



# SciMRC: Multi-perspective Scientific Machine Reading

---

LREC-COLING 2024

Xiao Zhang<sup>1</sup>, Heqi Zheng<sup>1</sup>, Yuxiang Nie<sup>1</sup>,  
Heyan Huang, Xian-Ling Mao



# Background

---

## \*Scientific Machine Reading Comprehension(SMRC)

\*Crucial for scientific research, education and industry

\*The amount of scientific literature has been soaring

\*Elevating the efficiency of machine scientific understanding becomes more essential



# Challenges

## \* Existing Research

- \* Has contributed significantly to this field
- \* Enhancing extracting information from scientific papers in response to questions
- \* Disregarded a **pivotal element**:
- \* **The inherent variation in comprehension levels** among readers then digesting the same text.



# Proposed Methods

## \* Multi-perspective Scientific Machine Reading Comprehension Dataset (SciMRC)

### \* Multi-perspective

\* Beginner Perspective

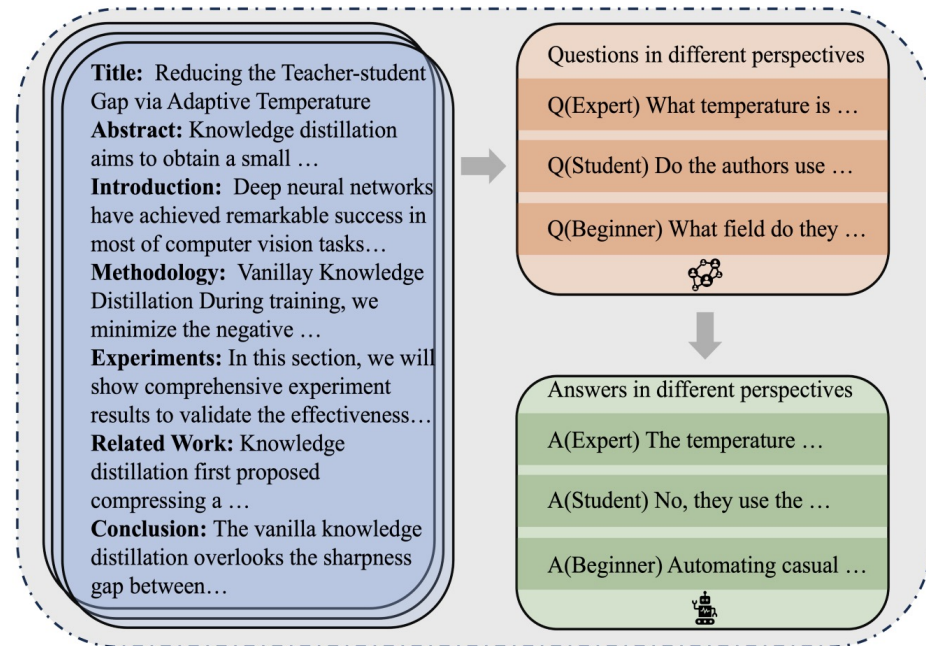
\* Student Perspective

\* Expert Perspective

### \* Collections

\* 741 Scientific Papers

\* 6,057 QA pairs



## \* Data Preparation

### \* Multiple Sources

- \* S2ORC, QASPER, Open-review

- \* Over 3000 papers

### \* Question-related survey

- \* Questionnaire for experts, students and beginners

  - \* Top 5 concerns questions when engaging in scientific papers

- \* Summarize 28 categories

### \* Answer Type

- \* Extractive, Generative, Yes | No, Unanswerable

- \* Should no supporting evidence be found within the paper, the answer type is categorized as Unanswerable.



## \* Multi-perspective Annotation Strategy

### \* Beginner's Perspective

- \* Annotators without domain-specific knowledge
- \* Craft questions and seek answers within the paper's full text, figures, and tables

### \* Student's Perspective

### \* Expert's Perspective



## \* Multi-perspective Annotation Strategy

- \* Beginner's Perspective

- \* Student's Perspective

- \* Leverage FacetSum dataset to get summarization of four key aspects in papers: purpose, method, finding and value.

- \* Annotators verify the accuracy of the generated abstracts and formulate questions based on them, subsequently searching for answers within the paper

- \* Expert's Perspective



## \* Multi-perspective Annotation Strategy

- \* Beginner's Perspective

- \* Student's Perspective

- \* Expert's Perspective

- \* Acquire related reviews from open-review.

- \* Annotators extract questions and answers from reviewer comments and author responses.





# Analysis of SciMRC

Type	Paper	Figure/Table	Question		Answer				Evidence
	Avg Paper Length	Avg Figure/Table Number	Avg Question Length	Avg Answer Length	Yes No	Generative	Extractive	Unanswerable	Avg Evidence Sentence Number
PERSPECTIVE									
BEGINNERS			10.0	17.2	331	754	2220	1	1.39
STUDENTS			9.8	11.7	266	340	1194	0	1.08
EXPERTS	3725.6	5.32	22.4	95.9	5	467	8	471	4.56
ALL			11.0	21.8	602	1561	3422	472	1.56

Table 2: Representative features from SciMRC categorized by different perspectives.

