promptly

AI can automate up to 80% of OMOP vocabulary mapping while maintaining controlled error.

Performance Evaluation of an AI-Based Vocabulary Mapping Platform for the OMOP Common Data Model

Background: Vocabulary mapping remains one of the most challenging and time-consuming steps in transforming real-world clinical data into the OMOP Common Data Model. Source terminologies are often heterogeneous, locally defined, and semantically ambiguous, limiting the scalability of traditional expert-driven workflows. Artificial intelligence offers an opportunity to enhance this process by combining probabilistic ranking with calibrated confidence, enabling selective automation while preserving semantic validity.

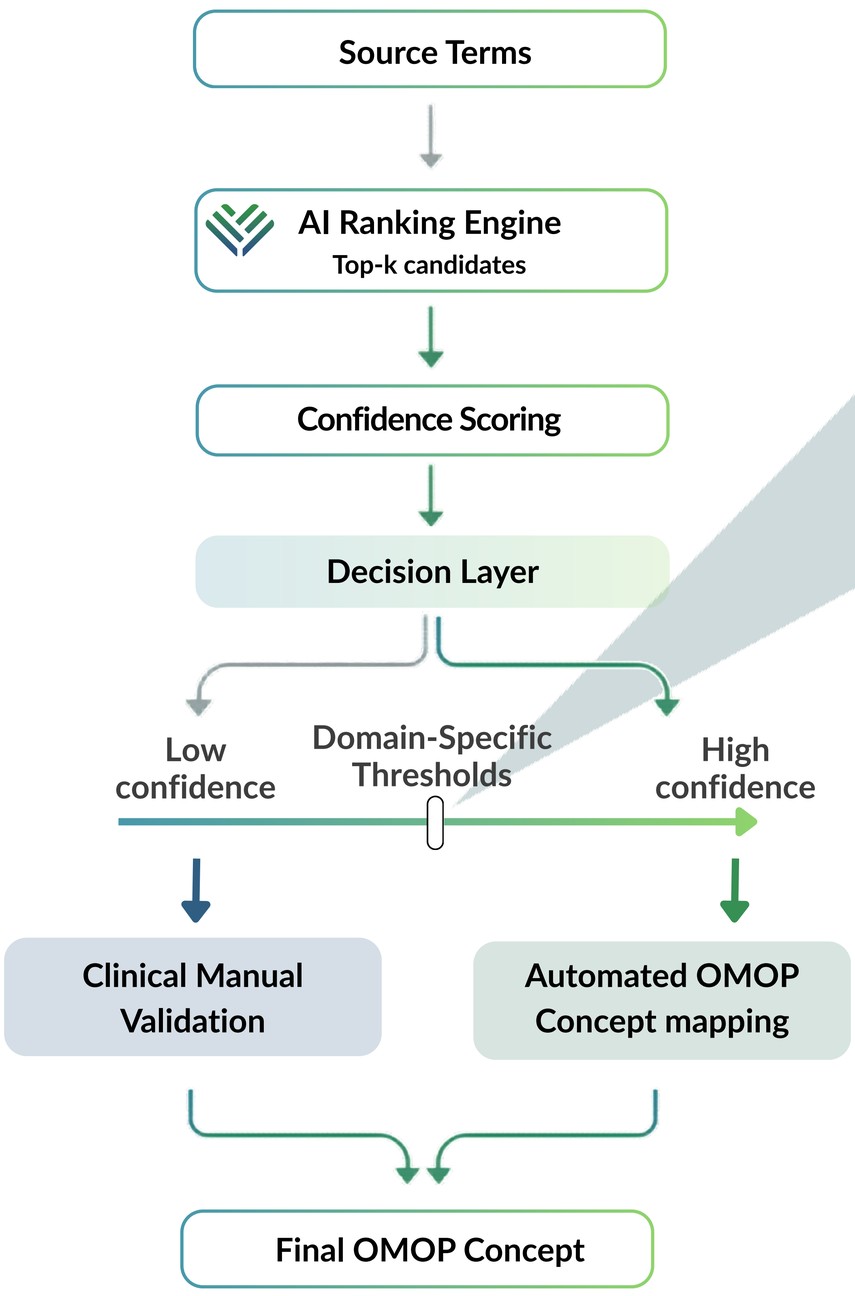

Figure 1. *AI-driven OMOP vocabulary mapping workflow*

We implemented an **AI-driven vocabulary mapping platform**[1] that generates ranked OMOP candidate concepts for each source term and assigns calibrated confidence scores based on semantic similarity and contextual alignment. Domain-specific thresholds determine whether mappings are accepted automatically or routed for expert validation. **Performance was evaluated** across Drug, Measurement, and Condition domains using **Recall@1/5/10, MRR@10, Automation Rate, and False Discovery Rate**.

