies were conducted after 2019, with fewer studies pre-2020 (Fig. 2).

Topics

Three thematic categories were identified across the final dataset: crime and policing (n=42), motor vehicle crashes and injuries (n=12), and administrative applications (n=7). All studies that did not specifically mention the topic of their studies were classified as administrative applications.

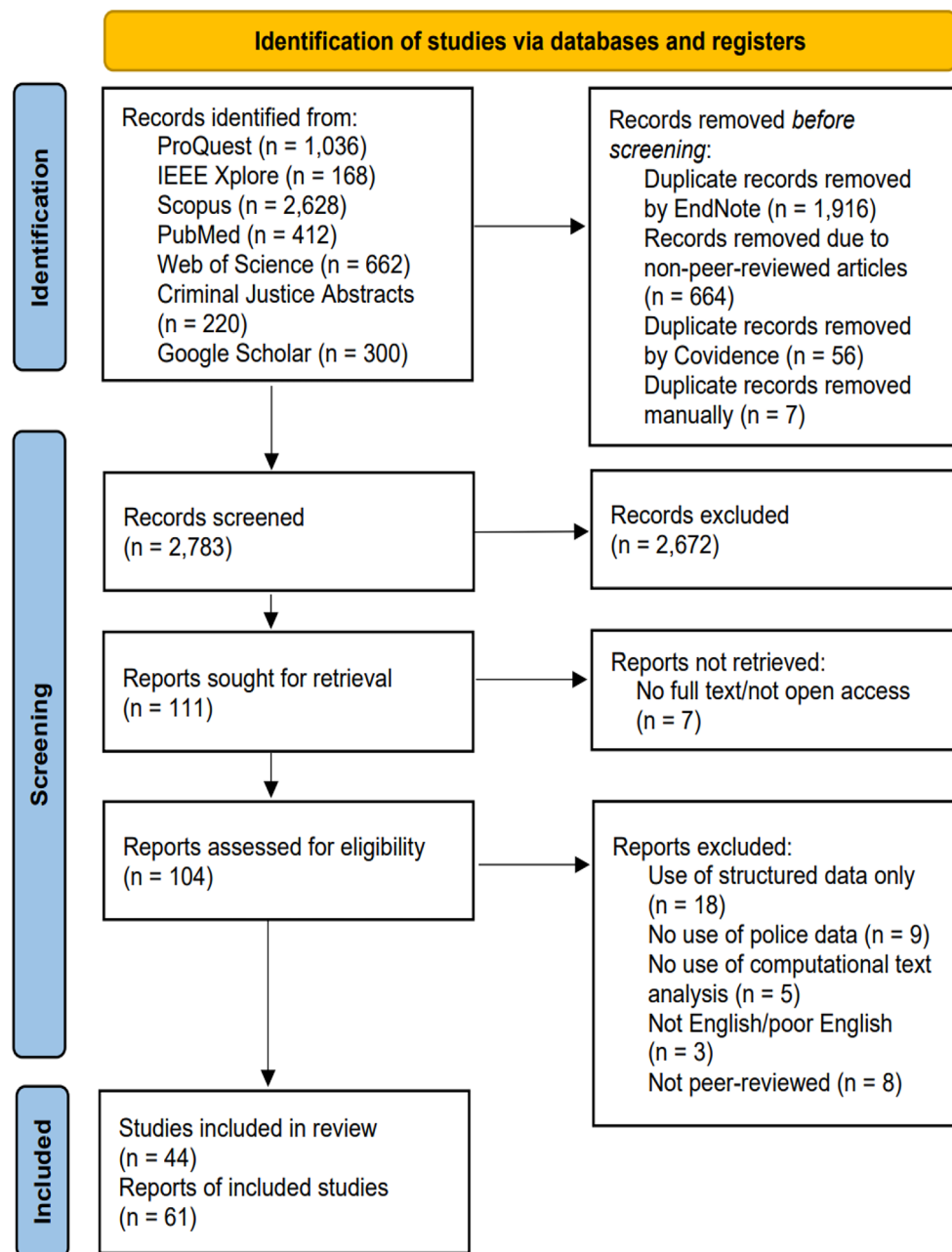


Fig. 1 PRISMA chart for scoping review of computational text analysis on unstructured police data (Page et al., 2021)

For crime and policing, the studies were distributed across several subtopics: 17 examined domestic violence (2021, 2022a, 2022b, 2024a, 2024b; Hodgkinson et al., 2023; Hwang et al., 2020; Karystianis et al., 2018, 2019, 2020; Poelmans, et al., 2009, 2010, 2011a, 2011b; Reutens et al., 2024; Wilson et al., 2022; Yu et al., 2021), 11 focused on policing practices (Baek et al., 2021; Basilio et al., 2020; Birks et al., 2020; Carnaz et al., 2021a, 2021b; Chen et al., 2004; Geurts et al., 2021; Haensch et al., 2024; Qiu et al., 2023; Quijano-Sanchez et al., 2018; Xing et al., 2024; Zhu & Xie, 2022), four examined mental health

(Attota et al., 2024; Brown et al., 2024; Karystianis et al., 2020, 2022a, 2022b), three addressed criminal investigations (Elyezjy et al., 2015; Helbich et al., 2013; Poelmans et al., 2012), two examined cybercrime (Kwon et al., 2024; Palad et al., 2020), two analyzed suicides (Goldstein et al., 2023; Zhou et al., 2023), and four focused on financial crime (Neumann et al., 2016), property crime (Navarrete et al., 2024), hate crime (Ortiz Salazar et al., 2024), and protests (Bekker, 2022).

For studies involving crashes and injuries, eight studies (Das et al., 2020a; Kwayu et al., 2020; Sayed et al.,

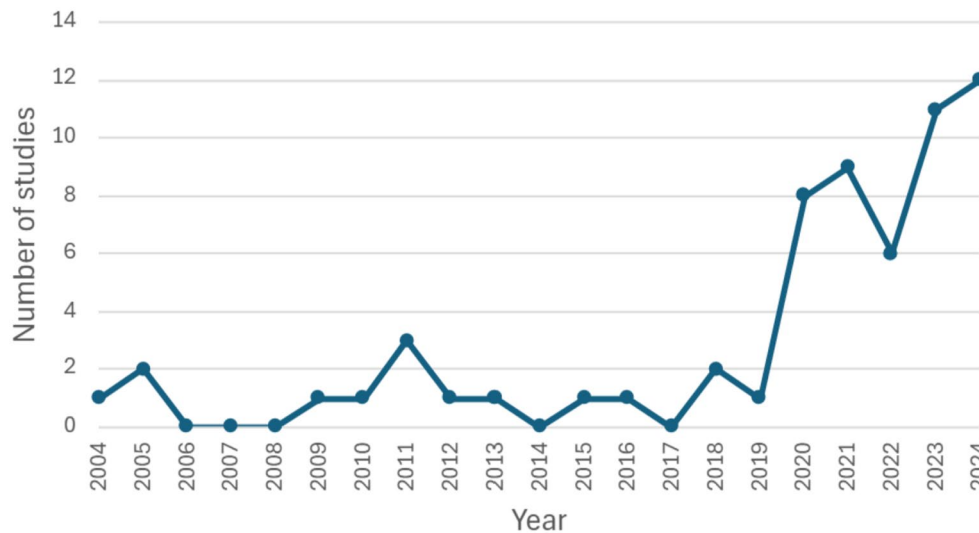


Fig. 2 Number of peer-reviewed, published studies ($n=61$) from 2004 to 2024 using computational text analysis methods on unstructured police data

2021; Brauner et al., 2022; Hosseini et al., 2022; Seo et al., 2023; Shipp et al., 2023; Bäumlner et al., 2024) focused on vehicle-related crashes, whereas four studies (Das et al., 2020b, 2023; Oliaee et al., 2024; Weng et al., 2023) addressed pedestrian-related incidents.

Within the administrative domain (Barros et al., 2023; Carnaz et al., 2021a, 2021b; Giannone et al., 2011; Juez-Hernandez et al., 2023; Li et al., 2005; Neves Oliviera et al., 2022; Wu & Pottenger, 2005), the primary focus was either on the development of tools designed to alleviate police workload or was unspecified. These tools did not exclusively cover criminal activity data, as police records also encompass various matters, such as missing person reports, lost item inquiries, and other administrative tasks, but remained mostly unspecified within the study description.

Country and affiliations

Most studies ($n=52$) were conducted in high-income countries. A majority (60/61) of first authors were university-affiliated with one affiliated with a police department. Most studies came from the United States ($n=21$) followed by Australia ($n=11$) and the Netherlands ($n=6$) (Table 2).

Datasets, methods, and evaluations used

The datasets utilized across the 61 studies were sorted alphabetically by country and classified according to police jurisdictional level: local, state, federal/national, and special units (Table 3).

A total of 23 studies reported 12 unique datasets of state police data (Karystianis et al., 2018, 2019, 2020; 2021; 2022a; 2022b; 2024a; 2024b; Hwang et al., 2020; Wilson et al., 2022; Reutens et al., 2024; Baumler et al., 2024; Hosseini et al., 2022; Zhu & Xie, 2022; Das et al.,

2020a, 2020b, 2023; Weng et al., 2023; Oliaee et al., 2024; Kwayu et al., 2020; Yu et al., 2021; Shipp et al., 2023; Sayed et al., 2021), 13 studies reported 12 unique datasets containing federal or national police agencies data (Baek et al., 2021; Barros et al., 2023; Bekker, 2022; Birks et al., 2020; Brauner et al., 2022; Geurts et al., 2021; Goldstein et al., 2023; Hodgkinson et al., 2023; Juez-Hernandez et al., 2023; Kwon et al., 2024; Quijano-Sanchez et al., 2018; Seo et al., 2023; Zhou et al., 2023), 16 studies reported 11 unique datasets contained local police departments data (Neves Oliviera et al., 2022; Qiu et al., 2023; Xing et al., 2024; Elyezjy et al., 2015; Poelmans et al., 2009, 2010, 2011a, 2011b, 2012; Langton et al., 2021; Attota et al., 2024; Brown et al., 2024; Wu & Pottenger, 2005; Haensch et al., 2024; Chen et al., 2004; Ortiz Salazar, 2024), one study from a judicial source (Navarrete et al., 2024), one study contained military police data (Basilio et al., 2020), one study was from the United States' Federal Bureau of Investigation (Helbich et al., 2013). Six studies used five unique datasets but did not specify their source (Carnaz et al., 2021a, 2021b; Giannone et al., 2011; Li et al., 2005; Neumann et al., 2016; Palad et al., 2020).

