

From Abstracts to Full Texts: The Impact of Context Positioning in LLM-Based Screening Automation

E. Sandner^{1,2}, M. Zeba², I. Jakovljevic¹, A. Simniceanu³, L. Fontana³, A. Henriques¹, A. Wagner¹, C. Gütl²

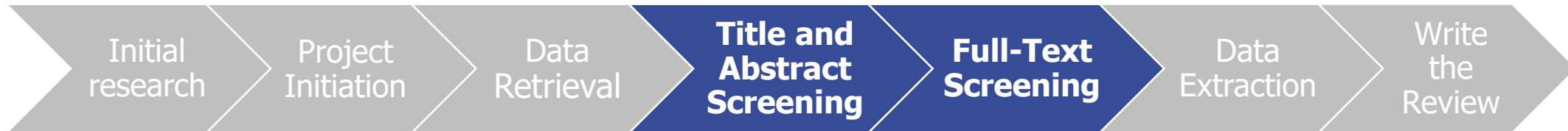
¹CERN, Geneva, Switzerland

²Graz University of Technology, Graz, Austria

³WHO, Geneva, Switzerland



Systematic Literature Review



Manual Literature Screening:

- Double Blinded Screening
- Most time-consuming aspect of SR process[2]
- Abandoned SRs most likely stopped in this phase[3]
- Yield Rate: ~ 1%
- ~95% excluded in TiAb screening

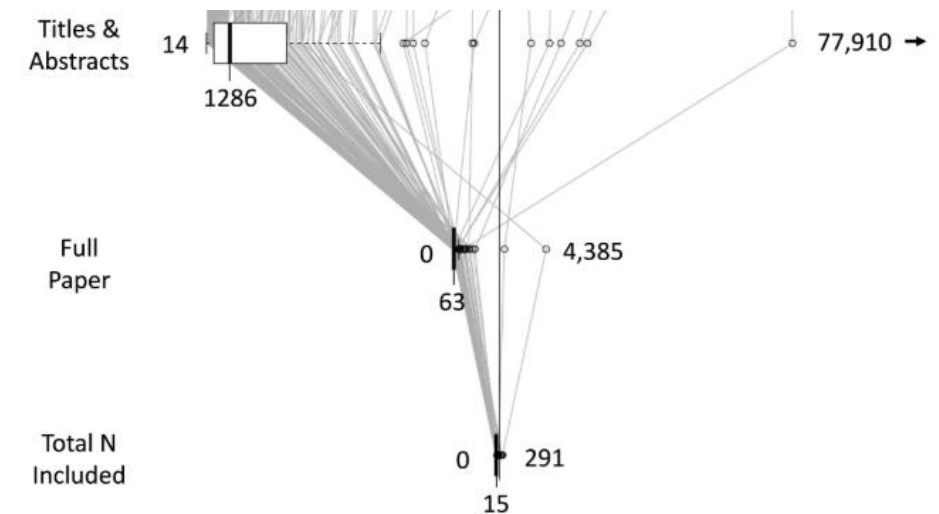


Figure from Borah et al. (2017)[1]

[1] R. Borah, A. W. Brown, P. L. Capers, and K. A. Kaiser, "Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry," *BMJ Open*, vol. 7, no. 2, p. e012545, 2017, British Medical Journal Publishing Group.
 [2] J. C. Carver, E. Hassler, E. Hernandez, and N. A. Kraft, "Identifying barriers to the systematic literature review process," in *2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2013, pp. 203-212.
 [3] M. R. Demetres, D. N. Wright, A. Hickner, C. Jedlicka, and D. Delgado, "A decade of systematic reviews: an assessment of Weill Cornell Medicine's systematic review service," *Journal of the Medical Library Association: JMLA*, vol. 111, no. 3, p. 728, 2023, Medical Library Association.

