

BioGraphia: A LLM-Assisted Biological Pathway Graph Annotation Platform

Xi Xu¹ Sumin Jo² Adam Officer³ Angela Chen⁴ Yufei Huang⁵ Lei Li¹

¹ Carnegie Mellon University, Language Technologies Institute

² Electrical and Computer Engineering, University of Pittsburgh

³ Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine

⁴ Carnegie Mellon University, Machine Learning Department

⁵ Medicine, University of Pittsburgh School of Medicine

{xixu, angelac3, leili}@cs.cmu.edu

{sumin.jo, ado81, yuh119}@pitt.edu

Abstract

Comprehensive pathway datasets are essential resources for advancing biological research, yet constructing these datasets is labor intensive. Recognizing the labor-intensive nature of constructing these critical resources, we present BioGraphia, a web-based annotation platform designed to facilitate collaborative pathway graph annotation. BioGraphia supports multi-user collaboration with real-time monitoring, curation, and interactive pathway graph visualization. It enables users to directly annotate the nodes and relations on the candidate graph, guided by detailed instructions. The platform is further enhanced with a large language model that automatically generates explainable and span-aligned pre-annotation to accelerate the annotation process. Its modular design allows flexible integration of external knowledge bases, and customization of the definition of annotation schema and, to support adaptation to other graph-based annotation tasks. Code is available at <https://github.com/LeiLiLab/BioGraphia>

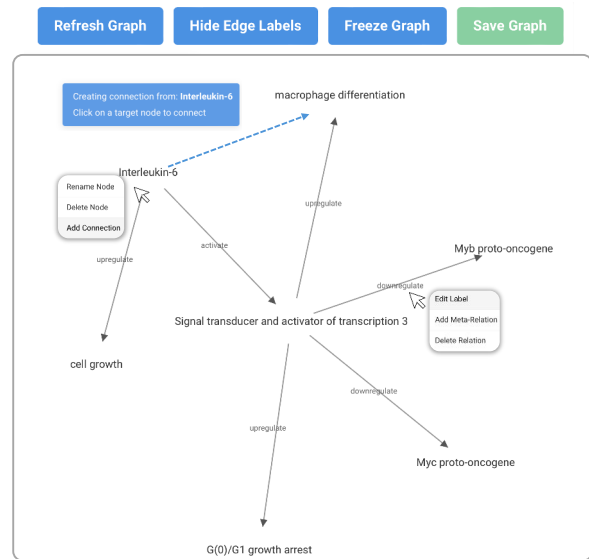


Figure 1: Traditional text-based annotation tools often lack a visual representation of the pathway, hindering both efficiency and accuracy. BioGraphia enhances pathway annotation through interactive graphical annotation interface. Annotators can edit elements and connections, improving efficiency and accuracy

1 Introduction

Biological pathways represent the intricate networks of molecular interactions that govern cellular processes. They describe the ordered series of molecular events, such as gene expression, protein modifications, and metabolic reactions, that underpin essential cellular functions. Despite the importance of pathway knowledge, extracting accurate and comprehensive pathway information from the vast and ever-growing biomedical literature remains a significant challenge.

Existing pathway extraction approaches, employed by DBs like INDRA (Bachman et al., 2023), KEGG (Kanehisa, 2000), and Reactome (Milacic et al., 2023), often rely on simplifying assumptions that limit their ability to capture true complexity of biological reactions. For instance, many methods focus on identifying binary relationships within

individual sentences or short text segments (Rindfleisch and Fiszman, 2003; Valenzuela-Escárcega et al., 2018; Allen et al., 2018), struggling to represent the multi-faceted and hierarchical nature of biological interactions. This often leads to incomplete or inaccurate pathway representations, as critical information about context dependent interactions are lost. Furthermore, integrating diverse evidence from multiple publications into a unified, coherent pathway representation is hampered by this lack of context, potentially resulting in contradictory or incomplete pathways (Sosa and Altman, 2022).

A major obstacle to developing and evaluating pathway extraction methods is the scarcity of high-quality ground truth data. However, manually annotating biological pathways from the literature is

a time-consuming and expensive process, requiring expert knowledge and careful attention to detail.