Four studies (Helbich et al., 2013; Li et al., 2005; Poelmans et al., 2012; Sayed et al., 2021) did not specify the tool(s) and performance metrics employed in their analyses. Most studies that reported their primary programming languages utilized Python ($n=9$) and R ($n=4$). Bidirectional Encoder Representations from Transformers (BERT) was the most cited pre-trained language model ($n=7$), while a number of studies ($n=6$) also incorporated libraries such as scikit-learn and natural language tool kit (NLTK). General Architecture for Textual Engineering (GATE) was used in 12 studies: 11 by Karystianis et al. and another by Elyezjy et al. (2015). Only three studies used LLMs (Goldstein et al., 2023;

Table 2 Included studies (n = 61) on computational text analysis on unstructured police data sorted by geographic location, economic classification, and first-author affiliation

Continent	Country of police data	Economic classification	First-author affiliation	Number of papers	Studies	
North America	United States (n = 21)	High-income country	Texas A&M University	5	Das et al., (2020a, 2020b); Shipp et al. (2023); Weng et al. (2023); Oliaea et al. (2024)	
			Kennesaw State University	2	Attota et al. (2024); Brown et al. (2024)	
			Lehigh University	2	Li et al. (2005); Wu & Pottenger (2005)	
			Georgia Institute of Technology	1	Zhu and Xie (2022)	
			Rowan University	1	Hosseini et al. (2022)	
			Seattle Police Department*	1	Ortiz Salazar et al. (2024)	
			Texas State University	1	Das et al. (2023)	
			Tufts University	1	Haensch et al. (2024)	
			University of Arizona	1	Chen et al. (2004)	
			University of Heidelberg**	1	Helbich et al. (2013)	
			University of Pennsylvania	1	Yu et al. (2021)	
			University of Utah	1	Goldstein et al. (2023)	
			University of Washington	1	Zhou et al. (2023)	
			University of Wisconsin-Milwaukee	1	Sayed et al. (2021)	
Oceania	Australia (n = 11)	High-income country	Western Michigan University	1	Kwayu et al. (2020)	
			University of New South Wales	10	Karystianis et al., (2021, 2022a, 2022b, 2024a, 2024b, 2018, 2019, 2020), Hwang et al. (2020), Reutens et al. (2024)	
Europe	The Netherlands (n = 6)	High-income country	Curtin University	1	Wilson et al. (2022)	
			KU Leuven**	5	Poelmans et al., (2009, 2010, 2011a, 2011b, 2012)	
	Spain (n = 3)	High-income country	Radboud University	1	Geurts et al. (2021)	
			Autonomous University of Madrid	1	Juez-Hernandez et al. (2023)	
			Carlos III University of Madrid	1	Quijano-Sánchez et al. (2018)	
	United Kingdom (n = 3)	High-income country	University of Malaga	1	Navarrete et al. (2024)	
			Manchester Metropolitan University	1	Langton et al. (2021)	
	Germany (n = 2)	High-income country	University of Leeds	1	Birks et al. (2020)	
			University of Leicester	1	Hodgkinson et al. (2023)	
			Dresden University of Technology	1	Bäumler et al. (2024)	
			Leibniz University Hannover	1	Brauner et al. (2022)	
			University of Evora	2	Carnaz et al., (2021a, 2021b)	
	Portugal (n = 2)	High-income country	University of Roma	1	Giannone et al. (2011)	
	Italy (n = 1)	High-income country	University of Koblenz**	1	Neumann et al. (2016)	
Asia	South Korea (n = 3)	High-income country	University of Koblenz**	1	Neumann et al. (2016)	
			Advanced Institute of Convergence Technology	1	Seo et al. (2023)	
			Electronics and Telecommunications Research Institute	1	Baek et al. (2021)	
	China (n = 2)	Upper middle-income country	University College London**	1	Kwon et al. (2024)	
			Nanjing Forest Police College	1	Qiu et al. (2023)	
	Philippines (n = 1)	Lower middle-income country	People's Public Security University of China	1	Xing et al. (2024)	
			Mindanao State University	1	Palad et al. (2020)	
	Palestine (n = 1)	Lower middle-income country	The Islamic University of Gaza	1	Elyezzy et al. (2015)	
	South America	Brazil (n = 3)	Upper middle-income country	Federal University of Campina Grande	1	Barros et al. (2023)

Table 2 (continued)

Continent	Country of police data	Economic classification	First-author affiliation	Number of papers	Studies
			Federal University of Ceara	1	Neves-Oliviera et al. (2022)
			Fluminense Federal University	1	Basilio et al. (2020)
Africa	South Africa (n = 1)	Upper middle-income country	University of Johannesburg	1	Bekker (2022)

*Non-university affiliation

**First-author affiliations geographically different than country of police data. University of Heidelberg—Germany; KU Leuven—Belgium; University of Koblenz—Germany; University College London—United Kingdom

Xing et al., 2024; Zhou et al., 2023). Most studies used datasets in English (n=37) followed by Dutch (n=6) (Online Resource 1).

Computational text analysis methods have been shown to yield promising results (>0.8 respective to their metrics) particularly when it comes to information extraction (Karystianis et al., 2018, 2019, 2024a; Xing et al., 2024; Giannone et al., 2011; Seo et al., 2023; Juez-Hernandez et al., 2023; Wu & Pottenger, 2005; Kwayu et al., 2020; Chen et al., 2004), identification (Attota et al., 2024; Brown et al., 2024; Elyezjy et al., 2015; Hwang et al., 2020; Karystianis et al., 2020, 2022a, 2024b; Qiu et al., 2023; Reutens et al., 2024; Wilson et al., 2022), classification (Attota et al., 2024; Brown et al., 2024; Das et al., 2023; Goldstein et al., 2023; Kwon et al., 2024; Poelmans et al., 2009, 2010, 2011a, 2011b), and prediction (Baek et al., 2021; Karystianis et al., 2021; Zhou et al., 2023). These methods are also performing well in data concordance (Karystianis et al., 2022b), anonymization and automatic mapping (Juez-Hernandez et al., 2023), estimating proportions (Langton et al., 2021), analysis and detection (Chen et al., 2004; Hodgkinson et al., 2023), and surveillance (Shipp et al., 2023).

Precision, recall, and F1-score (Ananiadou et al., 2006) were the most common evaluation metrics (n=23). Most studies demonstrated moderate to high performance (n=47; >0.5) across their metrics indicating general effectiveness. Only two studies (Goldstein et al., 2023; Zhou et al., 2023) had performance ranging from low to high (~0.25–0.90) depending on the task, highlighting the variability in model effectiveness across different use cases. Table 3 shows all the studies included in this scoping review categorized by dataset source and size.

Limitations and implications

A common limitation reported across many studies (n=19) was the insufficiency or inconsistency of the training data. Some studies (n=10) did acknowledge concerns regarding the generalizability of their findings, noting that the performance of their models may not translate effectively to unseen or future datasets. Additionally, a small number of studies (n=4) highlighted potential bias stemming from the variability in the quality of police reports as each jurisdiction has its own

reporting standards combined with the individual officer's reporting practice.

Despite these limitations, most of the studies (n=41) hinted at their proposed methods' practical value in supporting police investigations and surveillance efforts. The findings indicate that, while there are challenges related to data quality and generalizability, the integration of such tools into law enforcement data workflows may enhance investigative efficiency and support evidence-based decision making.

Ethics

None of the reviewed studies discussed any potential ethical implications of their work. Most studies (n=43; 70.5%) did not specify whether their research had received ethics approval. Only 15 studies (24.6%) reported ethics approval, and four (6.6%) stated the lack of ethics approval. One study (Goldstein et al., 2023) provided a rationale for the absence of their ethical approval, noting deidentified data on deceased individuals are not classified as human research.

Discussion

This scoping review examined published articles on the application of computational text analysis methods applied to unstructured police data. While the integration of machine learning and NLP tools is gaining traction with promising results, there is still room for further development in this area. The findings showed an increase in related publications from 2020, highlighting a growing interest in the automated processing of unstructured police data for several purposes ranging from extraction to prediction. This may reflect heightened awareness of the rich information contained in unstructured police data, policy changes, easier access to this type of data source or the appeal of applying emerging technologies such as multimodal LLMs (e.g., ChatGPT, Claude) that require much less technical expertise for their implementation. The surge in research outputs post-2019 could also signify a shift in academic and professional priorities, as more scholars and institutions engage with research in both the area of computational text analysis and law enforcement.

