

# Semantics for Science: Toward a Massive Scale Evidence-Centric Scientific Knowledge Graph

Mayank Kejriwal<sup>1</sup>

<sup>1</sup>University of Southern California, Los Angeles, CA, USA

## Abstract

Scientific knowledge is still communicated primarily through papers, while the actual objects that science accumulates are more granular: claims, methods, datasets, experimental traces, replications, contradictions, and theories. This mismatch has become more severe as scientific literature grows rapidly and large language models (LLMs) make the production and consumption of scientific text easier but not necessarily more reliable. In this blue-sky paper, we argue that semantic technologies should treat scientific assertions as first-class objects and support the creation of evidence-centric scientific knowledge graphs (KGs) directly from literature. We call this vision *semantics for science*. We argue that building such KGs is not simply a larger information extraction problem. It requires explicit representations of complex evidence traces, including provenance, uncertainty, scope conditions, temporal evolution, evidential support, and challenge relations. We further argue that the resulting infrastructure may require a reasoning engine for science that goes beyond conventional description logics entailment or graph traversal, combining symbolic semantics with defeasible, provenance-aware, and evidence-weighted reasoning. We outline a research agenda spanning representation languages, extraction pipelines, identity and alignment, contradiction handling, temporal revision, and benchmark design. Our claim is deliberately ambitious: if the Semantic Web community wants to shape how scientific knowledge is represented in the age of language models and AI agents, it should help move science from paper-centric publication systems toward computable, evidence-aware knowledge substrates.

## Keywords

scientific knowledge graphs, knowledge graphs and language models, scientific claim extraction, provenance, uncertainty-aware reasoning, evidence-centric knowledge bases

## 1. Introduction

Science does not ultimately advance by publishing well-formatted documents, but by changing what humanity is justified in believing strongly enough to build on [8]. In current practice, however, the dominant artifact of modern science remains the paper: a *narrative object* that bundles claims, evidence, methods, interpretation, caveats, and rhetoric into a unit optimized for communication and credit assignment. This was workable when the scientific literature was smaller and slower, but it is becoming increasingly obvious to all that the model is reaching its breaking point.

The scale of the problem is well documented. Research synthesis scholars describe the present as an age of abundant scholarly information in which the value of the literature depends critically on whether researchers can actually find and synthesize what matters [1]. The scientific publishing system itself is under strain from publication growth, review burden, and increasingly noisy quality signals [2]. In biomedicine alone, PubMed contains tens of millions of citations and continues to grow rapidly [3]. At the same time, LLMs are lowering the cost of interacting with this literature and, increasingly, of generating paper-like scientific prose. This makes the representational weakness of the current system more obvious: *text* is abundant, whereas *computable scientific knowledge* is not.

This paper proposes a blue-sky research agenda at the intersection of semantics, knowledge graphs (KGs), and scientific infrastructure. The central question can be stated in a deceptively simple manner: *can we automatically construct high-fidelity and trustworthy scientific KGs from the scientific literature?* Our answer is in the affirmative, but conditioned on the caveat that our target is *not* a conventional KG populated with decontextualized triples, which represents the dominant practice even today. Scientific

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

\*Corresponding author.

 0000-0001-5988-8305 (M. Kejriwal)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

knowledge is not just a set of facts. It is an evolving, evidence-bearing web of assertions whose status depends on factors like provenance, method, scope conditions, uncertainty, and time. Representing such knowledge faithfully is a blue-sky semantic challenge.

This motivates two theses. First, semantic technologies should help create *evidence-centric scientific KGs* in which the primary units extend beyond entities and relations to assertions backed by explicit evidence and contextual qualifiers. Second, such resources may require a *reasoning engine for science*: a reasoning layer able to operate over support, uncertainty, and revision, rather than only over static graph topology or ontological subsumption. Our proposal is deliberately provocative in the spirit of the SEMANTiCS community. We do not present a finished system. Instead, we argue that a major long-term opportunity for the semantics community is to help redesign the representational substrate of science itself.

