ORKGEx: Leveraging Language and Vision Models with Knowledge Graphs for Research Contribution Annotation

Hassan Hussein[1], Fahad Ahmed[2], Allard Oelen[1], Ralph Ewerth[2], Sören Auer[1]
[1]Data Science & Digital Libraries Research Group, [2]Visual Analytics Research Group
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany
{hassan.hussein, fahad.ahmed, allard.oelen, ralph.ewerth, auer}@tib.eu

***Abstract*—A major challenge in scholarly information retrieval is the semantic description of research contributions. While Generative Artificial Intelligence (AI) can assist in this regard, we need minimally invasive approaches for engaging users in the process. We introduce an innovative approach to annotating research articles directly within the browser and seamlessly integrate this approach with the Open Research Knowledge Graph (ORKG). This approach combines human intelligence with advanced neural and symbolic AI techniques to extract semantic research contribution descriptions and integrates the resulting AI-driven annotation tool within a web browser environment. Thus, we aim to facilitate user interaction and improve the creation and curation of scholarly knowledge. We evaluate the effectiveness of our neuro-symbolic approach through a comprehensive user study measuring the quality of the AI-assisted annotation process. Additionally, we illustrate the model's applicability and effectiveness with two use cases in sports analytics and environmental science.**

Keywords-Human; Multimodal AI; Knowledge Graphs; Human-Machine Collaboration.

I. INTRODUCTION

One of the main challenges in research is accurately describing scientific contributions in a way that captures their subtle meanings, contexts, and implications. It is necessary to ensure that these descriptions can be used across different domains and with various information retrieval systems. Achieving this level of precision and consistency is crucial for improving the discoverability, accessibility, and usability of research outputs, promoting collaboration among scientists, and advancing knowledge coherently and systematically.

Previous studies have highlighted these challenges in scholarly annotations. Nantke and Schlupkothen [1] provided a comprehensive analysis of the complexity of annotation practices in scholarly editions, while Howlett and Turner [2] demonstrated how inconsistent terminology across disciplines affects research discoverability. Larivière et al. [3] and Devriendt et al. [4] revealed significant challenges researchers face in standardizing contribution descriptions and receiving proper recognition across different platforms and domains. These challenges are further compounded by the increasing volume and complexity of scientific literature, as demonstrated by Chu and Evans [5] in their analysis of scientific field progression.

Scientific articles contain not only text but figures as well. However, figures often present complex knowledge that can be challenging to comprehend, requiring cognitive and perceptual efforts. Extracting and comprehending data from these figures is challenging due to their diverse formats and contexts. Compared to photos, chart figures offer other challenges as they incorporate text and visual elements, such as lines, bars etc., to show relationships, making accurate data extraction challenging. Meanwhile, traditional Optical Character Recognition (OCR) techniques are not effective for charts since they can only extract text. However, combining OCR techniques with image processing and computer vision techniques can improve data extraction, utlizing object detection techniques, however, they have shown less generalizability and support only a few chart types [6] . Cutting-edge Vision-Language Models (VLMs) and Multimodal Large Language Models (MLLMs) enhance conventional models by integrating vision-language learning, expanding their abilities to image data.

Utilizing Generative AI can streamline semantic description tasks by automating metadata generation and research findings summarization. However, it is crucial to develop user-friendly interfaces and tools that allow researchers to engage with AI systems actively. This collaboration enables researchers to provide feedback, corrections, and additional context where needed, ensuring the precision and relevance of semantic descriptions. In addition, it also establishes trust and transparency in AI-generated outputs. Furthermore, empowering users to participate in the semantic annotation process promises deeper comprehension of their research contributions and promotes knowledge exchange within the scientific community.

We introduce a novel approach to annotate research articles directly within the browser, seamlessly integrating with the Open Research Knowledge Graph (ORKG [7]) to improve the efficiency and effectiveness of knowledge curation and discovery within the research community. Researchers often face challenges in marking key concepts, relationships, and contributions in scholarly documents, which limits the semantic richness and machine-readability of research metadata. Our approach combines human intelligence with advanced AI techniques, utilizing web-based annotation tool to effortlessly annotate research articles. The approach includes multimodal AI systems leveraging language and vision-language models to provide comprehensive annotations. It features human-in-the-loop interaction, allowing researchers to highlight text and receive AI suggestions for relevant property names and data types. In addition, the approach automates the metadata extraction to reduce manual entry and improve accuracy. These annotations automatically integrate with the ORKG, enriching

 ISBN: 978-1-68558-243-2

its content with structured, machine-readable metadata and promoting greater interoperability and collaboration within the research community, ultimately advancing scientific knowledge in a more accessible and interconnected form. To evaluate the effectiveness of our hybrid neuro-symbolic approach in enhancing scholarly knowledge creation and curation, we are conducting a user study measuring the impact of AI assistance on annotation speed, accuracy, efficiency, trustworthiness, and user satisfaction. By analyzing the benefits and challenges of incorporating AI-driven tools into the research workflow, we aim to refine our methods to better meet the needs of the academic community and create a smooth, user-friendly experience that helps researchers produce and share high-quality, and well-annotated scientific knowledge.