Table 3 Included studies (n = 61) on computational text analysis on unstructured police data categorized by dataset source and size

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Karystianis et al. (2018)	New South Wales Police Force	Australia	State	492,393–526,787 domestic violence reports	GATE	Rule-based text mining	Extraction of mental health mentions	English	Precision: 0.85–0.98 Recall: 0.79–0.90 F1-score: 0.84–0.90	High
Karystianis et al. (2019)							Extraction of abuse types and victim injuries			
Karystianis et al. (2020)							Identification and analysis of mental health mentions			
Hwang et al. (2020)							Identification and analysis of domestic violence events involving people with autism			
Karystianis et al. (2021)					BERT		Predict domestic violence offenses			
Karystianis et al. (2022a)					GATE		Identification and analysis of abuse types, victim injuries, and mental health mentions			
Karystianis et al. (2022b)							Concordance of mental illness mentions from police narratives and hospital and emergency department records			
Wilson et al. (2022)							Identification and analysis of non-fatal strangulation in domestic violence events			
Karystianis et al. (2024b)							Identification and analysis of male victims in heterosexual relationships in domestic violence events			
Reutens et al. (2024)							Identification and analysis of domestic violence events involving people with dementia			
Karystianis et al. (2024a)				406,196 coercive control reports	GATE		Extraction of coercive control behaviors			

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Barros et al. (2023)	Brazilian Federal Police	Brazil	Federal/National	1000 documents	BERT BERTSUM	BERTSUM-ALD-MD	Information extraction and text summarization	Portuguese	ROUGE-1: 35.01 ROUGE-2: 21.65 ROUGE-L: 33.36	Moderate
Neves Oliveira et al. (2022)	Fortaleza Police	Brazil	Local	3083 annotated real narrative reports (393,550 tokens from police records)	Fast Text GLOVE Spacy Wang2Vec Word2Vec	Hierarchical entity-label disambiguation	Multi entity labelling: person (victim, parents, involved) and location (death, residence)	Portuguese	Macro-average: Precision: 0.66 Recall: 0.67 F1-score: 0.66	Moderate
Basilio et al. (2020)	Military Police of Rio de Janeiro	Brazil	Military Police	38,660 text files	J-Electre v.2 R	Latent Dirichlet Allocation with ELECTRE I	Topic modelling and knowledge discovery	English	N/A	N/A
Qiu et al. (2023)	A police station under the administration of a branch bureau	China	Local	23,678 data records	N/A	Backpropagation neural network model	Identification of ex-convict cases	Chinese	> 0.95 for precision, recall and f1-score	High
Xing et al. (2024)	Beijing Municipal Public Security Bureau	China	Local	110 emergency responses	Baichuan ChatGLM Llama Mistral Qwen Yi	Fine-tuned LLM with designed prompt	Information extraction	Chinese	Precision: 0.93 Recall: 0.92	High
Neumann et al. (2016)	Police interrogations	Countries in the European Union	Unspecified	Unspecified	AutoMAP ORA	Text mining and semantic network	Applying semantic network analysis on an existing money laundering case to understand complex relationships	English	N/A	N/A
Bäumler et al. (2024)	Police accident data from five German federal states	Germany	State	7622 road traffic accidents	BERT CatBoost	CatBoost for unprepared text features, BERT for augmented text features	Predicting 30 types of turning accidents	German	Unprepared text: CatBoost F1-score: 0.77 BERT F1-score: 0.76 Augmented text: CatBoost F1-score: 0.76 BERT F1-score: 0.77	Moderate
Brauner et al. (2022)	Police press releases from Presseportal.de	Germany	Federal/National	2560 press articles	BeautifulSoup ForceAtlas2 Gephi Simplemma TextBlob TextHero Tfidfvectorizer	Keyword extraction, network analysis, Louvain algorithm for clustering, and sentiment analysis	Identification and classification of causes of e-scooter crashes	German	Modularity: 0.118	Low

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Giannone et al. (2011)	Unspecified	Italy	Unspecified	96 documents	N/A	Bag-of-words kernel + word-sequence kernel	Relation mining of subjects	Italian	Accuracy: 0.88 Precision: 0.75 Recall: 0.85 F1-score: 0.80	Moderate
Elyezji et al. (2015)	Theft documents from police departments in the Gaza strip	Palestine	Local	Unspecified	GATE	Rule-based text mining	Criminal networks identification	Arabic	Precision: 0.84 Recall: 0.97 F1-score: 0.89	High
Palad et al. (2020)	Police incident reports	Philippines	Unspecified	49,822 words	Weka	J48 decision tree	Classification of online scam reports	Filipino	Accuracy: 0.67	Moderate
Camaz et al. (2021a, 2021b)	Unspecified	Portugal	Unspecified	163 documents (3 documents containing Apache Tika 4781 words from police reports)	Apache OpenNLP Apache Tika Newspaper3k Neo4j NLNet NLPPort	Semantic crime framework or SEMCrime	Extraction and classification of named-entities and relations in Portuguese, and graph representation	Portuguese	Mean precision: 0.81 Mean recall: 0.82 Mean f1-score: 0.73	Moderate
Bekker (2022)	South African Police Service IRIS	South Africa	Federal/National	156,000 police records	N/A	Gradient boosted model	Knowledge discovery of protest events	English and Afrikaans	Validation misclassification: 0.38 Validation Kolmogorov-Smirnov: 0.86	Moderate-high
Seo et al. (2023)	Daegu traffic crashes from the Korean Traffic Crash Analysis System	South Korea	Federal/National	2427 records	BERT	BERT	Information extraction on exact collision points and pre-crash vehicle maneuver	Korean	F1-score: 0.95–1.00	High
Baek et al. (2021)	Korea Information System of Criminal Justice Services	South Korea	Federal/National	5000 criminal cases	N/A	Convolutional neural networks-based prediction model	Crime type and risk level prediction	Korean	Accuracy: 0.91 Precision: 0.92 Recall: 0.82 F1-score: .84	High
Kwon et al. (2024)	Korean National Calls for Police Service	South Korea	Federal/National	15,000,000 call logs	R	Kernel SVM	Crime categories classification: sex crime, violent crime, fraud, juvenile delinquency, theft, others	Korean	Accuracy: 0.93 Precision: 0.97 Recall: 0.88 F1-score: 0.92	High

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Navarrete et al. (2024)	Police reports on property crimes from the Council of the Judiciary and the Dean of Judges of Malaga City	Spain	Special—Judiciary	Unspecified	N/A	Ontology, reasoner, and information extraction tool	Ontological reasoning to recommend applicable legal articles with confidence scores for judicial decision-making	Spanish	Fleiss Kappa: 0.471	Moderate
Juez-Hernandez et al. (2023)	Spanish National Police reports provided by Spanish National Office Against Hate Crimes	Spain	Federal/National	319 police reports	FlairNLP Transformers Tensorflow	model trained ad hoc in the FlairNLP framework, using a Bi-LSTM or GRU network with a CRF classifier combined with advanced embeddings otherwise also known as AGORA (Analysis and Geolocalization of Reports Anonymized)	Anonymization, information extraction, and automatic mapping	Spanish	Police domain f1-score: 0.93 Medical domain f1-score: 0.95	High
Quijano-Sánchez et al. (2018)	Spanish National Police	Spain	Federal/National	1122 robbery reports	N/A	Ridge Logistic Regression	Automated false report detection through probability	Spanish	Accuracy: 0.98 Precision: 0.78 F1-score: 0.78	Moderate
Poelmans et al., (2009, 2010, 2011a, 2011b)	Amsterdam-Amstelland Police	The Netherlands	Local	4814 police reports	N/A	Formal concept analysis and emergent self-organizing maps	Feature discovery for improved classification accuracy on domestic violence and non-domestic violence cases	Dutch	Accuracy: 0.89–0.90+	High
Poelmans et al. (2012)				266,157 reports	N/A	Formal concept analysis	Human trafficking suspects and victim identification		N/A	N/A
Geurts et al. (2021)	Dutch National Police	The Netherlands	Federal/National	116,680 victim records	N/A	Logistic regression with interaction term	Assessing the risk of repeat victimization	Dutch	AUC [95% CI] for severe violent crimes: 0.81 [0.80–0.83] AUC [95% CI] for less severe violent crimes: 0.77 [0.76–0.79]	Moderate

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Langton et al. (2021)	Greater Manchester Police	United Kingdom	Local	3,650,000 incidents	N/A	Convolutional neural network trained with human annotated data	Estimating the proportion and severity of calls-for-service involving persons with mentally ill health	English	Cross-validation with police incident logs and text mining algorithm indicate 90% agreement	High
Birks et al. (2020)	UK Police	United Kingdom	Federal/National	9621 residential burglary incidents	Gensim Mallet	Topic modelling using latent dirichlet allocation	Cluster identification of specific crime problems	English	N/A	N/A
Hodgkinson et al. (2023)	UK Police	United Kingdom	Federal/National	43,488 domestic abuse crimes	BERT Anomaly Mosum Package Transformers pytorch	Fine-tuned BERT model	Analyzing COVID-19's impact on domestic abuse trends	English	F1-Score: 0.84	High
Hosseini et al. (2022)	Alabama and Illinois statewide crash database	United States	State	421 crash reports	BERT Scikit-learn	BERT	Identification of wrong-way driving crashes	English	Average accuracy: 0.82 Average precision: 0.78 Average F1-score: 0.78	Moderate
Zhu and Xie (2022)	Atlanta Police Department	United States	State	10,056 crime incidents	NLTK	Spatio-temporal point process with a regularized restricted Boltzmann machine	Crime linkage detection	English	F1-score: 0.19–0.20	Low
Attota et al. (2024)	Cobb County Police	United States	Local	300 reports	Spacy Keras modal Tensorflow	BiLSTM network for narrative text, artificial neural network for categories, trainable layers for terms text	Behavioral health cues identification and classification	English	Accuracy: 0.85 F1-score: 0.85	High
Brown et al. (2024)						Experimental proposed model with embeddings and self-attention mechanism			Accuracy: 0.94 F1-score: 0.94	High

