

Mind Meets Machines?: AI Assisted Semantic Mapping of Clinical Concepts

Andrew S. Kanter^a, Jonathan Payne^b, Filipe Lopes^c, June Bronnert^d, Theresa Rihane^d

^aDepartment of Biomedical Informatics, Columbia University; ^bOpen Concept Lab; ^cOrango I/O (Health);

^dIntelligent Medical Objects, Inc. (IMO Health)

Background

The use of artificial intelligence to assist semantic mapping — the translation of local source concepts into standard reference vocabularies — is of growing interest across the OMOP/OHDSI community, where mapping is the difficult, time-consuming gateway step of every extract-transform-load (ETL) pipeline and is constrained by a global shortage of trained terminologists [1]. Whether AI can do this work reliably, and what role curated clinical interface terminologies play in making it reliable, remains poorly characterized [3]. We evaluate a multi-algorithm, AI-assisted mapping method against an ICD-11 target as a generalizable test case: ICD-11 is a demanding, heavily postcoordinated target [2], and the same pipeline transfers directly to mapping local concepts onto OMOP standard vocabularies. The bridge to OMOP is concrete — the Columbia International eHealth Laboratory (CIEL) dictionary is already represented in Athena as a non-standard vocabulary that cross-walks to standard OMOP concepts, and in operational ETL (e.g., the OHDSI Africa working group) source concepts are mapped to CIEL first and resolved with Usagi second [4,5]. A method that strengthens the CIEL-first step therefore sits upstream of, and complements, existing OHDSI tooling. The Open Concept Lab (OCL) is an open-source terminology server designated as a digital global good (DPGA registry ID: GID0091038) and the primary distribution medium for CIEL, itself a global good (DPGA registry ID: GID0090672); with several open-source, private, and not-for-profit partners, OCL has developed an AI-assisted mapping tool, OCL Mapper. This brief report describes the Mapper, the role of interface terminology in reaching standard codes, and an initial, pre-publication evaluation against ICD-11.

Methods

We assembled a validation set of 100 high-frequency clinical terms drawn at random from the IMO concept repository, deliberately excluding terms that are merely code descriptions. Two authors independently mapped each term to ICD-11 for Mortality and Morbidity Statistics (MMS) at maximum specificity, with a third author resolving disagreements; 68 of the 100 gold-standard codes are postcoordinated (a stem plus one or more X-extension or slash-compound codes), making this a deliberately demanding target. Each term was processed by OCL Mapper, which generates candidate ICD-11 targets from four algorithms: (1) the **CIEL Bridge**, a human-curated interface terminology that provides deterministic maps to ICD-11; (2) the WHO ICD-11 linearization search API; (3) the WHO ICD-11 autocode API; and (4) the **OCL ICD-11 Agent**, a custom "LLM-as-terminologist" that emulates a trained coder by constructing valid postcoordinated expressions through the WHO tooling rather than relying on free-text prompting. The candidate pools are then adjudicated by an "LLM-as-judge" **AI Assistant**, which returns a primary recommendation and up to two alternates, each labelled RECOMMENDED, CONDITIONAL, INSUFFICIENT, or REJECT with a confidence signal. Matches were scored strictly: a candidate counts only if it is string-identical to the gold standard after canonicalization, with postcoordination clusters compared as unordered stem and extension sets and residual codes (.Y/.Z) treated literally (Figure 1). Code normalization was performed on the post-coordinated ICD-11 codes to ensure that ordering of code components did not affect the match.

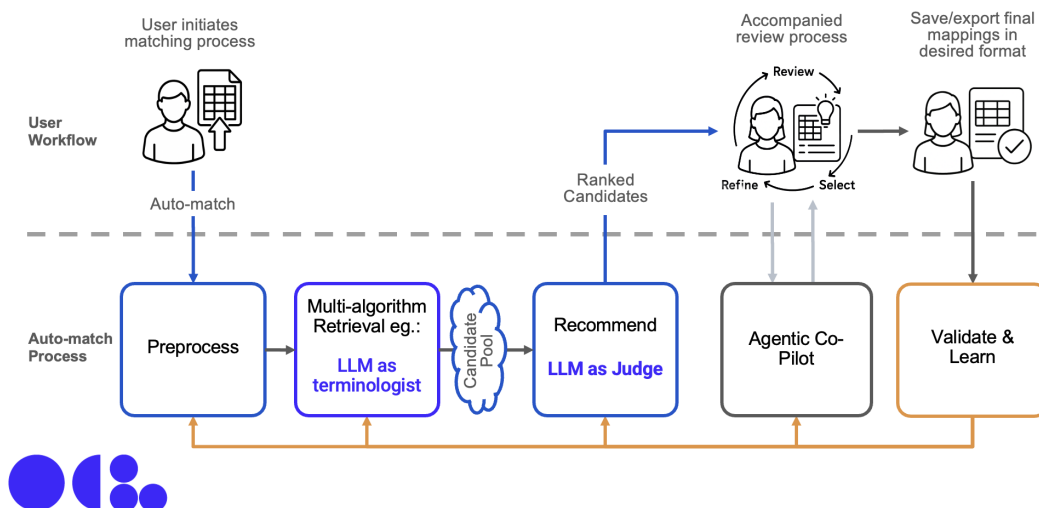


Figure 1. Overview of the OCL Mapper process, with AI components (LLM-as-terminologist and LLM-as-judge) highlighted in blue.

