AIDME: A Scalable, Interpretable Framework for AI-Aided Scoping Reviews

Michael Soprano, Sandip Modha, Kevin Roitero, Eddy Maddalena, Marco Viviani, Gabriella Pasi, Stefano Mizzaro

Quinto Convegno Nazionale CINI sull'Intelligenza Artificiale (Ital-IA 2025)

June 23, 2025 (Trieste, Italy)

UNIVERSITÀ DEGLI STUDI DI UDINE

hic sunt futura

Introduction and Motivation

- We had already completed **two** extensive literature reviews
  - The second strictly **adhered** to **PRISMA** guidelines
- In both cases, the process was **extremely time-consuming**
  - Screening **hundreds of papers**, weeks of manual curation
- We designed **AIDME** to add an **AI-assisted pre-screening step** before a full systematic review

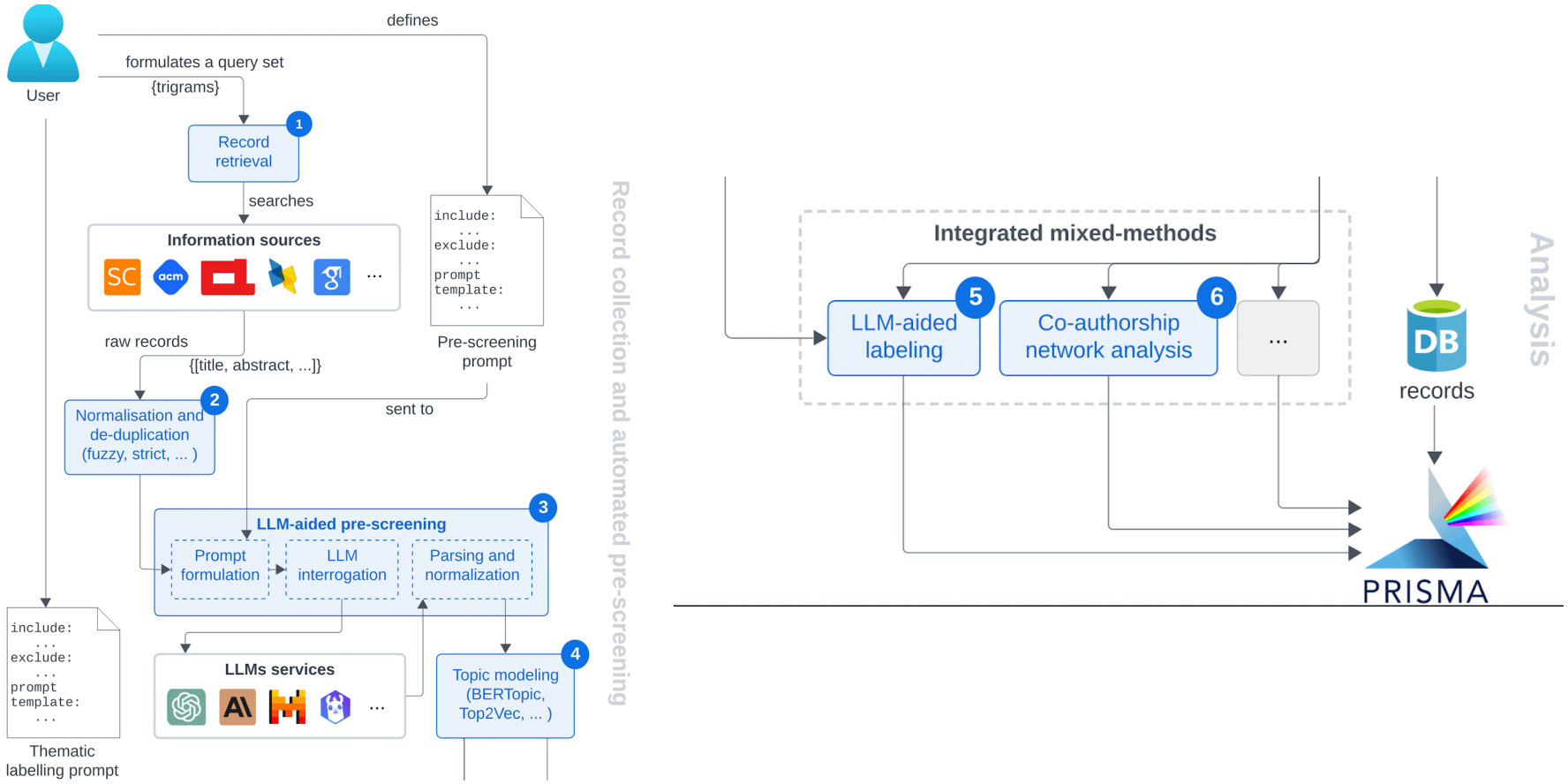
Why Automation is Needed

- **Rapid growth**: thousands of new studies yearly
  - "truthfulness assessment in misinformation"
    - → **25,000+** results on Google Scholar
- **High recall** needed to **avoid missing key studies**
- Manual screening is **tedious and error-prone**