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Weng et al. (2023)	Crash data from Austin, Dallas, San Antonio from Texas Department of Transportation's CRIS	United States	State	4442 crash narratives	BERT ktrain RoBERTa Transformers	Binary classification model	BERT model with model-agnostic meta learning algorithm	English	Accuracy: 0.44	Low
Das et al. (2023)						Pre-trained BERT base embedding network and fine-tuned linear transformation layer	Binary classification for intentionality and multiclass classification for pedestrian maneuver	English	Weighted average f1-score: 0.84 Macro average f1-score: 0.64	Moderate-high
Oliaee et al. (2024)						RoBERTa model with BCE loss function	Classification of crash types	English	Accuracy: 0.65 F1-score: 0.51	Moderate
Das et al. (2020a)	Dallas and Austin			90 from Dallas 295 from Austin	XGBoost	XGBoost	Classification of crash types	English	Up to: Sensitivity: 0.75 Specificity: 0.64 Accuracy: 0.70 Precision: 0.71 F1-score: 0.73	Moderate
Das et al. (2020b)	Police-reported crashes in Louisiana			652 crash records			Classification of number of vehicle involvements in a crash	English	Sensitivity: 0.70 Specificity: 0.74 Accuracy: 0.71 Precision: 0.85 F1-score: 0.77	Moderate
Wu & Pottenger (2005)	Fairfax County police incident reports, university crime reports, and patents	United States	Local	100 police incident reports	N/A	Semi-supervised active learning algorithm	Pattern discovery in information extraction	English	Up to 100% for precision, recall and F1-score	High
Helbich et al. (2013)	FBI Orion database accessed by Jennings Police Task Force	United States	Special	172 "Information Packages"	N/A	Unsupervised neural network and point pattern analysis	Knowledge discovery	English	N/A	N/A
Kwayu et al. (2020)	Michigan's crash reports	United States	State	12,671 crashes	R	Support vector machines using unigram and bigram features	Information extraction	English	Accuracy: 0.86 Area under ROC: 0.92	High

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Goldstein et al. (2023)	National Violent Death Reporting System Restricted Access Database	United States	Federal/National	1462 female firearm suicide cases	BeautifulSoup NLTK Scikit-learn	Support vector machine and gradient boosting models	Identification and classification off interpersonal disputes, problems with intimate partners, acute/chronic pain, and intimate partners and immediate family	English	F1-scores across five categories: 0.60–0.82	Moderate-High
Zhou et al. (2023)					ChatGPT FLAN-T5 FLAN-UL2	FLAN-UL2	Prediction of infrequent circumstances preceding female firearm suicides	English	F1-score: ~0.25–0.90 (presented in a graph)	Low–High
Haensch et al. (2024)	Northville Police Department	United States	Local	3800 printed pages	Fuzzywuzzy Geopandas Numpy OpenCV Rasterio Shapely Tesseract Optical Character Recognition engine	Comprehensive data science pipeline	Data science pipeline from collection of printed pdf pages to knowledge discovery	English	N/A	N/A
Yu et al. (2021)	Philadelphia Police Department	United States	State	220 domestic violence incidents	R	Proposed optimal matching algorithm	Integrate matching methods to combine quantitative and qualitative causal inference in English	English	N/A	N/A
Chen et al. (2004)	Phoenix Police Department; Tucson Police Department	United States	Local	36 from Phoenix PD 272 from Tucson PD	N/A	Deceptive identity detection	Named entity extraction, deceptive identity detection, and criminal network analysis	English	Training: 0.97 Testing: 0.94	High
Ortiz Salazar (2024)	Seattle Police Department	United States	Local	Unspecified	Scikit-learn XGBoost	XGBoost	Hate crime detection	English	Macro-average precision: 0.80 Macro-average f1-score: 0.79	Moderate

Table 3 (continued)

Studies	Dataset source	Country of dataset	Jurisdictions	Dataset size	Methods used	Best performing model	Study aim/model task	Language	Performance evaluation	Performance level
Shipp et al. (2023)	State-based crash records for Arkansas, Louisiana, New Mexico, Oklahoma, and Texas and fatal crash records from the National Highway Traffic Safety Administration's Fatality Analysis Reporting System	United States	State	1,024,767 narratives	NLTK Scikit-learn (Kim et al., 2021)	Bag of Keywords-support vector classifier (SVC), Bag of Keywords-Multinomial Naive Bayes, Bag of Words-SVC (Kim et al., 2021)	Injury surveillance	English	F1-score: 0.88, 0.89, 0.85 (Kim et al., 2021)	High
Li et al. (2005)	Unspecified	United States	Unspecified	Unspecified	N/A	Distributed higher order association rule mining	Information extraction within a distributed environment	English	N/A	N/A
Sayed et al. (2021)	Wisconsin Department of Transportation	United States	State	377,479 crash reports	N/A	Unigram + bigram noisy-OR classifier	Identification and analysis of misclassified work-zone crashes	English	N/A	N/A

Patterns across study themes, methods, and datasets

The included studies showed a focus on certain crime categories, particularly domestic violence ($n = 17$). However, this thematic imbalance was driven in part by a small number of prolific research groups, such as those led by Poelmans et al., (2009, 2010, 2011a, 2011b, 2012) and Karystianis et al., (2018, 2019, 2020, 2021, 2022a, 2022b, 2024a, 2024b; Hwang et al., 2020; Wilson et al., 2022; Reutens et al., 2024). While their contributions demonstrated the effectiveness of text mining in extracting insights from unstructured police narratives, they also revealed a concentration on a narrow set of topics both focusing almost exclusively on domestic violence. Conversely, there is a paucity of research on other crime types, including white-collar crime, drug offences, property offences, or other criminal activities. In addition, there have not been any computational text analysis efforts within the data of specialized agencies like border control or anti-corruption units to tackle immigration or corruption tasks. This imbalance points out the need for a broader application across different crime types and law enforcement agencies.

The demonstrated adaptability of computational text analysis suggests that automated approaches could be applied to a wider array of criminal activities or incidents with potentially promising performance. Still, reported strong performance may partly reflect publication bias, since studies with weaker or unsatisfactory results are less likely to appear in the literature. For example, Karystianis et al. (2019) identified abuse types from police narratives related to domestic violence but further expanded their methods to extract coercive control behaviors in slightly different datasets still stemming from the same law enforcement jurisdiction (i.e., New South Wales Police Force) with promising success (Karystianis et al., 2024a). Similarly, Das et al. (2020a) classified different crash types from police crash reports in Dallas and Austin, USA, using Extreme Gradient Boosted (XGBoost) models, but further expanded it to Louisiana police crash reports to classify the number of vehicles involved in a crash. Therefore, expanding the thematic scope of research would not only aid the generalizability of findings but ensure that technological benefits would be distributed more equitably across criminal justice systems.

Most studies were conducted within (or at least obtained data from) a small number of countries, particularly the United States, Australia, and the Netherlands. However, it is important to contextualize these findings. For instance, all 11 Australian studies (2021, 2022a, 2022b, 2024a, 2024b; Hwang et al., 2020; Karystianis et al., 2018, 2019, 2020; Reutens et al., 2024; Wilson et al., 2022) used two police datasets from one state (i.e., New South Wales). Similarly, the six Dutch studies (Geurts et al., 2021; Poelmans et al., 2009, 2010, 2011a, 2011b,

2012) originate from two data sources: The Amsterdam-Amstelland Police reports and Dutch National Police victim records. By contrast, the United States appear to have a broader adoption of access to police unstructured data with studies reporting the use of data from all levels of law enforcement including local, state and federal agencies, as well as specialized units such as the Federal Bureau of Investigation (Helbich et al., 2013). Interestingly, no published studies in the Nordic countries were found out in this review despite research suggesting leading outputs in epidemiological criminology (or justice health) research (Karystianis et al., 2022a, 2022b).

The remaining number of countries (e.g., South Africa, Philippines, Italy) and the lack of representation of other countries in the literature points to systemic barriers in police data use and sharing (Jones et al., 2024). Law enforcement agencies in many jurisdictions remain hesitant to open up their narrative data to researchers, even in anonymized form, due to privacy concerns, security risks, or institutional conservatism (Allen et al., 2023) employing strict access data protocols that do not allow for the flexibility of implementing complex computational text analysis pipelines. Consequently, much of the existing research relies on isolated collaborations between individual researchers and police departments, which, while valuable, are difficult to replicate, sustain or scale (Brown, 2018). This fragmented landscape underscores the need for both national and international frameworks and partnerships that facilitate responsible data sharing and encourage research in underrepresented regions.

Dataset sizes ranged from small (e.g., 36 narratives) to large scale (e.g., more than 15 million records). However, studies were not categorized according to size, as the volume and richness of the text within each record varied substantially. A smaller dataset with detailed narratives may offer as much analytical value as a much larger dataset with brief or sparse textual information. For example, Kwon et al. (2024) excluded approximately 1.8 million out of 15 million call log entries that contained fewer than nine Korean characters which rendered them insufficient for meaningful analysis, whereas Chen et al. (2004) who used a small sample of 36 narratives from the Phoenix Police Department to create a crime data mining framework to extract information showcased that using smaller and more detailed samples were sufficient in developing a working pipeline. These differences highlight the need to consider both quantity and quality of the police text when designing pipelines and applications of computational text analysis of unstructured police data, since limited text can limit the amount of training data while low-quality text can reduce generalizability and lead to noisy results (Sajith & Kathala, 2024). Studies in this field should report document information such as text length,

number of unique words, and stop word dictionary used for preprocessing for transparency and quality assurance (Hickman et al., 2022; Kwon et al., 2024).

However, while dataset quality and quantity are important considerations, future research should consider the context surrounding that data. As reviewed studies reported that data were imbalanced, non-generalizable, and possibly biased (Das et al., 2023; Goldstein et al., 2023; Karystianis et al., 2021), collecting information on how the data is developed would make their evidence more robust since it might include confounding variables that are potentially useful for future works. This highlights the importance of clearly reporting the data source, giving other researchers the opportunity to verify and reproduce the reported findings. We recommend future research to consult with police data experts within policing agencies to ensure accuracy in context reporting.