Toward this end, we introduce **BioGraphia**, a web-based system for pathway graph annotation. Similarly to previous annotation platforms (Yimam et al., 2013; Klie et al., 2018), BioGraphia supports the core functionalities required for collaborative annotation. It is immediately usable through any modern web browser and does not require local installation, greatly improving accessibility for annotators. For deployment, the system includes basic security features to ensure safe use in shared environments.

BioGraphia supports multiple user roles, including administrators, annotators, and curators, and provides built-in functionality for annotator management, interannotator agreement analysis, data curation, and progress monitoring. In addition, the platform offers quality control tools, supports the import of external knowledge bases, and allows annotators and administrators to share notes and comments within the system to facilitate collaborative curation. Beyond these standard features, BioGraphia introduces several novel capabilities:

- *Interactive Graphical Annotation Interface:* Users can directly annotate the pathway graphs generated by our system rather than identifying text spans in the source documents. This interface supports enriched graph representations and enables annotation of complex relations, such as meta-relations, directly on the graph structure, as shown in Figure 1.
- *LLM-assisted Explainable Span-Aligned Pre-Annotation:* Our integrated LLM module automatically generates candidate annotations with aligned spans and textual justifications, showing where in the text each relation was derived from and why. This functionality significantly improves annotation efficiency and consistency and reduces cognitive load for annotators.
- *Complex relation Annotation:* The platform supports the annotation of complex, multi-faceted relationships including biological context and hierarchical relations to provide richer semantic information.
- *Flexible and Extensible Design:* Users can easily extend the platform to integrate new knowledge sources, redefine annotation

schemas, and modify the LLM module to match the requirements of different annotation projects.

BioGraphia serves not only as a practical solution for pathway construction but also as a valuable resource for evaluating and refining knowledge graph construction methods. By providing high-quality benchmark annotations, BioGraphia facilitates the development of more accurate and comprehensive pathway extraction techniques, thereby supporting the creation of richer and more reliable biological knowledge graphs. Ultimately, this contributes to a deeper understanding of molecular mechanisms and accelerates the discovery of novel therapeutic interventions.

BioGraphia is released under the Apache 2.0 license, ensuring that the tool can be freely used and extended for future research and annotation projects without restriction.

In the following sections, we review related work in Section 2. Section 3 states the definition of Pathway Graph Annotation. Section 4 describes the system architecture and the functionality of each component and Section 5 reports results from our evaluation and presents example use cases. Finally, Section 6 concludes the paper and discusses future directions.

2 Related Work

Several annotation platforms have been developed for distributed annotation tasks, among which WebAnno (Yimam et al., 2013) and Inception (Klie et al., 2018) are the most widely used open-source solutions. However, both platforms are primarily tailored for linguistic annotation tasks and are limited in handling the complexity of biological pathway extraction. Specifically, their text span layer based annotation methods require annotators to spend significant time identifying relevant entities within text documents, making them inefficient for complex, graph-based biological annotation tasks.

GraphAnno (Gast et al., 2015) is an open-source annotation tool explicitly designed for knowledge graph construction. Nonetheless, GraphAnno’s command-line interface poses significant usability challenges, requiring considerable learning effort from annotators and thus restricting its scalability for large-scale annotation efforts. Recent advancements in LLMs have inspired novel approaches to annotation, such as ITAKE (Song et al.,

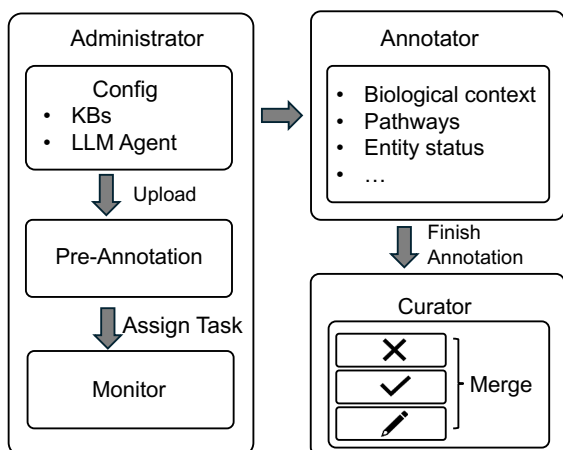


Figure 2: Workflow overview illustrating roles and responsibilities of the administrator, annotator, and curator in BioGraphia. The administrator configures the knowledge base and pre-annotation settings for the LLM module, uploads the materials to be annotated, and initiates pre-annotation. Afterward, the administrator assigns annotation tasks to annotators and monitors their progress. Once annotators complete their work, the curator reviews and merges annotations from different annotators.