Current 5-Tier TiAb Screening [4]

Instructions: ... Evaluate the paper based on the given criteria and assign it to one of the following groups:

1. Highly Relevant
2. Probably Relevant
3. Undecidable
4. Probably Irrelevant
5. Not Relevant

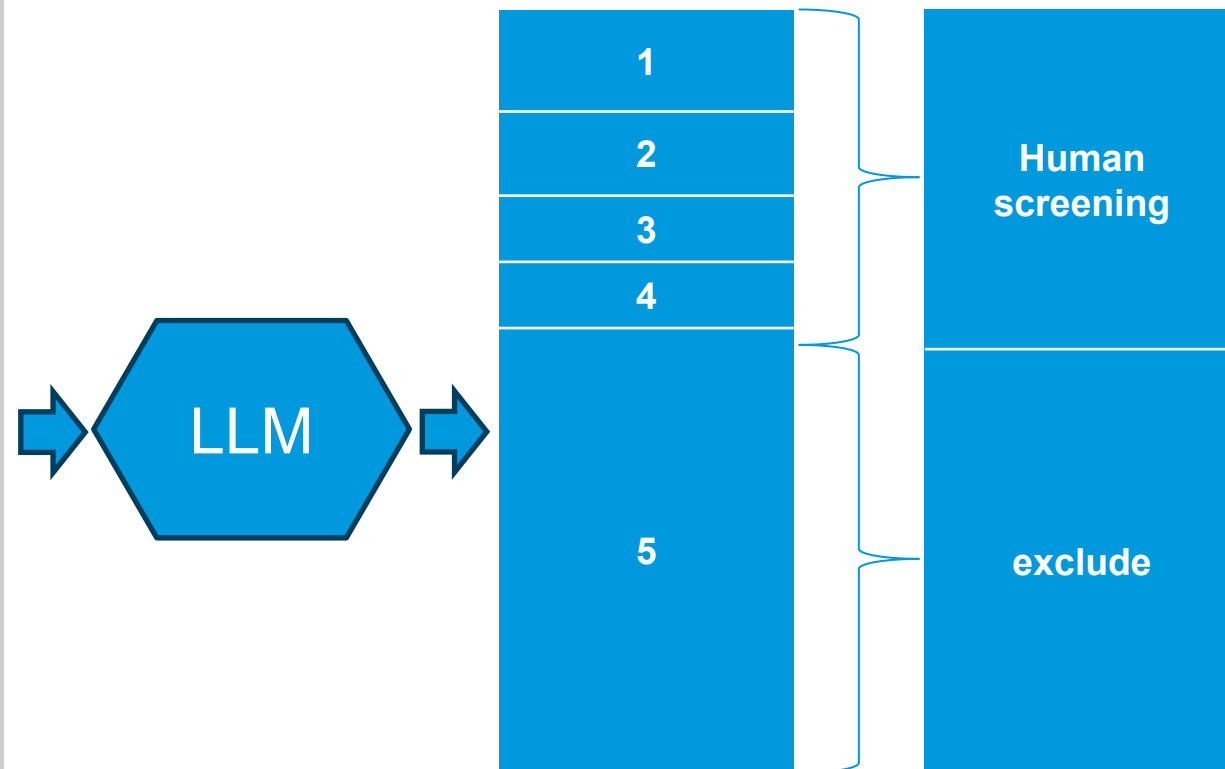
Only type the number of the selected group.

Title: {title}

Abstract: {abstract}

Inclusion criteria: {inclusion criteria}

Exclusion criteria: {exclusion criteria}



[4] Sandner, E., Hu, B., Simiceanu, A., Fontana, L., Jakovljevic, I., Henriques, A., ... & Gütl, C. (2024, November). **Screening Automation for Systematic Reviews: A 5-Tier Prompting Approach Meeting Cochrane's Sensitivity Requirement.** In 2024 2nd International Conference on Foundation and Large Language Models (FLLM) (pp. 150-159). IEEE