Computational tools

A trend observed across the included studies was the frequent use of widely available open-source tools, particularly Python. The most employed libraries and frameworks include BERT, NLTK, and scikit-learn (Barros et al., 2023; Goldstein et al., 2023; Zhu & Xie, 2022). These tools offer a robust foundation for various NLP tasks enabling researchers to perform tasks such as text classification, entity recognition, and topic modelling with ease (Koroteev, 2021; Varoquaux, 2015). Their application was noted to be incremental on unstructured police data which could indicate that accessing law enforcement narratives and implementing technical tools are not major barriers for research as they were a few years ago; rather, researchers appear to leverage accessible technologies to perform sophisticated analyses involving information extraction and unlock new insights in areas such as domestic violence and road safety.

Despite computational text analysis methods being applied in a diverse range of tasks (e.g., information extraction, identification, classification, prediction), they are heavily dependent on specific aims, algorithms, and dataset (Marciniak, 2016). Thus, it is challenging to determine which model is the most optimal for any given application. For instance, Kwayu et al. (2020) used support vector machines (SVM) for information extraction on crash reports and Xing et al. (2024) used fine-tuned LLM for information extraction on emergency responses, while Seo et al. (2023) achieved higher accuracy for information extraction on crash reports using BERT. However, it is not appropriate to claim that one model outperforms the other, as contextual variations including country, jurisdiction, dataset, language, task, and modeling approach make it difficult to assess the direct comparison of the suitability of their methods. Such processes would be better examined in future studies once more research

in this area has been conducted within the same context and constraints.

Despite the widespread use of these open-source tools, a noticeable gap exists in the adoption of more recent innovations, particularly LLMs. Only three studies (Goldstein et al., 2023; Xing et al., 2024; Zhou et al., 2023) incorporated LLMs in their methodologies, with this limited uptake likely reflecting a combination of technical and ethical concerns (e.g., retaining confidentiality and privacy of individuals noted in these narratives), insufficient data availability, and restrained research timeframe. LLMs such as ChatGPT (OpenAI et al., 2023) or LLaMa (Grattafiori et al., 2024) offer state-of-the-art performance across a range of NLP tasks, often requiring less domain-specific fine-tuning than previous models (e.g., BERT, NLTK) (Zhou et al., 2023). On the other hand, their deployment in sensitive domains like law enforcement raises significant concerns related to data security, model opacity, and potential for privacy breaches. The “black-box” nature of many LLM based applications, whereby internal decision-making processes are largely inscrutable, can be problematic when transparency and accountability are paramount (Rudin, 2019). Explainable artificial intelligence (XAI) might help mitigate this problem through interpretable models that could explain how models arrive at their predictions, but robust evaluation metrics are still necessary to improve XAI (Ali et al., 2023) as well as input from human expertise (e.g., human-in-the-loop) (Butcher et al., 2024). Nevertheless, understanding how these “advanced” models can reach specific conclusions is a necessary step in the law enforcement area to maintain boundaries between computational inference and human judgment, ensuring that humans make the final decision through justifiable means (Lukmanjaya et al., 2025).

Study limitations

This review highlights several methodological and reporting challenges from published research in this area. Several studies ($n = 18$) did not report any description of the tools used. While reporting concepts and approaches might be sufficient for replication in other fields, technical differences are particularly important in computational text analysis. Programming language differences (i.e., architectural differences between Python and R) might produce slightly different results due to varying algorithmic designs (Hill et al., 2024) or model stochasticity. Not reporting the programming languages or software used, or failing to provide the notebooks or scripts produced, potentially render results to be neither comparable nor replicable.

Furthermore, nine studies (including three with unsupervised methods) did not report evaluation metrics which decreases the reporting quality and relevance to

future works as their results are not measurable. While unsupervised tasks such as topic modelling, formal concept analysis, and network analysis might be justifiable to exclude a quantifiable evaluation method (Basilio et al., 2020; Helbich et al., 2013; Poelmans et al., 2012), providing an evaluation method would make the results comparable with future research. A key challenge in this research area is that confidentiality concerns might limit the method and evaluation reporting. However, while such confidentiality concerns may prevent full disclosure of certain data elements, omitting key methodological details (e.g., tools, evaluation metrics) diminishes reproducibility and impedes cumulative progress in this research area (Øby, 2025; Semmelrock et al., 2025).

Finally, ethical approvals and implications were not reported in most studies included in this review. This absence could have negative implications if research in this area is directly translated to social outcomes or if the models developed are used in any type of production environment. However, that does not necessarily imply that research in this field did not seek to obtain the appropriate ethics approvals through approved channels; it may simply reflect an omission in reporting. Nevertheless, having ethic approvals and implications for such data ensures that sensitive research is less biased and impartial.

Limitations

This scoping review comes with a few limitations. Firstly, our study included only peer-reviewed articles, excluding grey literature such as PhD theses, government reports, internal police reports, and conference proceedings. This exclusion may have led to the omission of valuable insights or emerging evidence not published in academic journals, which could explain why Nordic countries were not represented in our final dataset. Secondly, while the initial search on the ProQuest database yielded 1305 records, only 1036 records were extractable and suitable for screening, potentially affecting the coverage of this review. Thirdly, the literature was sourced from only seven academic databases. Although these databases were selected for their relevance and coverage, it is possible that additional studies in other databases were missed. Fourthly, 1/3 of the included studies came from three author-groups (Karystianis et al., Poelmans et al., and Das et al.) which might skew this scoping review's methodology and conclusions, and despite a research intake in this area from 2020, there is still a lack of applying computational text analysis on police unstructured data. The classification of the studies' topics across three categories might also have an effect on the result reporting. Potentially, breaking them down into different categories might show different insights.

In addition, the classification of performance into low (0–0.5), moderate (0.51–0.8), and high (0.81–1.0) categories was defined by the authors, which may introduce subjectivity and limit generalizability. The search query focused on “police” data and not “crime” records. While the exclusion of the term “crime” may result in the omission of some relevant studies, its inclusion was tested which led to the retrieval of a large volume of records beyond the scope of this review which often reported gathering “crime data” from various sources (e.g., news sites, social media, etc.) that are not directly from police. The inter-rater reliability also shows a Cohen's Kappa of 0.49 with moderate agreement, indicating that the screeners (WL, GK) might be inconsistent in selecting the criteria, but were mitigated with discussion with a third author (CH). Finally, LLM terminologies were not included in the search strategy due to variability and their potential inclusion of studies with “generative” aims which were also out of scope for this review (e.g., LLMs in predictive policing based on non-police data).

Future implications

Despite the studies' limitations, the overall performance of the models used across jurisdictions was reported to be moderate to high, indicating the viability and effectiveness of computational text analysis in policing applications. The predominance of open-source tools suggests that technological barriers to entry are relatively low, allowing for a broader base of researchers, especially those in resource-constrained settings, to participate in this emerging field. Future research on computational text analysis of police unstructured data should also explore other topics other than domestic violence and crashes, and work towards building the required infrastructure to access data from untapped law enforcement agencies such as anti-corruption and border-control units.

Applying computational text analysis on police data highlights the promising potential of this research area and the importance of establishing data-sharing agreements between policing jurisdictions and academic institutions. While providing access to sensitive data carries privacy and security risks, future works could benefit in the long-term if implemented appropriately (Lukmanjaya et al., 2025). Such initiatives could mitigate the under-reporting problems and reveal unknown patterns, ultimately enhancing resource efficiency and informing evidence-based decision-making.

One promising area for future exploration is the use of LLMs as an alternative to traditional rule-based and supervised machine learning models, which often require significant domain expertise and manual effort to build and maintain. LLMs offer potentially higher performance and lower learning curves using prompt engineering and

transfer learning, thus making them more accessible to cross-disciplinary research teams and law enforcement personnel (Brown et al., 2020). However, realizing this potential will require rigorous evaluation, stakeholder consultation, and the development of best practice tailored to sensitive criminal justice contexts (Council on Criminal Justice, 2024; Lukmanjaya et al., 2025). Moving forward, researchers should strive to strike a balance between privacy protection and transparency by clearly documenting their analytical workflows, anonymization techniques, performance benchmarks and ethical considerations.

To responsibly harness the capabilities of LLMs in law enforcement contexts, future research must establish and adhere to clear protocols for data protection, explainability and bias mitigation (Jobin et al., 2019; Raji et al., 2020). This includes developing strategies to evaluate model outputs rigorously and ensure they align with legal and ethical standards and domain expertise. Researchers should also explore the use of hybrid models that combine the interpretability of rule-based systems with the flexibility of machine learning. Establishing safeguards could facilitate the ethical use of LLMs, enabling researchers and law enforcement agencies to benefit from cutting edge NLP technology while minimizing associated risks (Zhang et al., 2023). Reporting transparency of data and model evaluation is also key to developing robust systems from police unstructured data.

Conclusion

This scoping review examined 61 studies that described the application of computational text analysis on unstructured police records. A total of 44 unique police datasets were identified across the included studies, ranging from very small (e.g., 36 narratives) to large scale (e.g., 15 million records). Our findings demonstrate that research interest in this area has increased from 2020 onwards, with most topics revolving around crime, motor vehicle crashes, and aiding police in administrative tasks. Most studies used programming languages such as Python and R with BERT-based models for text analysis generating moderate to high performances in a variety of tasks that include information extraction and text classification. Ethical implications were omitted by most studies, lacking any reference to ethics approval. This scoping review highlights the need to develop data-sharing frameworks between police agencies and researchers to support transparency, reproducibility, and ethical research practices.