The article is structured as follows: Firstly, we review the related work in Section II. In Section III we discuss our approach. Furthermore, in Section IV, we explain our implementation through two use cases. Moreover, in Section V we discuss the users' feedback of our implementation. Additionally, in Section VI, we examine the methodological limitations and challenges in evaluating annotation systems, particularly focusing on the complexities of measuring annotation efficiency and system accuracy. Finally, in Section VII, we conclude and discuss the future work

II. Related Work

Over the last decade, Knowledge Graphs (KG) have emerged as a standard solution and a prominent research trend in academia and industry [8]. KGs [9] organize data in a structured and interconnected manner, simplifying access and analysis of complex datasets. Despite ongoing efforts to enhance the data quality of KGs, significant gaps in entities and relationships persist [9]. In our earlier work [10][11], we focused on improving reproducibility and addressing data inconsistencies, but these approaches often still require substantial manual intervention. Furthermore, the scholarly ecosystem remains predominantly document-centric [12], which we have previously argued hinders reproducibility and complicates the peer review process.

Current document processing approaches struggle with capturing complex document structures [13], while the integration of multimodal content, such as figures and tables, remains a significant challenge [14]. These limitations underscore the need for more comprehensive approaches that can handle both structural and multimodal aspects of scholarly documents while maintaining semantic relationships. This highlights the urgent need for new methods to manage and review the large volume [15] of published articles.

Consequently, the task of semantic description and extraction of scholarly knowledge became a primary focus of research, especially with the rise of advanced AI techniques. Current tools like SciBERT [16] and BERT [17] for scholarly texts have shown promise in text annotation. However, they often struggle with multimodal content integration, particularly in combining textual and visual information for comprehensive knowledge extraction. In this context, neural-symbolic approaches have emerged, integrating neural networks with the symbolic reasoning capabilities of KGs to enhance data quality [18]. These approaches foster comprehensive Knowledge Graph Reasoning (KGR) throughout the life cycle of KGs, offering improvements in scalability and interpretability. For instance, Knowledge Enhanced Graph Neural Networks (KeGNN) incorporate domain knowledge into graph neural networks through knowledge enhancement layers [19]. Additionally, a neural-symbolic system for KG entailment has been proposed, using abstract, generic symbols to discover entailment patterns [20]. However, neural-symbolic integration encounters challenges in effectively integrating robust learning and expressive reasoning under uncertainty [21].

Human-in-the-loop (HITL) computing is essential for developing effective AI systems by integrating human and machine intelligence [22][23]. This synergy leads to accurate results, fosters transparency, builds trust and demonstrates human control over AI [22][24][23]. Designing HITL systems enhances user experience and fosters new interactions between humans and AI [22][25]. HITL can also improve the interpretation of multimodal data, such as images and text, by enhancing collaboration and reducing measurement uncertainty [26]. Studies show that multimodal approaches integrating multiple modalities and involving human interaction have significantly enhanced multimedia comprehension [27]. To address these challenges, there is a pressing need to design intuitive user interfaces for effective human-AI collaboration.

TABLE I. ANALYZING KEY ANNOTATION TOOLS: INSIGHTS FROM AN ORKG COMPARISON

Tool Name	Focus	Type	Limitations	AI Utilization
LightTag	Text	Web-based	Limited features	No
GATE Teamware	NLP oriented	Web-based	Complex interface	Limited
VIA	Audio, Image, Video	Desktop	Limited to media files	Limited
UAM	Text	Desktop	Limited functionality	No
Anafora	Text	Web-based	Requires user configuration	No
YEDDA	Text	Desktop	Limited features	No

As shown in Table I, based on our comparative analysis [28], the available tools lack effective HITL integration. They have various weaknesses, including limited features, and complex interfaces, and require significant Natural Language Processing (NLP) expertise. Additionally, these tools often fail to fully utilize AI capabilities and do not support simultaneous annotation of text, figures, and other media types. This limitation hinders comprehensive data annotation and interpretation. In addition, information in the scholarly articles is mostly derived from their text. However, figures within these articles contain significant data that can be extracted and utilized [29].

Conventionally, OCR has been widely used to extract text from images, including scanned documents [30]. Nevertheless, when applied to charts and non-textual figures, OCR falls short due to the complexity and non-textual nature of the content [30][31]. Charts and plots often contain non-textual elements like data points, lines, bars, etc. Hybrid-OCR models also fall short to perform well on chart figures [32][33]. To solve the challenges of information extraction from figures, state-of-the-art Vision Transformers (ViT) [34], Vision Language Models (VLMs) such as UniChart [35], and LLaVa-NeXT [36] (LLaVa 1.6) have shown great results on the figure classification, chart-to-table and figure summarization tasks respectively.

In conclusion, while knowledge graphs maintain immense promise for organizing scholarly data, they demand significant effort to create and maintain. Neural-symbolic approaches offer significant advancements but also encounter specific challenges. Therefore, incorporating human expertise with multi-modal AI systems presents a viable pathway to bridge these gaps, enabling more effective and efficient scholarly knowledge management.

