# Lab Shared Folder – Project References and Analyses

This repository organizes reference files, analyses, and project notes related to clause-level classification and graph-based methods for patient-provider communication.

## Folder Structure

### **EPPCminerProject**

* **Yi-Chun’s Files**
  + All files moved to Rimi\_Project\_References

### **PVminderProject**

* **Yi-Chun’s Files**
  + None currently

### **Rimi\_Project\_References**

#### **MeetingNotes/**

* Five files containing initial discussions (May 11–21, 2025).  
  Document early design choices, annotation issues, and task framing.

#### **Proposal and Annotation Analysis**

* **2025May\_Annotation\_Analysis.pdf** – Analysis of annotation consistency, span statistics, and label distributions.
* **2025Healthcare\_Proposal\_Ideas\_with\_Scenario.pdf** – Drafted proposal ideas with scenarios.
* **2025May\_Anonotation\_Analysis.pdf** – Duplicate/alternate annotation analysis draft.

#### **Graph-based Methods/**

* **Ambiguity Analyses**
  + **Code/Subcode/Combined per-label disagreement** – Per-label metrics showing how often a label’s nearest neighbors belong to different classes. High disagreement suggests fragile or ambiguous labels.
  + **Ambiguous label pairs** – Table of label pairs frequently confused in nearest-neighbor space. Helps identify overlapping categories or unclear boundaries.
  + **Conflict samples** – Concrete example pairs of texts and labels where neighbors disagree, useful for qualitative error analysis.
  + **Label disagreement summary** – Aggregate metrics that highlight which labels have the weakest separability.
* **Cross\_clause\_splits/**
  + **Span\_flags** – Flags annotations whose spans extend across clauses or appear too long.
  + **No\_cross\_clause** – Subset of annotation data where spans do *not* cross clause boundaries.
  + **Only\_cross\_clause** – Subset of annotation data where spans *do* cross clause boundaries.
* **2025\_Simple\_ML\_Testing.pdf**  
  Results of ablation tests with simple machine learning baselines (logistic regression, bag-of-words). Compares span-based vs. text-based representations for goal-oriented intent labels.
* **2025\_Graph\_States.pdf**  
  Graph-theoretic measures (e.g., energy, Weisfeiler-Lehman test scores) computed across syntactic, semantic, and narrative graph abstractions. Shows whether graph structure adds new information beyond text embeddings.

#### **Proposal Reference Collections**

* **1\_Visual Narrative for Communication Efficiency/**  
  References compiled when preparing the proposal on visual narrative methods.
* **2\_Value Aligned Learning Experiences/**  
  References compiled when preparing the proposal on value-aligned learning scenarios.