



International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

CaseGuard – Domestic Violence Case Pattern Analyzer: An AI-Assisted Case Management and Decision-Support System for Legal-Aid Organizations

Subramanian E^{#1}, Siju Wilson P^{#2}, Sreejaa RT^{#3}, Sreya P^{#4}, Thaneesha R^{#5}, Vikas S^{#6}

^{#1} Assistant Professor, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: esubramaniancse@siet.ac.in

^{#2} Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: sijuwilsonp24cs@srishakthi.ac.in

^{#3} Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: sreejaart24cs@srishakthi.ac.in

^{#4} Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: sreyp24cs@srishakthi.ac.in

^{#5} Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: thaneeshar24cs@srishakthi.ac.in

^{#6} Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India. Email: vikass24cs@srishakthi.ac.in

ABSTRACT

The proposed CaseGuard – Domestic Violence Case Pattern Analyzer is a web-based, AI-assisted case management and decision-support system that helps legal-aid organizations, lawyers, and case workers organize and review domestic violence cases. Case files are typically assembled from narrative statements, intake forms, and incident records, and professionals must read, compare, and consolidate this material manually before they can recognise recurring behaviour or situations needing urgent attention. The system converts unstructured statements and intake information into structured case records, identifies indicators such as threats, intimidation, isolation, and repeated incidents, and presents severity-related and escalation-related information for professional review. A centralized dashboard supports case creation, review, and status monitoring; a lawyer brief generator produces structured, editable, and printable case summaries; and a rule-based Safe Action Navigator presents action guidance, evidence checklists, urgency information, and referral options. Authentication, role-based access control, input validation, and audit logging support confidentiality and accountability, and the modular architecture allows each component to be tested and improved independently. The current prototype uses a mock analysis module and rule-based guidance, so its outputs are preliminary and are not validated risk predictions; the paper therefore also states the limitations of the prototype and a validation protocol covering extraction accuracy, expert agreement, fairness, and usability. CaseGuard is designed to support, not replace, professional judgment.

Keywords: Domestic Violence, Case Management, Decision Support, Natural Language Processing, Pattern Analysis, Severity and Escalation Assessment, Lawyer Brief Generation, Safe Action Navigator, Role-Based Access Control, Audit Logging, Legal Aid.

I. INTRODUCTION

Domestic violence is a widespread public-health and legal problem. Global estimates indicate that a substantial proportion of women experience physical or sexual violence from an intimate partner during their lifetime [1], [2]. Legal-aid organizations, advocacy groups, and social support services are often the first formal point of contact for affected individuals, and the case workers and lawyers in these organizations must understand what has happened, how often, and how serious the situation is before they can recommend a course of action.

The difficulty lies in the form of the information. A case is rarely recorded as clean, structured data. It is spread across written statements, intake forms, timelines, and supporting documents, and describes multiple incidents that include threats, intimidation, isolation, restrictions on communication, and other forms of coercive control. Individual incidents can look unrelated when read separately, yet taken together they may show a repeated pattern or an increasing level of concern. Research on coercive control shows that abuse is frequently a sustained pattern of domination rather than a set of isolated events [3], [4], [5], and known risk factors for severe or lethal violence have been documented [6], [7]. Validated instruments such as the Danger Assessment [8] and brief screening tools [9], including versions adapted to specific cultural settings [10], help professionals structure risk assessment,

but they presuppose that the relevant facts have already been identified and organized. Extracting those facts from narrative records remains manual, time-consuming work.

The proposed CaseGuard system addresses this challenge by connecting unstructured narrative with structured case information through a single web-based platform. Authorized users can enter intake data and written statements, review AI-assisted findings on incidents and indicators, generate a lawyer brief, and consult a rule-based Safe Action Navigator, all from one centralized dashboard. The system is intended for authorized professionals; it does not make legal decisions and does not replace independent risk assessment.

A key feature of the platform is its human-in-the-loop design. Findings are linked to the statements from which they were derived, generated analysis is displayed separately from reported information, and every brief is editable before use. This provides transparency and allows professionals to verify, correct, or reject any automated output, so that the final record reflects their judgment rather than the output of the software.

