Towards a Human-Computer Collaborative Scientific Paper Lifecycle: A Pilot Study and Hands-On Tutorial

Qingyun Wang[1], **Carl Edwards**[1], **Heng Ji**[1], **Tom Hope**[2,3]
[1] University of Illinois at Urbana-Champaign [2] Allen Institute for Artificial Intelligence (AI2)
[3] The Hebrew University of Jerusalem
{qingyun4,cne2,hengji}@illinois.edu,{tomh}@allenai.org

Abstract

Due to the rapid growth of publications varying in quality, there exists a pressing need to help scientists digest and evaluate relevant papers, thereby facilitating scientific discovery. This creates a number of urgent questions; however, computer-human collaboration in the scientific paper lifecycle is still in the exploratory stage and lacks a unified framework for analyzing the relevant tasks. Additionally, with the recent significant success of large language models (LLMs), they have increasingly played an important role in academic writing. In this **cutting-edge** tutorial, we aim to provide an all-encompassing overview of the paper lifecycle, detailing how machines can augment every stage of the research process for the scientist, including scientific literature understanding, experiment development, manuscript draft writing, and finally draft evaluation. This tutorial is devised for *researchers interested in this rapidly-developing field of NLP-augmented paper writing*. The tutorial will also feature a session of hands-on exercises during which participants can guide machines in generating ideas and automatically composing key paper elements. Furthermore, we will address current challenges, explore future directions, and discuss potential ethical issues. A toolkit designed for human-computer collaboration throughout the paper lifecycle will also be made publically available. The tutorial materials are online at https://sites.google.com/view/coling2024-paper-lifecycle/.

Keywords: Paper Lifecycle Assistant, LLM, Human-Computer Collaboration

1. Introduction

Scientists are experiencing information overload (Landhuis, 2016) due to the rapid growth of scientific literature. From March 13, 2020, to June 2, 2022, during the COVID-19 pandemic, more than a million articles related to the coronavirus were published (Wang et al., 2020a). However, scientists peruse only about 300 papers annually (Van Noorden, 2014). Simply put, when writing new articles, scientists cannot review all related papers. Beyond this, another major obstacle is the quality of the papers. While many research articles, especially preprints, provide new perspectives for researchers, they also duplicate findings, spread misinformation, and show disagreements among themselves (Wang et al., 2021b). This can cause a seemingly paradoxical increase of misinformation in scientific dissemination as the number of papers increases (Casigliani et al., 2020). Finally, field-specific language can be a barrier to scientific communication (Han et al., 2018; Lucy et al., 2023). For example, Glasziou et al. (2020) shows that collaboration and communication for research were extremely limited during the early stage of COVID-19, causing massive waste in research.

To address these pressing issues, researchers are developing AI methods to mitigate distorted scientific dissemination, generate new research directions, and ultimately draft papers. The recent dramatic advances in large language models raise the tantalizing prospect that such a capability is within reach. For example, researchers have tested ChatGPT (OpenAI, 2023) in writing essays (Stokel-Walker, 2022), research papers (Conroy, 2023c), or even grant applications (Park, 2023). According to a Nature postdoc survey (Nordling, 2023), 31% of respondents use AI chatbots in their work. Despite such popularity, fundamental challenges remain for this vision to materialize. Even with the assistance of search engines, LLMs sometimes generate fake references with incorrect metadata or cite papers that do not exist (Conroy, 2023b). Additionally, LLMs tend to generate papers with extensive plagiarism (Anderson et al., 2023) and inaccurate results (Hosseini et al., 2023).

To address those challenges, we will explore the following questions in this tutorial:

- Why do we care about AI-assisted literature review?
- How can humans leverage computers to evaluate the quality of scientific papers?
- How can AI facilitate new scientific ideas?
- How can we address the ethical issue of large language models in the paper lifecycle?

Specifically, we will offer a comprehensive introduction to recent techniques for a series of tasks involved in the paper lifecycle. To begin with, we will divide the paper lifecycle into four parts: the scientific literature review, hypothesis generation and experiments, paper drafting, and paper evaluation. Furthermore, we will engage the audience

LREC-COLING 2024 Tutorials, pages 56–67

in a hands-on Google Colab project to write and evaluate a new paper draft assisted by LLMs. We will also concurrently discuss ethical concerns in the field throughout the tutorial and include a specific section for ethical concerns. Finally, we will discuss the remaining challenges and future directions for the AI-assisted scientific paper lifecycle. We will construct a toolkit and related papers for the AI-assisted scientific paper lifecycle on GitHub.

2. Target Audience

The tutorial will be accessible to all NLP researchers who wish to develop NLP methods for the scientific paper lifecycle. While no specific background knowledge is required, having a basic understanding of pretrained language models, graph neural networks, and other basic deep learning technologies would be helpful. We expect around **50** to **100** participants based on the popularity.

