DISCUSSION 2025
MAY

Revision history

Updated date	Note
2025.04.30	Initial discussion
~~2025.05.01~~	~~Potential NSF calls info (only info collection)~~
2025.05.02	About the project
2025.05.09	Potential tracks
2025.05.11	Potential methods and initial plans
2025.05.11	Potential early venues

Understanding of Lab Project & My Framing (1/2)

Topic:

Mining and Modeling Patient-Provider Communication Through Annotated Text Data

Objective (Lab-wide):

- Define and refine a nested CODE schema (categories and subcategories) for patient-provider messaging.

- Develop and compare automated annotation methods (LLM, BERT, others) to support scalable labeling.

- Use the resulting system to understand communication patterns, patient values, and outcomes.

💡 Prepared based on Week 1 discussions to ensure alignment and clarify where my own methods may contribute uniquely.

Understanding of Lab Project & My Framing (2/2)

Observed Research Questions

- How can we define a taxonomy of communication intents and values from unstructured text?

- What methods are most effective for automating annotation: LLMs, BERT, topic modeling, etc.?

- Can we support flexible annotation (merging, splitting CODEs) over time?

- How do different modeling approaches compare in accuracy and interpretability?

💡 Prepared based on Week 1 discussions to ensure alignment and clarify where my own methods may contribute uniquely.

Potential tracks

Motivation

- Current CODE system has **nested categories** with semantic overlaps.

- Manual annotation is **time-consuming and rigid**, making automation difficult.

- Existing models (LLM, BERT) focus on accuracy but **lack reasoning structure**.

My Proposed Direction

Goal:

- Build a graph-based framework to support semantic annotation reasoning—from high-level intents to fine-grained subcategories.

Key Ideas:

- Represent sentences, CODEs, and their relations as a **knowledge graph.**

- Apply **graph neural networks (GNNs)** or hybrid models for flexible classification.

- Explore **hierarchical decision layers** for CODE/sub-CODE structure.

Why This Matters

- Novel method distinct from other team directions.

- Leverages my expertise in narrative modeling and graph-based reasoning.

Collaboration Considerations

- I can explore graph-based modeling without overlapping current directions (LLM / BERT pipelines).

- Open to integrating my method into shared benchmarks later.

- Interested in long-term extensions (e.g., interaction simulation, narrative structure).

Prototype Idea

Prototype Idea – Graph-Based Annotation Reasoning (Mini Version), 1/2

Goal

- Test whether a graph-based structure can support reasoning for CODE prediction in unstructured patient-provider messages.

What I Plan to Build (Mini Version)

- Construct a partial graph from unlabeled messages:

- Nodes: message, speaker role, extracted concepts (e.g., "pain", "medication"), emotional cues.

- Edges: mentions, expresses, role, similar_to (if embeddings used).

- Try rule-based reasoning and/or a small GNN model to predict CODEs.

Prototype Idea – Graph-Based Annotation Reasoning (Mini Version), 2/2

Why This Matters

- Adds a structural reasoning layer missing from current BERT/LLM pipelines.

- Supports semantic flexibility in annotation (e.g., similar messages, soft CODE grouping).

- Aligns with my prior work on knowledge graphs and narrative structure.

Next Steps

- Start with ~100 annotated messages.

- Build simple graph-based classifier (baseline).

- Compare to flat classifiers (e.g., BERT).

Possible venues

Possible venues for early feedback

Workshop (Full Name) @ ECAI ()	Focus	Deadline(s)	
HC@AIxIA + HYDRA 2025 International Joint Workshop of Artificial Intelligence for Healthcare and HYbrid Models for Coupling Deductive and Inductive ReAsoning	AI in healthcare; hybrid symbolic + neural reasoning	June 15 / July 20 / Aug 25	Direct match for graph-based CODE reasoning
AIBio 2025 1st Workshop on Artificial Intelligence for Biomedical Data	Applied AI for biomedical and clinical datasets	May 20	Relevant to modeling on patient-provider data
EXPLIMED 2025 *2nd Workshop on Explainable Artificial Intelligence for the Medical Domain*	Transparency and interpretability in medical AI	Est. June (TBD)	Graph reasoning can support explainable labeling

Other info for references

Possible NSF calls

NSF CISE Core Programs (NSF 24-589)

Focus: Foundational research in AI, computing, information science, and cyberinfrastructure, link:

- **Relevant Tracks** (via CISE/IIS):
  - *Robust Intelligence (RI)* – AI and reasoning systems
  - *Human-Centered Computing (HCC)* – human-AI interaction, multimodal interfaces
- **Funding Levels**:
  - Small: ≤ $600K, 3 years (submit anytime)
  - Medium: $600K–$1.2M, 4 years (Oct 1–23 annually)