Extending 5-Tier for Full-Text Screening

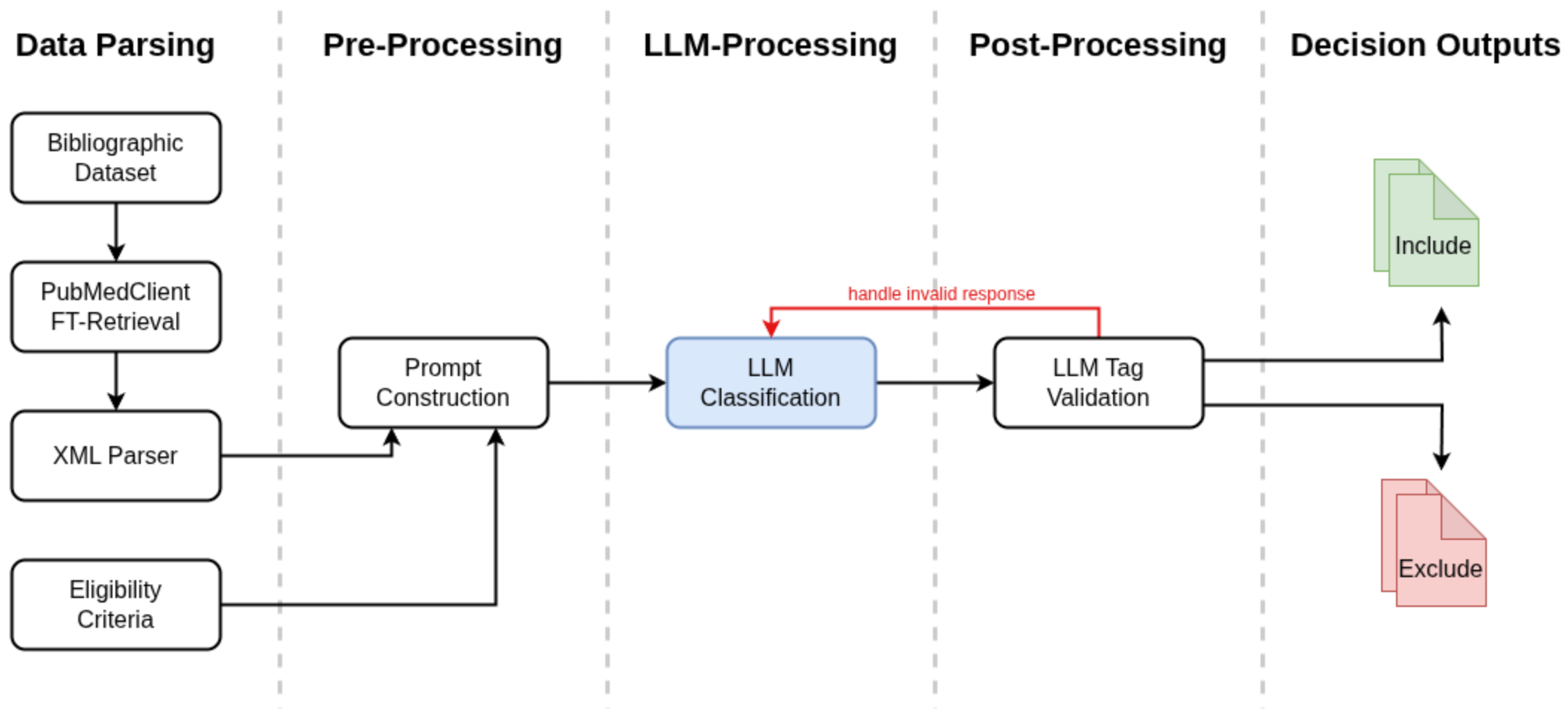
- Title and Abstract (TiAb) might not offer enough information for screening decision
- Quality of TiAb highly depends on author
- Full-Text (FT) has all information from paper, thus assumption that screening decisions get better
- Attention to 'lost-in-the-middle' phenomenon for long-context
- Evaluation on locally hosted Llama 3.1 8B
- Main Research Question: „Does providing the FT as additional context during screening yield higher sensitivity and greater workload reduction compared to LLM-powered TiAb screening?“

Prompt Structure

- Adjusted System Prompt: Exchanged terms ‘title and abstract’ with ‘full-text’
- **Title and Abstract Approach (TiAb)** (Baseline)
Existing 5-Tier TiAb screening
- **Full-Text 1 Approach (FT1)**
Simple extension of 5-Tier TiAb screening
- **Full-Text 2 Approach (FT2)**
Addresses ‘lost-in-the-middle’ phenomenon by repetition of eligibility criteria

TiAb	FT1	FT2
System Prompt	Adjusted System Prompt	Adjusted System Prompt
Title and Abstract	Full-Text	Eligibility Criteria
Eligibility Criteria	Eligibility Criteria	Full-Text
		Eligibility Criteria