2024), which integrates LLM module and ModeIOps into the annotation workflow to enhance annotation quality and efficiency. However, ITAKE focuses solely on traditional NLP annotation tasks such as named entity recognition and focus on supporting conventional neural models and LLMs with basic prompting strategies, which are insufficient for complex biological pathway annotation requirements. Additionally, ITAKE is not designed for graph-based annotation workflows and is not yet available as an open-source platform. Among these tools, BioGraphia represents a modern, open-source graph annotation platform that supports multi-user collaborative environments with role-based access control, real-time collaboration features, and comprehensive project management capabilities for large-scale biological pathway annotation projects.

3 Task Definition

We formalize the task of Pathway Graph Annotation as the extraction and representation of structured biological pathway knowledge from unstructured text sources, such as scientific papers, laboratory experiment reports, and other scientific documentation. Formally, given a textual document D , the goal is to extract a pathway graph $G_D = (C, N, E)$.

Biological Context (C) defines the context in which the pathway is described. Specifically, the context C is represented as a tuple (c_{name}, c_{type}) , where:

- c_{name} explicitly identifies the biological context (e.g., a specific disease or cell type) if provided.
- c_{type} specifies the category of the context.¹

Nodes (N) represent entities identified from document D . Each node $n \in N$ is characterized by:

- t_n : the category of the node, selected from a predefined taxonomy².
- s_n : the context-specific status of the entity, explicitly stated in document D .

Edges (E) are directed relations representing explicit regulatory interactions between nodes, extracted from D . Each edge $e \in E$ is a tuple $(n_{head}, n_{tail}, r, M)$, where:

- $n_{head}, n_{tail} \in N$ are the source and target nodes of the relation.
- r is a label describing the nature of relation.
- M is an optional set of meta-relations with each meta-relation structured as $(r_{meta}, n_{metaTarget})$, where r_{meta} describes the type of secondary effect and $n_{metaTarget}$ is the target node of the secondary effect. Meta-relations are restricted to represent a molecular event where a primary relation between two nodes has indirect or secondary effects to another node.

4 BioGraphia System Overview

BioGraphia incorporates three distinct user roles: Administrator, Annotator, and Curator. The overall workflow begins with administrators configuring project parameters, proceeds with annotators performing the detailed annotation tasks, and concludes with curators reviewing and finalizing annotations. An overview of this workflow is presented in Figure 2. Each component is elaborated upon in the following subsections.

¹The set of context types includes Disease, Cell/Tissue, Pathological/Environmental Stress, Undefined/Generic.

²Gene, Protein, Biological Process, Chemical Compound, RNA Modification, Medicine or Drug, Disease, Clinical Phenotype

