**Background:**

Immune-checkpoint inhibitors (ICIs) are associated with unique immune-related adverse events (irAEs). Comprehensive exploration of irAEs in the real-world setting has been limited due to unstructured nature of clinical notes in EHRs necessitating significant manual chart review. We tested the capabilities of Large-Language Models to extract specific irAEs from unstructured clinical notes.

**Methods:**

Eight- hundred seventeen cases were manually reviewed by clinicians and annotated for irAEs at the patient level. We focused on colitis, as the in this use-case because it had the highest prevalence in our database. The de-identified notes were cleaned and pre-processed. A classification model was used as a first layer of the pipeline to select sentences with expressions related to colitis. A BioClinicaBERT model was fine-tuned to extract words representing colitis-related drugs or terms. Next, we fine-tuned an intermediate annotator model to construct a training corpus from our dataset. The annotator model was trained to extract general drugs and AEs from clinical notes using Medline ADE corpus. The resulting model from the intermediate annotation step was further fine-tuned on the irAE clinical notes. The output of the model was the presence or the absence of a mention of colitis in the note with a highlighting of the words mentioning the events and the drugs.

**Results: We leveraged 817 manually annotated clinical notes. XX patients developed any grade irAEs and xx patient developed colitis. XX samples were used for training and xx for testing.** The classification model of sentence mentioning colitis or not achieved an F1-score of 0.77 and an accuracy of 0.84. The BioClinicalBERT model achieved an F1-score of = 0.91 and an accuracy of xx after being fine-tuned on irAE clinical notes.

**Conclusion:**

We developed a pipeline to automatically extract drugs and irAEs from clinical notes of 1400 patients treated with ICI. The combination of BioClinicalBERT and Medline ADE corpus showed are promising large-language models to extract clinical attributes from unstructured clinical notes. Further investigations are needed to test the model on larger dataset and other irAEs.