The platform is built as a modular web application with authentication, role-based access control, input validation, and audit logging. The analysis component is exposed through a fixed interface, so that the mock implementation used in the current prototype can later be replaced by an evaluated NLP model without changes to the other modules. This paper presents the design and implementation of CaseGuard, states the status and limitations of the prototype openly, and proposes a validation protocol for moving from a mock analysis module to evaluated components. The main contributions are a modular architecture that integrates intake, analysis, assessment, brief generation, and action guidance in one workflow; a centralized dashboard and an editable lawyer brief generator; a privacy-conscious design with traceability between reported information and generated analysis; and a transparent evaluation protocol.

The remainder of the paper is organized as follows. Section II reviews related work. Section III describes the methodology and architecture. Section IV details the implementation of each module. Section V summarizes the advantages of the approach. Section VI reports the prototype status, limitations, and evaluation protocol. Sections VII and VIII present the conclusion and future work.

II. LITERATURE SURVEY

AI and NLP for Case Narratives Natural Language Processing (NLP) techniques such as text classification, entity extraction, and semantic analysis are widely used to derive structure from free text [11]. Distributed word representations [12], recurrent networks [13], deep learning more generally [14], and attention-based transformer models [15] such as BERT [16] have progressively improved the ability of systems to represent context. Topic models [17] and classical classifiers such as support vector machines [18] and random forests [19] remain useful baselines, particularly where labeled data are limited. Domestic violence statements contain descriptions of incidents, threats, emotional abuse, and controlling behaviour in ordinary language, which makes them a natural, though sensitive, target for these techniques.

Legal NLP and Documentation Surveys of legal artificial intelligence describe tasks such as legal judgment prediction, document summarization, and information retrieval [20], and domain-adapted language models have been proposed for legal text [21]. These works support the premise that summarization and structured presentation can reduce documentation effort. CaseGuard applies the same premise to lawyer brief generation while keeping the professional in control of the final text.

Incident and Pattern Extraction A single statement may describe several events at different times. Information-extraction techniques can identify incident descriptions, threats, intimidation, isolation, and recurring behaviours, and comparing these across statements can surface patterns that would otherwise require extensive manual comparison. The value of such extraction depends on traceability: reviewers must be able to see which part of the original statement supports each extracted item.

Severity and Escalation Assessment Risk assessment in domestic violence has been studied extensively, and validated instruments exist for lethality and severe-violence risk [8], [9], [10]. Case-control evidence shows that specific circumstances are associated with elevated risk [6]. These instruments are intended for use by trained professionals. AI-assisted decision support can organize the relevant indicators and highlight information needing attention, but should not be presented as a definitive prediction of future violence.

Fairness, Bias, and Accountability Algorithmic risk tools can produce unequal error rates across groups [22], and a widely cited study found that a deployed health-care algorithm exhibited substantial racial bias because of the proxy label it was trained to predict [23]. Surveys describe multiple sources of bias in data and models [24], and model documentation practices such as model cards have been proposed to communicate intended use and limitations [25]. These findings are directly relevant to any system that labels the severity of a person's situation, and they motivate the evaluation protocol in Section VI.

Centralized Case Management and Safe Action Guidance When documents and incident details are held in separate places, professionals find it hard to obtain a complete overview of a case. A centralized dashboard addresses this by organizing records through a consistent interface. Safe action guidance, including evidence checklists and referral information, is commonly delivered through predefined rules, which are transparent and can be reviewed by domain experts before use.

Gaps Identified in Existing Systems Validated risk instruments assume that facts are already organized; legal NLP research largely addresses other document types; and general case management tools do not connect narrative analysis, brief generation, and safety guidance. Support for traceable, human-reviewable analysis of domestic violence narratives within a single workflow also remains limited. CaseGuard targets this gap by combining these functions in one privacy-conscious platform, while stating clearly that its analysis component is not yet validated.

Summary The literature establishes that individual technologies for NLP, risk assessment, legal document processing, and case management have matured, but their integration into a holistic, accountable decision-support platform for domestic violence casework remains an open area. CaseGuard distinguishes itself by unifying structured case intake, source-linked analysis, brief generation, and rule-based guidance under a single access-controlled interface, with an explicit protocol for validating the analysis component.

