

Any-to-Any Terminology Mapping: AI-assisted mapping of local terminologies

Julio Cesar Barbour Oliveira ¹, Vinicius Anjos de Almeida ¹, Gabriela Naomi Ishibe ¹, Giovanna Bertelli de Gouveia ¹, Kalil Vitor Marto Bernal ¹, Karen Barbour Oliveira ¹, Rodrigo Gonçalves Queijo ¹, Vinícius Goularte-Silva ¹, Adriano Yudi Tugava ², Ana Paula dos Santos Cardoso ², Ariella Aro ², Bruno de Andrade Zanesco ², Elene Paltrinieri Nardi ², Évellin do Espírito Santo ², Gabriela Boscachi Flavio ², Gustavo Moreira Rodrigues ², Ian Bezerra ², Ingrid Mayara Benachio ², Letícia Lanzillotta de Mello ², Luís Guilherme Espinosa e Souza ², Maria Beatriz Magalhães Scabim ², Maria Letícia Maeus Trindade ², Marina Zindel da Cruz ², Rhuan Favero Costa ², Verônica Costa Biselli ², Victoria Bueno de Araújo ², Daniela Mayumi Usuda Prado Rocha ¹

¹ Precision Data, São Paulo-SP, Brazil

² Universidade de São Paulo, São Paulo-SP, Brazil

Background

Mapping national healthcare terminologies to international standard vocabularies is a critical step for enabling interoperability, data reuse, and participation in distributed research networks such as those built on the OMOP Common Data Model (CDM) [1,2]. The OMOP CDM, maintained by the Observational Health Data Sciences and Informatics (OHDSI) community, standardises observational health data using controlled vocabularies — including SNOMED CT as the primary clinical reference terminology — to support large-scale federated analyses [2,3]. In Brazil, the Terminologia Unificada da Saúde Suplementar (TUSS), published by the Agência Nacional de Saúde Suplementar (ANS), encodes healthcare procedures in the supplementary health sector and lacks a systematic mapping to OMOP standard concepts [4]. Achieving this mapping is labour-intensive, requiring expert review to assess semantic equivalence between source codes and SNOMED CT concepts. Artificial intelligence (AI) may accelerate this workflow by proposing candidate mappings, but its performance must be evaluated within human-in-the-loop review processes to ensure accuracy and scientific rigour. This study evaluates the Any-to-Any Terminology Mapping (AATM) Python library to support collaborative mapping of TUSS procedure terms to SNOMED CT.

Methods

TUSS organises surgical and therapeutic procedures by anatomical region into a hierarchical code system regulated by ANS [4]. A subset of 500 TUSS terms was submitted to the automated terminology mapping pipeline of the AATM Python library, an open-source package providing tools for automatic terminology mapping that integrates with the OMOP CDM [5]. The standard AATM pipeline was applied and comprised: (i) automatic translation of source terms using the "gemini-2.5-flash" large language model [5]; (ii) semantic information retrieval using the "embeddinggemma-300M" embedding model [6]; and (iii) final concept selection via the "first-result-selector" option, which returns the top-ranked candidate from the retriever. Each candidate is assigned a semantic similarity score derived from the retrieval step. The reference concept database was drawn from OMOP standard vocabularies and filtered to include SNOMED CT concepts from the Procedure, Observation, and Condition domains.

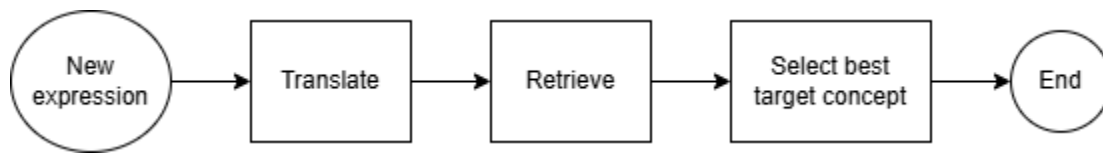


Figure 1. A simple visual representation of the automatic terminology mapping pipeline used from the AATM library.

