

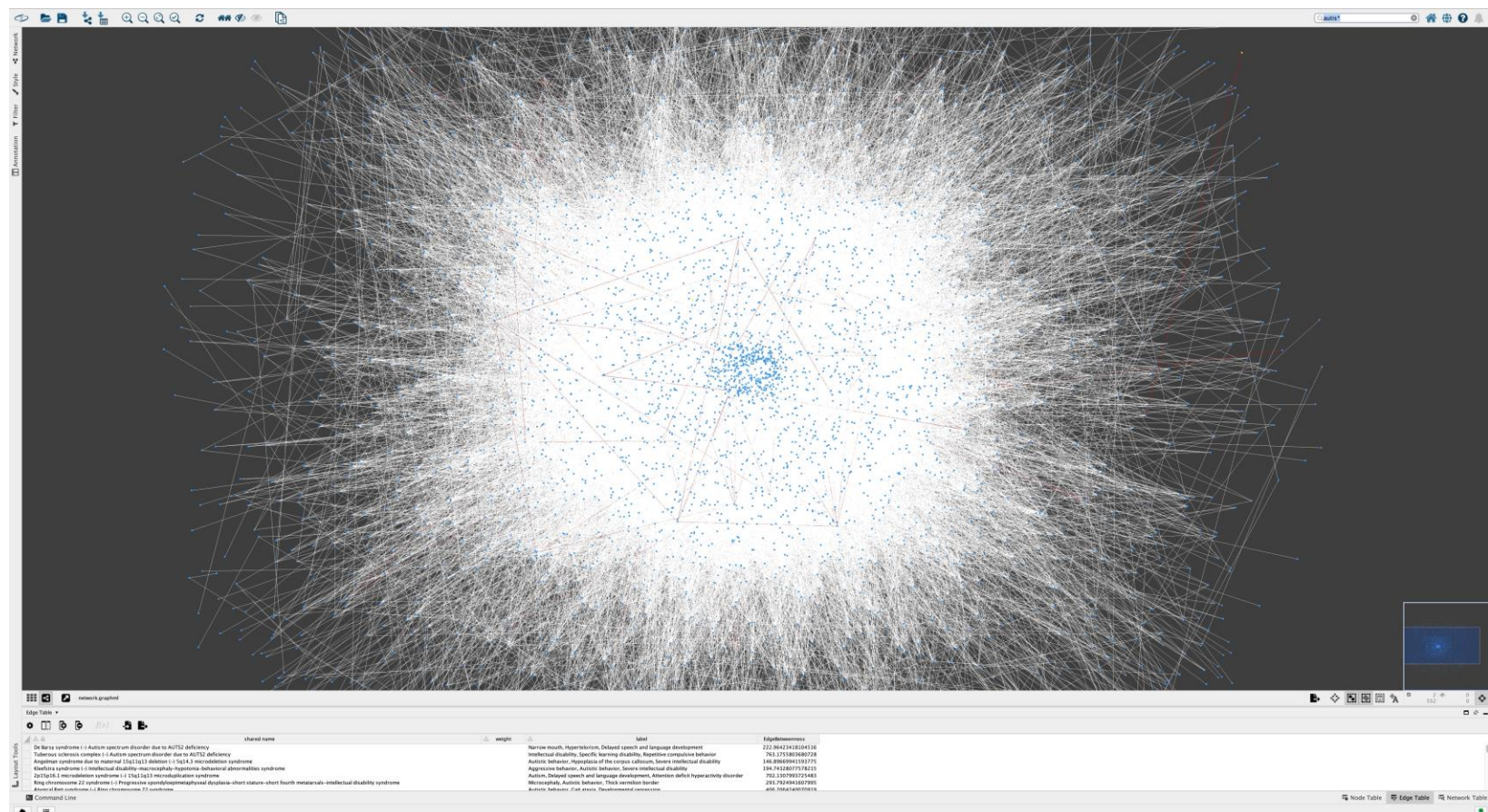
Rare Disease Discovery

Adam Cross^{1,2}, John Wu³, Lang Cao³, Jimeng Sun³

UICOMP¹, Jump Simulation², UIUC³

PROBLEM TRYING TO SOLVE

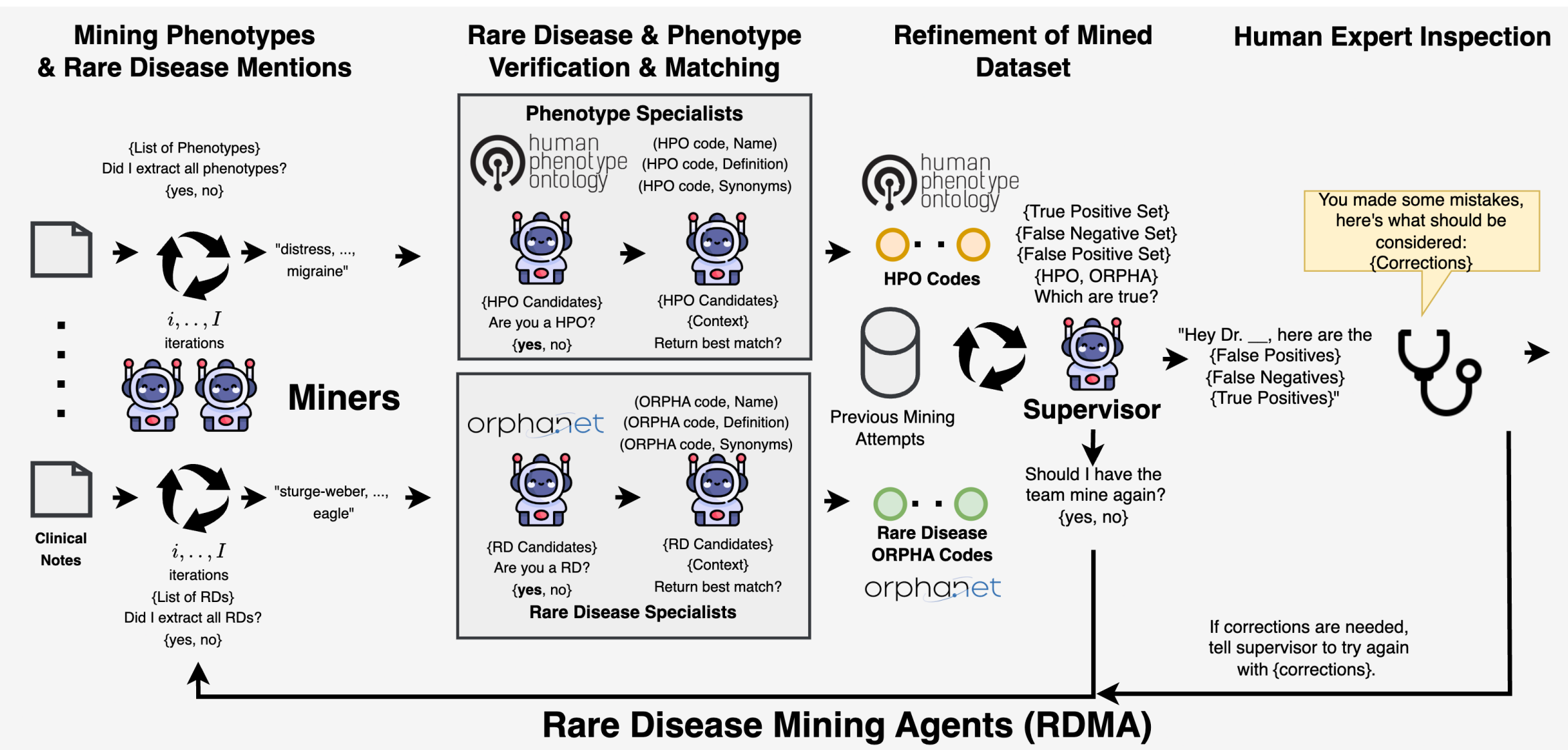
- In aggregate, 1 in 10 people has a rare disease globally
- Over 9000 rare diseases exist, and physicians cannot know them all
- Providers do not always recognize when a patient needs genetic testing
- Genetic testing is costly
- Referrals take time (months to years)