3. Outline – The Paper Lifecycle [210]

3.1. Background and Motivation [15]

We will begin the tutorial with a comprehensive overview of the scientific paper lifecycle by showcasing various applications in accelerating scientific discovery (Gil, 2022; Birhane et al., 2023), including scientific literature review, scientific hypothesis generation, experiment development, paper draft generation, and draft evaluation. Specifically, we will focus on the recent trend of applying LLMs in academic writing, briefly discussing the benefits and potential ethical concerns of this approach.

3.2. Scientific Literature Review [40]

Scientific Knowledge Base Construction [20] We will introduce scientific LLMs, which usually focus on domain-adaptive pre-training (Phan et al., 2021; Scao et al., 2022; Hong et al., 2023). Working with these general model architectures as tools, we will describe why knowledge graphs are still necessary in the LLM era by providing cases where LLMs fail due to a lack of structured knowledge. Then, we will focus on how various scientific information extraction (IE) tasks are formulated (Hou et al., 2019; Cohan et al., 2019; Jain et al., 2020; Cattan et al., 2021; Panapitiya et al., 2021; Shen et al., 2022; Song et al., 2023). Finally, we will discuss how researchers can improve the knowledge base quality and utilize those tools to enhance the paper reading experience (Fok et al., 2023).

Retrieving Relevant Information [20] Given the exponential growth of papers and the language barrier between different disciplines, scientists need effective ways to retrieve relevant papers. Specifically, we will provide real-world examples of information retrieval (IR) in the Covid-19 (Wang et al., 2020a). Then, we will comprehensively introduce the tasks in scientific information retrieval. Further, we will cover existing methods and applications of scientific information retrieval by categorizing them into four major types, including scientific paper retrieval (Hongwimol et al., 2021), paper relationship discovery (Luu et al., 2021), scientific evidence extraction (Li et al., 2021), and scientific dataset recommendation (Viswanathan et al., 2023). Finally, we will discuss how information retrieval can be used for downstream tasks related to the paper lifecycle, including scientific idea discovery (Hope et al., 2020), and scientific fact-checking (Wang et al., 2023). We will also discuss the potential risks of incorrect information retrieval results.

3.3. Hypothesis Generation and Experiments [25]

Generating Research Directions [20] Since we have built the knowledge base and retrieved relevant papers based on certain topics, we will then present automatic scientific hypotheses generation, the goal of which is to suggest potential research directions for researchers. We will start by showing drug repurposing for COVID-19 as a real-world application of scientific hypothesis generation (Hope et al., 2020; Zhang et al., 2021a; Wang et al., 2021b). We will then give an overview of literature-based research direction discovery (Henry and McInnes, 2017; Hope et al., 2023). After that, we will show how to effectively utilize existing literature and a knowledge base to discover new scientific directions (Wang et al., 2019; Krenn et al., 2023). Lastly, we will discuss the ethical considerations for scientific hypothesis discovery, including usage requirements, potential risks, and system performance limitations.

LLMs as Experimental Agents [5] In this paragraph, we will discuss several real-world applications of using LLMs for experimental agents, including experimental planning and scientific reasoning techniques. By integrating external knowledge bases and domain-specific tools, LLMs can help experts by formulating synthesis procedures (Bran et al., 2023), editing drugs (Liu et al., 2023), analyzing prediction results (Kumar et al., 2023), or even automating experiments (Wierenga et al., 2023). Currently, this direction remains highly exploratory.

3.4. Hands-on Paper Draft Assistant [50]

We will lead a hands-on exercise session using Google Colab, an important component of our tutorial. We will start by providing attendees with a group of seed terms as starting topics and their background knowledge (i.e., background sentences, knowledge graphs, and citation networks). The goal of this practice is first to generate new research ideas about these seed terms and finally to generate key elements of a paper, including a title, an abstract, and a related work section for these topics. Every attendee will initially brainstorm the most effective strategies to generate new hypotheses from the given input. They will later design a pipeline to write key elements of the paper, given the generated hypotheses and background knowledge. We will also ask participants to evaluate the generation quality from multiple perspectives including automatic and human evaluation.

Because writing code from scratch is time-consuming, we will let participants choose from pre-installed state-of-the-art hypotheses and paper generation frameworks. We will also provide them with prepaid accounts and corresponding datasets. By the end of this session, audiences will understand how to build systems for hypothesis generation and paper writing, be familiar with methods prevalent in the realm of automatic scientific paper writing, and know evaluation methods for paper generation. We will release a toolkit on GitHub.

3.5. Drafting a Paper [20]

In this part, we will divide the process of scientific paper writing into several components. We will first review available related work generation frameworks which utilize pretrained language models and graph neural networks (Lu et al., 2020; Ge et al., 2021). Next, we will dive into a more challenging aspect of paper writing: generating paper abstracts based on titles and knowledge graphs (Koncel-Kedziorski et al., 2019; Wang et al., 2019). We will also explore the generation of other paper components, including claim generation (Wright et al., 2022), definition generation (August et al., 2022), table captioning (Chen et al., 2021), and figure captioning (Hsu et al., 2021). Lastly, we will discuss human-AI collaborative writing (Lee et al., 2022).

