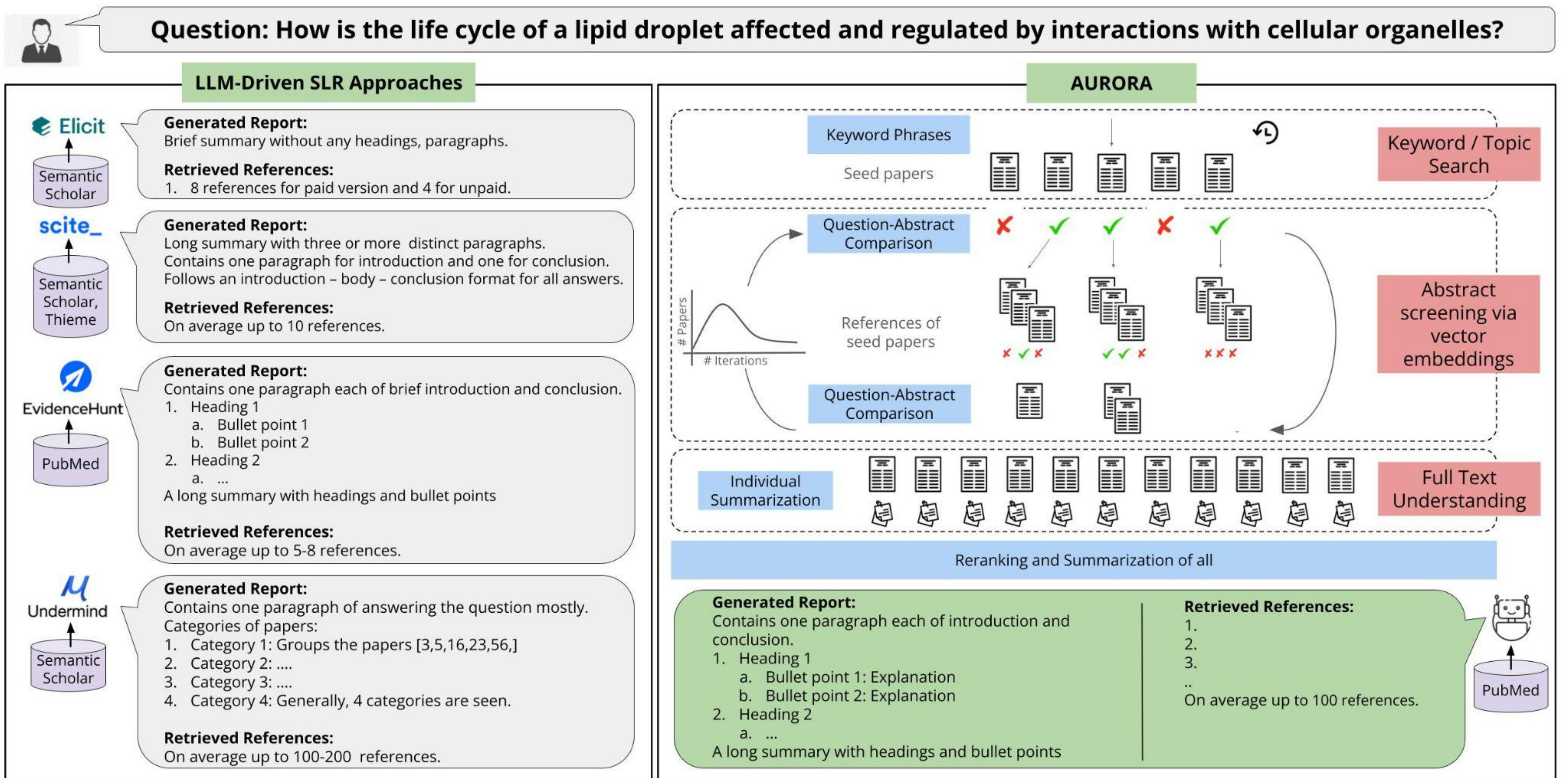