III. AI-Driven Multimodal Annotation Approach for Scholarly Knowledge Graphs

We propose a novel approach that integrates AI multimodality (language and vision-language models) with neuro-symbolic methods, involving human-in-the-loop interaction to semi-automatically create and curate scholarly knowledge graphs with a minimally invasive user experience.

Unlike existing annotation tools that typically focus on either text-only annotation or require extensive manual intervention, our approach uniquely combines multimodal AI capabilities to handle both textual and visual content simultaneously. This integration, coupled with automatic metadata extraction and research field detection, significantly reduces the annotation workload compared to traditional tools while maintaining high accuracy through human verification.

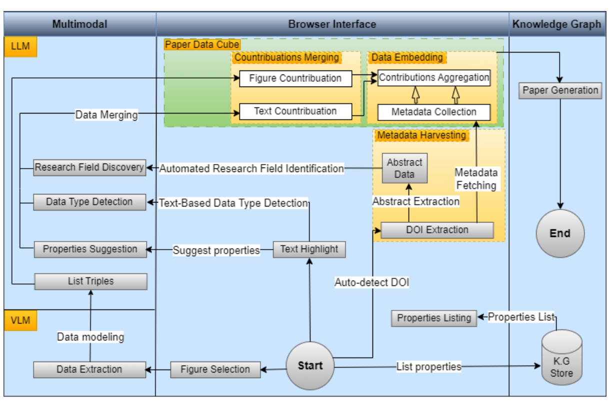

Figure 1. Illustration of the synergistic workflow integrating a multi-model AI system, knowledge graph, and human intelligence in a browser-based annotation system.

Figure 1 demonstrates our minimally invasive approach, meaning it requires minimal user intervention, to reduce user workload. We automatically extract the DOI and metadata from the annotated article, eliminating manual entry. Furthermore, we use AI Multimodal (LLM, and VLM) and knowledge graph capabilities to analyze the content of the article, particularly the abstract keywords, to automatically detect the research field and match it with predefined research fields in ORKG. When users highlight text, the LLM(For the text annotation, we use the gpt-3.5-turbo, as it the most capable model for such task) suggests relevant property names (e.g., "method" or "results") and automatically detects the data type (e.g., text, resource, integer). Meanwhile, we automatically list some most used properties for the user from the ORKG.

Our approach combines vision-language models to provide comprehensive annotations through prompt engineering by integrating multi-modal capabilities (explained in subsection III-A). Users can direct the LLM to process and annotate text, and the VLM for figures within research articles, identifying key concepts, methods, results, and discussions. In addition, the VLM analyzes visual figures within the article, to label significant elements and data points. To ensure coherent interpretation, our system creates connections between the extracted visual data and the surrounding textual content while enabling user verification. This is achieved by presenting the extracted data alongside the original figures, allowing users to cross-match and verify the accuracy of the interpretation

This approach allows the AI multimodal to interpret textual and visual data, thereby providing a holistic annotation of the research article. By integrating these advanced components, our approach furnishes a powerful and minimally invasive tool for researchers to efficiently annotate their articles, leveraging both AI multimodal and KG for accurate and context-aware annotations across multiple modalities.

A. Knowledge Extraction from Figures

We have developed an end-to-end pipeline which utilizes a fine-tuned ViT model and pretrained VLMs for figure classiication and chart comprehension tasks such as summarization and chart-to-text. We fine-tuned the ViT [34] model on the DocFigure [37] dataset to classify scientific figures. The fine-tuned model shows approx 50% accuracy on DocFigure [37]. The UniChart [35] is instruction-tuned on a variety of chart comprehension tasks and has shown improved results compared to other SOTA models. LLaVa-Next [36] has shown promising results for general figure summarization task.

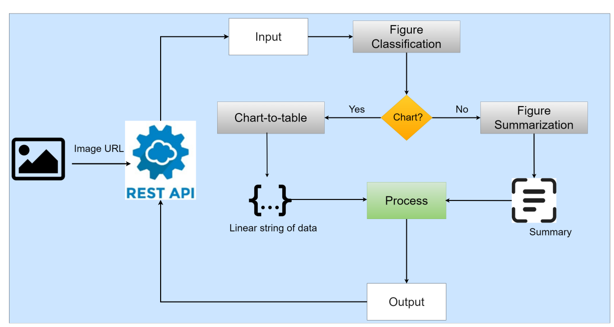

Figure 2. End-to-end information extraction from figures pipeline and REST API.

Figure 2 shows an end-to-end inference pipeline that can be accessed through a REST API. The plugin sends the URL of the selected figure to the REST API. The image is preprocessed before it is passed to the figure classification model. The classification output is then used to determine whether the

 ISBN: 978-1-68558-243-2

figure is a chart/plot or some other type of figure. In case of chart figures, it is used to extract the data table using the UniChart while for non-chart/plot figures the LLaVa-NeXT is used to summarize what the figure actually is about. The outputs are then processed and returned to the REST API caller in the form of a JSON response.

In summary, our approach combines AI Multimodal language and visual models with neuro-symbolic techniques and human-in-the-loop interaction to facilitate the creation and curation of scholarly knowledge graphs with minimal invasive effort. The approach aims to automate the extraction of DOIs and metadata, reducing manual data entry. Using large foundation models and knowledge graph techniques, the approach facilitates analyzing article content to identify and categorize research fields and matches them to predefined ORKG fields.