3.6. Paper Review and Ethics (45)

Automatic Scientific Reviewing [15] An important step in the process of scientific writing is evaluating paper quality to prevent distorted scientific dissemination. Due to the rapid growth in the number of paper submissions, the quality of peer reviews has become a widely discussed topic, as shown in Section 5.3 of Rogers et al. (2023). Therefore, we will present an automatic scientific review assistant to alleviate this issue. We will first demonstrate current progress in automatic scientific review (Yuan et al., 2022). We will then divide the scientific review process into two tasks: peer-review score prediction (Kang et al., 2018) and review comment generation (Wang et al., 2020b). We will also focus on knowledge-guided review score prediction and review comment generation (Yuan and Liu, 2022). Finally, we will discuss automatic peer review in the era of LLMs (Liu and Shah, 2023; Zeng et al., 2023), which includes error detection, checklist verification, paper recommendation, and corpus comparison (Zhong et al., 2023).

Scientific Fact-Checking [15] We will start this section by introducing the danger of misinformation in scientific publications during the COVID-19 pandemic (Nelson et al., 2020). Additionally, language models tend to generate non-factual content (Maynez et al., 2020). We will also outline the importance of scientific fact-checking and highlight its difference from general fact-checking. Then, we will cover current scientific fact-checking datasets (Wadden et al., 2020; Sarrouti et al., 2021) and potential approaches (Zhang et al., 2021b; Yu et al., 2022) for this task. Finally, we will focus on the existing papers on human-centered fact-checking (Glockner et al., 2022; Juneja and Mitra, 2022) and try to adapt them to the scientific domain.

Ethics Concerns in the LLM Era [15] We will recap the increasing trend of using LLMs in academic writing. We will discuss the benefits of LLMs for scholarly publishing, including performing straightforward but time-consuming tasks (Conroy, 2023c) and improving equity in science (Lund et al., 2023). We will then address its risks and ethical concerns by showing a paper (Ayache and Omand, 2022) generated by GPT3 (Brown et al., 2020) as an example. Based on that paper, we will highlight potential issues, including incorrect reference (Conroy, 2023b), extensive plagiarism (Anderson et al., 2023), accuracy concerns (Hosseini et al., 2023), and equity concerns due to its subscription fee. Further, we will show the current challenges in AI-generated research paper detection (Gao et al., 2023). We will also include potential solutions for detecting AI-generated text (Crothers et al., 2023), such as watermarking LLMs (Kirchenbauer et al., 2023), writing style analysis (Ma et al., 2023).

3.7. Open Questions [15]

At the end of the tutorial, we will first discuss recent exploratory work. We will discuss making scientific ideas more accessible to the general public with text style transfer (Dangovski et al., 2021; Goldsack et al., 2022; Fatima and Strube, 2023). We will conclude the tutorial by presenting the remaining challenges and future directions, including 1) multimodal analysis of formulas, tables, figures, and citation networks, 2) multimodal scientific hypothesis generation, and 3) automatic verification of the new hypothesis.

4. Diversity Considerations

The methods introduced in this tutorial can help mitigate the language barrier in interdisciplinary science communication. We will cover a broad diversity of methods and applications in different domains. The methods we introduced are mostly domain/language-agnostic. Therefore, they can apply to different domains with various languages. We estimate that only 15-20% of the work will involve one of the four presenters. The papers we discussed in the tutorial are produced by authors from a variety of backgrounds. Our diverse tutorial team represents two universities (UIUC and HUJI) and originates from three geographically distant countries (across China, Israel, and the U.S.). Their seniority varies, ranging from junior/senior Ph.D. students to assistant/full professors, and the team includes a female researcher. Our presenters will promote our tutorial on social media to help diversify our audience participation.

5. Reading List

- Related Tutorials (Jiang and Shang, 2020; Chen et al., 2022; Asai et al., 2023)
- General Guideline (Gil, 2022; Yuan et al., 2022; Birhane et al., 2023; Lund et al., 2023)
- Survey Papers (Li and Ouyang, 2022; Vladika and Matthes, 2023; Hope et al., 2023)
- Scientific IE (Luan et al., 2018; Jain et al., 2020; Shen et al., 2022; Song et al., 2023)
- Scientific IR (Wang et al., 2020c)
- Review Generation (Yuan and Liu, 2022)
- Hypothesis Generation (Krenn et al., 2023)
- Paper Draft Generation(Wang et al., 2021a)

6. Presenters

Qingyun Wang is a Ph.D. student in the Computer Science Department at UIUC. His research lies in controllable knowledge-driven natural language generation, focusing on NLP for scientific discovery. He served as a PC member for multiple conferences including ICML, ACL, ICLR, NeurIPS, etc. He previously entered the finalist of the first Alexa Prize competition. He received the NAACL-HLT 2021 Best Demo Reward. He has experience presenting a tutorial at EMNLP 2021.

Carl Edwards is a Ph.D. student in the Computer Science Department at UIUC. Broadly, he is interested in information extraction, information retrieval, text mining, representation learning, and multimodality. Particularly, he is interested in applying these to the scientific domain to accelerate scientific discovery. His current work focuses on integrating natural language and molecules, especially using multimodal representations.