Conceptual Architecture



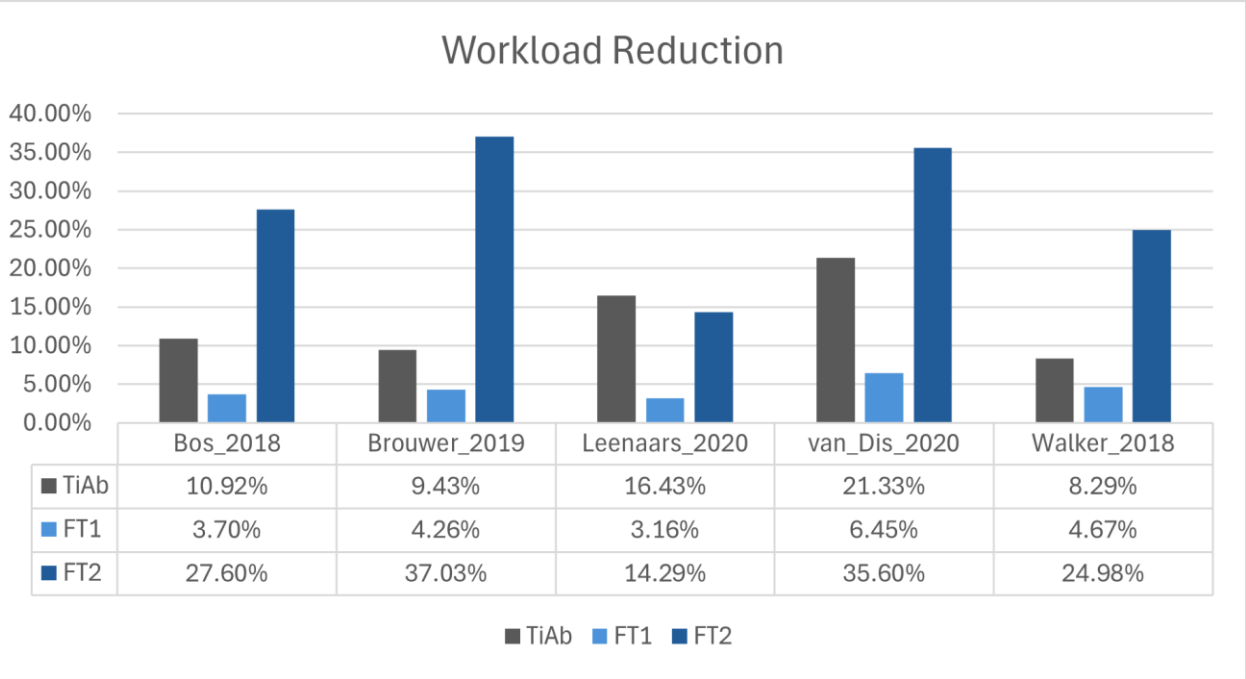
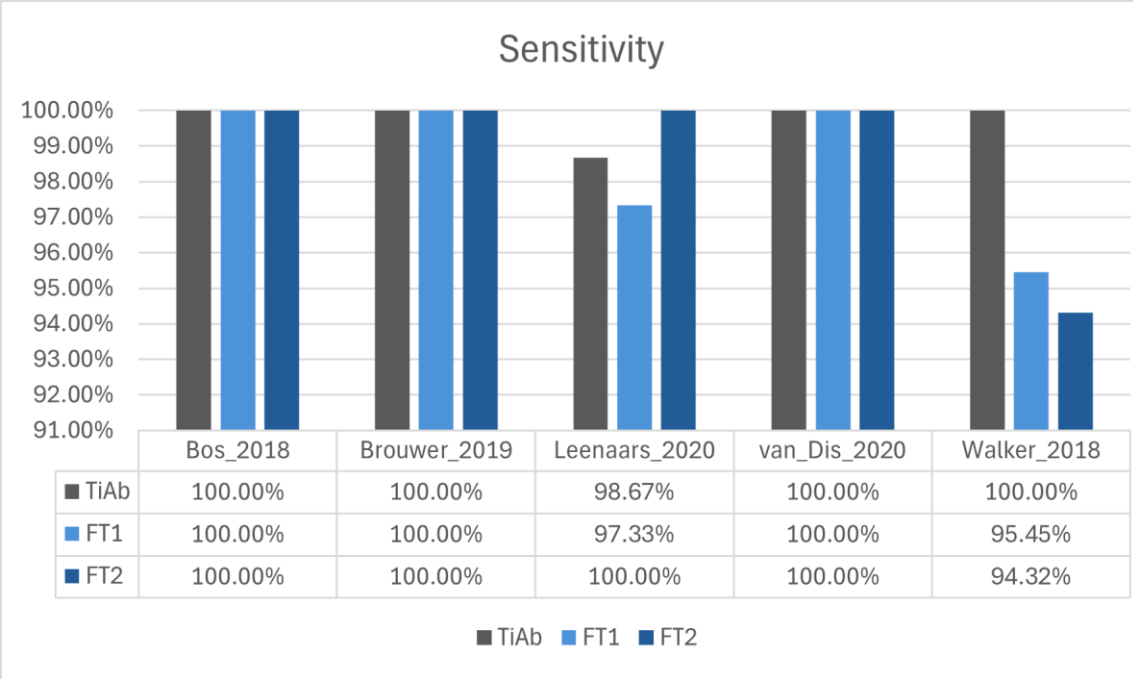
Datasets

