Extracting Clinical Entities from Discharge Notes Using a Pretrained BERT Model

For this task, I explored how a pretrained medical language model could identify and extract clinical terms from discharge summaries. I used a Hugging Face model: (NeverLearn/Medical-NER-finetuned-ner) built on top of BERT and optimized for Named Entity Recognition in clinical contexts. After loading the model and tokenizer, I ran discharge note texts through the pipeline and obtained predicted entities labeled with categories like diagnosis, symptom, treatment, medication, and follow-up action.

The model did a solid job pulling out relevant terms from the text. For example, in the sentence "*Patient showed improvement. Prescribed antibiotics for 5 days,*" it correctly flagged "*antibiotics*" as a medication and "*5 days*" as a duration. Similarly, phrases like "*follow-up scan next month*" were accurately categorized as clinical events and dates. While some entity boundaries and confidence scores weren't perfect, a good number of predictions were reasonably aligned with what a clinician might highlight.

That said, using a general-purpose model, even one fine-tuned for medical NER, comes with caveats. One key risk is entity ambiguity. For instance, "stable" might describe a condition or a patient's emotional state depending on context. Hallucination is another issue, especially if the model overextends its predictions into nearby text that doesn't actually contain clinical meaning. Also, since this model wasn't trained on my specific data, it may miss more institution-specific or nuanced medical terminology.

In terms of practical implications, this approach could be a useful first step in automating discharge summary parsing, perhaps as part of a larger pipeline for risk prediction or care coordination. The results are interpretable and actionable enough to help flag follow-up actions or extract problem lists without deep NLP customization.

With more time and more data, I'd fine-tune the model on a curated, labeled subset of our actual discharge notes to better adapt it to local clinical language. I'd also explore ensemble approaches or compare against other models like Flan-T5 or BioBART to see how generation-based models handle ambiguous phrases. Finally, evaluating performance against clinician-annotated ground truth would help surface blind spots and improve reliability for downstream use.