Data Privacy and Ethical Considerations Domestic violence records contain sensitive personal statements and identifying details, and unauthorized disclosure can create additional risk for affected individuals. Established practice includes authentication, role-based access control [26], restricted permissions, and activity logging. Encryption, retention policies, and controlled information sharing are additional safeguards required for real deployment. Ethical frameworks recommend transparent data usage policies, informed consent, and de-identification of any data used for research, ensuring that affected individuals retain control over their information.

III. PROPOSED METHODOLOGY

CaseGuard follows a modular, human-in-the-loop design in which case management, information collection, AI-assisted analysis, severity and escalation assessment, lawyer brief generation, and action guidance operate as separate but connected components. The modules communicate through defined interfaces so that each can be developed, tested, and replaced independently. The methodology follows an iterative software development approach, with each module developed and validated independently before integration into the full workflow.

The development process begins with a requirement analysis phase derived from the workflow of legal-aid professionals. Key functional requirements include: secure creation and retrieval of case records; collection of information through intake forms and written statements; identification of incidents, threats, intimidation, isolation, and recurring behaviour; presentation of severity and escalation concerns for review; generation of editable, printable lawyer briefs; action guidance, evidence checklists, and referrals; and access control, audit logging, and administrative oversight.

The user enters case information through an intake form or provides a written statement. The analysis module processes the content and organizes findings into a structured representation of incidents, indicators, and timeline entries. Findings are linked to the originating statement so that reviewers can verify them against the source rather than rely on automated interpretation. The assessment component then derives review-priority information from the structured findings, and results are made available on the dashboard, from which authorized users can open the case, edit the lawyer brief, and consult the Safe Action Navigator.

The data model comprises seven core entities: the case, the statement, the incident, the detected indicator, the assessment result, the generated brief, and the audit event. Keeping incidents and indicators linked to their source statements preserves the relationship between an incident and its surrounding context and allows comparison of timelines across statements.

The analysis module is designed to detect indicator categories (repeated incidents, threats, intimidation, isolation and restrictions on communication, and other controlling behaviours) and to measure their recurrence and change over time across the available statements. The assessment component aggregates these results into review-priority labels that point reviewers to cases that may need prompt attention. These labels are decision-support signals; they are not probabilities of future violence. Indicator categories are to be aligned with those documented in validated instruments [8], [9] and reviewed by domain experts.

The Safe Action Navigator is a rule base in which predefined conditions on case attributes map to guidance, evidence checklists, urgency-related information, and referral options. Rules are explicit and can be inspected and revised by domain experts without changing other modules. Output is advisory: immediate safety needs and legal requirements must be assessed by qualified professionals or emergency services.

Security and ethics are treated as design requirements. Access is restricted to authenticated users according to role; sensitive operations are written to an audit log; inputs are validated; and security middleware reduces common application risks. Ethical safeguards include professional review of all generated output, display of generated analysis separately from original statements, editable briefs, and an explicit statement of intended use and limitations.

The testing methodology encompasses functional testing of each module and end-to-end testing of the workflow from case entry to guidance, together with checks of role-restricted access and input validation. Results from each phase inform iterative refinements to both the modules and the user interface. Evaluation of the analysis component against expert-annotated data is deferred to the protocol described in Section VI.

IV. SYSTEM IMPLEMENTATION

The implementation of CaseGuard translates the proposed methodology into a web-based application. The system is organized into eight modules: Authentication and Access Control, Case Intake and Management, AI-Assisted Analysis, Severity and Escalation Assessment, Dashboard, Lawyer Brief Generator, Safe Action Navigator, and Administrative Monitoring and Audit Logging. Each module is implemented independently and integrated through defined interfaces.

Authentication and Access Control Module Users sign in with credentials, and permissions are assigned by role so that access to case records and generated documents is limited according to responsibilities. Input validation and security middleware protect the application interface.