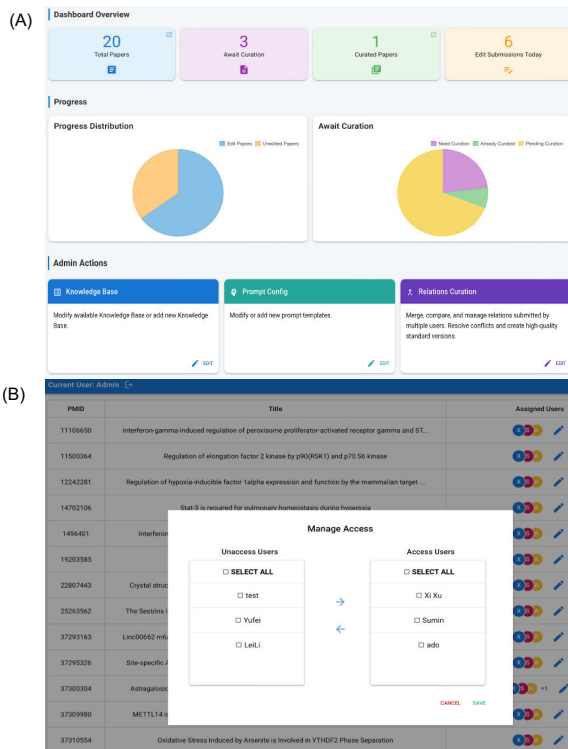


Figure 3: Administrator project management interface. Panel (A) shows how the administrator can monitor the progress of annotation across tasks. Panel (B) displays the interface for assigning annotation tasks to different annotators; each annotator can only view the tasks assigned to them. The administrator can also configure the knowledge base and pre-annotation LLM module settings.

4.1 Project Management

Administrators are responsible for initializing annotation projects. Initially, administrators configure the pre-annotation LLM module and integrate structured knowledge bases, which are imported in JSON format. The administrator assigns each knowledge base to specific annotation types, enabling annotators to access relevant contextual information during their tasks. Additionally, administrators can import scientific papers that need to be annotated individually or in batches for pre-annotation processing. Once our LLM module completes the pre-annotation process, administrators assign annotation tasks to individual annotators as shown in Figure 3.

Annotators then gain exclusive access to the assigned papers through their user-specific dashboards. Administrators also maintain oversight through real-time monitoring capabilities, tracking annotation progress and assessing task completion status to ensure efficiency and productivity.



Figure 4: Annotator workspace in BioGraphia. Panel (A) shows the structured table view, while panel (B) presents the interactive pathway graph. Annotators can annotate nodes, relations, meta-relations, and biological contexts either through the graph interface or the table view. Clicking a relation highlights its supporting textual evidence. Hovering over a node displays its detailed attributes. Annotations are guided by predefined biological knowledge bases.

4.2 Pre-annotated by Pathway LLM

The Pathway LLM employs carefully designed Chain-of-Thought (CoT) few-shot prompting (Wei et al., 2023) crafted by domain experts. These prompts clearly outline annotation criteria and provide illustrative examples detailing intermediate reasoning steps necessary for accurate annotation. Upon generating preliminary annotations, the Pathway LLM module parses the output to produce structured JSON files. Subsequently, these files are processed through our span-alignment tool, which identifies corresponding textual spans from the source documents, ensuring alignment between annotations and textual evidence. Additionally, the system integrates a graph visualization tool, producing interactive span-aligned pathway graphs. The modular architecture of the Pathway LLM module enables customization, allowing users to tailor the LLM module's configuration to suit project-specific annotation requirements.

4.3 Multi-user Annotation

Annotators access their tasks through a dedicated workspace as shown in Figure 4. Upon selecting a specific paper, annotators review the detailed pre-annotations provided by the Pathway LLM module. Pre-annotations include comprehensive biological

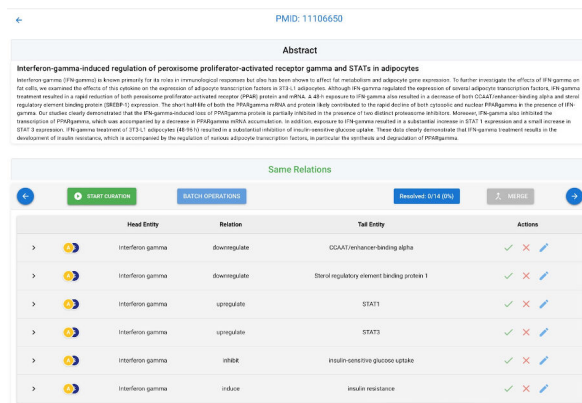


Figure 5: Curation interface of BioGraphia. Curators review completed annotations, resolve conflicts among annotators, and finalize pathway annotations.

context, entity names, node types, and context-specific statuses, assuming an initial positive state unless explicitly stated otherwise.