Heng Ji is a professor at the Computer Science Department of UIUC, and Amazon Scholar. She is a leading expert on multimodal multilingual information extraction. She has coordinated the NIST TAC Knowledge Base Population task since 2010. She has served as the PC Co-Chair of many conferences including NAACL-HLT2018 and AACL-IJCNLP2022 and has presented many tutorials. She is elected as NAACL secretary 2020-2023. Her research interests broadly cover information extraction and NLP for Science, particularly in leveraging NLP for drug discovery.

Tom Hope is an assistant professor at the School of Computer Science and Engineering of HUJI, and a research scientist at AI2. He develops artificial intelligence methods that augment and scale scientific knowledge discovery by harnessing vast repositories of scientific knowledge. His work has received four best paper awards, appeared in top venues, and received coverage from Nature and Science. He was awarded the 2022 Azrieli Early Career Faculty Fellowship, and was a member of the KDD 2020 Best Paper Selection Committee.

7. Other Tutorial Information

All tutorial materials are publicly available at `https://sites.google.com/view/coling2024-paper-lifecycle/`.

8. Ethics Statement

The methods we introduce in the tutorial aim to provide investigative leads for a scientific domain. The final results are not intended to be used without human review. We emphasize that the tools introduced in tutorials are designed to assist human scientists. The identified research directions and the process should be evaluated by trained researchers to ensure ethical outcomes. Because many methods are built on top of pretrained large language models, those systems may exhibit bias due to their pretraining dataset. This tutorial also provides opportunities to discuss the ethical considerations when designing and using those methods and provides a specific section to discuss ethical considerations related to LLMs. Most training sets for these methods are written in English, which might alienate readers historically underrepresented in the NLP domain.

Acknowledgements

This work is supported by the Molecule Maker Lab Institute: an AI research institute program supported by NSF under award No. 2019897, by DOE Center for Advanced Bioenergy and Bioproducts Innovation U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DESC0018420, by U.S. the AI Research Institutes program by National Science Foundation and the Institute of Education Sciences, Department of Education through Award No. 2229873 - AI Institute for Transforming Education for Children with Speech and Language Processing Challenges, and by AI Agriculture: the Agriculture and Food Research Initiative (AFRI) grant no. 2020-67021- 32799/project accession no.1024178 from the USDA National Institute of Food and Agriculture. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of, the National Science Foundation, the U.S. Department of Energy, and the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

9. Bibliographical References

2023. Overcoming the language barrier in science communication. *Nature Reviews Bioengineering*, 1(5):305–305.

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.

Nash Anderson, Daniel L Belavy, Stephen M Perle, Sharief Hendricks, Luiz Hespanhol, Evert Verhagen, and Aamir R Memon. 2023. Ai did not write this manuscript, or did it? can we trick the ai text detector into generated texts? the potential future of chatgpt and ai in sports & exercise medicine manuscript generation.

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Akari Asai, Sewon Min, Zexuan Zhong, and Danqi Chen. 2023. Retrieval-based language models and applications. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 6: Tutorial Abstracts)*, pages 41–46, Toronto, Canada. Association for Computational Linguistics.

Tal August, Katharina Reinecke, and Noah A. Smith. 2022. Generating scientific definitions with controllable complexity. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8298–8317, Dublin, Ireland. Association for Computational Linguistics.

Eliot H Ayache and Conor Omand. 2022. Generating scientific articles with machine learning. *Machine Learning Repository*, arXiv:2203.16569.

Abeba Birhane, Atoosa Kasirzadeh, David Leslie, and Sandra Wachter. 2023. Science in the age of large language models. *Nature Reviews Physics*, 5(5):277–280.

Daniil A Boiko, Robert MacKnight, and Gabe Gomes. 2023. Emergent autonomous scientific research capabilities of large language models. *Chemical Physics Repository*, arXiv:2304.05332.

Andres M Bran, Sam Cox, Andrew D White, and Philippe Schwaller. 2023. Chemcrow: Augmenting large-language models with chemistry tools. *Chemical Physics Repository*, arXiv:2304.05376.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel Weld. 2020. TLDR: Extreme summarization of scientific documents. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4766–4777, Online. Association for Computational Linguistics.

Virginia Casigliani, Francesca De Nard, Erica De Vita, Guglielmo Arzilli, Francesca Maria Grosso, Filippo Quattrone, Lara Tavoschi, and Pierluigi Lopalco. 2020. Too much information, too little evidence: is waste in research fuelling the covid-19 infodemic? *BMJ*, 370.

Arie Cattan, Sophie Johnson, Daniel S Weld, Ido Dagan, Iz Beltagy, Doug Downey, and Tom Hope. 2021. Scico: Hierarchical cross-document coreference for scientific concepts. In *3rd Conference on Automated Knowledge Base Construction*.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Hong Chen, Hiroya Takamura, and Hideki Nakayama. 2021. SciXGen: A scientific paper dataset for context-aware text generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1483–1492, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Muhao Chen, Lifu Huang, Manling Li, Ben Zhou, Heng Ji, and Dan Roth. 2022. New frontiers of information extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, pages 14–25, Seattle, United States. Association for Computational Linguistics.

