Summary:

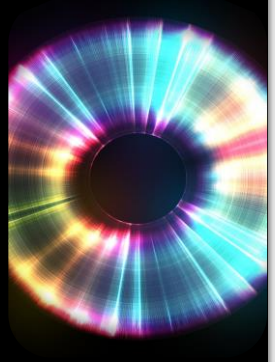
Drug Prescription Practices in Discharge Notes

Chart reviews are a commonplace procedure to maintain quality, compliance, ethics, and consistency of patient medical records. While lots of information can be extracted from structured fields, free-text notes, written by the clinician, can offer invaluable new, unique, or additional insights.

In this case, Natural Language Processing models, pre-trained for clinical text analysis, have been applied to identify and extract drug names, with their corresponding dose/route/frequency from a discharge note. This pipeline can be used for

1. Extracting the target combination: drug/dose/route/frequency to review the clinician's prescription practices.
2. Comparing the drug record from discharge notes to the one from the Medications section of the chart for quality and compliance.
3. Re-purposing the pattern to extract other target entities from the notes such as complaints, tests, diseases.

Repository on GitHub: https://github.com/olga12kz-DS/NER_Discharge_Note/tree/main



Data preparation: extraction, formatting, and processing

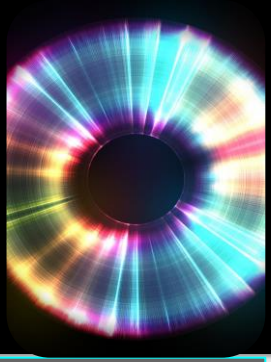
Extract “Discharge Notes” field from the full dataset using list **comprehension and regex pattern.**

Clean up the text by removing unnecessary characters and extra spaces using **regex**.

Compare performance of the pre-trained models on processing time, entities detected (unique vs. shared).

Create **pipelines** to process the text and **display results** in structured format.

Building an optimal model



Three pre-trained models and medspaCy library were tested and compared:

1. en_core_sci_sm:
   - light-weight, optimized for speed rather than accuracy.

2. en_core_sci_md:
   - medium-size, balanced between speed and accuracy.

3. **en_ner_bc5cdr_md:**
   - **medium-size, trained on BC5CDR corpus, biomedical and scientific texts, optimized for Name Entity Recognition for diseases and chemicals.**

4. MedSpaCy:
- Python library built on spaCy for processing clinical and medical text.

Best Performance | Example of Output

Combination of en_ner_bc5cdr and regex pattern matching

Drug	Dose	Route	Frequency
Morphine	4mg	IV	x4
Ondansetron	4mg	IV	x2
Lorazeman	0.5mg	IV	x1
Ceftriaxone	2g	IV	x1