## 2. Papers Are the Wrong Unit of Computation

Researchers often ask questions such as: What do we really know about a pathway? Which findings about a benchmark are robust across evaluation settings? Where are the strongest contradictions in a literature? What evidence would most efficiently change current belief about a mechanism? Such questions are not naturally answered by retrieving a ranked list of papers. Instead, we need to operate and reason over claim-level objects, their evidential relationships, and their contexts. The modern paper is poorly suited to this task, even though, at small scale, it serves the important purpose of communicating the necessary knowledge to the human for said reasoning. A single paper packages many distinct scientific contributions into a single narrative object, containing observational (and both explicit and implicit causal) claims, null results, measurement details, speculative interpretations, and references to prior work. Humans can often unpack this structure, but machines struggle. As a result, each new reader, reviewer, graduate student, or LLM pipeline is forced to reconstruct a local model of the field from prose.

*Metascience* has repeatedly highlighted the consequences. Improving the reliability and efficiency of science requires work not only on producing studies but also on dissemination, reproducibility, and incentives [5]. Disagreement in the scientific literature is widespread enough to have become an object of explicit measurement [6]. Even systematic reviews, often treated as the gold standard for synthesis, show non-trivial variability under replication because of incomplete reporting and reviewer judgment [7]. In short, the bottleneck is not just having too little information but also the mismatch between the artifact science produces and the knowledge science seeks to accumulate.

## 3. A Shift toward Evidence-Centric Scientific KGs

We propose a shift from paper-centric to evidence-centric scientific knowledge graphs (KGs). The key idea is that the primary object in a scientific KG should be a *scientific assertion* linked to explicit evidence, rather than a bare (subject–predicate–object) triple detached from its scientific context.

We use the name *ClaimRxiv* for this vision. Its key idea is not merely to host papers with richer metadata, but make scientific assertions first-class public objects that can be submitted, revised, and recombined across documents. In such a system, papers remain important as narrative containers for methods, interpretation, and credit, but they are no longer the only computationally meaningful unit. Instead, ClaimRxiv would comprise an assertion-centric representation in which each important scientific claim is linked to its supporting evidence, its scope conditions. It would also contain critical details on the claims’ provenance and their current epistemic status. Table 1 summarizes the minimal components that such an assertion-centric representation should capture.

There is already encouraging precedent for fragments of this vision. The Gene Ontology (GO) captures traceable annotation statements, and GO-CAM extends such annotations into structured models of biological function [9, 10]. Nanopublications and related provenance-focused work offer compact claim-level publication units [11, 12]. The Evidence Graph Ontology models scientific findings as defeasible

**Table 1**

Core components of an assertion-centric scientific representation.

| Component                       | Role in the representation                                               | Example                                                                       |
|---------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Assertion content               | What exactly is being claimed.                                           | “Protein A binds Protein B.”                                                  |
| Scope conditions                | The benchmark, instrument, or conditions under which the claim holds.    | “Observed in human T cells under assay S at pH 7.4.”                          |
| Evidence links                  | The textual spans, tables, figures, or supporting external resources.    | “Supported by Fig. 2, Table 1, and dataset D123.”                             |
| Provenance                      | Who stated the claim, where, when, and based on which process or source. | “Reported by Smith et al. in Nature 2024”                                     |
| Uncertainty and confidence      | Whether the claim is tentative, well established, or disputed.           | “Confidence: moderate; single-lab finding not yet independently replicated.”  |
| Support and challenge structure | What corroborates the claim and what contests it.                        | “Supported by two follow-up studies, challenged by one negative replication.” |
| Temporal evolution              | How the claim changes as new evidence appears.                           | “Upgraded from tentative to established after three later replications.”      |

claims embedded in support and challenge structures [13]. Scientific information extraction work has also begun moving beyond shallow entity-relation triples toward richer claim graphs, as illustrated by SciClaim [14]. Recent systems such as ProvCaRe show how provenance metadata from millions of biomedical articles can be transformed into semantic provenance graphs [15]. But despite such progress, these efforts continue to remain fragmented across communities and domains. What is missing is a general semantic agenda for scientific KGs that unifies claims, evidence, and reasoning. What we are advocating for in ClaimRxiv is a large-scale effort along the lines of services like arXiv, which was instrumental in increasing the reach of research and spurring discussions of scientific consensus that do not rely as rigidly on peer-review gatekeeping [4].

## 4. Representing ClaimRxiv as a KG

To make the blue-sky proposal more concrete, it is useful to ask what ClaimRxiv would actually look like as a KG. Our view is that it should not be modeled as a document repository with RDF metadata attached after the fact. It should instead be modeled as an *assertion graph* in which scientific assertions are first-class nodes connected to entities, contexts, evidence objects, and higher-order epistemic relations.

