

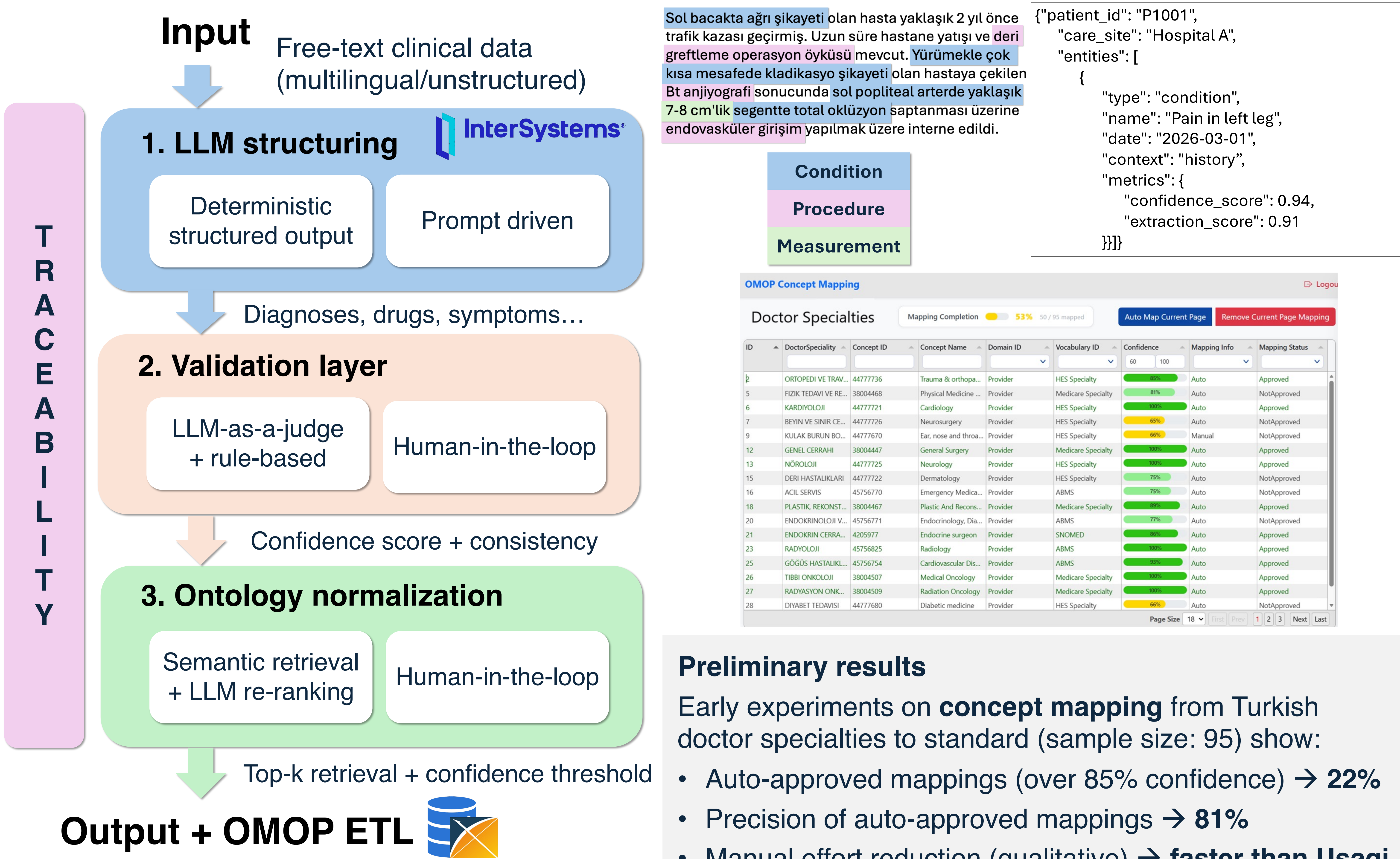
Explainable, Governance-Ready Clinical NLP Pipeline for OMOP-CDM Using Locally Deployed LLMs

Alisson Licon¹, Muhammad Waseem¹, Alex Loleit¹
¹Traverse Health

Clinical free-text notes contain **rich patient information** but remain difficult to **reuse for large-scale research, limiting integration** into standardized frameworks such as OMOP-CDM. While NLP methods have improved clinical information extraction, traditional rule-based systems lack flexibility, and modern machine learning approaches often face **challenges in interpretability, scalability, and cross-lingual adaptation**. Mapping local clinical concepts —especially from non-English data— to standardized vocabularies remains a critical bottleneck.

We propose an **explainable, end-to-end clinical NLP pipeline** that transforms free-text into standardized, research-ready data aligned with OMOP-CDM.

A preliminary implementation has been developed, focusing on the ontology normalization stage. In parallel, an LLM-based structuring module has been implemented to convert free-text into structured representations, currently undergoing validation.



Preliminary results

Early experiments on **concept mapping** from Turkish doctor specialties to standard (sample size: 95) show:

- Auto-approved mappings (over 85% confidence) → **22%**
- Precision of auto-approved mappings → **81%**
- Manual effort reduction (qualitative) → **faster than Usagi**

Early experiments on **LLM structuring** of Turkish doctor notes show:

- Processing time → **~3–4 sec / note (local LLM)**
- Real-time structuring + storage**

Conclusion

We present an **explainable, end-to-end framework** for structuring clinical free text into OMOP-aligned data. The pipeline demonstrates the potential to enhance standardization, **transparency**, and scalable research. Local LLMs ensure **data governance and accessible deployment**

Future work will focus on **full pipeline integration, large-scale validation, and benchmarking** against existing OMOP mapping approaches

