A101 · Transferring Disease Knowledge from Biomedical Literature to Longitudinal Clinical Records for Inborn Error of Immunity Phenotyping

Presenter: Mansooreh Ahmadian — University of Colorado Anschutz Medical Campus

Authors: Mansooreh Ahmadian, Zhixin Lun, Pia J. Hauk, Sara J. Deakyne Davies, Todd Miller, Jordan Abbott, Elena W Y Hsieh

Session: Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Electronic health record (EHR)-linked biobanks require accurate phenotyping for cohort selection from longitudinal records. We developed a biobank of 123 patients with inborn errors of immunity (IEIs) and non-IEI controls, comprising more than 23,000 clinical notes. IEIs have heterogeneous, overlapping phenotypes, and diagnosis often requires integrating years of clinical evidence. General-purpose large language models (LLMs) identified IEI cases but struggled with complex patients, particularly those with similar autoimmune disorders.

We propose a hybrid framework combining longitudinal LLM reasoning with specialized IEI models trained using ontology-based distant supervision. We use the International Union of Immunological Societies (IUIS) classification to define the IEI disease set and its Online Mendelian Inheritance in Man (OMIM) identifiers to connect diseases to ontology concepts. Ontology names, synonyms, and lexical variants provide distant supervision for IEITagger, an IEI-specific tagger combining lexical matching with PubMedBERT fine-tuned for concept recognition and normalization. IEITagger identifies IEI mentions in literature and extracts disease-labeled snippets, which train disease-specific classifiers. Disease mentions are masked and phenotypically related non-IEI diseases are used as hard negatives to promote learning of phenotypic context rather than terminology alone.

The specialized models are integrated with LLM-derived longitudinal patient representations to predict IEI status and rank candidate diagnoses. The hybrid framework improves IEI classification over general-purpose LLM prompting, particularly in ambiguous cases with overlapping autoimmune and inflammatory phenotypes. These preliminary results demonstrate that disease-specific knowledge learned from biomedical literature is transferable to longitudinal clinical records, supporting rare-disease phenotyping without manually annotated training data and more precise cohort discovery.