- Title
- Abstract
- Full-Text (retrieved)
- Ground-truth decision (FT consensus)
- Eligibility Criteria
- Data source: Synergy Dataset Collection [5]

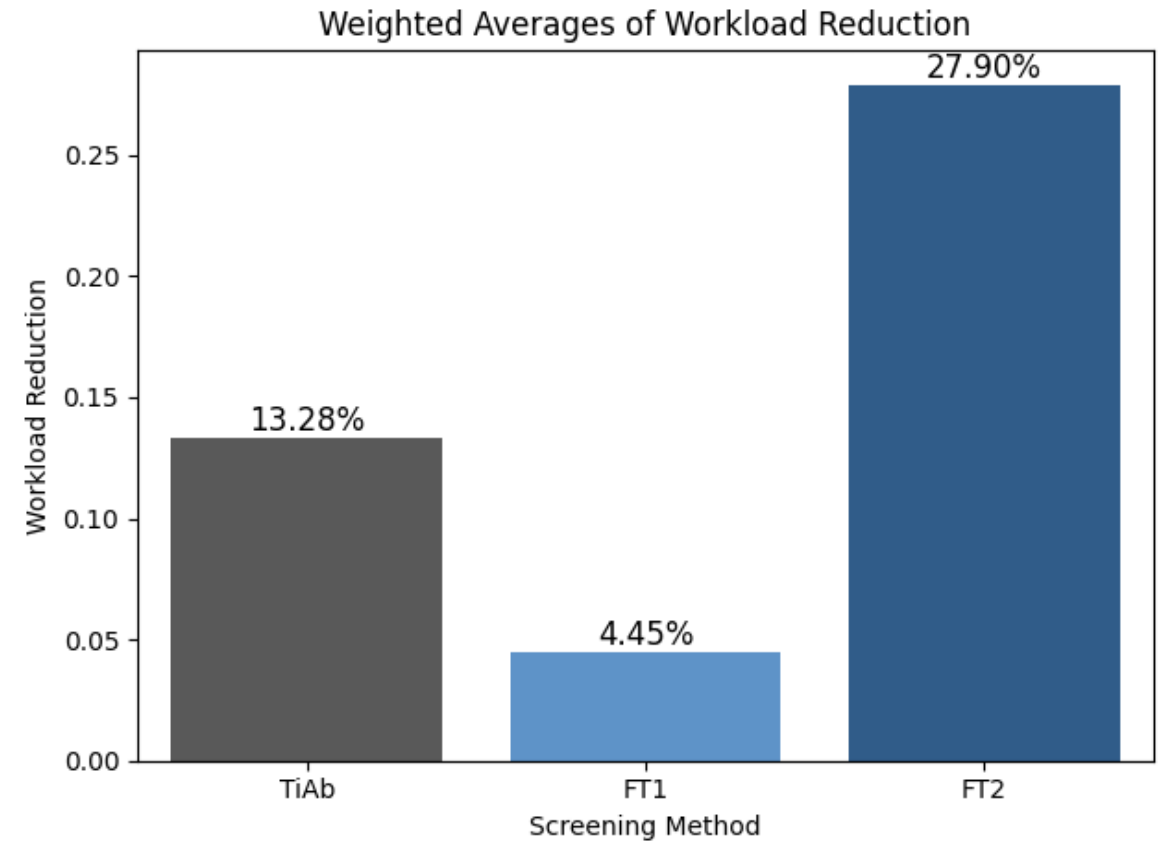
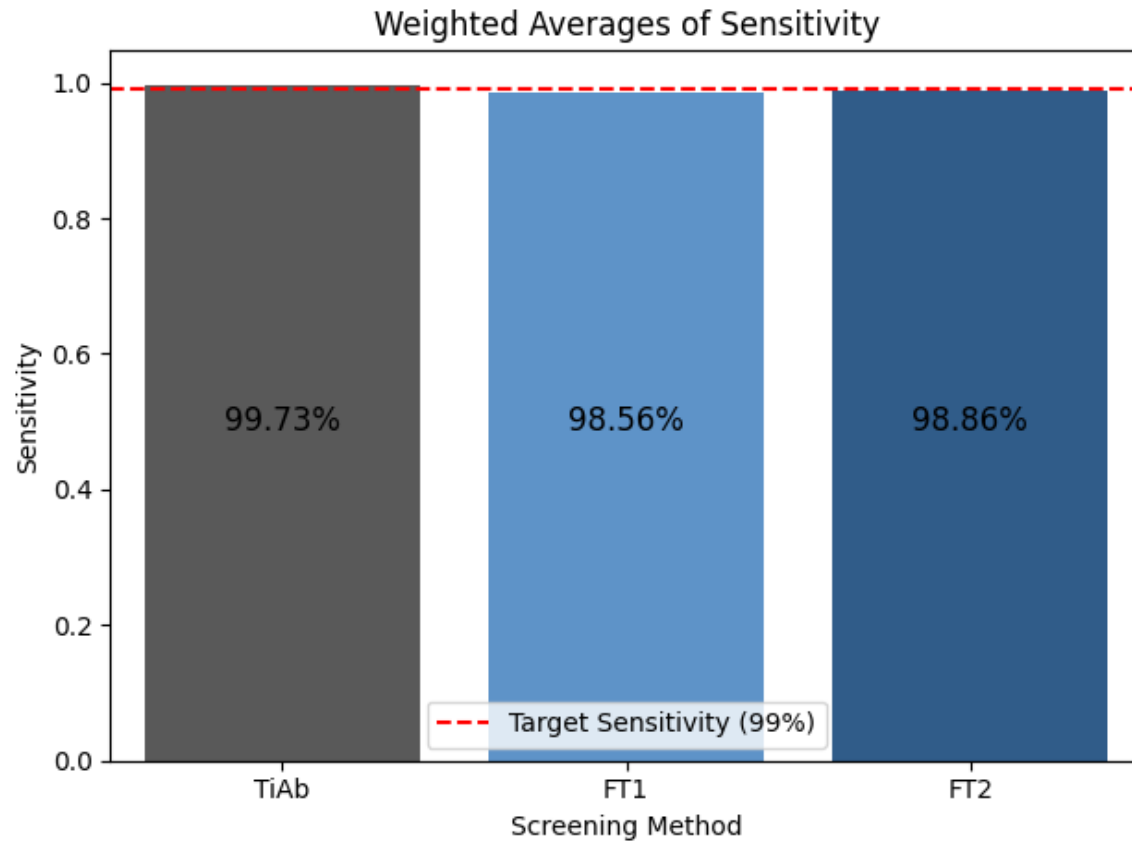
Dataset	Domain	No. of Records	No. of Includes
Bos_2018	Medicine	1.163	5
Brouwer_2019	Psychology, Medicine	6.482	11
Leenaars_2020	Medicine	791	75
van_Dis_2020	Psychology, Medicine	1.753	15
Walker_2018	Biology, Medicine	3.234	88
Total		13.423	194