Dimitrios Christofidellis, Giorgio Giannone, Jannis Born, Ole Winther, Teodoro Laino, and Matteo Manica. 2023. Unifying molecular and textual representations via multi-task language modelling. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 6140–6157. PMLR.

Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Dan Weld. 2019. Pretrained language models for sequential sentence classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3693–3699, Hong Kong, China. Association for Computational Linguistics.

Gemma Conroy. 2023a. How chatgpt and other ai tools could disrupt scientific publishing. *Nature*, 622(7982):234–236.

Gemma Conroy. 2023b. Scientific sleuths spot dishonest chatgpt use in papers. *Nature*.

Gemma Conroy. 2023c. Scientists used chatgpt to generate an entire paper from scratch—but is it any good? *Nature*, 619(7970):443–444.

Evan N. Crothers, Nathalie Japkowicz, and Herna L. Viktor. 2023. Machine-generated text: A comprehensive survey of threat models and detection methods. *IEEE Access*, 11:70977–71002.

Tamara Czinczoll, Helen Yannakoudakis, Pushkar Mishra, and Ekaterina Shutova. 2022. Scientific and creative analogies in pretrained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2094–2100, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Rumen Dangovski, Michelle Shen, Dawson Byrd, Li Jing, Desislava Tsvetkova, Preslav Nakov, and Marin Soljačić. 2021. We can explain your research in layman's terms: Towards automating science journalism at scale. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12728–12737.

Luigi De Angelis, Francesco Baglivo, Guglielmo Arzilli, Gaetano Pierpaolo Privitera, Paolo Ferragina, Alberto Eugenio Tozzi, and Caterina Rizzo. 2023. Chatgpt and the rise of large language models: the new ai-driven infodemic threat in public health. *Frontiers in Public Health*, 11:1166120.

Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. 2022. Translation between molecules and natural language. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 375–413, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Mehwish Fatima and Michael Strube. 2023. Cross-lingual science journalism: Select, simplify and rewrite summaries for non-expert readers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1843–1861, Toronto, Canada. Association for Computational Linguistics.

Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti Hearst, Andrew Head, and Daniel S Weld. 2023. Scim: Intelligent skimming support for scientific papers. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, page 476–490, New York, NY, USA. Association for Computing Machinery.

Jacob G. Foster, Andrey Rzhetsky, and James A. Evans. 2015. Tradition and innovation in scientists' research strategies. *American Sociological Review*.

Giacomo Frisoni, Miki Mizutani, Gianluca Moro, and Lorenzo Valgimigli. 2022. BioReader: a retrieval-enhanced text-to-text transformer for biomedical literature. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5770–5793, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Catherine A Gao, Frederick M Howard, Nikolay S Markov, Emma C Dyer, Siddhi Ramesh, Yuan Luo, and Alexander T Pearson. 2023. Comparing scientific abstracts generated by chatgpt to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine*, 6(1):75.

Yubin Ge, Ly Dinh, Xiaofeng Liu, Jinsong Su, Ziyao Lu, Ante Wang, and Jana Diesner. 2021. BACO: A background knowledge- and content-based framework for citing sentence generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1466–1478, Online. Association for Computational Linguistics.

Yolanda Gil. 2022. Will ai write scientific papers in the future? *AI Magazine*, 42(4):3–15.

Paul P Glasziou, Sharon Sanders, and Tammy Hoffmann. 2020. Waste in covid-19 research. *BMJ*, 369.

Max Glockner, Yufang Hou, and Iryna Gurevych. 2022. Missing counter-evidence renders NLP fact-checking unrealistic for misinformation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5916–5936, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tomas Goldsack, Zhihao Zhang, Chenghua Lin, and Carolina Scarton. 2022. Making science simple: Corpora for the lay summarisation of scientific literature. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10589–10604, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Chi Han, Qifan Wang, Wenhan Xiong, Yu Chen, Heng Ji, and Sinong Wang. 2023. Lm-infinite: Simple on-the-fly length generalization for large language models. *Machine Learning Repository*, arXiv:2308.16137.

Paul KJ Han, Brian J Zikmund-Fisher, Christine W Duarte, Megan Knaus, Adam Black, Aaron M Scherer, and Angela Fagerlin. 2018. Communication of scientific uncertainty about a novel pandemic health threat: ambiguity aversion and its mechanisms. *Journal of health communication*, 23(5):435–444.

Sam Henry and Bridget T McInnes. 2017. Literature based discovery: models, methods, and trends. *Journal of biomedical informatics*, 74:20–32.

Zhi Hong, Aswathy Ajith, James Pauloski, Eamon Duede, Kyle Chard, and Ian Foster. 2023. The diminishing returns of masked language models to science. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1270–1283, Toronto, Canada. Association for Computational Linguistics.