At a minimum, such a graph would contain five kinds of nodes:

1. *Assertion nodes*, each representing a canonicalized scientific claim;
2. *Entity and concept nodes*, examples including genes, proteins, diseases, materials, tasks, benchmarks, methods, datasets, theorems, and other domain objects already familiar from KG practice;
3. *Context nodes*, which capture scope conditions such as organism, assay, sample size regime, benchmark setting, hardware configuration, or theorem assumptions;
4. Evidence nodes, representing textual passages, tables, figures, datasets, protocols, and experimental results from which an assertion is derived; and
5. Provenance and actor nodes, representing authors, papers, venues, institutions, timestamps, curation agents, and extraction processes.

The key modeling decision is that an assertion node should mediate between domain entities and epistemic structure. For example, a conventional biomedical KG may directly encode a triple such as

*Protein-A interactsWith Protein-B.* A ClaimRxiv-style graph would instead create an assertion node corresponding to a more scoped statement such as: “Protein A binds Protein B under assay condition S with affinity range R in organism O.” That assertion node would then link outward to the participating entities, to one or more evidence nodes extracted from literature, to contextual qualifiers, and to provenance records identifying who made the claim and on what basis. The graph therefore treats a claim as a semantic object that can itself be annotated and reasoned over.

This design suggests a hyper-relational or statement-centric graph model rather than a simple triple store. RDF reification, singleton properties, named graphs, nanopublication patterns, RDF-star style annotations, or property graph formalisms could all serve as partial implementation strategies. The important point is not the exact encoding choice, but the *semantic commitment*: the relation of interest is often not between two entities directly, but between two entities as asserted in a particular *context* with *particular support*. In that sense, ClaimRxiv would be closer to a graph of scientific discourse acts and evidence-bearing assertions than to a conventional factual knowledge base.

Once modeled this way, several valuable graph patterns become possible. Assertions can be connected by *supports*, *challenges*, *refines*, *specializes*, *generalizes*, *replicates*, and *supersedes* relations. Evidence nodes can be weighted by evidential quality or study design, enabling neighborhood analyses that distinguish a claim with many weak supports from a claim supported by fewer but higher-quality studies. Context nodes make it possible to detect that two apparently contradictory assertions are actually about different populations or conditions. Provenance nodes allow users to query not only what is claimed, but who claims it, when the claim entered the literature, and whether it was machine-extracted, author-submitted, or curator-verified.

This graph perspective also clarifies why ClaimRxiv is different from a generic literature graph. A citation graph connects documents. A scholarly metadata graph connects papers, authors, venues, and topics. A domain knowledge graph connects entities through typed relations. ClaimRxiv would connect all three layers through an assertion-centered middle layer. That middle layer is what lets us answer questions such as: which benchmark claim is most contested, which theorem depends on a lemma later weakened by a counterexample, which molecular interaction has only low-quality support, or which assertions in a field have converged across multiple labs and methods. In other words, the graph is not merely a storage mechanism, but the substrate for scientific memory.

An especially important design question concerns identity. When should two textual claims map to the same assertion node? The answer cannot rely on string similarity alone. We need graph-based canonicalization procedures that combine time-honored graph matching techniques like ontology alignment, entity resolution, and context matching. Two claims may share the same core predicate but differ in assay conditions, effect size direction, evaluation metric, or population. Claim identity must therefore be treated as a systematic equivalence problem over both relational content and scope conditions. In a mature ClaimRxiv graph, canonical assertion nodes might coexist with more local claim mentions, allowing systems to preserve textual diversity while still constructing shared hubs for scientific synthesis.

Representing ClaimRxiv as a KG opens the door to new forms of graph-native computation. Community detection could reveal schools of thought or clusters of mutually reinforcing claims. Path analysis could trace how a benchmark result depends on datasets, evaluation protocols, and prior assumptions. Belief revision operators could update local confidence landscapes as new evidence nodes are added. LLMs could use the graph as a grounding substrate for question answering or critique generation, but the graph would also discipline the LLM by forcing outputs to attach to explicit assertions, evidence records, and provenance trails. This is precisely why the proposal is interesting for the Semantic Web: it invites the community not merely to build better extraction pipelines, but to design a new kind of KG whose semantics are aligned with the natural structure of science itself.