IV. A Browser Based Interface Implementation:

Our focus is on creating a browser-based extension aims to streamline the workflow of researchers. The extension promotes collaboration between humans and AI Multimodal, to reduce the time and effort researchers spend on selecting properties and identifying relevant research fields. Additionally, it allows researchers to extract data from figures with just few clicks, making the annotation process more efficient for both figures and text. In this section, we showcase two use cases to illustrate our approach. Use case IV-A demonstrates text annotation, while use case IV-B focuses on annotating figures.

A. Sports Analytics Domain Use Case:

To demonstrate our extension's text annotation capabilities, we analyzed a recent paper by Settembre et al. [38]. This work was selected for its dense scientific terminology, making it an ideal test case for evaluating our AI's effectiveness in identifying data types and suggesting appropriate property names. Figure 3 demonstrates our streamlined text annotation pipeline, which consists of four key steps.:

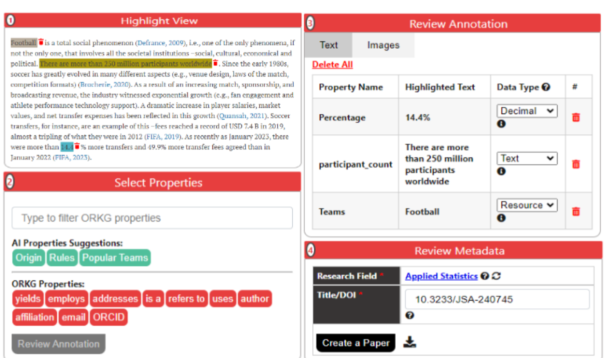

Figure 3. Streamlined Text Annotation Pipeline: From Highlight text to Property Selection, Annotation Review, Datatype Detection, and Metadata Evaluation.

- **1. Highlight View**: This view shows the standard browser interface, permitting researchers to read their selected paper. Users can highlight the text and then click the browser extension to display a list of properties as explained in the next step.
- **2. Select Properties**: In this interface, AI-suggested properties are highlighted in green. Using a predefined prompt template 1, the LLM recommends properties based on highlighted text. Additionally, top properties from the ORKG are listed as red-colored buttons. Users can further search for properties by entering names in a designated search box. Once the highlighted text is associated with a property, the "Review Annotation" button becomes active, allowing users to proceed to the next step.

> **AI Prompt 1:** Act as an ORKG researcher. Given the selected text: (Highlighted Text), identify the most suitable property names. Provide a JSON array with a maximum of four concise property names. Format the response as follows: ['property1', 'property2', 'property3', 'property4']. 1

- **3. Review Annotation**: In this step, the extension sends automatically prompt 2 to the LLM to determine if the highlighted text can be classified as a resource, aligning with the concept of linked data [39]. The annotated data is displayed in a tabular format, with the data type decided by the LLM's findings and accompanied by an explanation, viewable by hovering over the info icon. Users can fully control annotations, including deleting any or all annotated text.

> **AI Prompt 2:** Act as an ORKG researcher. Given the selected text: (Highlighted Text), decide if it should be treated as a resource. Provide your decision and explain why the text should or should not be treated as a resource. 2

- **4. Review Metadata**: The paper's DOI is extracted using a regular expression, typically the first DOI at the top of the paper. Then, metadata is harvested through the Semantic Scholar API service [40]. Concurrently, the LLM analyzes the paper's abstract using a prompt 3 to identify top three relevant research fields. Users can toggle the LLM results by clicking the refresh icon for accurate research field suggestions.

> **AI Prompt 3:** Act as an ORKG researcher. Given the following abstract: (Abstract), identify the three most relevant research fields. Provide a JSON array of concise research field names without values or additional information. Format the response as follows: ["research field1", "research field2", "research field3"]. 3

The extension compiles all metadata, research fields, and highlighted text into a single object for submission to the ORKG to create the paper entry, combining all highlighted text as a single contribution. Additionally, users can download the paper object by clicking the icon next to the "Create a Paper" button, which promotes data interoperability.

B. Environmental Science Domain Use Case:

For our second use case, we evaluated our visual model's annotation capabilities using an Environmental Science paper by Darian-Smith [41]. Specifically, we analyzed Figure 2 from their work, which presents "The Economist: Democracy Index 2023: Age of Conflict" (page 3), as demonstrated in Figure 4. This figure was selected for its complex line chart visualization, which provides an excellent test case for our multimodal annotation approach.:

1) **Selecting Figure and Reviewing the Annotation:** The process starts with the user selecting the figure they want to annotate. When they activate the Chrome extension, an interface showing the selected images will appear. The user can then click on the image annotation view, which will open in a new window.
2) **Image Annotation Review**: After the separate window opens, a request is sent to the VLM to extract key data from the figure. The extracted data is then sent to the LLM using a prompt 4 to format it into triples. Once formatted, the data is displayed to the user, as illustrated in step two. The user can edit, add, or remove the triples as needed.