Pollawat Hongwimol, Peeranuth Kehasukcharoen, Pasit Laohawarutchai, Piyawat Lertvittayakumjorn, Aik Beng Ng, Zhangsheng Lai, Timothy Liu, and Peerapon Vateekul. 2021. ESRA: Explainable scientific research assistant. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 114–121, Online. Association for Computational Linguistics.

Tom Hope, Doug Downey, Daniel S Weld, Oren Etzioni, and Eric Horvitz. 2023. A computational inflection for scientific discovery. *Communications of the ACM*, 66(8):62–73.

Tom Hope, Jason Portenoy, Kishore Vasan, Jonathan Borchardt, Eric Horvitz, Daniel Weld, Marti Hearst, and Jevin West. 2020. SciSight: Combining faceted navigation and research group detection for COVID-19 exploratory scientific search. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 135–143, Online. Association for Computational Linguistics.

Mohammad Hosseini, Lisa M Rasmussen, and David B Resnik. 2023. Using ai to write scholarly publications.

Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, and Debasis Ganguly. 2019. Identification of tasks, datasets, evaluation metrics, and numeric scores for scientific leaderboards construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5203–5213, Florence, Italy. Association for Computational Linguistics.

Ting-Yao Hsu, C Lee Giles, and Ting-Hao Huang. 2021. SciCap: Generating captions for scientific figures. In *Findings of the Association for*

Computational Linguistics: EMNLP 2021, pages 3258–3264, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yue Hu and Xiaojun Wan. 2014. Automatic generation of related work sections in scientific papers: An optimization approach. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1624–1633, Doha, Qatar. Association for Computational Linguistics.

Sarthak Jain, Madeleine van Zuylen, Hannaneh Hajishirzi, and Iz Beltagy. 2020. SciREX: A challenge dataset for document-level information extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7506–7516, Online. Association for Computational Linguistics.

Meng Jiang and Jingbo Shang. 2020. Scientific text mining and knowledge graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 3537–3538, New York, NY, USA. Association for Computing Machinery.

Prerna Juneja and Tanushree Mitra. 2022. Human and technological infrastructures of fact-checking. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2).

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6:391–406.

Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz. 2018. A dataset of peer reviews (PeerRead): Collection, insights and NLP applications. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1647–1661, New Orleans, Louisiana. Association for Computational Linguistics.

John Kirchenbauer, Jonas Geiping, Yuxin Wen, Jonathan Katz, Ian Miers, and Tom Goldstein. 2023. A watermark for large language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 17061–17084. PMLR.

Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. 2019. Text Generation from Knowledge Graphs with Graph Transformers. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2284–2293, Minneapolis, Minnesota. Association for Computational Linguistics.

Mario Krenn, Lorenzo Buffoni, Bruno Coutinho, Sagi Eppel, Jacob Gates Foster, Andrew Gritsevskiy, Harlin Lee, Yichao Lu, João P. Moutinho, Nima Sanjabi, Rishi Sonthalia, Ngoc Mai Tran, Francisco Valente, Yangxinyu Xie, Rose Yu, and Michael Kopp. 2023. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nature Machine Intelligence*.

Varun Kumar, Leonard Gleyzer, Adar Kahana, Khemraj Shukla, and George Em Karniadakis. 2023. Mycrunchgpt: A chatgpt assisted framework for scientific machine learning. *Machine Learning Repository*, arXiv:2306.15551.

Esther Landhuis. 2016. Scientific literature: Information overload. *Nature*, 535(7612):457–458.

Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery.

Xiangci Li, Gully Burns, and Nanyun Peng. 2021. Scientific discourse tagging for evidence extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2550–2562, Online. Association for Computational Linguistics.

Xiangci Li and Jessica Ouyang. 2022. Automatic related work generation: A meta study. *Computation and Language Repository*, arXiv:2201.01880.

Ryan Liu and Nihar B Shah. 2023. Reviewergpt? an exploratory study on using large language models for paper reviewing. *Computation and Language Repository*, arXiv:2306.00622.

Shengchao Liu, Jiongxiao Wang, Yijin Yang, Chengpeng Wang, Ling Liu, Hongyu Guo, and Chaowei Xiao. 2023. Chatgpt-powered conversational drug editing using retrieval and domain feedback. *Chemical Physics Repository*, arXiv:2305.18090.

Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld. 2020. S2ORC: The semantic scholar open research corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4969–4983, Online. Association for Computational Linguistics.

Yao Lu, Yue Dong, and Laurent Charlin. 2020. Multi-XScience: A large-scale dataset for extreme multi-document summarization of scientific articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8068–8074, Online. Association for Computational Linguistics.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3219–3232, Brussels, Belgium. Association for Computational Linguistics.

Li Lucy, Jesse Dodge, David Bamman, and Katherine Keith. 2023. Words as gatekeepers: Measuring discipline-specific terms and meanings in scholarly publications. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6929–6947, Toronto, Canada. Association for Computational Linguistics.