Results

Across the 100 terms, the candidate pool contained the correct ICD-11 code for 71, and the single best candidate across the four algorithms was correct for 69. The OCL AI Assistant's primary recommendation was strictly correct for 59 terms, and its top three (primary or alternate) for 70. When the answer was present in the pool the AI Assistant identified it 99% of the time (70 out of 71), locating the current limit at retrieval rather than judgment. The four algorithms proved complementary rather than redundant: for 23 of the 71 correctly-mapped terms (32%), only a single algorithm surfaced the correct candidate, so no individual source was sufficient on its own. The CIEL interface terminology was the strongest single contributor — correct at rank one for 51 terms — and, more tellingly, the least replaceable: it uniquely surfaced the correct code for 14 terms that no other algorithm found, far more than any other source. When the bridge surfaced the answer, the AI Assistant selected it as the top recommendation in 47 of 55 cases (85%). The CIEL Bridge algorithm properly identified 20 additional source terms but the CIEL-mapped ICD-11 codes differed from the validation set. Five percent were incorrect and 15% were semantically similar but not identical to the validated code. Postcoordination demands were substantial — for example, "severe sepsis" maps to 1G40&XS65, and right- versus left-sided "upper-quadrant pain" to MD81.10&XK9K versus MD81.10&XK8G — and both the curated bridge and the agent contributed postcoordinated candidates (Figure 2). Because CIEL is a non-standard Athena vocabulary that cross-walks to standard OMOP concepts, these maps feed OMOP ETL directly. These are preliminary results on a small, strict-only sample; a fuller evaluation is in preparation.

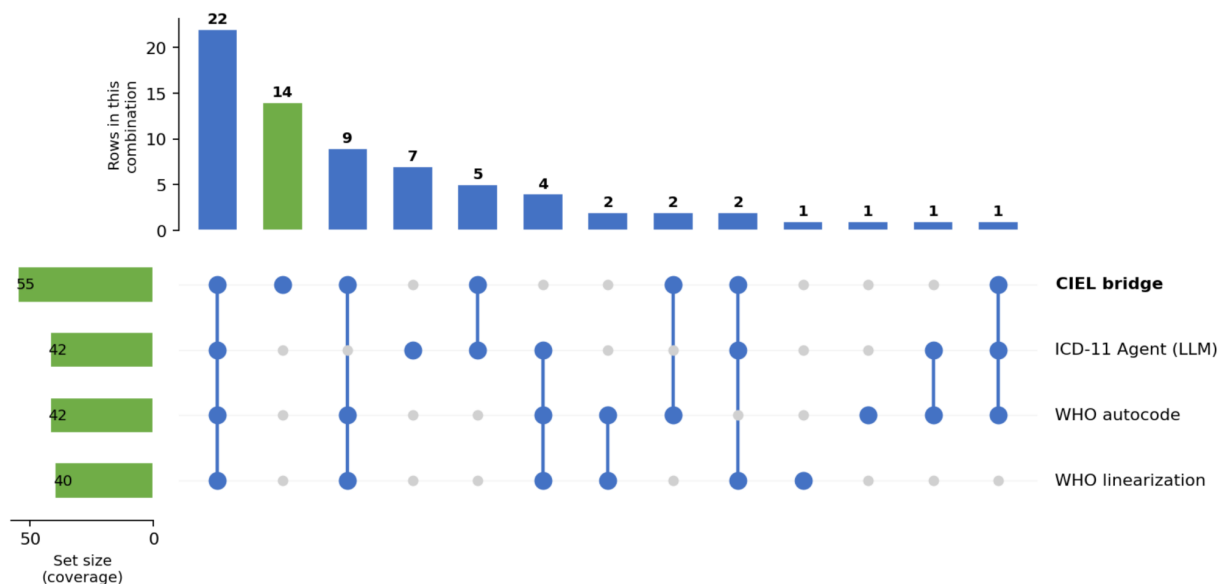


Figure 2. Candidate-pool coverage & unique contribution by algorithm. Green bar = CIEL's 14 rows found by no other algorithm. (UpSet plot; n = 100, strict matching).

Conclusions

This initial evaluation suggests that AI can meaningfully assist ICD-11 mapping, but that its reliability rests on the breadth and quality of the candidate pool it is given. Two capabilities stand out as complementary. A human-curated interface terminology — CIEL — was the strongest and least replaceable single source, underscoring that curated terminology infrastructure remains central even as AI advances. At the same time, an agentic "LLM-as-terminologist" that constructs postcoordinated expressions is a promising, emerging contributor that can in principle reach codes beyond any fixed retrieval pool; it complements, rather than replaces, the curated maps. Human mapping is redacted but still required to achieve 100% coverage. Because the OCL Mapper method is vocabulary-agnostic, the same pipeline applies to mapping local concepts onto OMOP standard vocabularies, where a CIEL-first, Usagi-second workflow positions it as complementary to existing OHDSI tooling [4,5]. For settings with limited terminologist capacity, such assistance is a practical accelerator. OCL Mapper is open source, and we welcome collaboration with the OHDSI community on its evaluation and use.

References

1. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574–578. PMID: 26262116.
2. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. *BMC Med Inform Decis Mak.* 2021;21(Suppl 6):206. doi:10.1186/s12911-021-01534-6.
3. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc.* 2006;13(3):277–288. PMID: 16501181.
4. Observational Health Data Sciences and Informatics. The Book of OHDSI. 2021.

<https://ohdsi.github.io/TheBookOfOhdsi/>

5. Usagi: mapping source codes to OMOP standard vocabularies. OHDSI.
<https://ohdsi.github.io/Usagi/>