> **AI Prompt 4:** Act as a researcher and transform the following JSON object (Image Annotation Data) into subject-predicate-object form. Return the data in a JSON object, strictly in the following format: "data": ["subject": "subject value", "predicate": "predicate name", "object": "object value"] Ensure that each entry includes both the subject value and the predicate name. 4

3) **Review Metadata** This step is similar to step 4 in use case IV-A. The key discrepancy here is that each figure annotation is considered a new contribution when added to the ORKG.

Figure 4. Automatic extraction of key data points from a line chart is enabled by a Visual Language Model (VLM) integrated within a Chrome extension.

In our study, we demonstrated the capabilities of our extension in two different fields: sports analytics and environmental science. In sports analytics IV-A, our extension efficiently annotated text using LLM for data type detection, property suggestion, and research field discovery. It integrated with ORKG to streamline the process. In the field of environmental science IV-B, our extension adeptly annotated figures such as line charts using advanced visual and language models. It was able to extract and format the data into triples. Both use cases highlighted the extension's versatility, accuracy, and its ability to assist researchers in annotating various types of research artifacts. It is important to highlight that, the extension permits users to annotate text and figures simultaneously in a single batch process. However, we demonstrated these capabilities separately to emphasize their individual functionalities and benefits.

V. User Evaluation

The evaluation of user feedback serves to systematically validate the performance and effectiveness of our Chrome extension across several key dimensions. Our evaluation is presented in five subsections, structured as follows:

A. Evaluation Design

We conducted a comprehensive user study with 11 participants recruited through professional networks, including postdoctoral researchers, PhD students, and software developers. Of these, 82% had prior ORKG experience, providing valuable feedback on enhancing the extension based on their familiarity with the ORKG pain points. Since the extension is not yet available on the Chrome Web Store, we asked the participants to install TeamViewer software to access our local machine and evaluate the extension remotely. After a brief demo of the extensions' capabilities, participants independently evaluated it. They then filled out a feedback form for analysis

B. Statistical Analysis Methods

All evaluations used a 5-point Likert scale (>3 indicating agreement). The core statistical measures are including sample mean and standard deviation [42], and 95% confidence intervals via t-distribution [43]

C. Results and Analysis

TABLE II. SPEED AND EFFICIENCY METRICS

Feature	Mean $\pm$ SD	95% CI	Key Finding
Figure Triples	4.27 $\pm$ 0.65	(3.83, 4.71)	Highly efficient
Property Suggestions	3.73 $\pm$ 1.14	(2.97, 4.49)	Variable performance
Overall Speed	4.82 $\pm$ 0.39	(4.56, 5.08)	Significant gain

TABLE III. RELIABILITY AND PERFORMANCE ANALYSIS

Metric	Mean $\pm$ SD	Distribution	95% CI
Data Type Detection	4.09 $\pm$ 0.67	5*:27.3% 4*:54.5% 3*:18.2%	(3.64, 4.54)
Field Classification	4.00 $\pm$ 0.74	5*:27.3% 4*:45.5% 3*:27.3%	(3.50, 4.50)
Metadata Extraction	4.27 $\pm$ 0.86	5*:54.5% 4*:18.2% 3*:27.3%	(3.69, 4.85)

D. Visual Analysis

Figure 5 demonstrates that 82% of participants reported accelerated annotation compared to traditional interface, with 90% effectiveness in figure-based triple extraction and 73% accuracy in field detection and metadata harvesting. The human-AI collaboration metrics (Figure 6) reveal 82% user trust in AI-generated content and 73% accuracy perception, with 82% rating the AI content as transparent and explainable. The user experience assessment (Figure 7) indicates high satisfaction levels with 73% reporting adequate control, 90% ease of learning, and 90% positive overall experience.

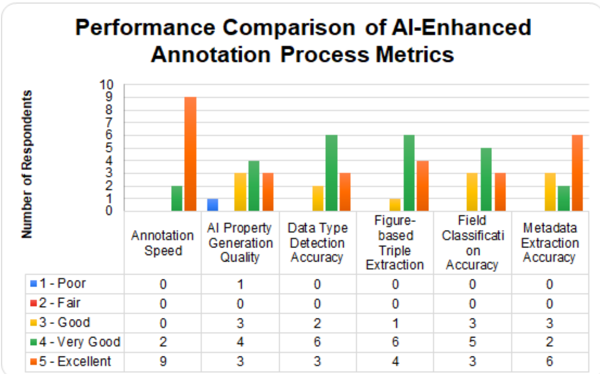

Figure 5. AI-enhanced annotation process metrics showing effectiveness across tasks (Likert scale 1-5). Notable results: figure-based triple extraction (90% positive) and metadata harvesting (73% accuracy).