Brady D Lund, Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. Chatgpt and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5):570–581.

Kelvin Luu, Xinyi Wu, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, and Noah A. Smith. 2021. Explaining relationships between scientific documents. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2130–2144, Online. Association for Computational Linguistics.

Yongqiang Ma, Jiawei Liu, Fan Yi, Qikai Cheng, Yong Huang, Wei Lu, and Xiaozhong Liu. 2023. Ai vs. human–differentiation analysis of scientific content generation. *Computation and Language Repository*, arXiv:2301.10416.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.

Taylor Nelson, Nicole Kagan, Claire Critchlow, Alan Hillard, and Albert Hsu. 2020. The danger of misinformation in the covid-19 crisis. *Missouri Medicine*, 117(6):510.

Linda Nordling. 2023. How chatgpt is transforming the postdoc experience. *Nature*, 622:655 – 657.

OpenAI. 2023. Gpt-4 technical report. *ArXiv*, abs/2303.08774.

Gihan Panapitiya, Fred Parks, Jonathan Sepulveda, and Emily Saldanha. 2021. Extracting material property measurement data from scientific articles. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5393–5402, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Joo-Young Park. 2023. Could chatgpt help you to write your next scientific paper?: concerns on research ethics related to usage of artificial intelligence tools. *Journal of the Korean Association of Oral and Maxillofacial Surgeons*, 49(3):105–106.

Juan Manuel Parrilla. 2023. Chatgpt use shows that the grant-application system is broken. *Nature*.

Long N Phan, James T Anibal, Hieu Tran, Shaurya Chanana, Erol Bahadroglu, Alec Peltekian, and Grégoire Altan-Bonnet. 2021. Scifive: a text-to-text transformer model for biomedical literature. *Computation and Language Repository*, arXiv:2106.03598.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.

Anna Rogers, Marzena Karpinska, Jordan Boyd-Graber, and Naoaki Okazaki. 2023. Program chairs' report on peer review at acl 2023. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages xl–lxxv, Toronto, Canada. Association for Computational Linguistics.

Frank Rosenblatt. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Mobashir Sadat and Cornelia Caragea. 2022. SciNLI: A corpus for natural language inference on scientific text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7399–7409, Dublin, Ireland. Association for Computational Linguistics.

Mourad Sarrouti, Asma Ben Abacha, Yassine Mrabet, and Dina Demner-Fushman. 2021. Evidence-based fact-checking of health-related claims. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3499–3512, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *Computation and Language Repository*, arXiv:2211.05100.

Zejiang Shen, Kyle Lo, Lucy Lu Wang, Bailey Kuehl, Daniel S. Weld, and Doug Downey. 2022. VILA: Improving structured content extraction from scientific PDFs using visual layout groups. *Transactions of the Association for Computational Linguistics*, 10:376–392.

Yu Song, Santiago Miret, and Bang Liu. 2023. MatSci-NLP: Evaluating scientific language models on materials science language tasks using text-to-schema modeling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3621–3639, Toronto, Canada. Association for Computational Linguistics.

C Stokel-Walker. 2022. Ai bot chatgpt writes smart essays—should professors worry?[published online ahead of print december 9, 2022]. *Nature News*.

Chris Stokel-Walker. 2023. Chatgpt listed as author on research papers: many scientists disapprove. *Nature*, 613(7945):620–621.

Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. Galactica: A large language model for science. *Computation and Language Repository*, arXiv:2211.09085.

Ana Sabina Uban, Cornelia Caragea, and Liviu P. Dinu. 2021. Studying the evolution of scientific topics and their relationships. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1908–1922, Online. Association for Computational Linguistics.

Richard Van Noorden. 2014. Scientists may be reaching a peak in reading habits. *Nature*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems (Neruips)*. Curran Associates, Inc.

Vijay Viswanathan, Luyu Gao, Tongshuang Wu, Pengfei Liu, and Graham Neubig. 2023. DataFinder: Scientific dataset recommendation from natural language descriptions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10288–10303, Toronto, Canada. Association for Computational Linguistics.

Vijay Viswanathan, Graham Neubig, and Pengfei Liu. 2021. CitationIE: Leveraging the citation graph for scientific information extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 719–731, Online. Association for Computational Linguistics.

Juraj Vladika and Florian Matthes. 2023. Scientific fact-checking: A survey of resources and approaches. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6215–6230, Toronto, Canada. Association for Computational Linguistics.

David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. 2020. Fact or fiction: Verifying scientific claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7534–7550, Online. Association for Computational Linguistics.

David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Iz Beltagy, Lucy Lu Wang, and Hannaneh Hajishirzi. 2022. SciFact-open: Towards open-domain scientific claim verification. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4719–4734, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Gengyu Wang, Kate Harwood, Lawrence Chillrud, Amith Ananthram, Melanie Subbiah, and Kathleen McKeown. 2023. Check-COVID: Fact-checking COVID-19 news claims with scientific evidence. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 14114–14127, Toronto, Canada. Association for Computational Linguistics.

Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Michael Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey A. Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Nancy Xin Ru Wang, Christopher Wilhelm, Boya Xie, Douglas M. Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. 2020a. CORD-19: The COVID-19 open research dataset. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online. Association for Computational Linguistics.

Qingqin Wang, Yun Xiong, Yao Zhang, Jiawei Zhang, and Yangyong Zhu. 2021a. Autocite: Multi-modal representation fusion for contextual citation generation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, WSDM '21, page 788–796, New York, NY, USA. Association for Computing Machinery.

Qingyun Wang, Lifu Huang, Zhiying Jiang, Kevin Knight, Heng Ji, Mohit Bansal, and Yi Luan. 2019. PaperRobot: Incremental draft generation of scientific ideas. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1980–1991, Florence, Italy. Association for Computational Linguistics.

Qingyun Wang, Manling Li, Xuan Wang, Nikolaus Parulian, Guangxing Han, Jiawei Ma, Jingxuan Tu, Ying Lin, Ranran Haoran Zhang, Weili Liu, Aabhas Chauhan, Yingjun Guan, Bangzheng Li, Ruisong Li, Xiangchen Song, Yi Fung, Heng Ji, Jiawei Han, Shih-Fu Chang, James Pustejovsky, Jasmine Rah, David Liem, Ahmed ELsayed, Martha Palmer, Clare Voss, Cynthia Schneider, and Boyan Onyshkevych. 2021b. COVID-19 literature knowledge graph construction and drug repurposing report generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 66–77, Online. Association for Computational Linguistics.

Qingyun Wang, Qi Zeng, Lifu Huang, Kevin Knight, Heng Ji, and Nazneen Fatema Rajani. 2020b. ReviewRobot: Explainable paper review generation based on knowledge synthesis. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 384–397, Dublin, Ireland. Association for Computational Linguistics.

Xuan Wang, Yingjun Guan, Weili Liu, Aabhas Chauhan, Enyi Jiang, Qi Li, David Liem, Dibakar Sigdel, John Caufield, Peipei Ping, and Jiawei Han. 2020c. EVIDENCEMINER: Textual evidence discovery for life sciences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 56–62, Online. Association for Computational Linguistics.

Rick P. Wierenga, Stefan M. Golas, Wilson Ho, Connor W. Coley, and Kevin M. Esvelt. 2023. Pylabrobot: An open-source, hardware-agnostic interface for liquid-handling robots and accessories. *Device*, 1(4):100111.

Dustin Wright, David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Isabelle Augenstein, and Lucy Lu Wang. 2022. Generating scientific claims for zero-shot scientific fact checking. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2448–2460, Dublin, Ireland. Association for Computational Linguistics.

Xinyu Xing, Xiaosheng Fan, and Xiaojun Wan. 2020. Automatic generation of citation texts in scholarly papers: A pilot study. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6181–6190, Online. Association for Computational Linguistics.

Michihiro Yasunaga, Jungo Kasai, Rui Zhang, Alexander R Fabbri, Irene Li, Dan Friedman, and Dragomir R Radev. 2019. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7386–7393.

Michihiro Yasunaga, Jure Leskovec, and Percy Liang. 2022. LinkBERT: Pretraining language models with document links. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8003–8016, Dublin, Ireland. Association for Computational Linguistics.

Yue Yu, Chenyan Xiong, Si Sun, Chao Zhang, and Arnold Overwijk. 2022. COCO-DR: Combating distribution shift in zero-shot dense retrieval with contrastive and distributionally robust learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1462–1479, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Weizhe Yuan and Pengfei Liu. 2022. Kid-review: Knowledge-guided scientific review generation with oracle pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11639–11647.

Weizhe Yuan, Pengfei Liu, and Graham Neubig. 2022. Can we automate scientific reviewing? *Journal of Artificial Intelligence Research*, 75:171–212.

Qi Zeng, Mankeerat Sidhu, Hou Pong Chan, Lu Wang, and Heng Ji. 2023. Meta-review generation with checklist-guided iterative introspection. *Computation and Language Repository*, arXiv:2305.14647.

Rui Zhang, Dimitar Hristovski, Dalton Schutte, Andrej Kastrin, Marcelo Fiszman, and Halil Kilicoglu. 2021a. Drug repurposing for covid-19 via knowledge graph completion. *Journal of Biomedical Informatics*, 115:103696.

Zhiwei Zhang, Jiyi Li, Fumiyo Fukumoto, and Yanming Ye. 2021b. Abstract, rationale, stance: A joint model for scientific claim verification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3580–3586, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ruiqi Zhong, Peter Zhang, Steve Li, Jinwoo Ahn, Dan Klein, and Jacob Steinhardt. 2023. Goal driven discovery of distributional differences via language descriptions. *Computation and Language Repository*, arXiv:2302.14233.

Chenguang Zhu, Yichong Xu, Xiang Ren, Bill Yuchen Lin, Meng Jiang, and Wenhao Yu. 2022. Knowledge-augmented methods for natural language processing. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 12–20, Dublin, Ireland. Association for Computational Linguistics.