

THE ECOSYSTEM OF EVIDENCE

Navigating the future
of the ecosystem of evidence

Artificial intelligence and digital transformation
in research and healthcare

11th EBHC International Conference
Taormina, 21st – 24th October 2026

#EBHC2026

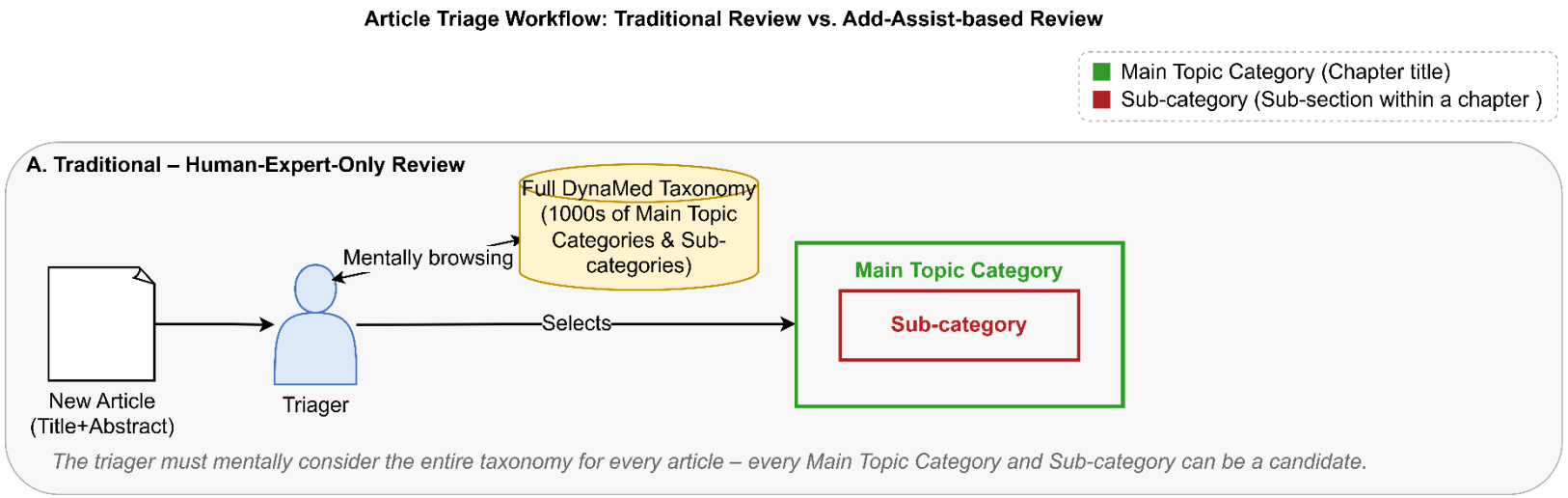
AI-Assisted Topic Suggestion System Used in Triaging Clinical Articles for Updating Evidence-Base

Development and External Evaluation

Ashirbani Saha
McMaster University, Canada
Co-authors: Mehrdad Eshraghi Dehaghani, Rick Parrish, Tamara Navarro-Ruan, Cynthia Lokker, Alfonso Iorio

30th GIMBE 1996-2026 EVIDENCE FOR HEALTH

Background and aims



Health Information Research Unit (HiRU), McMaster University supports a literature Surveillance System for DynaMed

Title: Twice-Yearly Lenacapavir for HIV Prevention in Men and Gender-Diverse Persons.

Abstract: BACKGROUND: Twice-yearly subcutaneous lenacapavir has been shown to be efficacious for prevention of HIV infection in cisgender women.....

Selected Topic Placement: **Preexposure Prophylaxis (PrEP) for HIV>PrEP Efficacy>Lenacapavir**

Aim: Develop and externally evaluate an AI-assisted topic-title suggestion system for triaging clinical articles across a broad range of clinical topics.

