

Peer Review

Review of: "Decoding Complexity: CHPDA – Intelligent Pattern Exploration with a Context-Aware Hybrid Pattern Detection Algorithm"

Zahra Amiri¹

1. Ivy College of Business, Iowa State University, United States

The proposed system presents a well-structured approach to detecting and managing sensitive information by integrating regex-based pattern recognition, AI-driven named entity recognition (NER), and validation functions. The use of Google RE2 for regex matching ensures efficiency and security, while the secondary AI-powered rescan enhances accuracy by identifying entities that do not adhere to fixed patterns. The discussion on exact pattern matching using the Aho-Corasick algorithm is insightful, particularly in highlighting its ability to process multiple patterns simultaneously with linear time complexity. However, the explanation of computational trade-offs, such as increased memory usage, could be further elaborated. Similarly, the validation techniques for PII and PHI, including Luhn's algorithm and country-specific format checks, strengthen the system's robustness, but a more detailed exploration of real-time verification methods could improve clarity. The introduction of a proximity-based confidence scoring mechanism is a strong approach to reducing false positives, yet the rationale behind parameter choices, such as distance thresholds and weighting factors, could be more explicitly discussed.

The performance evaluation demonstrates the system's scalability, achieving high accuracy while maintaining low false positive rates. The comparison of regex libraries and exact match algorithms provides valuable insights into their trade-offs, though including standard deviations and more extensive performance benchmarks would enhance reliability. The system's efficiency on ARM-based devices suggests its suitability for lightweight environments, but further details on multi-user performance and real-time adaptability would strengthen its practical applicability. While the paper

successfully highlights the advantages of integrating AI with regex-based detection, the discussion on future improvements—particularly in terms of multilingual support, real-time adaptability, and enhanced data privacy mechanisms—could be expanded. Overall, the study presents a scalable and efficient framework for sensitive data detection, offering a strong foundation for future advancements in AI-driven data security solutions.

Declarations

Potential competing interests: No potential competing interests to declare.