OMOP Domain	Optimal Threshold	False Discovery Rate (FDR)	Automated Mappings (%)	Manual Validation Required (%)
Drugs	79,0%	13,4%	79,2%	20,8%
Measurement	87,5%	13,4%	81,0%	19,0%
Condition	90,5%	15,0%	70,5%	29,5%

Figure 2. *Comparative Summary Across OMOP Domains*

Domain-specific threshold analysis revealed a consistent trade-off between automation coverage and semantic accuracy across OMOP domains. Measurements achieved the highest automated mapping rate while maintaining low semantic error, whereas Conditions required stricter thresholds due to greater ambiguity. Across all domains, ranking metrics remained robust, indicating that non-automatically accepted mappings were still frequently retrieved among top-ranked candidates, thereby supporting efficient expert validation.

OMOP Domain	Recall@1 (%)	Recall@5 (%)	Recall@10 (%)	MRR@10
Condition	51,85	62,57	64,85	0,565
Procedure	47,20	68,44	72,57	0,559

Figure 3. *Recall and Ranking Performance by OMOP Domain*

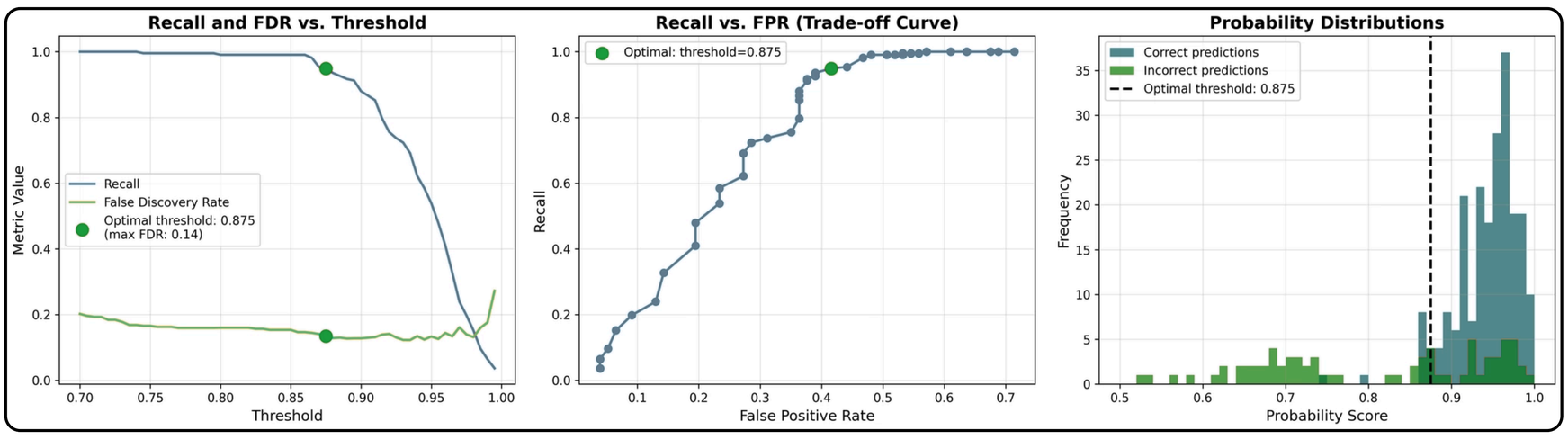

Figure 4. *Threshold analysis for the Measurement domain*

Limitations: This evaluation represents an initial assessment of the platform and is based on a relatively limited dataset. Performance was assessed within selected OMOP domains, and results may differ in other domains with distinct semantic characteristics. In addition, no direct external benchmark comparison with established mapping tools was performed in this study.

Conclusion: AI-driven vocabulary mapping can substantially reduce the manual burden of OMOP ETL by automating a large proportion of mappings while maintaining controlled semantic error. By combining probabilistic ranking, calibrated confidence thresholds, and human-in-the-loop validation, this approach enables scalable, reliable, and efficient standardization of heterogeneous clinical data into the OMOP Common Data Model.

OHDSI

Miguel Ferreira[1,2], Paulo Ferreira[1], Miguel Rebelo[1], Catarina Pires[1], Luís Magalhães[1]

[1]Promptly Health
[2]Faculty of Medicine - University of Porto

promptly