Annotators can interact directly with the graphical representation of the annotations; clicking on graph relations highlights corresponding textual evidence supporting the annotation. Annotators may refine annotations by editing node labels, relations, supporting text, and by introducing additional relations, including complex meta-relations. Annotators can also annotate biological contexts strictly based on predefined knowledge bases, ensuring consistency and accuracy. We also have table view for annotation.

Moreover, annotators can document comments and annotations-specific notes, visible to both administrators and other annotators, so that annotator and administrator can collaborative resolve potential issues and continuous improvement. Upon completion of their tasks, annotators submit their annotations for subsequent curation, initiating the final review stage performed by curators.

4.4 Curation

We utilize an approach analogous to version control merging, as shown in Figure 5 systematically identifying annotations that have unanimous agreement as well as annotations where discrepancies exist between annotators. Relations agreed upon by all annotators are automatically highlighted, while conflicting annotations are clearly flagged. Curators then carefully review these flagged annotations, deciding to accept, reject, or modify them to ensure accuracy and coherence. To maintain robust quality control, the curation phase is initiated only after an annotation instance has been independently

completed by at least two annotators.

4.5 Ensuring system and data security

System security BioGraphia incorporates a number of comprehensive security measures to protect sensitive data and ensure a safe collaboration environment for users. We implement role-based access control that specifies permissions between administrators, annotators, and curators. This ensures that users can only modify and access data appropriate for their assigned roles. We also incorporate user authentication through bcrypt and session management through JSON Web Tokens (JWT), maintaining secure authentication states throughout user interactions within the platform. To prevent malicious inputs, such as injection attacks, BioGraphia incorporates data validation for all user-submitted content.

5 Evaluation

We evaluated BioGraphia by comparing its performance against traditional manual annotation methods and the widely used annotation platform, INCEPTION, using a controlled annotation experiment.

5.1 Experimental Setup

We selected 160 abstracts related to biological pathways for the annotation study. We recruited 15 annotators with biomedical backgrounds from Carnegie Mellon University and the University of Pittsburgh. Annotators were compensated for their participation, and no conflicts of interest existed between annotators and authors.

Annotators received two warm-up examples with feedback to familiarize themselves with each annotation method. Subsequently, each annotator was randomly assigned to use the three annotation methods—Manual annotation, INCEPTION, and BioGraphia—in a randomized order to minimize learning and ordering effects. For each annotation method, annotators were given four abstracts to annotate.

To ensure consistent task complexity across annotation tools and annotators, abstracts were assigned based on the Corrected Type-Token Ratio (CTTR) (Guiraud, 1954) score, a quantitative measure reflecting text complexity. We use lexicalrichness³ for CTTR score calculation. We employed

³<https://github.com/LSYS/lexicalrichness>

stratified folding to evenly distribute the complexity of the abstracts, ensuring a balanced and fair comparison among methods.

To streamline the annotation process and maintain quality control, we developed an experimental website featuring annotation training materials and built-in quality checks. A sample of the website can be found in <https://t23qu3znup3a.share.zrok.io/#/register>.

5.2 Efficiency and Cognitive Load

We evaluated annotation efficiency by measuring the average task completion time across all annotators for each annotation method. BioGraphia achieved an average completion time of 21 minutes, whereas manual annotation took an average of 38 minutes, and INCEpTION took 49 minutes. This significant reduction in completion time indicates that BioGraphia substantially improves annotation efficiency.

Additionally, we assessed cognitive load using the NASA Task Load Index (NASA-TLX)(Hart, 2006), a widely recognized subjective workload assessment tool. The NASA-TLX evaluates various dimensions of perceived workload, including mental demand, physical demand, temporal demand, performance, effort, and frustration levels, providing a weighted overall workload score.

BioGraphia received a NASA-TLX weighted workload score of 23, indicating a low cognitive demand. In contrast, manual annotation and INCEpTION scored significantly higher, with scores of 63 and 72 respectively, both indicating moderate to high cognitive demands. This result demonstrates that tools designed for traditional NLP task annotation and manual methods significantly increase mental stress when applied to complex biological annotation tasks, whereas BioGraphia maintains a notably lower workload.