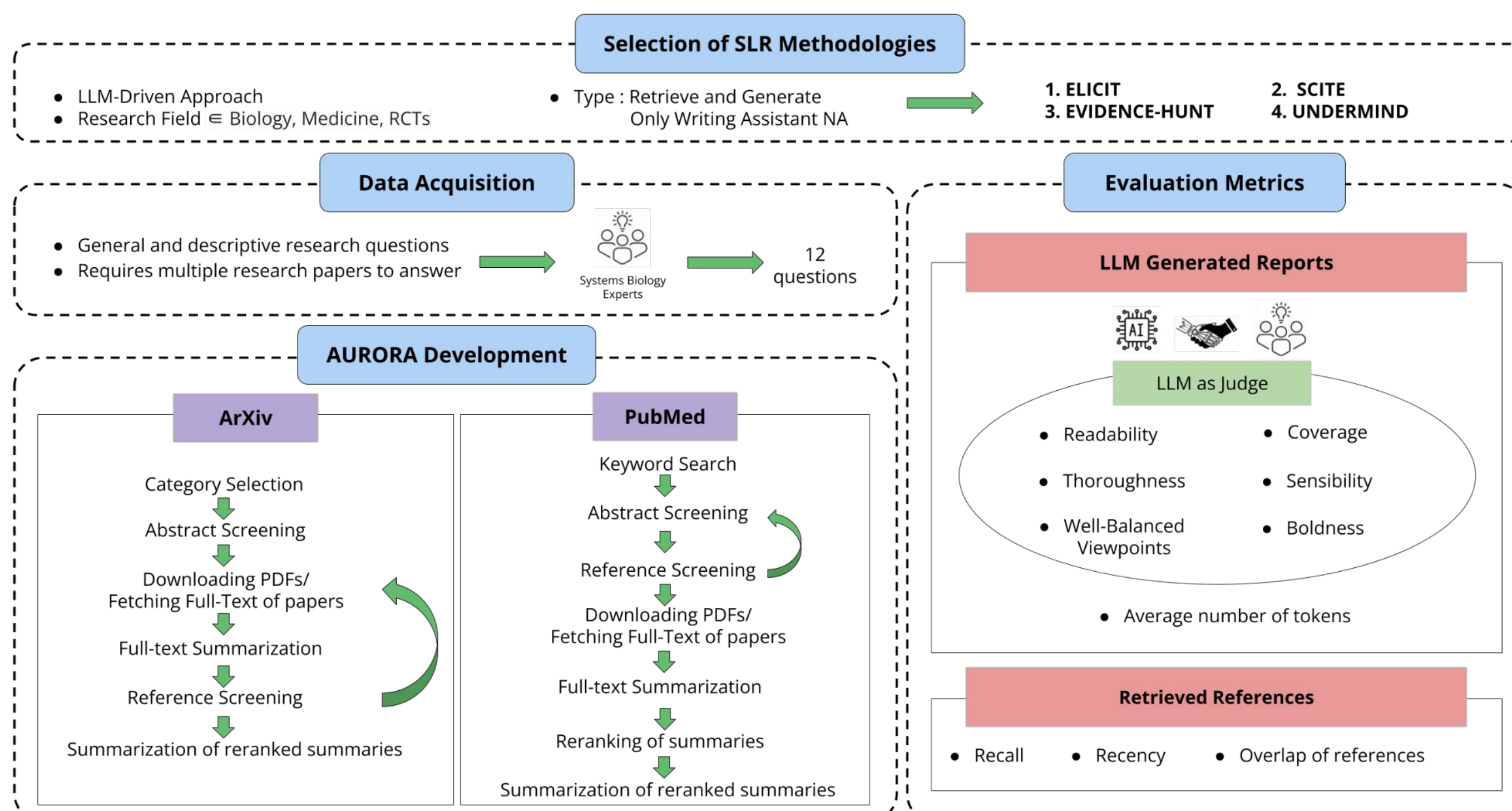


