

Peer Review

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

Samir Singh¹

1. Emerging Business Opportunity, Advantech (Taiwan), Taipei, Taiwan

- While the authors have clearly put significant work into benchmarking different regex engines and string matching algorithms, I found myself wanting more substance when it came to the AI component. They keep referencing this hybrid AI+Regex approach but never really explain how their NER model works, what data it was trained on, or any of the technical details that would let someone reproduce their work.
- The performance metrics look promising at first glance (91.6% F1 score), but without comparing CHPDA to any existing tools or approaches in the market, it's hard to know if this is actually an improvement over what's already available. Is this better than what Microsoft, Google, or open-source alternatives offer?
- I was also concerned about scalability. They show linear scaling with dataset size, which sounds fine until you consider enterprise environments dealing with petabytes of data. Would this approach still be feasible? I didn't find its explanation in this paper.
- The proximity scoring idea seems interesting, but it feels like they came up with it based on intuition rather than solid theoretical grounds. It would be more promising if there were more justifications for why this particular approach works.
- On the positive side, their benchmarking of regex engines is thorough, and the Aho-Corasick algorithm explanation is well done. The paper also honestly acknowledges limitations like multilingual support.

I think this work has potential, but in its current form, it leaves too many questions unanswered for me to be fully convinced. With some more details on the AI implementation and direct comparisons to existing solutions, it could be much stronger.

Declarations

Potential competing interests: No potential competing interests to declare.