Our Goals

1. **Reduce manual effort** while keeping **human oversight**
2. Ensure **transparency**: all decisions are **visible and auditable**
3. Provide a **domain-agnostic** solution for any rapidly expanding research field
    - Case study: **evaluation of truthfulness assessment in fact-checking**

The AIDME Framework



Record retrieval → De-dup → LLM pre-screen → Topic modeling → LLM labeling → Network analysis

Steps 1–3: Retrieval and Pre-Screen

1. **Retrieve records** using structured trigram queries
   - Scopus, ACM DL, ACL, DBLP
2. **Normalize and de-duplicate**: exact and fuzzy matching
3. **LLM-assisted pre-screening**
   - `GPT-4o-mini` with a **recall-oriented** prompt

Steps 4–6: Mapping and Analysis

1. **Topic modeling** to uncover **thematic clusters**
   - `BERTopic` + `MPNet` provide the best performance
2. **LLM-aided thematic labeling**: classify each record as `Human-based`, `Automated`, or `Combination`
3. **Co-authorship network**: visualize **collaboration patterns**

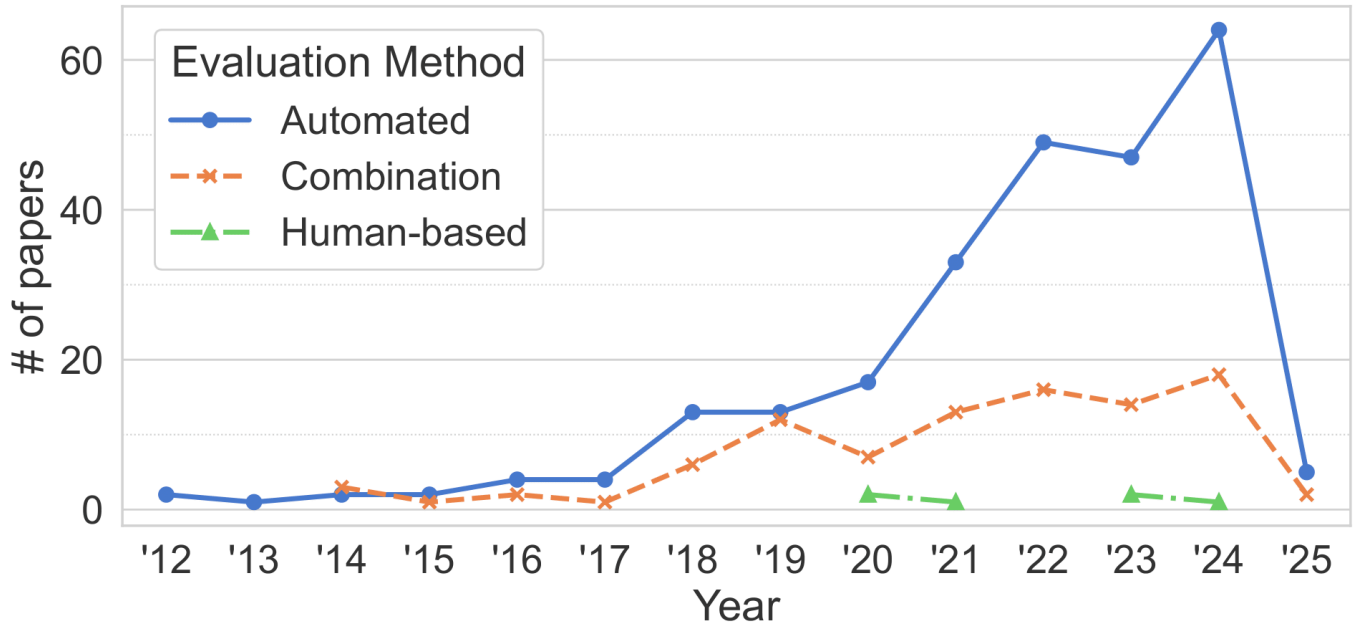
RQ1 – Efficiency and Scalability

- Definition: “record” = **title + abstract** (PRISMA terminology)
- **AI-assisted pre-screening** reduces workload by **97%**
  - Estimated reviewer time saved: **110–220 hours**