Our study introduces AURORA, a novel approach designed to overcome limitations in LLM-driven literature reviews by balancing reference retrieval and report quality for complex, cross-domain field of systems biology.



AURORA: Automated Understanding and Review Of Research Articles

End-to-End Framework For Evaluation

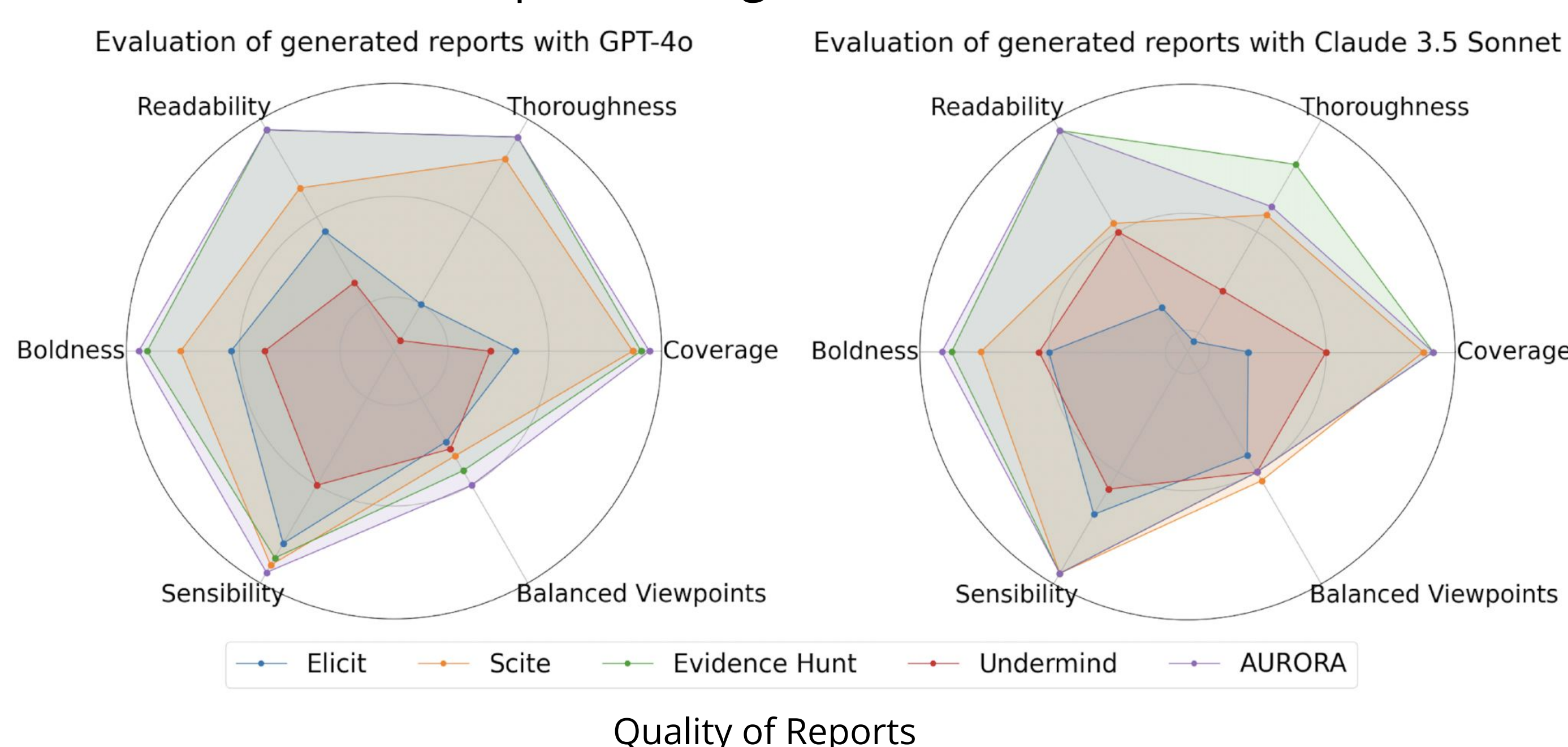


Framework for comparing Systematic Literature Reviews

Evaluation and Observations

Generated Report

Evidence Hunt and AURORA scores a tie, followed by Scite. Elicit and Undermind are underperforming.



Generated References

Recency ∈ [2022, 2023, 2024]