Abbreviations

BERT	Bidirectional encoder representations from transformers
BERTSUM	Bidirectional encoder representations from transformers for summarizations

BERTSUM-ALD-MD	BERTSUM arbitrarily long document with multi-domain
Bi-LSTM	Bidirectional long-short term memory
BOCSAR	Bureau of crime statistics and research
CATBOOST	Categorical boosting
FLAN-T5	Fine-tuned language net on text-to-text transfer transformers
FLAN-UL2	Fine-tuned language net on unifying language learning
GATE	General architecture for textual engineering
GLoVE	Global vectors for word representations
LLMs	Large language models
PDF	Portable document format
PRISMA-SCR	Preferred reporting items for systematic reviews and meta-analyses@extension for scoping review
RoBERTa	Robustly optimized BERT pretraining approach
SVC	Support vector classifier
SVM	Support vector machine
NLP	Natural language processing
NLTK	Natural language tool kit
ORA	Organizational risk analyzer
XAI	Explainable artificial intelligence
XGBOOST	Extreme gradient boosting
WEKA	Waikato environment for knowledge analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40163-026-00272-2>.

Additional file 1.

Acknowledgements

The authors would like to thank Jennifer Whitfield for her valuable assistance and expert consultation in developing the search strategy for this scoping review.

Author contributions

WL contributed to study initialization, literature review, method implementation, result analysis, result interpretation, manuscript initialization, and revision. CH was responsible for result analysis, result interpretation, and manuscript revision. TB was responsible for result interpretation, manuscript revision, and study supervision. GK was responsible for study conception and initialization, method implementation, result interpretation, manuscript revision, and study supervision. All authors contributed to the manuscript and approved the submitted version.

Funding

Open Access funding enabled and organized by CAUL and its Member Institutions

Data availability

All data are publicly available.

Declarations

Competing interests

The authors declare no conflicts of interest related to this study.

Author details

¹School of Population Health, University of New South Wales, Level 5, 55 Botany St, Randwick, Sydney, NSW 2031, Australia

²University of Salzburg, Salzburg, Austria

³City St George's, University of London, London, UK

Received: 6 October 2025 / Accepted: 21 January 2026

Published online: 15 February 2026

References

- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Diaz-Rodriguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- Allen, C. I., Payne, S. H., & Valentine, J. L. (2023). Ethical data sharing in forensic research. *Forensic Science International: Synergy*, 6, 100322. <https://doi.org/10.1016/j.fsfsyn.2023.100322>
- Ananiadou, S., Kell, D. B., & Tsujii, J. (2006). Text mining and its potential applications in systems biology. *Trends in Biotechnology*, 24(12), 571–579. <https://doi.org/10.1016/j.tibtech.2006.10.002>
- Attota, D. C., Azmee, A. A., Khan, M. A. A. H., Pei, Y., Thomas, D., & Nandan, M. (2024). Semantic learning and attention dynamics for behavioral classification in police narratives. *Smart Health*, 32, Article 100479. <https://doi.org/10.1016/j.suhl.2024.100479>
- Baek, M., Park, W., Park, J., Jang, K., & Lee, Y. (2021). Smart policing technique with crime type and risk score prediction based on machine learning for early awareness of risk situation. *IEEE Access*, 9, 131906–131915. <https://doi.org/10.1109/ACCESS.2021.3112682>
- Barros, T. S., Pires, C. E. S., & Nascimento, D. C. (2023). Leveraging BERT for extractive text summarization on federal police documents. *Knowledge and Information Systems*, 65, 4873–4903. <https://doi.org/10.1007/s10115-023-01912-8>
- Basilio, M., Brum, G., & Pereira, V. (2020). A model of policing strategy choice: The integration of the Latent Dirichlet Allocation (LDA) method with ELECTRE I. *Journal of Modeling in Management*. <https://doi.org/10.1108/JM2-10-2018-0166>
- Bäumler, M., & Prokop, G. (2024). Predicting the type of road traffic accident for test scenario generation. *IEEE Access*, 12, 28950–28971. <https://doi.org/10.1109/ACCESS.2024.3367744>
- Bekker, M. (2022). Better, faster, stronger: Using machine learning to analyse South African police-recorded protest data. *South African Review of Sociology*, 52(1), 4–23. <https://doi.org/10.1080/21528586.2021.1982762>
- Birks, D., Coleman, A., & Jackson, D. (2020). Unsupervised identification of crime problems from police free-text data. *Crime Science*, 9, Article 18. <https://doi.org/10.1186/s40163-020-00127-4>
- Brauner, T., Heumann, M., Kraschewski, T., Prahlow, O., Rehse, J., Kiehne, C., & Breitter, M. H. (2022). Web content mining analysis of e-scooter crash causes and implications in Germany. *Accident Analysis and Prevention*, 178, 106833. <https://doi.org/10.1016/j.aap.2022.106833>
- Brown, R. (2018). Understanding law enforcement information sharing for criminal intelligence purposes. Australian Insititute of Criminology, 566. Retrieved from https://www.aic.gov.au/sites/default/files/2020-05/ti_le_information_sharing_2018.pdf
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Amodei, D. (2020). Language Models are Few-Shot Learners. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.
- Brown, M., Azmee, A. A., Khan, M. A. A. H., Thomas, D., Pei, Y., & Nandan, M. (2024). Adaptive attention-aware fusion for human-in-the-loop behavioral health detection. *Smart Health*, 32, 100475. <https://doi.org/10.1016/j.suhl.2024.100475>
- Bureau of Crime Statistics and Research (BOCSAR). (2025). Using Crime Statistics. NSW Government, Retrieved from https://bocsar.nsw.gov.au/statistics-dashboards/crime-and-policing/using_crime_statistics.html
- Butcher, B., Zilka, M., Hron, J., Cook, D., & Weller, A. (2024). Optimising human-machine collaboration for efficient high-precision information extraction from text documents. *ACM Journal on Responsible Computing*, 1(2), 1–27. <https://doi.org/10.1145/365259>
- Carnaz, G., Antunes, M., & Nogueira, V. B. (2021a). An annotated corpus of crime-related Portuguese documents for NLP and machine learning processing. *Data*, 6(7), 71. <https://doi.org/10.3390/data6070071>
- Carnaz, G., Nogueira, V. B., & Antunes, M. (2021b). A graph database representation of Portuguese criminal-related documents. *Informatics*, 8(2), 37. <https://doi.org/10.3390/informatics8020037>
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, 37(4), 50–56. <https://doi.org/10.1109/MC.2004.1297301>
- Clarivate. (2023). *EndNote* (Version 20.6) [Reference management software]. Retrieved from <https://endnote.com/>
- Council on Criminal Justice. (2024). The Implications of AI for Criminal Justice. Retrieved from <https://counciloncj.org/the-implications-of-ai-for-criminal-justice/>
- Covidence systematic review software. (2025). Veritas Health Innovation, Melbourne, Australia. Retrieved from <https://www.covidence.org>
- Dakalbab, F., Talib, M. A., Waraga, O. A., Nassif, A. B., Abbas, S., & Nasir, Q. (2022). Artificial intelligence & crime prediction: A systematic literature review. *Social Sciences & Humanities Open*, 6(1), 100342. <https://doi.org/10.1016/j.ssaho.2022.100342>
- Das, S., Dutta, A., Dey, K., Jalayer, M., & Mudgal, A. (2020b). Vehicle involvements in hydroplaning crashes: Applying interpretable machine learning. *Transportation Research Interdisciplinary Perspectives*, 6, 100176. <https://doi.org/10.1016/j.trip.2020.100176>
- Das, S., Le, M., & Dai, B. (2020a). Application of machine learning tools in classifying pedestrian crash types: A case study. *Transportation Safety and Environment*, 2(2), 106–119. <https://doi.org/10.1093/tse/tdaa010>
- Das, S., Oliaee, A. H., Le, M., Pratt, M. P., & Wu, J. (2023). Classifying pedestrian maneuver types using the Advanced Language Model. *Transportation Research Record: Journal of the Transportation Research Board*, 2677(7), 599–611. <https://doi.org/10.1177/03611981231155187>
- Elyezji, N. T., & Elhalees, A. M. (2015). Investigating crimes using text mining and network analysis. *International Journal of Computer Applications*, 126, 19–25. <https://doi.org/10.5120/ijca2015906134>
- Federal Bureau of Investigation. (2025). Crime data explorer: Crime Trends. U.S. Department of Justice. Retrieved from <https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/crime-trend>
- Geurts, R., Raaijmakers, N., Delsing, M. J. M. H., Spapens, T., Wientjes, J., Willems, D., & Scholte, R. H. J. (2021). Assessing the risk of repeat victimization using structured and unstructured police information. *Crime & Delinquency*, 69(9), 1736–1757. <https://doi.org/10.1177/00111287211047533>
- Giannone, C., Basili, R., Naggari, P., & Moschitti, A. (2011). Supervised semantic relation mining from linguistically noisy text documents. *International Journal on Document Analysis and Recognition*, 14, 213–228. <https://doi.org/10.1007/s10032-010-0138-0>
- Golder, S., Xu, D., O'Connor, K., Wang, Y., Batra, M., & Hernandez, G. G. (2025). Leveraging natural language processing and machine learning methods for adverse drug event detection in electronic health/medical records: A scoping review. *Drug Safety*, 48, 321–337. <https://doi.org/10.1007/s40264-024-01505-6>
- Goldstein, E. V., Mooney, S. J., Takagi-Stewart, J., Agnew, B. F., Morgan, E. R., Haviland, M. J., Zhou, W., & Prater, L. C. (2023). Characterizing female firearm suicide circumstances: A natural language processing and machine learning approach. *American Journal of Preventive Medicine*, 65(2), 278–285. <https://doi.org/10.1016/j.amepre.2023.01.030>
- Grattafiori, A. et al. (2024). The Llama 3 Herd of Models. *ArXiv*. Retrieved from <https://arxiv.org/abs/2407.21783>
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, 10(9), e0138237. <https://doi.org/10.1371/journal.pone.0138237>
- Haensch, A., Gordon, D., Knudson, K., & Cheng, J. (2024). A multi-method data science pipeline for analyzing police service. *The American Statistician*, 79(1), 91–101. <https://doi.org/10.1080/00031305.2024.2374275>
- Helbich, M., Hagenauer, J., Leitner, M., & Edwards, R. (2013). Exploration of unstructured narrative crime reports: An unsupervised neural network and point pattern analysis approach. *Cartography and Geographic Information Science*, 40(4), 326–336. <https://doi.org/10.1080/15230406.2013.779780>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 144–146. <https://doi.org/10.1177/1094428120971683>
- Hill, C., Du, L., Johnson, M., & McCullough, B. D. (2024). Comparing programming languages for data analytics: Accuracy of estimation in Python and R. *Wires Data Mining and Knowledge Discovery*, 14(3), e1531. <https://doi.org/10.1002/widm.1440>
- Hodgkinson, S., Dixon, A., Halford, E., & Farrell, G. (2023). Domestic abuse in the Covid-19 pandemic: Measures designed to overcome common limitations of trend measurement. *Crime Science*, 12(1), 12. <https://doi.org/10.1186/s40163-023-00190-7>
- Hosseini, P., Khoshirsat, S., Jalayer, M., Das, S., & Zhou, H. (2022). Application of text mining techniques to identify actual wrong-way driving (WWD) crashes in police reports. *International Journal of Transportation Science and Technology*. <https://doi.org/10.1016/j.ijst.2022.12.002>