TOMORROW

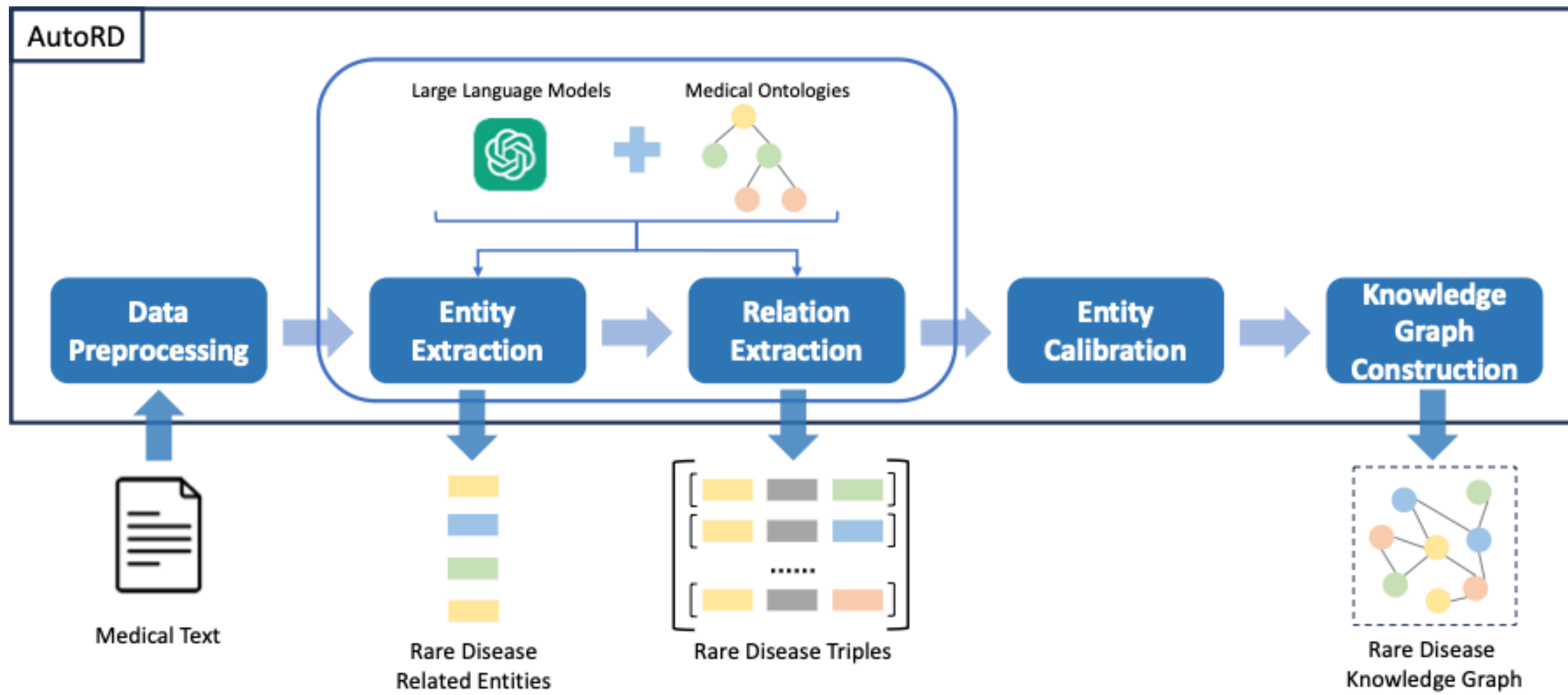
OUTCOME/GOALS/IMPACT TO PATIENT

- Goal 1: create a system of finding patients with known rare diseases in the electronic health record (cannot simply use ICD-9/10 codes because there are none for most rare diseases)
- Goal 2: Build a pipeline to identify patients at high risk for undiagnosed rare diseases
- Goal 3: Provide this information to physicians and test real-world performance like precision, recall, F1, sensitivity, specificity, and user feedback
- Goal 4: Refine, iterate, and disseminate for multi-site studies



JOURNEY TO GET THERE/PLANNED JOURNEY

- Project has spanned 2.5 years with \$300,000 in internal funding
- Initial phase involved the development of a rare disease knowledge graph and advanced natural language processing
- Current phase leverages large language models and AI agents to mine for potential phenotypes
- Presenting current work at PAS 2025 and submitted abstract to AMIA 2025
- Manuscript published in JMIR Medical Informatics
- Second manuscript pending



DIRECT IMPACT TO PATIENT/FAMILIES

- Patients with known rare diseases are more readily and comprehensively extracted from electronic health records
- Patients at high risk for an undiagnosed rare disease are discovered
- Relevant disease information and model reasoning are provided to the clinician
- The clinician decides whether to pursue additional testing, referral, or treatment