Case Intake and Management Module Authorized users create case records, enter intake-form data and written statements, and record incident details and timeline entries. The module maintains the status of each case and serves as the foundation on which all other modules operate.

AI-Assisted Analysis Module This module examines the submitted content and returns structured findings: incidents, indicators such as threats, intimidation, and isolation, and recurring elements across statements. In the current prototype it is a mock implementation that exposes the same interface a trained model would use. Its outputs therefore demonstrate the data flow and user interface and should not be read as validated machine-learning predictions. Because the interface is fixed, an evaluated NLP component can later replace the mock without changes to the other modules.

Severity and Escalation Assessment Component The component organizes the findings into severity indicators and escalation-related information for the dashboard, helping users prioritize which cases to examine first.

Dashboard The dashboard is the primary interface. It lists cases with status indicators, shows incident timelines, severity labels, and analysis summaries, and links to generated briefs and support guidance. It is designed to reduce complexity and let users review records without searching through multiple documents.

Lawyer Brief Generator The generator consolidates statements, incident chronology, reported threats, and key observations into a structured summary. The brief is presented in an editable and printable form so that the professional can verify, correct, and extend it before use.

Safe Action Navigator The navigator evaluates the predefined rules against the case information and presents applicable action guidance, evidence checklists, urgency-related information, and referral options.

Administrative Monitoring and Audit Logging Relevant system activities, such as sign-in, case access, and document generation, are recorded to support accountability and oversight, and administrators can monitor the use of the platform.

During system integration, all individual modules were combined into a single working platform. This involved linking internal components, testing cross-module data flow between intake, analysis, dashboard, brief generation, and guidance, and verifying that case information passed correctly between modules.

The testing and debugging phase followed. Functional checks covered case creation, retrieval, analysis, brief editing, and role-restricted access. Input-validation errors identified during testing were corrected to achieve a stable prototype.

V. ADVANTAGES

1. Structured Case Understanding

Organized Findings: Unstructured statements and forms are organized into incidents, indicators, and timelines.

Pattern Visibility: Repeated behaviour and escalating concerns are easier to see than when each document is read in isolation.

Consistent Representation: A common data model gives every case the same structure, which supports comparison across statements.

2. Reduced Administrative Effort

Centralized Information: Consolidating case information in one place reduces repetitive manual organization.

Starting-Point Briefs: Generated briefs give professionals an editable draft instead of a blank page.

Measurement Pending: The size of the time saving has not yet been measured and is included in the evaluation protocol (Section VI).

3. Traceability and Transparency

Source-Linked Findings: Findings remain connected to the original statements.

Separation of Content: Generated material is kept distinct from reported information.

Verification Over Reliance: Reviewers can check each item against its source instead of relying on automated interpretation.

4. Human-Centred Decision Support

Advisory Output: The system presents priority and guidance information while leaving assessment and legal decisions to qualified professionals.

Editable Documents: Briefs can be corrected and extended before use.

Stated Limits: Intended use and limitations are stated explicitly to prevent over-reliance.

5. Confidentiality and Accountability

Controlled Access: Authentication and role-based access control restrict access to authorized users.

Input Protection: Input validation and security middleware reduce common application risks.

Audit Trail: Audit logging records sensitive activity for oversight.

6. Modularity and Extensibility

Fixed Interfaces: The mock analysis component can be replaced by validated models without redesigning the platform.

Editable Rules: New guidance rules can be added and revised by domain experts without changing other modules.

Integration Ready: Brief templates and integrations can be added as requirements evolve.

7. Consistent, Accessible Review

Single Dashboard: A single dashboard supports follow-up, documentation review, and administrative oversight across cases.

Status Monitoring: Case status indicators make it easier to see which cases need attention.

Reduced Searching: Users review records without searching through multiple documents.

VI. RESULTS AND ANALYSIS

This section reports what the current prototype demonstrates, what it does not, and how it should be evaluated. No quantitative performance claims are made, because the analysis module is a mock and no labeled evaluation dataset has been used. The system was exercised through functional testing of each module and of the integrated workflow.