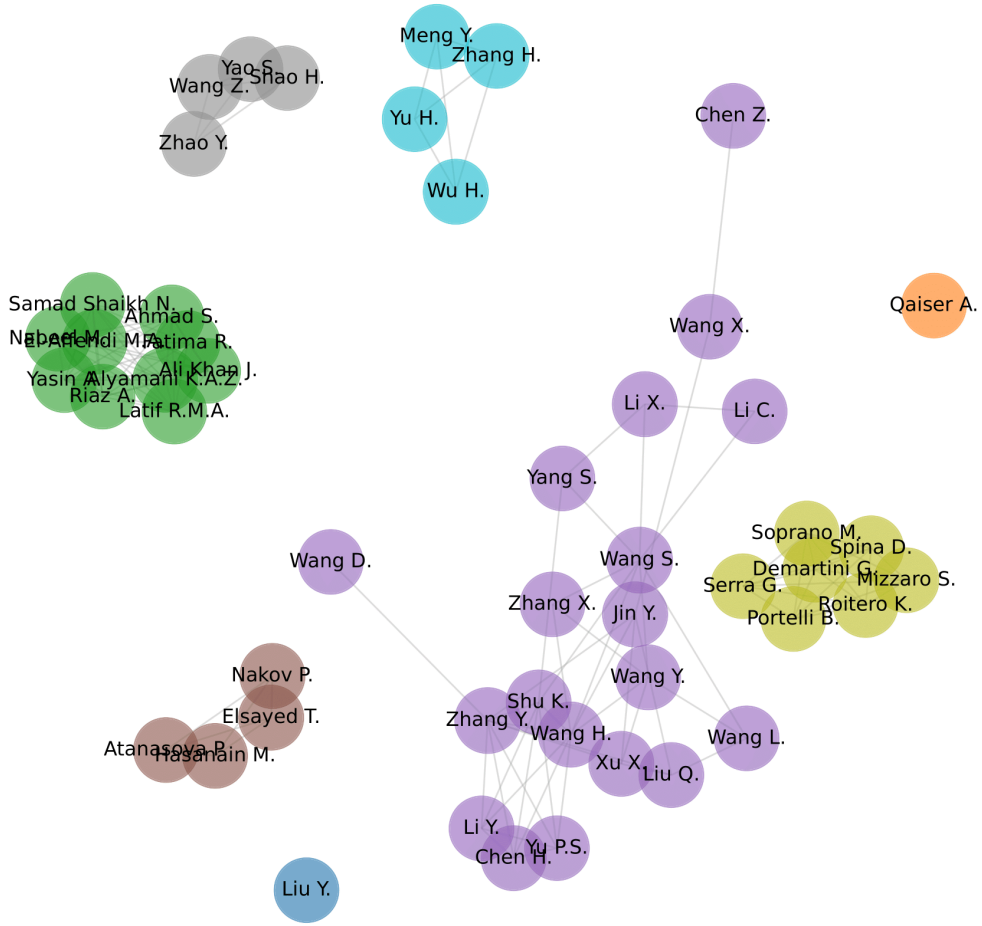
Stage	Records left	% of initial
Raw retrieval	~14,000	100%
After normalization and de-duplication	12,259	88%
After LLM pre-screening	**497**	**3%**

RQ2 – Case Study

- **Automated methods** dominate (84%), primarily ML/NLP
- **Combination methods** (14%), integrating human judgment and automated models, are growing
- **Human-only studies** (2%) are increasingly rare



Collaboration Landscape



- Major **China/US cluster**
- **Qatar/Bulgaria hub** bridging communities
- **Italy–Australia collaboration**
  - Udine + Queensland

Louvain-detected communities in the co-authorship network of the 50 most collaborative authors

Key Takeaways

- **AIDME** integrates **AI-based pre-screening** into review workflows
- Aligns with **PRISMA 2020** guidelines for **transparency** and **reproducibility**
- Delivers **transparent**, **auditable** outputs at every stage
- Reduces **manual effort**, enabling focus on **synthesis** and **insights**

Future Directions

- Extend to **full-text analysis** and **living reviews**
- Integrate **citation graph analytics** (e.g., **temporal PageRank**)
- Customize **prompts** for diverse research domains
- **Evaluate false negatives** and perform the subsequent full systematic review

Acknowledgments

- Partially supported by:
  - **PRIN 2022 Project** — *MoT: The Measure of Truth*
    - Evaluation-centered Machine-Human Hybrid Framework for Assessing Information Truthfulness

Thank you!

- **Repository:** https://osf.io/8t27c/ (or QR code)
  - Results, prompts, and additional figures
- **Contact:** michael.soprano@uniud.it
  - Feedback and collaborations welcome!