We argue that it is feasible to automatically construct such richly represented scientific KGs from literature, but the technical challenge is subtler than that implicated by standard KG construction. Recent work shows rapid progress in extracting scientific claims, evidence structures, and KGs from text. Transformer-based models have been used to extract fine-grained claim graphs with qualifiers and evidence attributes [14]. LLM-assisted pipelines are beginning to construct evidence-centric biomedical

KGs from full text while preserving study context and quantitative support [16]. Other systems focus on ontology and KG generation in scientific domains using open LLMs [17] and multi-agent pipelines [18, 19, 20].

Nevertheless, some critical limitations persist, most important of which is that most approaches still collapse scientific findings into flat triples too early. This loses study design, quantitative effect sizes, temporal context, and negative evidence. There is also far too much focus on just extraction quality in the literature: the representation must support downstream use by humans and machines in tasks such as contradiction detection, evidence comparison, and what-if reasoning about the effect of future experiments. Finally, existing pipelines often treat corpora as static, even though scientific knowledge evolves and conflicting evidence must be integrated over time. Emerging temporally aware systems such as MedKGent make this limitation explicit by modeling incremental knowledge accretion and conflict resolution [21].

So the problem is not merely whether a KG can be automatically populated, but whether automatically constructed representations can preserve enough *epistemic structure* to be useful for science rather than merely for document indexing. A scientifically meaningful KG must retain distinctions between observation and interpretation, strong and weak evidence, and broad claims vis-a-vis those that only hold under narrow conditions. If those distinctions are flattened away during extraction, the resulting graph may still support search and summarization, but it will not support serious scientific reasoning. Our goal here is not just scalable KG population, but faithful epistemic compression.

## 5. Do We Need a Separate Reasoning Engine for Science?

Standard semantic reasoning is indispensable, but - we argue - insufficient. Ontological entailment, consistency checking, and graph traversal do not by themselves tell us what to believe when evidence conflicts, when results are scoped to different experimental settings, or when confidence should change as new studies accumulate. Closer study suggests that scientific reasoning has distinctive properties:

- it is **defeasible**: claims can be weakened or overturned;
- it is **provenance-sensitive**: who asserted something, using what method, matters;
- it is **uncertainty-aware**: confidence is graded, not binary;
- it is **contextual**: apparently contradictory results may apply to different settings;
- it is **temporal**: evidence arrives over time and changes belief states;
- it is often **argumentative**: support and challenge relations among claims matter.

These characteristics suggest that a scientific reasoning engine must minimally combine several traditions that are often treated separately: semantic web reasoning, argumentation, uncertainty representation, and statistical or evidential aggregation. The EVI ontology already points in this direction by treating scientific findings as defeasible claims embedded in support and challenge graphs [13]. Provenance-oriented extensions of nanopublications likewise point toward explicit modeling of multi-source trust and aggregated knowledge provenance [12].

One possible design is a layered reasoning architecture. A lower layer would handle classical symbolic tasks: identity resolution, type checking, and constraint validation. Above that, an evidence layer would aggregate support and challenge relations, taking into account source provenance, study design, recency, and replication. A temporal revision layer would manage belief updates over time. Finally, an interaction layer could couple LLMs with structured semantic state, using the graph not as a static memory dump but as a substrate for grounded question answering, critique, and hypothesis generation.

Our design here should not be misunderstood to be a call to replace logic with heuristics, let alone as a proposal to hand scientific judgment over to LLMs. What we are advocating for is to build semantics-native infrastructures in which symbolic structure and evidential reasoning can work together. It will require both novel research agendas, as well as investments in engineering and tooling. There is very little work currently, either in the Semantic Web or beyond, on how we can reconcile scientific

reasoning (which involves some degree of abductive and subjective interpretation) with other types of non-monotonic logics. In contrast, much discourse on this has been published in the philosophical literature. Building useful computational systems by drawing on such discourse remains an interesting direction for future work.

## 6. A Research Agenda for the Semantics Community

In considering these challenges, we outline six research problems that could define a long-term agenda for semantics for science.

**Representational Primitives for Scientific Assertions.** We need interoperable semantic models for assertions, evidence items, challenge relations, scope conditions, and evidential quality. Existing ontologies provide useful pieces, but there is no broadly adopted representation that captures the full life cycle of scientific claims across domains. Without common primitives, scientific KGs will remain siloed and difficult to align across fields. A mature ClaimRxiv-like infrastructure would need assertion schemas that are expressive enough to preserve epistemic nuance, but standardized enough to support cross-domain integration and reuse.