The status of each module in the current prototype is as follows. The authentication, role-based access control, and audit logging module is implemented and has been functionally exercised, although an independent security review is still pending. The case intake and management module is implemented and functionally exercised. The AI-assisted analysis module is a mock implementation, and no accuracy claims are made for it. The severity and escalation component is derived from the mock output, so it is not validated and must not be read as a risk prediction. The lawyer brief generator is implemented in editable and printable form, but the quality of its content has not yet been evaluated. The Safe Action Navigator is rule-based, and its rules and referral information require expert review and verification. The dashboard and monitoring module is implemented, and its usability has not yet been evaluated.

The end-to-end workflow, from case entry through analysis, dashboard review, brief generation, and guidance, runs in the prototype, and role-based access and audit logging operate on that workflow. The analysis output is placeholder content intended to exercise the interface and demonstrate the data flow.

The prototype has the following limitations. First, the analysis and severity outputs are preliminary and must not be used as risk assessments. Second, rule content and referral information have not been verified against current local legal-aid resources. Third, security features depend on correct configuration, deployment, and maintenance, and the platform has not undergone independent penetration testing. Fourth, narrative statements may be incomplete, translated, or reflect trauma-related memory gaps, and automated extraction may mishandle such text. Finally, no evaluation with real case data has been performed, and none should be performed without ethical approval, informed consent, and de-identification.

Before any analysis component is used in practice, it should be evaluated against expert-annotated data. Indicator and incident extraction should be assessed with precision, recall, and F1 [27] per indicator category against expert-annotated statements, and timeline construction should be assessed for completeness and ordering accuracy against expert-built timelines. Severity and escalation labels should be compared with structured instrument scores and panel review using a chance-corrected agreement statistic [29] and ROC analysis [28], with recall for high-urgency cases prioritized because missed indicators and false alarms have different consequences. Error costs should be discussed with domain experts.

Fairness should be examined by comparing error rates across subgroups on a stratified evaluation set, reporting metrics per subgroup as well as overall. The lawyer brief should be reviewed by practising lawyers for accuracy, completeness, and the editing effort required. Usability can be measured with the System Usability Scale [30], task time, and error counts among representative users, and the time saved in brief preparation can be measured in a controlled comparison with manual preparation.

Overall, the prototype shows that the proposed architecture can connect intake, analysis, documentation, and guidance in a single controlled workflow, and that the fixed analysis interface allows a validated model to be introduced later. The principal open question is not the engineering of the workflow but the validity and fairness of the analysis component, and the protocol above is intended to address it. Model documentation [25] stating intended use, data, and known limitations should accompany any future model.

VII. CONCLUSION

In conclusion, the CaseGuard project presents an AI-assisted case management and decision-support system for domestic violence cases in legal-aid settings. By combining secure case intake, structured analysis, severity and escalation information, lawyer brief generation, rule-based safe action guidance, and audit logging in a modular workflow, the system demonstrates a structured approach to organizing and reviewing sensitive case information and reducing manual organization effort.

The current prototype uses a mock analysis module and rule-based guidance, so its analysis outputs are preliminary and are not validated risk assessments. The contribution of this work is therefore the architecture, the workflow, and the safeguards, together with a protocol for validating the analysis component. The system is intended to support professional review, confidentiality, and responsible handling of sensitive information, and it does not replace legal judgment, trained support workers, or emergency services.

From a technical standpoint, the modular architecture and fixed module interfaces position CaseGuard well for future enhancement. Overall, CaseGuard provides a transparent, accountable, and extensible foundation on which validated analysis components can be developed and evaluated with legal-aid professionals.

VIII. FUTURE WORK

The CaseGuard system has been developed with a modular architecture that enables expansion. The key future enhancement is the integration of validated NLP methods for incident extraction, indicator identification, and timeline construction. These methods should be evaluated according to the validation protocol in Section VI before use in real assessments, with source-linked evidence for every extracted item.

Severity and escalation assessment methods should be developed with qualified domain experts, aligned with validated instruments, tested on appropriate datasets, and assessed for bias. Results should be presented as decision support rather than definitive conclusions. Planned improvements to lawyer briefs include customizable templates, improved incident chronology, structured evidence references, and a clearer separation between reported information and verified facts.