- Hwang, Y. I., Zheng, L., Karystianis, G., Gibbs, V., Sharp, K., & Butler, T. (2020). Domestic violence events involving autism: A text mining study of police records in New South Wales, 2005–2016. *Research in Autism Spectrum Disorders*, 78, 101634. <https://doi.org/10.1016/j.rasd.2020.101634>
- Jaradat, S., Acharya, N., Shivshankar, S., Alhadidi, T. I., & Elhenawy, M. (2025). AI for data quality auditing: Detecting mislabeled work zone crashes using large language models. *Algorithms*, 18(6), 317. <https://doi.org/10.3390/a18060317>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. <https://doi.org/10.1038/s42256-019-0088-2>
- Joh, E. E. (2018). Artificial Intelligence and Policing: First Questions, 41 *SEATTLE U. L. REV.*, 1139.
- Jones, H., Bandyopadhyay, A., Kennedy, N., Brophy, S., Evans, J., Bellis, M., Rowe, B., McInerney, C., & Moore, S. (2024). Investigating methods of sharing data between police, health, education, and social services: Semi-structured interviews with police service areas in Wales. *Police Journal: Theory, Practice and Principles*, 97(4), 731–741. <https://doi.org/10.1177/0032258X231220864>. (Original work published 2024).
- Joorabchi, A., English, M., & Mahdi, A. E. (2016). Text mining stackoverflow: An insight into challenges and subject-related difficulties faced by computer science learners. *Journal of Enterprise Information Management*, 29(2), 255–275. <https://doi.org/10.1108/JEIM-11-2014-0109>
- Juez-Hernandez, R., Quijano-Sánchez, L., Liberatore, F., & Gómez, J. (2023). AGORA: An intelligent system for the anonymization, information extraction and automatic mapping of sensitive documents. *Applied Soft Computing*, 145, 110540. <https://doi.org/10.1016/j.asoc.2023.110540>
- Karystianis, G., Adily, A., Schofield, P. W., Greenberg, D., Jorm, L., Nenadic, G., & Butler, T. (2019). Automated analysis of domestic violence police reports to explore abuse types and victim injuries: Text mining study. *Journal of Medical Internet Research*, 21(3), e13067. <https://doi.org/10.2196/13067>
- Karystianis, G., Adily, A., Schofield, P., Knight, L., Galdon, C., Greenberg, D., Jorm, L., Nenadic, G., & Butler, T. (2022a). Automatic extraction of mental health disorders from domestic violence police narratives: Text mining study. *Journal of Medical Internet Research*, 20(9), e11548. <https://doi.org/10.2196/11548>
- Karystianis, G., Adily, A., Schofield, P. W., Wand, H., Lukmanjaya, W., Buchan, I., Nenadic, G., & Butler, T. (2022a). Surveillance of domestic violence using text mining outputs from Australian police records. *Frontiers in Psychiatry*, 12, 787792. <https://doi.org/10.3389/fpsy.2021.787792>
- Karystianis, G., Cabral, R. C., Adily, A., Lukmanjaya, W., Schofield, P., Buchan, I., Nenadic, G., & Butler, T. (2022b). Mental illness concordance between hospital clinical records and mentions in domestic violence police narratives: Data linkage study. *JMIR Formative Research*, 6(10), e39373. <https://doi.org/10.2196/6/39373>
- Karystianis, G., Cabral, R. C., Han, S. C., Poon, J., & Butler, T. (2021). Utilizing text mining, data linkage and deep learning in police and health records to predict future offenses in family and domestic violence. *Frontiers in Digital Health*, 3, 602683. <https://doi.org/10.3389/fgdh.2021.602683>
- Karystianis, G., Chowdhury, N., Sheridan, L., Reutens, S., Wade, S., Allnut, S., Kim, M., Poynton, S., & Butler, T. (2024a). Text mining domestic violence police narratives to identify behaviours linked to coercive control. *Crime Science*, 13, Article 2. <https://doi.org/10.1186/s40163-024-00200-2>
- Karystianis, G., Kita, S., Lerigo, F., Sheridan, L., & Butler, T. (2024b). Characteristics of adult male victims in intimate heterosexual relationships from domestic violence police narratives. *Crime Science*, 13, 28. <https://doi.org/10.1186/s40163-024-00228-4>
- Karystianis, G., Simpson, A., Adily, A., Schofield, P., Greenberg, D., Wand, H., Nenadic, G., & Butler, T. (2020). Prevalence of mental illnesses in domestic violence police records: Text mining study. *Journal of Medical Internet Research*, 22(12), e23725. <https://doi.org/10.2196/23725>
- Kim, J., Trueblood, A. B., Kum, H. C., & Shipp, E. M. (2021). Crash narrative classification: Identifying agricultural crashes using machine learning with curated keywords. *Traffic Injury Prevention*, 22(1), 74–78. <https://doi.org/10.1080/15389588.2020.1836365>
- Koroteev, M. V. (2021). BERT: A review of applications in natural language processing and understanding. Retrieved from <https://arxiv.org/pdf/2103.11943>
- Kuang, D., Brantingham, P. J., & Bertozzi, A. L. (2017). Crime topic modeling. *Crime Science*, 6, 12. <https://doi.org/10.1186/s40163-017-0074-0>
- Kwayu, K. M., Kwigizile, V., Zhang, J., & Oh, J.-S. (2020). Semantic N-Gram Feature Analysis and Machine Learning-Based Classification of Drivers' Hazardous Actions at Signal-Controlled Intersections. *Journal of Computing in Civil Engineering*, 34(4). [https://doi.org/10.1061/\(ASCE\)CP:1943-5487.0000895](https://doi.org/10.1061/(ASCE)CP:1943-5487.0000895)
- Kwon, D., Borrión, H., & Wortley, R. (2024). Measuring cybercrime in calls for police service. *Asian Journal of Criminology*, 19, 329–351. <https://doi.org/10.1007/s11417-024-09432-2>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Langton, S., Bannister, J., Ellison, M., Haleem, M. S., & Krzemieniewska-Nandwani, K. (2021). Policing and mental ill-health: Using Big Data to assess the scale and severity of, and the frontline resources committed to, mental ill-health-related calls-for-service. *Policing: A Journal of Policy and Practice*, 15(3), 1963–1976. <https://doi.org/10.1093/policing/15.3>
- Li, S., Wu, T., & Pottenger, W. M. (2005). Distributed higher order association rule mining using information extracted from textual data. *ACM SIGKDD Explorations Newsletter*, 7(1), 26–35. <https://doi.org/10.1145/1089815.1089820>
- Liu, Y., Guo, Q., Yang, C., & Liao, Y. (2025). TIPS: Tailored information extraction in public security using domain-enhanced large language model. *Computers, Materials & Continua*, 83(2), 2555–2572. <https://doi.org/10.32604/cmc.2025.060318>
- Lukmanjaya, W., Butler, T., Cox, S., Perez-Concha, O., Bromfield, L., & Karystianis, G. (2025). Leveraging AI to investigate child maltreatment text narratives: Promising benefits and addressable risks. *JMIR Pediatrics and Parenting*, 8, e73579. <https://doi.org/10.2196/73579>
- Marciniak, D. (2016). Computational text analysis: Thoughts on the contingencies of an evolving method. *Big Data & Society*. <https://doi.org/10.1177/2053951716670190>
- Metreau, E., Young, K. E., Eapen, S. G. (2024). World Bank Country Classifications by Income Level for 2024–2025. World Bank Blogs. Retrieved from <https://blogs.worldbank.org/en/opendata/world-bank-country-classifications-by-income-level-for-2024-2025>
- Mishra, S., Misra, A. (2017). Structured and unstructured big data analytics. In 2017 international conference on current trends in computer, electrical, electronics and communication (CTCEEC), Mysore, India, 2017, pp. 740–746. <https://doi.org/10.1109/CTCEEC.2017.8454999>
- Moses, L. B., & Chan, J. (2016). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing and Society*, 28(7), 806–822. <https://doi.org/10.1080/10439463.2016.1253695>
- Navarrete, F., Garrido, Á. L., Bobed, C., Atencia, M., & Vallecillo, A. (2024). Ontology-driven automated reasoning about property crimes. *Business & Information Systems Engineering*. <https://doi.org/10.1007/s12599-024-00886-3>
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543. <https://doi.org/10.1287/mksc.1120.0713>
- Neubauer, L., Straw, I., Mariconti, E., & Tanczer, L. M. (2023). A systematic literature review of the use of computational text analysis methods in intimate partner violence research. *Journal of Family Violence*. <https://doi.org/10.1007/s10896-023-00517-7>
- Neumann, M., & Sartor, N. (2016). A semantic network analysis of laundering drug money. *Journal of Tax Administration*, 2(1), 73–94.
- Neves Oliveira, B. S., de Fernans Oliveira, A., de Monteiro Lira, V., da Linhares Coelho Silva, T., & de Fernans Macêdo, J. A. (2022). HELD: Hierarchical entity-label disambiguation in named entity recognition task using deep learning. *Intelligent Data Analysis*, 26(3), 637–657. <https://doi.org/10.3233/IDA-205720>
- Øby, E. (2025). Assessing transparency and methodological precision in variable measurement within organizational research: Implications for validity. *Quality and Quantity*, 59(Suppl 1), 489–506. <https://doi.org/10.1007/s11135-024-01991-x>
- Octoman, O., Cox, S., Arney, F., Chong, A., & Tucker, E. (2023). Narrative and fixed-field data: Are we underestimating the risk of family and domestic violence? *Child Abuse Review*, 32(4), e2811.
- Oliaee, A. H., Das, S., & Le, M. (2024). Automating Pedestrian crash typology using transformer models. *Transportation Research Record: Journal of the Transportation Research Board*. <https://doi.org/10.1177/03611981241260691>
- OpenAI et al. (2023). ChatGPT (GPT-4) [Large language model]. Retrieved from <http://arxiv.org/pdf/2303.08774>
- Ortiz Salazar, A. (2024). Detecting hate crimes through machine learning and natural language processing. *Police Practice and Research*. <https://doi.org/10.1080/15614263.2024.2397363>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lahu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