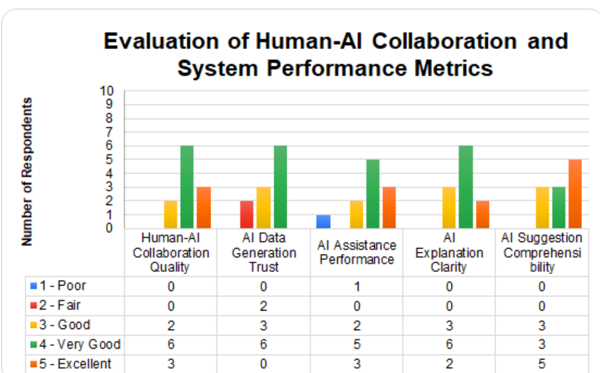

Figure 6. Human-AI interaction quality metrics showing strong user trust (82%) and AI suggestion comprehensibility (73%).

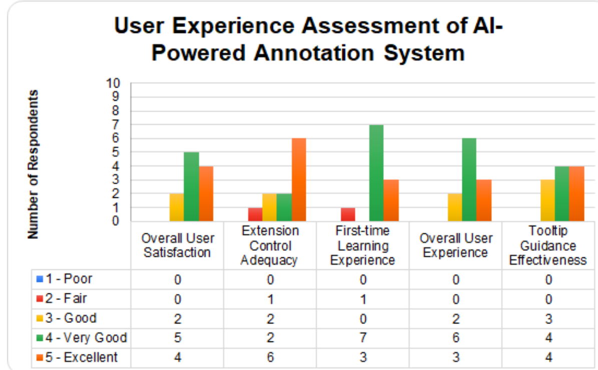

Figure 7. User experience metrics showing strong control (73%), ease of use (90%), and overall satisfaction (90%).

E. Qualitative Analysis

The ORKGEx extension received positive feedback for its ease of use, AI-generated suggestions, and ability to save time during the annotation process. Participants found the tool intuitive and efficient, with many noting that it significantly improved their workflow compared to the other annotaion tools. The AI-generated property suggestions were particularly appreciated, though some users suggested providing more context (e.g., surrounding paragraphs) to enhance relevance. Tooltips were deemed helpful but could be improved with better visibility, such as larger fonts and more noticeable colors. Overall, the extension was seen as a valuable addition to the research community, with potential for further refinement. A summary of the key findings, along with supporting quotes from participants, is provided in Table IV.

The evaluation demonstrates the extension's effectiveness in enhancing annotation processes. Key findings include:

- **Strong user satisfaction:** with the extension, as evidenced by a high mean rating for overall satisfaction (4.18 ± 0.75)
- **Robust performance:** in AI-assisted features, particularly in metadata extraction (mean: 4.27 ± 0.86) and data type detection (mean: 4.09 ± 0.67).
- **Positive user feedback:** highlighting the extension's ease of use, efficiency, and potential to streamline the annotation process.

VI. Limitation

When considering our evaluation methodology, we identified several limitations despite the positive results. While our user study provided valuable perception metrics and validation data, there remain inherent challenges in conducting a fully comprehensive evaluation:

1) **Performance measurement limitations**: Although we obtained significant user-reported speed improvements (4.82 ± 0.39), establishing absolute performance baselines remains challenging due to:
   - Annotator's expertise and familiarity with the domain
   - Paper length and structural complexity
   - Research field and complexity of the scientific content
   - Number and complexity of figures
2) **System evaluation challenges**: We need to do a more comprehensive evaluation to:
   - Larger-scale comparison between manual and automated annotations
   - Controlled environment for quantitative time measurements
   - Standardized test sets with varying complexity levels
3) **Sample diversity**: Although our evaluation with 11 participants (82% with ORKG experience) provided statistically significant results, it is important to note that the small sample size may limit the generalizability of our findings.

TABLE IV. SUMMARY OF KEY FINDINGS WITH SUPPORTING QUOTES

Aspect	Positive Feedback	Constructive Feedback	Supporting Quote
Ease of Use	Intuitive and easy to learn.	Some technical issues and limited testing time affected usability.	*"The tool efficiently harnesses AI capabilities to suggest relevant scientific data in relation to the given user input."* (Masters Student)
AI Suggestions	Helpful in simplifying the annotation process.	Needs more context (e.g., surrounding paragraphs) for better relevance.	*"AI-generated property suggestions needed more context about the selected text."* (Frontend Developer)
Tooltips	Provide helpful information.	Need better visibility (e.g., larger fonts, catchy colors).	*"Tooltips provide helpful information, therefore they have to be represented in a more obvious way (customizing tooltips with a bigger font and a catchy color would be a great addition)."* (Backend Developer)
Metadata Extraction	Generally accurate.	Some participants did not closely examine this feature.	*"I didn't have the option to skip some questions in the evaluations because I didn't closely examine aspects like metadata extraction."* (Frontend Developer)
Overall Satisfaction	High satisfaction with the extension's ability to improve annotation efficiency.	Some participants found it difficult to assess certain features due to limited testing.	*"Overall, I see a huge potential in having such an extension to bring annotated data to the knowledge graph. Keep it up!"* (Frontend Developer)

Broader validation across different expertise levels and larger sample sizes would strengthen our results.

These limitations highlight the complexity of evaluating annotation tools in real-world scenarios, where controlled experiments must be balanced against practical usability and varied user needs.