Methods

- **Dataset for development (internal validation):**

23 periodic snapshots (Nov 2024 – Apr 2025)

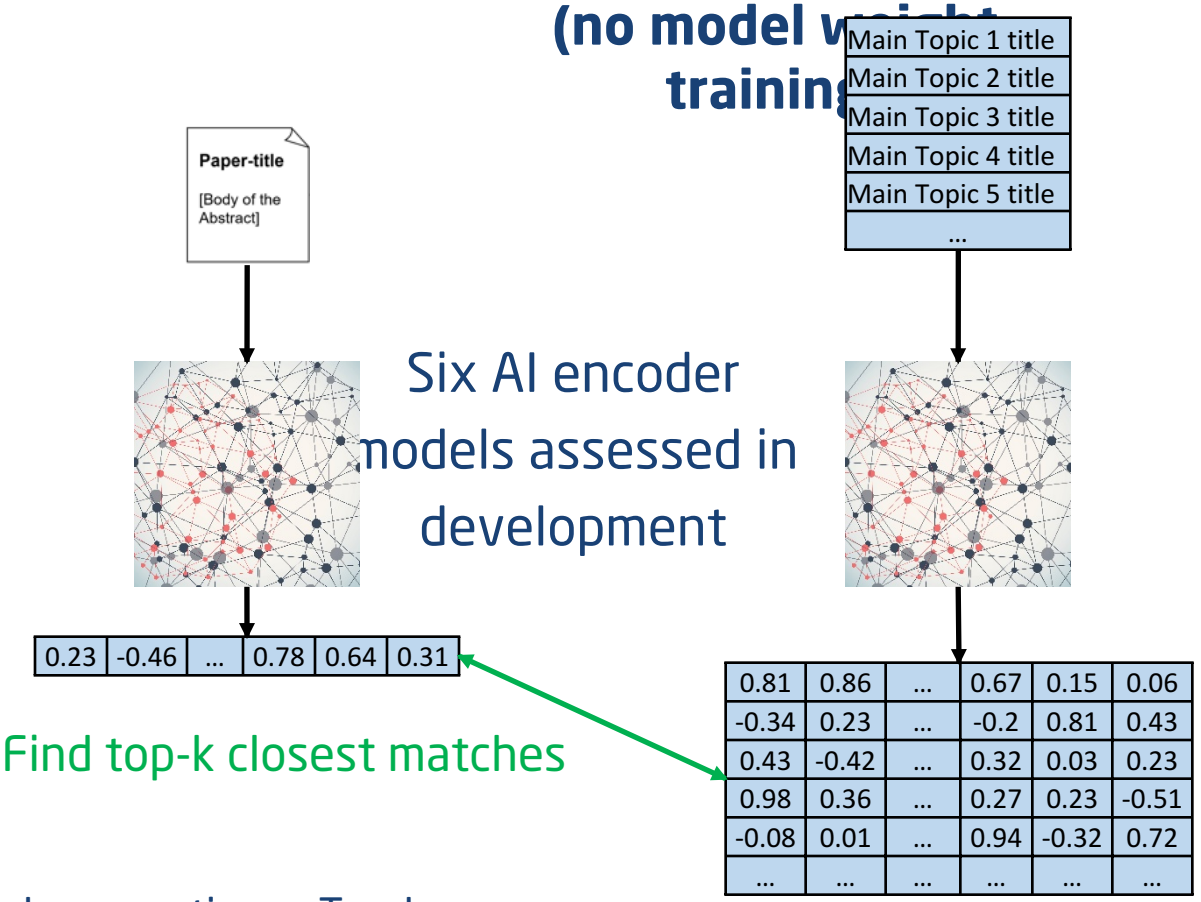
- 1,476 title-abstract pairs (of a particular type) annotated with triager-selected main topic categories.
- 50% held-out per snapshot for internal testing.

- **Dataset for testing (external validation):**

All snapshots (May 2025 - Feb 19, 2026);

- 3,703 title-abstract pairs (all types) annotated with triager-selected main topic categories.