**Extraction Beyond Triples.** Scientific information extraction should target evidence-centric objects rather than only triples. This includes extracting quantitative effects, methodological qualifiers, negative findings, and conditions - possibly bounded by uncertainty qualifier - from text, tables, and figures. In practice, this means building multimodal extraction pipelines that can move fluidly between prose, figures, supplementary material, and structured data. It also means evaluating systems on whether they preserve the semantics of scientific findings, not merely whether they recover the right subject, predicate, and object.

**Identity, Alignment, and Canonicalization.** Scientific KGs must decide when two textual claims are the same assertion under different wording, when one is a specialization of another, and when they are genuinely different because of hidden scope conditions. This requires new work on canonicalization and claim identity, not just entity linking. The challenge is especially acute in science because small contextual changes can alter meaning dramatically, even when surface forms look similar. A useful system must therefore combine linguistic normalization with context-aware alignment methods.

**Contradiction, Disagreement, and Revision.** We need representations and algorithms for distinguishing direct contradiction from contextual variation, and for tracking how communities' confidence in claims changes over time. This implies maintaining scientific KGs as living structures rather than as static snapshots extracted from a corpus at one moment in time. It also raises governance questions: when should a claim be marked as disputed, when should confidence be revised, and how should conflicting evidence be exposed to users without oversimplification?

**Reasoning Benchmarks for Science.** Benchmarking should move beyond extraction accuracy toward question-level utility. Can a system answer what is known, what is contested, and what evidence would most efficiently resolve uncertainty? Can it surface contradictions with provenance? Can it explain why two seemingly similar claims should not be merged? These benchmarks should ideally be designed with real scientific workflows in mind, such as literature review, experiment prioritization, or post-publication critique. The key evaluation target is not only correctness in isolation, but usefulness for epistemic tasks that matter in practice.

**Human-Centered Scientific Knowledge Interfaces.** Scientific KGs will only matter if they are usable by scientists. Interfaces should allow users to inspect evidence trails, compare conflicting studies, query for scope conditions, and understand why a system assigned a confidence level or merged

two assertions. Scientists will need interaction paradigms that support both high-level synthesis and low-level auditability, since trust depends on being able to drill down into the supporting record. In this sense, interface design is far from peripheral but occupies a central role in determining whether evidence-centric scientific KGs can become part of everyday research practice.

## 7. ClaimRxiv as a Blue-Sky Research Program for Scientific KGs

We emphasize that the ClaimRxiv proposal differs from mainstream KG construction primarily by treating the scientific assertion *vis-a-vis* the entity or the document, as the primary semantic unit. But it also emphasizes evidence and challenge structure rather than focusing strongly on traditional information extraction. It also aims to develop a dedicated reasoning layer whose semantics are tuned to the epistemology of science rather than to generic database facts. In that sense, ClaimRxiv is not just an application domain for existing semantic methods but as a test case for whether such methods can (at scale) support a fundamentally richer representation of scientific knowledge.

ClaimRxiv is both high-risk and high-gain. The risk is obvious: the problem is hard, interdisciplinary, and unlikely to yield a universal schema or one-shot solution. Scientific language is messy, domain-specific, multimodal, and revisionary. Different fields carry different standards of evidence and even different notions of what counts as a claim. The gain, however, is correspondingly large. If successful, a ClaimRxiv-like infrastructure could transform literature search, review synthesis, contradiction detection, and AI-assisted scientific discovery. The Semantic Web community is especially well placed to lead this effort given its deep expertise in knowledge representation, and reasoning.

## 8. Conclusion

We have argued for a new research direction: semantics for science. The central premise is that science needs better computational substrates than papers alone. Semantics for science also raises interesting questions - that even philosophers of science are grappling with - about what artifacts science *should* produce for the age of AI. The Semantic Web community can play a consequential role in this movement by helping decide the future of scientific infrastructure - one in which assertions, evidence, and challenge become first-class semantic citizens. Papers would remain valuable narratives. But the cumulative memory of science should become computable.