VII. Conclusion and Future Work

In this paper, we proposed a novel approach to enhance the semantic description of research articles with web-based annotation integrated with the Open Research Knowledge Graph. Our evaluation demonstrated strong performance across multiple metrics, with notably high user satisfaction (90%) and effective AI-assisted features. However, it is important to note that the 90% user satisfaction is based on a sample size of 11 participants, and while statistically significant, it may not fully represent the broader user base.

In the future, we plan to explore advanced techniques in computer vision and natural language processing, such as transformer-based models for figure extraction and fine-tuned language models for research describtion, to improve the extraction and interpretation of complex figures within scholarly articles. To ensure real-world applicability, we will investigate scalability challenges specific to browser-based annotation, such as optimizing the extension's performance with large documents, and managing the synchronization of annotations across multiple users and sessions.

Additionally, we plan to address critical deployment challenges including: ensuring consistent annotation behavior across various browser versions, and maintaining responsive performance while processing complex figures and metadata.

While our current evaluation provides statistically significant insights into user experience and measured benefits, future work will expand evaluation scope with larger, more diverse research communities. To mitigate potential evaluation biases, we will employ systematic sampling methods to ensure balanced representation across different academic disciplines, experience levels, and institutional backgrounds. This expanded evaluation will include both ORKG experts and novices, enabling us to understand how the extension performs across different expertise levels and varied real-world academic workflows. Additionally, we will focus on expanding support for complex data types and interactive elements, such interactive tables, to further enhance the extension's utility across diverse use cases.

The source code of the implementation is available under: https://github.com/Webo1980/ORKGEx

References

[1] J. Nantke and F. Schlupkothen, *"Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization"*. Berlin: De Gruyter, 2020. DOI: 10.1515/9783110689112.

[2] K. Howlett and E. Turner, "Research on the benefits of nature to people: How much overlap is there in citations and terms for 'nature' across disciplines?", *People and Nature*, Dec. 2023.

[3] V. Larivière, D. Pontille, and C. Sugimoto, "Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT)", *MIT Press - Journals*, Jan. 2021. DOI: 10.1162/qss_a_00097.

[4] T. Devriendt, P. Borry, and M. Shabani, "Credit and recognition for contributions to data-sharing platforms among cohort holders and platform developers in Europe : Interview study", *JMIR Publications Inc.*, Jan. 2022. DOI: 10.2196/25983.

[5] J. Chu and J. Evans, "Slowed canonical progress in large fields of science", *Proceedings of the National Academy of Sciences*, Dec. 2022. DOI: 10.1073/pnas.2021636118.

[6] J. Luo *et al.*, "Chartocr: Data extraction from charts images via a deep hybrid framework", in *Proceedings of the IEEE/CVF*

 ISBN: 978-1-68558-243-2

winter conference on applications of computer vision, 2021, pp. 1917–1925. DOI: 10.1109/WACV48630.2021.00196.

[7] S. Auer *et al.*, *"open research knowledge graph"*, https://orkg.org, Accessed: 2025.02.02.

[8] Y. Kong *et al.*, "Bolt defect classification algorithm based on knowledge graph and feature fusion", *Energy Reports*, vol. 8, pp. 856–863, 2022.

[9] C. Peng *et al.*, "Knowledge graphs: Opportunities and challenges", *Artificial Intelligence Review*, vol. 56, pp. 13 071–13 102, 2023. DOI: 10.1007/s10462-023-10465-9.

[10] H. Hussein *et al.*, "Increasing reproducibility in science by interlinking semantic artifact descriptions in a knowledge graph", in *Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration*, D. H. Goh, S.-J. Chen, and S. Tuarob, Eds., Singapore: Springer Nature Singapore, 2023, pp. 220–229, ISBN: 978-981-99-8088-8. DOI: https://doi.org/10.1007/978-981-99-8088-8_19.

[11] H. Hussein *et al.*, "KGMM - a maturity model for scholarly knowledge graphs based on intertwined human-machine collaboration", in *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries*, Y.-H. Tseng, M. Katsurai, and H. N. Nguyen, Eds., Cham: Springer International Publishing, 2022, pp. 253–269, ISBN: 978-3-031-21756-2.

[12] M. Stocker *et al.*, "Fair scientific information with the open research knowledge graph", *FAIR Connect*, 2023.

[13] D. Wang *et al.*, "Docgraphlm: Documental graph language model for information extraction", *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pp. 14 137–14 151, 2023. DOI: https://doi.org/10.1145/3539618.3591975.

[14] L. Wang *et al.*, "Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific pdf accessibility for blind and low vision users", *ArXiv*, vol. abs/2105.00076, 2021.

[15] J. Greenberg *et al.*, "Knowledge organization systems: A network for AI with helping interdisciplinary vocabulary engineering", *Cataloging & Classification Quarterly*, vol. 59, no. 8, pp. 720–739, Nov. 17, 2021, ISSN: 0163-9374, 1544-4554. DOI: 10.1080/01639374.2021.1995918.

[16] L. Beltagy, K. Lo, and A. Cohan, "Scibert: Pretrained contextualized embeddings for scientific text", *CoRR*, vol. abs/1903.10676, 2019. arXiv: 1903.10676.