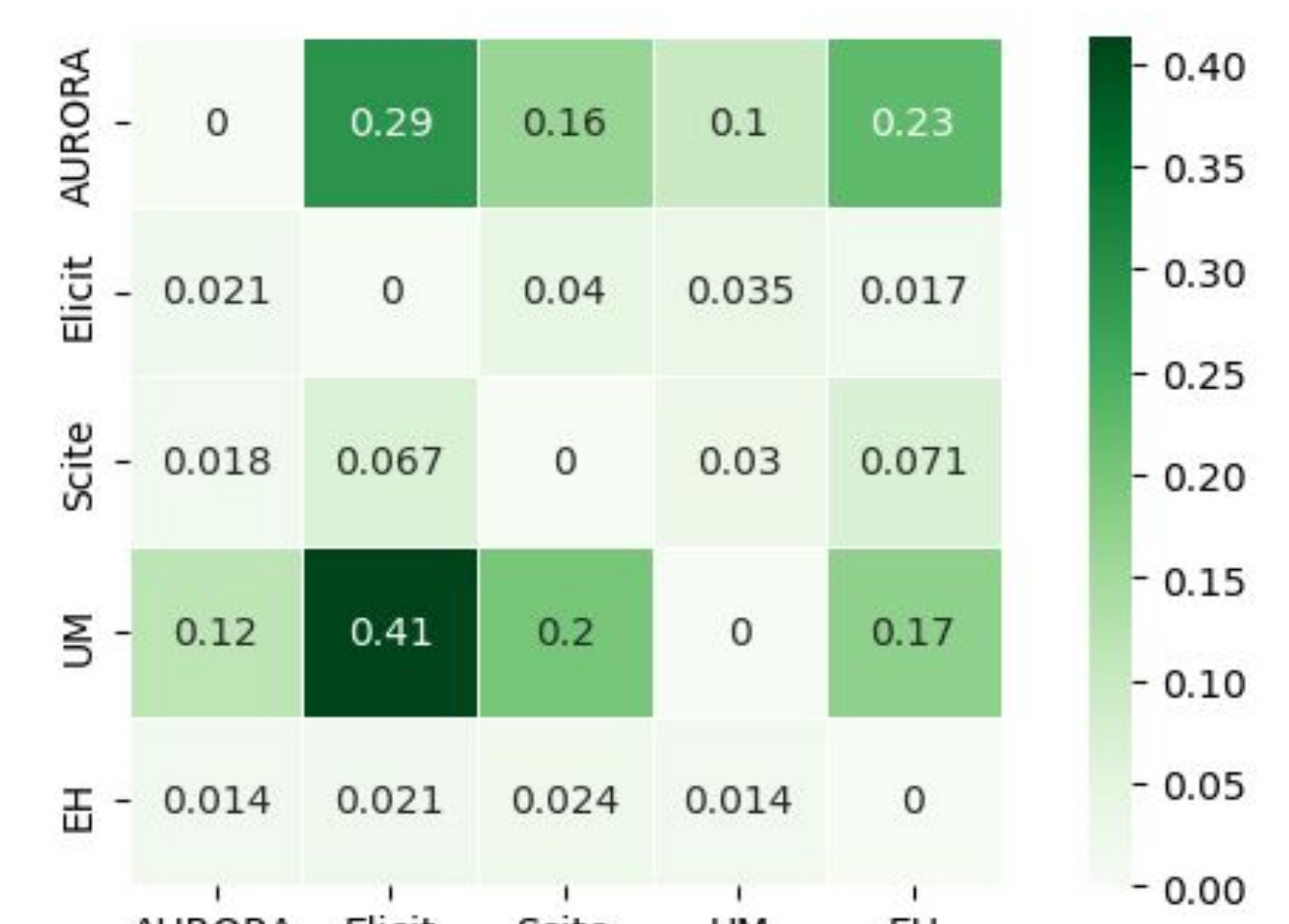
For both metrics, AURORA performs well to others after Undermind. Undermind uses Semantic Scholar including PubMed as a subset.

$$\text{Overlap of A} = \frac{\# \text{References } A \cap B}{\text{References of B}}$$

Undermind, followed by AURORA shows high overlap of references. Interestingly, other approaches like Elicit, Scite, Evidence Hunt shows significantly low overlap, indicating that they fail to find the popular papers for a research question.

SLR Approach	Articles (million)	Recall (%)	Recency (avg. \#)
Elicit	200	4.46	0.75
Scite	200	6.64	3.00
EvidenceHunt	37	3.88	2.17
Undermind	200	47.13	17.41
AURORA	37	47.98	7.25

Quality of References



Observations

- Undermind – Great Reference Finder!
- Evidence Hunt – Good summaries!
- AURORA – Good summaries, Good reference Finder!

Dependency over Repository Metadata

Repositories like ArXiv and Pubmed differ in availability of references in their respective metadata. ArXiv does not have references while Pubmed have references in its metadata. These differences doesn't allow for the unification of the repositories