## Declaration on Generative AI

During the preparation of this work, the author(s) used the GRAIL LaTeX agent available at <https://grailai.io> in order to provide minor writing and formatting help. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

## References

- [1] Gusenbauer M. The age of abundant scholarly information and its synthesis- A time when 'just google it' is no longer enough. *Res Synth Methods*. 2021 Nov;12(6):684-691. doi: 10.1002/jrsm.1520. Epub 2021 Sep 7. PMID: 34378322; PMCID: PMC9291810.
- [2] Hanson, Mark A., Barreiro, Pablo Gómez, Crosetto, Paolo and Brockington, Dan. The strain on scientific publishing. *Quant Sci Stud*. 2024;5(4):823-843. doi: 10.1162/qss\_a\_00327.
- [3] National Library of Medicine. MEDLINE/PubMed Production Statistics [Internet]. Bethesda (MD): U.S. National Library of Medicine; 2024 [cited 2026 May 28]. Available from: [https://www.nlm.nih.gov/bsd/medline\\_pubmed\\_production\\_stats.html](https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html).
- [4] Maganur, S., and Kejriwal, M. (2026). Structural shifts in institutional participation and collaboration within the AI arXiv preprint research ecosystem. arXiv preprint arXiv:2602.03969.

- [5] Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1(1), 0021.
- [6] Lamers WS, van Eck NJ, and Waltman L. Meta-Research: Investigating disagreement in the scientific literature. *eLife*. 2021;10:e72737. doi: 10.7554/eLife.72737.
- [7] Hamilton, D. G., McKenzie, J. E., Nguyen, P. Y., Rethlefsen, M. L., McDonald, S., Brennan, S. E., ... Page, M. J. (2025). Evaluation of the replicability of systematic reviews with meta-analyses of the effects of health interventions. *medRxiv*, 2025-07.
- [8] Heisenberg, W. (2025). Physics and philosophy: The revolution in modern science.
- [9] The Gene Ontology Consortium. The Gene Ontology Resource [Internet]. 2024 [cited 2026 May 28]. Available from: <https://geneontology.org/>.
- [10] Thomas PD, Hill DP, Mi H, et al. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. *Nat Genet*. 2019 Oct;51(10):1429-1433. doi: 10.1038/s41588-019-0500-1.
- [11] Groth P, Gibson A, and Velterop J. The anatomy of a nanopublication. *Inf Serv Use*. 2010;30(1-2):51-56. doi: 10.3233/ISU-2010-0613.
- [12] Menotti, L., Marchesin, S., Giachelle, F., and Silvello, G. (2025). Provenance-driven nanopublications: representing source lineage and trust networks for multi-source assertions. *International Journal on Digital Libraries*, 26(4), 24.
- [13] Manir SA and Clark T. EVI: Evidence Graph Ontology [Internet]. 2020 [cited 2026 May 28]. Available from: <https://fairscape.github.io/EVI/index.html>.
- [14] Magnusson I and Friedman S. Extracting Fine-Grained Knowledge Graphs of Scientific Claims: Dataset and Transformer-Based Results. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*; 2021. p. 3581-3593. doi: 10.18653/v1/2021.emnlp-main.381.
- [15] Sahoo, S. S., Valdez, J., Rueschman, M., and Kim, M. (2019). Semantic provenance graph for reproducibility of biomedical research studies: Generating and analyzing graph structures from published literature. *Studies in health technology and informatics*, 264, 328.
- [16] Zong, C., Lv, S., Xue, S. T., Zheng, H., Wan, J., and Zhang, L. (2026). Building evidence-based knowledge bases from full-text literature for disease-specific biomedical reasoning. *arXiv preprint arXiv:2603.28325*.
- [17] Oarga, A., Hart, M., Bran, A. M., Lederbauer, M., and Schwaller, P. (2026). Scientific knowledge graph and ontology generation using open large language models. *Digital Discovery*, 5(3), 1269-1279.
- [18] Lu, Y., Wu, W., Zhao, X., Peng, R., and Wang, J. (2026). Karma: Leveraging multi-agent llms for automated knowledge graph enrichment. *Advances in Neural Information Processing Systems*, 38, 56421-56453.
- [19] Pan S, Luo L, Wang Y, Chen C, Wang J, and Wu X. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans Knowl Data Eng*. 2024;36(7):3580-3599. doi: 10.1109/TKDE.2024.3352100.
- [20] Yang Z, Yuan S, Shao Z, et al. A review on synergizing knowledge graphs and large language models. *Computing*. 2025;107:143. doi: 10.1007/s00607-025-01499-8.
- [21] Zhang, D., Wang, Z., Li, Z. Z., Yu, Y., Jia, S., Dong, J., ... Song, L. (2025). Medkgent: A large language model agent framework for constructing temporally evolving medical knowledge graph. *arXiv preprint arXiv:2508.12393*.