Human reviews were distributed among 21 reviewers. The consensus was defined by having 2 independent annotations that agreed on the SNOMED CT code selected and on the mapping relation between the TUSS term and the selected SNOMED CT code. Reviewers had access to a pre-selected SNOMED CT code and mapping relation from the AATM pipeline. They were able to accept the AI-suggested concept or substitute an alternative concept. Mapping relations consisted of the following categories, as defined by OMOP annotation guidelines: “equal”, “equivalent”, “wider”, “narrower”, “inexact”, or “unmatched”.

The AATM project page is available at <https://precisiondata.github.io/aatm/>.

Results

The annotation dataset included 500 TUSS procedure terms evaluated for mapping to SNOMED CT. At least one active peer response was available for 499 terms, yielding 980 independent annotations. Overall, 486 terms (97.2%) were mapped to at least one SNOMED CT code, while 13 (2.6%) were classified as unmatched. Most terms were mapped to a single code, with 28 (5.6%) having more than one code in the consolidated annotations.

Peer consensus was reached for 298 terms (59.6%). Another 183 (36.6%) remained awaiting approval after disagreement, 18 (3.6%) received only one annotation, and 1 (0.2%) remained pending. The most frequent consolidated mapping relation was “wider” (228; 45.6%), followed by “equivalent” (140; 28.0%) and “equal” (64; 12.8%).

AI participation through AATM was substantial. In the consolidated results, annotators accepted the AI suggestion for 322 terms (64.4%) and altered it for 177 (35.4%). At the response level, AI suggestions were accepted in 676 annotations (69.0%) and altered in 304 (31.0%).

AI acceptance varies by mapping relation. Alteration rates were lowest for “equal” mappings (8.8%) and higher for “wider” (43.3%) and “unmatched” responses (90.0%), indicating that AI support was most reliable for exact mappings and least reliable when no adequate SNOMED CT target was available. AI alteration was also more frequent among terms awaiting approval than among terms that reached consensus (35.0% vs. 28.9%).

The main source of disagreement was not the selection of different SNOMED CT targets, but the assignment of different semantic relations to the same primary target. Same-target/different-relation disagreements accounted for 102 terms (55.7% of disagreements), while different primary target selection accounted for 57 (31.1%). If same-target/different-relation cases are considered agreement on code selection, consensus on the selected SNOMED CT code increases from 298 terms (59.6%) to 400 (80.0%), and unresolved disagreements decrease from 183 to 81. This also

shifts AI acceptance from 71.1% to 75.0% among consensus terms and from 52.5% to 9.9% among remaining disagreement cases, suggesting that future AATM improvements should focus on mapping relation classification.

Disagreement Categories

Unique TUSS terms reconstructed from active responses

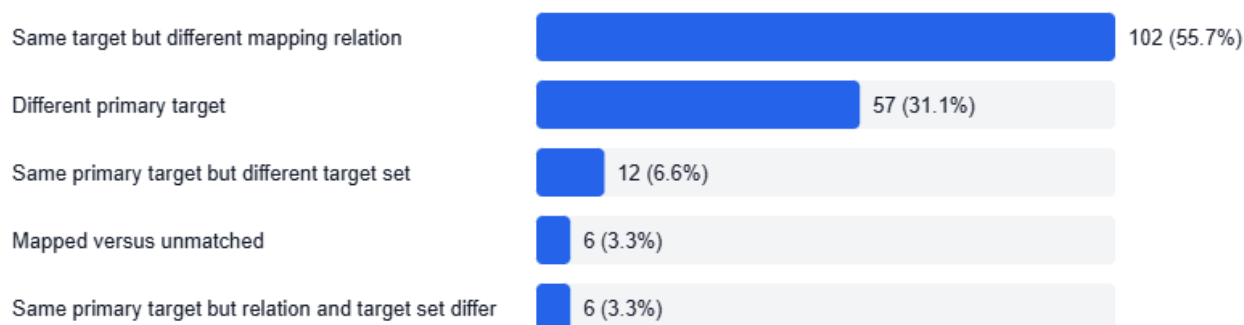


Figure 2. Distribution of disagreement categories among TUSS-to-SNOMED CT mappings. Horizontal bars show the number of TUSS terms in each mutually exclusive disagreement category. Percentages represent the proportion of terms with disagreements (n = 183). Same-target/different-relation cases indicate agreement on the primary SNOMED CT concept but disagreement on the mapping relation.

