



OHDSI Workgroup Objectives and Key Results (OKR)

NLP Workgroup

Hua Xu



NLP Workgroup Purpose

NLP WG exists to promote the use of textual information in electronic health records (EHRs) , to facilitate the generation of evidence for observational studies.

- Develop **standard representations** for clinical text and NLP output data
- Build **methods and tools** to facilitate textual data processing
- Conduct **cross-institutional studies** and disseminate **best practice of using textual data** for real world evidence generation



NLP 2025 Objectives and Key Results

Objective 1: Conduct multi-site clinical studies that utilize both structured and textual data

Key results

1. Characterizing the Anticancer Treatment Trajectory and Pattern in Patients Receiving Chemotherapy (Oncology) – annotation guideline development
2. Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study (Psychiatry) – revising the protocol
3. Social Determinants of Health and Treatment Outcomes in Type 2 Diabetes: A Multi-Site Analysis of LEGEND-T2DM study – finalizing SDoH factors and subset of Treatment-Comparator-Outcome pairs

Objective 2: Open source LLMs for information extraction from clinical notes

Key results

1. Development of Kiwi - An LLM-based Clinical Information Extraction System
2. Expanding Kiwi for open-source community development

Objective 3: Knowledge dissemination

Key results

1. Monthly presentations on NLP advances (~ 30 attendees on average in 2024)
2. Book of OHDSI – NLP Chapter Textbook -> Cookbook