The Safe Action Navigator can be expanded with verified legal-aid resources, support-service directories, and location-appropriate referrals, and with more context-sensitive evidence checklists. It must prioritize confidentiality and user safety and avoid assuming that a single course of action suits every situation.

Security and privacy enhancements include stronger encryption, secure document handling, configurable data-retention policies, consent management, finer-grained access controls, auditable access histories, and regular security testing, including independent penetration testing before any real deployment.

Finally, longer-term platform goals include improved document processing, incident timeline visualization, configurable assessment indicators, customizable dashboards and reporting, multilingual support, and integration with authorized legal-aid services, guided by feedback from legal-aid professionals.

IX. REFERENCES

- [1]. Devries, K. M., Mak, J. Y. T., García-Moreno, C., et al. (2013). The global prevalence of intimate partner violence against women. *Science*, 340(6140), 1527–1528. <https://doi.org/10.1126/science.1240937>
- [2]. García-Moreno, C., Jansen, H. A. F. M., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *The Lancet*, 368(9543), 1260–1269. [https://doi.org/10.1016/S0140-6736\(06\)69523-8](https://doi.org/10.1016/S0140-6736(06)69523-8)
- [3]. Johnson, M. P. (2006). Conflict and control: gender symmetry and asymmetry in domestic violence. *Violence Against Women*, 12(11), 1003–1018. <https://doi.org/10.1177/1077801206293328>
- [4]. Myhill, A. (2015). Measuring coercive control: what can we learn from national population surveys? *Violence Against Women*, 21(3), 355–375. <https://doi.org/10.1177/1077801214568032>
- [5]. Stark, E. (2007). *Coercive Control: How Men Entrap Women in Personal Life*. Oxford University Press.
- [6]. Campbell, J. C., Webster, D., Koziol-McLain, J., et al. (2003). Risk factors for femicide in abusive relationships: results from a multisite case control study. *American Journal of Public Health*, 93(7), 1089–1097. <https://doi.org/10.2105/AJPH.93.7.1089>
- [7]. Stith, S. M., Rosen, K. H., Middleton, K. A., et al. (2000). The intergenerational transmission of spouse abuse: a meta-analysis. *Journal of Marriage and Family*, 62(3), 640–654. <https://doi.org/10.1111/j.1741-3737.2000.00640.x>
- [8]. Campbell, J. C., Webster, D. W., & Glass, N. (2009). The Danger Assessment: validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence*, 24(4), 653–674. <https://doi.org/10.1177/0886260508317180>
- [9]. Snider, C., Webster, D., O'Sullivan, C. S., & Campbell, J. (2009). Intimate partner violence: development of a brief risk assessment for the emergency department. *Academic Emergency Medicine*, 16(11), 1208–1216. <https://doi.org/10.1111/j.1553-2712.2009.00457.x>
- [10]. Sabri, B., Campbell, J. C., Khan, N. A., et al. (2024). Assessing risk for severe domestic violence and related homicides perpetrated by partners and in-laws: adapted danger assessments for women in abusive relationships in India. *BMC Public Health*, 24, Article 1870. <https://doi.org/10.1186/s12889-024-19364-2>

- [11]. Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 345–420. <https://doi.org/10.1613/jair.4992>
- [12]. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- [13]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [14]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- [15]. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- [16]. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- [17]. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- [18]. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- [19]. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [20]. Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). How does NLP benefit legal system: a summary of legal artificial intelligence. *Proceedings of ACL 2020*, 5218–5230.
- [21]. Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: the muppets straight out of law school. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2898–2904.
- [22]. Chouldechova, A. (2017). Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- [23]. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.
- [24]. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), Article 115.
- [25]. Mitchell, M., Wu, S., Zaldivar, A., et al. (2019). Model cards for model reporting. *Proceedings of FAT* '19*, 220–229.
- [26]. Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *IEEE Computer*, 29(2), 38–47.
- [27]. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.
- [28]. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [29]. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- [30]. Brooke, J. (1996). SUS: a quick and dirty usability scale. In P. W. Jordan et al. (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.