[5] De Bruin, J., Ma, Y., Ferdinands, G., Teijema, J., & Van de Schoot, R. (2023). SYNERGY – Open machine learning dataset on study selection in systematic reviews (Version 1) [Data set]. DataverseNL. <https://doi.org/10.34894/HE6NAQ>

Results per Dataset



Weighted Average per Approach



Conclusion and Limitations

- Additional context in the form of full-text does not increase sensitivity
- Significant reduction of workload (FT2), with only missing > 99% sensitivity in 1 out of 5 datasets
- More extensive evaluation dataset needed
- Limited amount of open-access papers for FT retrieval
- More advanced FT screening approaches needed by only including selected chapters or guiding LLM during screening with enhanced focus on selected chapters

From Abstracts to Full Texts: The Impact of Context Positioning in LLM-Based Screening Automation

Contributors:

- **Elias Sandner (CERN, TU Graz)**
- **Marko Zeba (TU Graz)**
- Igor Jakovljevic (CERN)
- Alice Simiceanu (WHO)
- Luca Fontana (WHO)
- Andre Henriques (CERN)
- Andreas Wagner (CERN)
- Christian Gütl (TU Graz)

Supplementary Material:

<https://doi.org/10.5281/zenodo.16419873>

Contact:

E-Mail: elias.sandner@cern.ch