NSF HNDS (NSF 23-568)

Focus: Human behavior via network/data science + infrastructure to support SBE research, link:

- **Relevant Tracks**:
  - HNDS-R (Core Research) – theory building via network/data science (by permission)
  - HNDS-I (Infrastructure) – tools and datasets to enable SBE discovery
- Deadlines:
  - HNDS-I: Feb 6 & Aug 7
  - HNDS-R: Jan 9 & Jul 10 (permission required for direct submission)

Additional info

Eligibility (J-1 Postdoc at Yale):
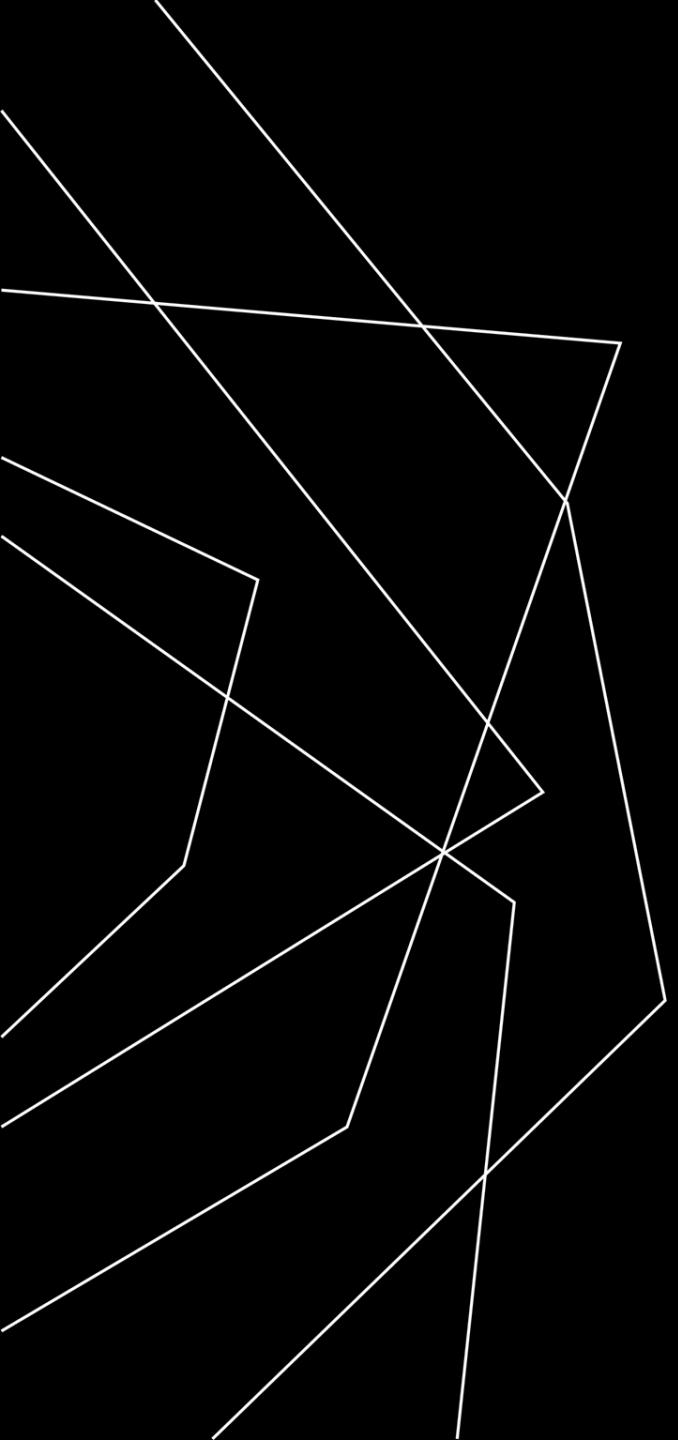
- ❌ Cannot serve as PI

- ✅ Can join as Co-PI or Senior Personnel with Yale's approval

- NSF allows participation if affiliated with a U.S. institution

My Contribution:

- Expertise: Multimodal AI, Knowledge Graphs, Narrative Modeling

Discussion Points:

- Do you see potential to align this with your current research in healthcare data?

- Could we co-develop a proposal under one of the two NSF programs?

- Would you consider leading as PI, with me supporting as Co-PI or key personnel?

THANK YOU

Yi-Chun (Rimi) Chen

984-500-8488

ychen74@alumni.ncsu.edu