

In our experiments with two named entity recognition (NER) systems, denoted as System A and System B, both models demonstrate strong overall performance, achieving high precision, recall, and F1-score. System B consistently outperforms System A in terms of accuracy, demonstrating an accuracy of 99% compared to 98% for System A. Notably, both systems excel in recognizing common entities such as persons, organizations, and locations, with F1-scores reaching or exceeding 99%.

However, challenges are observed in the recognition of less frequent entities, where precision and recall values vary.

System B exhibits improved performance in handling these less frequent entities, which shows its ability to generalize across a wider range of entity types. Despite these promising results, there might be a couple of limitations in our approach. First, we did not account for the imbalance in entity occurrences. Further research might include techniques such as oversampling to deal with this. Moreover, we only fine-tuned one type of BERT model: although powerful in NER tasks, its robustness and performance should be tested against cutting-edge models, including newer Large Language Models.