Furthermore, BioGraphia exhibited low variance in task completion times, with a standard deviation of only 3.3 minutes (Coefficient of Variation = 14.35%). This consistency further supports BioGraphia's effectiveness in providing a stable and predictable annotation experience.

BioGraphia enhances annotation efficiency and reduces cognitive load compared to traditional manual annotation and existing annotation platforms like INCEpTION. The platform's intuitive design and interactive annotation interface effectively support annotators, making complex biological pathway annotation tasks more manageable and consis-

tent.

6 Conclusion and future work

In this paper, we presented BioGraphia, a web-based annotation platform tailored specifically for distributed pathway graph annotation tasks. BioGraphia integrates collaborative functionalities with real-time monitoring, interactive graph visualization, and LLM-assisted Explainable Span-aligned Pre-Annotation powered by large language models. These features show improvement in annotation efficiency and quality, addressing key limitations found in existing platforms. Its flexible modular architecture further allows seamless integration of external knowledge graphs and customization of annotation schemas, broadening its applicability across various biological annotation scenarios.

Future work will focus on further improving BioGraphia by systematically collecting and incorporating annotator feedback. We plan to enhance the user interface design and provide more comprehensive instructions and guidelines, making the platform more intuitive and scalable for large-scale annotation tasks. Additional efforts will be directed toward refining the design to facilitate annotators in effectively updating aligned text spans. We will also enhance the monitoring system to offer administrators improved oversight capabilities. Finally, we aim to build and maintain a high-quality, continuously updated dataset for biological pathways, incorporating advanced features for dataset management, visualization, inference support, and ongoing updates.

References

- James Allen, Omid Bahkshandeh, William De Beaumont, Lucian Galescu, and Choh Man Teng. 2018. Effective broad-coverage deep parsing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
- John A Bachman, Benjamin M Gyori, and Peter K Sorger. 2023. Automated assembly of molecular mechanisms at scale from text mining and curated databases. *Molecular Systems Biology*, 19(5):e11325.
- Volker Gast, Lennart Bierkandt, and Christoph Rzym-ski. 2015. [Creating and retrieving tense and aspect annotation with GraphAnno, a lightweight tool for multi-level annotation](#). In *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, London, UK. Association for Computational Linguistics.

- Pierre Guiraud. 1954. Les caractères statistiques du vocabulaire: essai de méthodologie. (*No Title*).
- Sandra G Hart. 2006. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA.
- M. Kanehisa. 2000. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30.
- Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. [The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation](#). In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.
- Marija Milacic, Deidre Beavers, Patrick Conley, Chuqiao Gong, Marc Gillespie, Johannes Griss, Robin Haw, Bijay Jassal, Lisa Matthews, Bruce May, Robert Petryszak, Eliot Ragueneau, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Ralf Stephan, Krishna Tiwari, Thawfeek Varusai, Joel Weiser, Adam Wright, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D’Eustachio. 2023. [The reactome pathway knowledgebase 2024](#). *Nucleic Acids Research*, 52(D1):D672–D678.
- Thomas C. Rindflesch and Marcelo Fiszman. 2003. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, 36(6):462–477.
- Jiahe Song, Hongxin Ding, Zhiyuan Wang, Yongxin Xu, Yasha Wang, and Junfeng Zhao. 2024. [ITAKE: Interactive unstructured text annotation and knowledge extraction system with LLMs and ModelOps](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 326–334, Bangkok, Thailand. Association for Computational Linguistics.
- D. N. Sosa and R. B. Altman. 2022. Contexts and contradictions: a roadmap for computational drug repurposing with knowledge inference. *Brief Bioinform*, 23(4).
- Marco A Valenzuela-Escárcega, Özgün Babur, Gus Hahn-Powell, Dane Bell, Thomas Hicks, Enrique Noriega-Atala, Xia Wang, Mihai Surdeanu, Emek Demir, and Clayton T Morrison. 2018. Large-scale automated machine reading discovers new cancer-driving mechanisms. *Database*, 2018: bay098.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. [Chain-of-thought prompting elicits reasoning in large language models](#).
- Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. [WebAnno: A flexible, web-based and visually supported system for distributed annotations](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia, Bulgaria. Association for Computational Linguistics.