[17] S. Bano and S. Khalid, "Bert-based extractive text summarization of scholarly articles: A novel architecture", in *2022 International Conference on Artificial Intelligence of Things (ICAIoT)*, 2022, pp. 1–5. DOI: 10.1109/ICAIoT57170.2022.10121826.

[18] H. Chen *et al.*, "Neural symbolic reasoning with knowledge graphs: Knowledge extraction, relational reasoning, and inconsistency checking", *Elsevier BV*, Sep. 2021. DOI: 10.1016/j.fmre.2021.08.013.

[19] L. Werner *et al.*, "Knowledge Enhanced Graph Neural Networks for Graph Completion", *2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, Mar. 2023.

[20] M. Ebrahimi *et al.*, "Neuro-Symbolic Deductive Reasoning for Cross-Knowledge Graph Entailment", in *AAAI Spring Symposium Combining Machine Learning with Knowledge Engineering*, Oct. 2021.

[21] R. Manhaeve *et al.*, "Neuro-symbolic ai = neural + logical + probabilistic ai", in *Neuro-Symbolic Artificial Intelligence*, Jan. 2021. DOI: 10.3233/FAIA210354.

[22] C. Emmanouilidis and S. Waschull, "Human in the Loop of AI Systems in Manufacturing", *NOW PUBLISHERS INC*, Nov. 2021. DOI: 10.1561/9781680838770.ch9.

[23] J. Ostheimer, "Human-in-the-loop Computing : Design Principles for Machine Learning Algorithms of Hybrid Intelligence", *Linnéuniversitetet, Institutionen för informatik (IK)*, Jan. 2019.

[24] C. Emmanouilidis *et al.*, "Human in the AI Loop in Production Environments", *Springer Science and Business Media LLC*, Aug. 2021. DOI: 10.1007/978-3-030-85910-7_35.

[25] K. Tsiakas and D. Murray-Rust, "Using human-in-the-loop and explainable AI to envisage new future work practices", *Association for Computing Machinery (ACM)*, Jan. 2022. DOI: 10.1145/3529190.3534779.

[26] J. Effendi *et al.*, "Multimodal Chain: Cross-Modal Collaboration Through Listening, Speaking, and Visualizing", *Institute of Electrical and Electronics Engineers (IEEE)*, vol. 9, pp. 70 286–70 299, 2021. DOI: 10.1109/ACCESS.2021.3077886.

[27] T. Gong *et al.*, "Multimodal-gpt: A vision and language model for dialogue with humans", *arXiv*, 2023.

[28] H. Hussein *et al.*, *"comparative analysis of annotation tools: Focus, methods, and user suitability"*, https://orkg.org/comparison/R1354448, Accessed: 2025.02.02, DOI: 10.48366/R1354448.

[29] G. Park, J. Rayz, and L. Pouchard, "Figure descriptive text extraction using ontological representation", *eprint arXiv:2208.06040*, Aug. 2020. DOI: 10.48550/arXiv.2208.06040.

[30] J. Ghorpade-Aher *et al.*, "Text retrieval from natural and scanned images", *International Journal of Computer Applications*, vol. 133, no. 8, pp. 10–12, Jan. 2016, ISSN: 0975-8887. DOI: 10.5120/ijca2016907840.

[31] G. Skitalinskaya and N. Düpont, "Ocr report", *Deutsche Forschungsgemeinschaft (DFG)*, 2021. DOI: 10.26092/elib/1517.

[32] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer", *Journal of machine learning research*, vol. 21, no. 140, pp. 1–67, 2020.

[33] M. Zhou *et al.*, "Enhanced chart understanding in vision and language task via cross-modal pre-training on plot table pairs", *arXiv preprint arXiv:2305.18641*, 2023.

[34] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale", *arXiv preprint arXiv:2010.11929*, 2020.

[35] A. Masry *et al.*, "Unichart: A universal vision-language pretrained model for chart comprehension and reasoning", *arXiv preprint arXiv:2305.14761*, 2023.

[36] H. Liu *et al.*, *Llava-next: Improved reasoning, ocr, and world knowledge*, https://llava-vl.github.io/blog/2024-01-30-llava-next/, Accessed: 2025.02.02, 2024.

[37] K. Jobin, A. Mondal, and C. Jawahar, "Docfigure: A dataset for scientific document figure classification", in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, IEEE, vol. 1, 2019, pp. 74–79. DOI: 10.1109/ICDARW.2019.00018.

[38] M. Settembre *et al.*, "Factors associated with match outcomes in elite european football – insights from machine learning models", *IOS Press*, 2024. DOI: 10.3233/JSA-240745.

[39] W. W. W. C. (W3C), *Linked data*, https://www.w3.org/wiki/LinkedData, Accessed: 2025.02.02.

[40] Allen Institute for AI, *Semantic Scholar API*, https://www.semanticscholar.org/product/api, Academic search and discovery platform, Accessed: 2025.02.02.

[41] E. Darian-Smith, "The challenge of political will, global democracy and environmentalism", *Environmental Policy and Law*, vol. 54, pp. 117–126, Jan. 2024. DOI: 10.3233/EPL-239023.

[42] S. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press, 2014.

[43] Student, "The probable error of a mean", *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908, ISSN: 00063444, 14643510.