Semantic retrieval system (no model weight training)



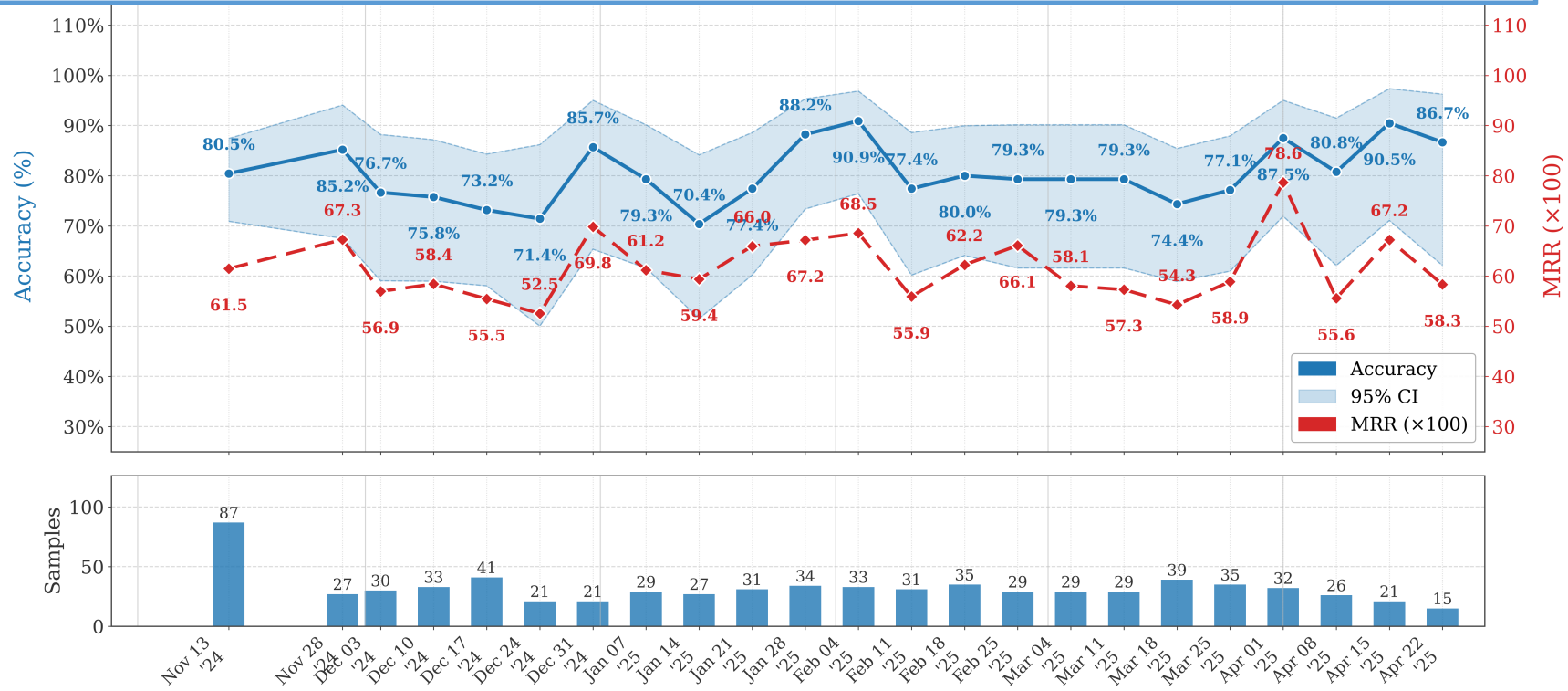
- **Evaluation:**

Correctness: Appearance of correct topic-title among the top k suggestions - Top-k accuracy

Ranking: Closeness of the correct suggestion to rank 1 on an average – Mean Reciprocal Rank (MRR)

Results

Internal Validation on Periodic Snapshots



Best model: NovaSearch/jasper_en_vision_language_v1 at k=5, Median top-5 accuracy 79.39% [IQR 7.49]; MRR 0.59 [IQR 0.09].

External Validation

Model matched topic-titles for 78.8% of the 3,703 title-abstract pairs.

Rank distribution:

Rank 1: 49.18%

Rank 2: 14.39%

Rank 3: 7.13%

Rank 4: 4.67%

Rank 5: 3.43%

Limits and conclusions

- **Conclusions:** The AI-assisted system matched ~80% of triager-selected topic-titles, demonstrating strong promise for scaling and accelerating clinical literature triage.
- **Impact:** Automating topic-title suggestions reduces manual effort and supports HiRU's mission to keep its evidence-base continuously up to date.
- **Limitations:** System does not include the most recent language encoders; ongoing experiments aim to improve accuracy and suggestion ranking.
- **Next steps:** Integrate newer encoder models; expand evaluation across more clinical domains; explore other learning paradigms to adapt to evolving topic-title lists.