- Palad, E. B. B., Burden, M. J. F., Torre, C. R. D., & Uy, R. B. C. (2020). Performance evaluation of decision tree classification algorithms using fraud datasets. *Bulletin of Electrical Engineering and Informatics*, 9(6), 2518–2525. <https://doi.org/10.1159/eei.v9i6.2630>
- Parker, S. (2025). Supervised Natural Language Processing classification of violent death narratives: Development and assessment of a compact Large Language Model. *JMIR AI*, 4, e68212. <https://doi.org/10.2196/68212>
- Poelmans, J., Elzinga, P., Ignatov, D. I., & Kuznetsov, S. O. (2012). Semi-automated knowledge discovery: Identifying and profiling human trafficking. *International Journal of General Systems*, 41(8), 774–804. <https://doi.org/10.1080/03081079.2012.721662>
- Poelmans, J., Elzinga, P., Viaene, S., & Dedene, G. (2010). Curbing domestic violence: Instantiating C-K theory with formal concept analysis and emergent self-organizing maps. *Intelligent Systems in Accounting, Finance & Management*, 17, 167–191. <https://doi.org/10.1002/isaf.319>
- Poelmans, J., Elzinga, P., Viaene, S., & Dedene, G. (2011a). Formally analysing the concepts of domestic violence. *Expert Systems with Applications*, 38(4), 3116–3130. <https://doi.org/10.1016/j.eswa.2010.08.103>
- Poelmans, J., Elzinga, P., Viaene, S., Van Hulle, M. M., & Dedene, G. (2009). Gaining insight in domestic violence with Emergent Self Organizing Maps. *Expert Systems with Applications*, 36(9), 11864–11874. <https://doi.org/10.1016/j.eswa.2009.04.027>
- Poelmans, J., Van Hulle, M. M., Viaene, S., Elzinga, P., & Dedene, G. (2011b). Text mining with emergent self organizing maps and multi-dimensional scaling: A comparative study on domestic violence. *Applied Soft Computing*, 1(4), 3870–3876. <https://doi.org/10.1016/j.asoc.2011.02.026>
- Qiu, M., Zhang, X., & Wang, X. (2023). An ex-convict recognition method based on text mining. *International Journal of Security Networks*, 18(1), 10–18. <https://doi.org/10.1504/ijns.2023.129905>
- Quijano-Sánchez, L., Liberatore, F., Camacho-Collados, J., & Camacho-Collados, M. (2018). Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police. *Knowledge-Based Systems*, 149, 155–168. <https://doi.org/10.1016/j.knosys.2018.03.010>
- Rahwan, I., Cebrian, M., Obradovich, N., et al. (2019). Machine behaviour. *Nature*, 568, 477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Geburu, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, pp. 33–44. <https://doi.org/10.1145/3351095.3372873>
- Ramon-Gonen, R., Dori, A., & Shelly, S. (2023). Towards a practical use of text mining approaches in electrodiagnostic data. *Scientific Reports*, 13, 19483. <https://doi.org/10.1038/s41598-023-45758-0>
- Reutens, S., Karystianis, G., Withall, A., & Butler, T. (2024). Characteristics of domestic violence perpetrators with dementia from police records using text mining. *Frontiers in Psychiatry*, 15, 1331915. <https://doi.org/10.3389/fpsy.2024.1331915>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models. *Nature Machine Intelligence*, 1, 206–215.
- Sajith, A., Kathala, K. C. R. (2024). Is training data quality or quantity more impactful to small language model performance? *ArXiv*, arXiv:2411.15821.
- Sarzaeim, P., Mahmoud, Q. H., Azim, A., Bauer, G., & Bowles, I. (2023). A systematic review of using machine learning and natural language processing in smart policing. *Computers*, 12(12), 255. <https://doi.org/10.3390/computers12120255>
- Sayed, M. A., Qin, X., Kate, R. J., Anisuzzaman, D. M., & Yu, Z. (2021). Identification and analysis of misclassified work-zone crashes using text mining techniques. *Accident Analysis & Prevention*, 159, 106211. <https://doi.org/10.1016/j.aap.2021.106211>
- Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, 46, e70002. <https://doi.org/10.1002/aaai.70002>
- Seo, Y., Park, J., Oh, G., Kim, H., Hu, J., & So, J. (2023). Text classification modeling approach on imbalanced-unstructured traffic accident descriptions data. *IEEE Open Journal of Intelligent Transportation Systems*, 4, 955–965. <https://doi.org/10.1109/OJITS.2023.3335817>
- Shah, T., Allen, C., Ibrahim, A., Kefalas, H., & Stevens, B. (2024). The use of unstructured data to study police use of force. *Chance*, 37(4), 18–23. <https://doi.org/10.1080/09332480.2024.2434437>
- Shipp, E. M., Trueblood, A. B., Kum, H. C., Perez, M., Vasudeo, S., Sinha, N., Pant, A., Wu, L., & Ko, M. (2023). Using motor vehicle crash records for injury surveillance and research in agriculture and forestry. *Journal of Safety Research*, 86, 21–29. <https://doi.org/10.1016/j.jsr.2023.06.004>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garrity, C., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–543. <https://doi.org/10.7326/M18-0850>
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn: Machine learning without learning the machinery. *GetMobile*, 19(1), 29–33. <https://doi.org/10.1145/2786984.2786995>
- Vidler, T., McGarry, K., Baglee, D. (2023). Text mining legal documents for Cla use extraction. In: The 19th international conference on data science (ICDATA'23), 24–27 Jul 2023, Las Vegas, USA.
- Wang, Z., Xu, S., & Zhu, L. (2018). Semantic relation extraction aware of N-gram features from unstructured biomedical text. *Journal of Biomedical Informatics*, 2018, 59–70. <https://doi.org/10.1016/j.jbi.2018.08.011>
- Weng, Y., Das, S., & Paal, S. G. (2023). Applying few-shot learning in classifying pedestrian crash typing. *Transportation Research Record*, 677(8), 563–572. <https://doi.org/10.1177/03611981231157393>
- Wilson, M., Spike, E., Karystianis, G., & Butler, T. (2022). Nonfatal strangulation during domestic violence events in New South Wales: Prevalence and characteristics using text mining study of police narratives. *Violence against Women*, 28(10), 2259–2285. <https://doi.org/10.1177/10778012211025993>
- Wu, T., & Pottenger, W. M. (2005). A semi-supervised active learning algorithm for information extraction from textual data. *Journal of the American Society for Information Science*, 56, 258–271. <https://doi.org/10.1002/asi.20119>
- Xing, X., & Chen, P. (2024). Entity extraction of key elements in 110 police reports based on large language models. *Applied Sciences*, 14(17), 7819. <https://doi.org/10.3390/app14177819>
- Yu, R., Small, D. S., Harding, D., Avelanes, J., & Rosenbaum, P. R. (2021). Optimal matching for observational studies that integrate quantitative and qualitative research. *Statistics and Public Policy*, 8(1), 42–52. <https://doi.org/10.1080/2330443X.2021.1919260>
- Zhang, J., Ji, X., Zhao, Z., Hei, X., Choo, K. K. R. (2023). Ethical considerations and policy implications for large language models: Guiding responsible development and deployment. *ArXiv*. <https://arxiv.org/pdf/2308.02678>.
- Zhou, W., Prater, L. C., Goldstein, E. V., & Mooney, S. J. (2023). Identifying rare circumstances preceding female firearm suicides: Validating a large language model approach. *JMIR Mental Health*, 10, e49359. <https://doi.org/10.2196/49359>
- Zhu, S., & Xie, Y. (2022). Spatiotemporal-textual point processes for crime linkage detection. *The Annals of Applied Statistics*, 16(2), 1151–1170.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.