AI Accept/Change Pattern by Relation

Active responses; bar length is 100% within each relation

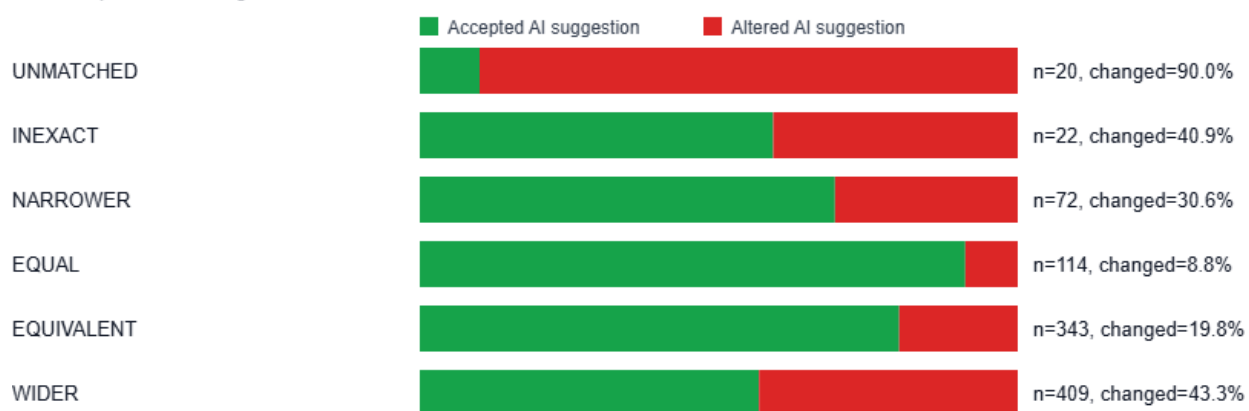


Figure 3. AI suggestion acceptance and modification by mapping relation. Stacked horizontal bars show accepted and altered AI suggestions across individual annotation responses. AI modification was most frequent among “unmatched” responses (90.0%) and least frequent among “equal” mappings (8.8%), indicating that AI suggestions were more often retained for exact mappings and more often modified for absent, broader, or less direct mappings.

Conclusion

AATM supported efficient human-in-the-loop mapping of Brazilian TUISS procedure terms to SNOMED CT, with AI suggestions accepted in most annotation responses and consensus reached for 59.6% of terms. The results indicate that AI can reduce reviewer workload, particularly for exact mappings, while expert validation remains essential for broader, ambiguous, or unmatched cases. By identifying the mapping relations where AI suggestions most often fail, this study also highlights clear priorities for improving AATM in future initiatives. Combining multilingual semantic retrieval with collaborative expert review, AATM offers a scalable and open approach for expanding OMOP-compatible vocabulary mappings across diverse health systems.

References

1. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553–64.
2. Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc*. 2024;31(3):583–90.
3. Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform*. 2006;121:279–90.
4. Agência Nacional de Saúde Suplementar (ANS). Terminologia Unificada da Saúde Suplementar (TUISS) [Internet]. Rio de Janeiro: ANS; 2013 [cited 2026 Jun 5]. Available from: <https://www.ans.gov.br/a-ans/sala-de-noticias-ans/operadoras-e-servicos-de-saude/2010-rol-de-procedimentos-e-terminologia-unificada-da-saude-suplementar>
5. Comanici G, Bieber E, Schaekermann M, Pasupat I, Sachdeva N, Dhillon I, et al. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. *arXiv [Preprint]*. 2025. <https://doi.org/10.48550/ARXIV.2507.06261>
6. Vera HS, Dua S, Zhang B, Salz D, Mullins R, Panyam SR, et al. EmbeddingGemma: Powerful and Lightweight Text Representations. *arXiv [Preprint]*. 2025. <https://doi.org/10.48550/ARXIV.2509.20354>