1	Methods	2	Experimental results	3	Dataset/code	4	Experimental settings
5	Model architecture	6	Experimental analysis	7	Related work	8	Background
9	Baseline	10	Motivation	11	Contribution	12	Innovation point
13	Future work	14	Research field	15	Research background	16	Evaluation Metric
17	Insufficiency of previous work	18	Experimental conditions	19	Assumption	20	Model transferability
21	Limitations	22	Innovation	23	Thesis Outline	24	Landing Application
25	Case Study	26	Publication Details	27	Complexity	28	Whether to propose a new task

Table 3: All the 28 question categories in SciMRC



# Analysis of SciMRC

1	Methods	2	Experimental results	3	Dataset/code	4	Experimental settings
5	Model architecture	6	Experimental analysis	7	Related work	8	Background
9	Baseline	10	Motivation	11	Contribution	12	Innovation point
13	Future work	14	Research field	15	Research background	16	Evaluation Metric
17	Insufficiency of previous work	18	Experimental conditions	19	Assumption	20	Model transferability
21	Limitations	22	Innovation	23	Thesis Outline	24	Landing Application
25	Case Study	26	Publication Details	27	Complexity	28	Whether to propose a new task

Table 3: All the 28 question categories in SciMRC



Figure 2: The distribution of question categories



# Experiments

## \* Modeling

- \* Employ T5 and LED as backbone models

## \* Evaluation

- \* Accuracy for Yes|No
- \* Rouge-L for the remains

## \* Settings

- \* B-formed: The model is trained using the beginner's perspective training data.
- \* S-formed: The model is trained using the student's perspective training data.
- \* E-formed: The model is trained using the expert's perspective training data.
- \* BS-formed: The model is trained using both beginner's and student's perspective training data.
- \* BE-formed: The model is trained using both beginner's and expert's perspective training data
- \* SE-formed: The model is trained using both student's and expert's perspective training data.
- \* BSE-formed: The model is trained using all training data.
- \* Per-formed: This setting involves the use of BSE-formed model and E-formed model in known question perspectives settings. The BSE-formed model is used for evaluating the expert perspective, while the BSE-formed model is used for evaluating student and beginner p



# Experiment

## \* Experimental Results

\* Utilize perspective's prior knowledge would be effective

Model	Dev				Test			
	Beginner	Student	Expert	Overall	Beginner	Student	Expert	Overall
T5-B	24.77 $\pm$ 0.27	39.50 $\pm$ 0.72	11.29 $\pm$ 0.45	26.47 $\pm$ 0.23	25.60 $\pm$ 0.34	41.51 $\pm$ 0.44	11.39 $\pm$ 0.20	29.03 $\pm$ 0.13
T5-S	24.32 $\pm$ 0.11	45.08 $\pm$ 0.98	10.64 $\pm$ 0.28	27.80 $\pm$ 0.40	24.68 $\pm$ 0.31	44.16 $\pm$ 0.59	9.97 $\pm$ 0.13	29.16 $\pm$ 0.33
T5-E	9.74 $\pm$ 0.16	17.03 $\pm$ 0.57	<b>14.76</b> $\pm$ 0.20	13.11 $\pm$ 0.18	11.12 $\pm$ 0.21	18.19 $\pm$ 0.66	<b>15.66</b> $\pm$ 0.19	13.79 $\pm$ 0.34
T5-SE	23.50 $\pm$ 0.33	45.17 $\pm$ 0.28	13.34 $\pm$ 0.07	28.09 $\pm$ 0.08	24.26 $\pm$ 0.29	43.34 $\pm$ 0.84	13.97 $\pm$ 0.43	29.12 $\pm$ 0.21
T5-BE	25.18 $\pm$ 0.49	39.65 $\pm$ 0.47	13.64 $\pm$ 0.71	27.21 $\pm$ 0.46	25.94 $\pm$ 0.02	41.50 $\pm$ 0.82	14.68 $\pm$ 0.61	29.61 $\pm$ 0.30
T5-SB	26.06 $\pm$ 0.54	45.93 $\pm$ 0.87	11.24 $\pm$ 0.16	29.12 $\pm$ 0.41	<b>26.46</b> $\pm$ 0.15	<b>46.27</b> $\pm$ 0.22	10.92 $\pm$ 0.09	30.99 $\pm$ 0.21
T5-BSE	<b>26.87</b> $\pm$ 0.89	<b>46.86</b> $\pm$ 0.72	13.60 $\pm$ 0.24	30.25 $\pm$ 0.70	26.34 $\pm$ 0.49	45.28 $\pm$ 0.24	14.53 $\pm$ 0.28	31.05 $\pm$ 0.23
Per-T5	<b>26.87</b> $\pm$ 0.89	<b>46.86</b> $\pm$ 0.72	<b>14.76</b> $\pm$ 0.20	<b>30.45</b> $\pm$ 0.60	26.34 $\pm$ 0.49	45.28 $\pm$ 0.24	<b>15.66</b> $\pm$ 0.19	<b>31.18</b> $\pm$ 0.23
LED-B	25.51 $\pm$ 1.61	32.1 $\pm$ 2.06	10.02 $\pm$ 0.29	24.25 $\pm$ 1.33	25.02 $\pm$ 1.21	33.15 $\pm$ 0.67	11.36 $\pm$ 0.23	26.15 $\pm$ 0.93
LED-S	22.55 $\pm$ 2.26	41.93 $\pm$ 6.73	9.16 $\pm$ 0.5	25.77 $\pm$ 3.11	21.64 $\pm$ 1.95	44.05 $\pm$ 5.12	9.69 $\pm$ 0.47	27.47 $\pm$ 2.72
LED-E	6.80 $\pm$ 0.49	7.75 $\pm$ 0.47	14.62 $\pm$ 0.32	8.76 $\pm$ 0.45	7.22 $\pm$ 0.53	7.98 $\pm$ 0.75	<b>14.77</b> $\pm$ 0.11	8.27 $\pm$ 0.53
LED-SE	24.57 $\pm$ 0.35	45.16 $\pm$ 1.03	13.29 $\pm$ 0.34	28.55 $\pm$ 0.48	23.79 $\pm$ 0.27	47.42 $\pm$ 1.23	12.52 $\pm$ 0.48	30.00 $\pm$ 0.14
LED-BE	27.40 $\pm$ 0.94	32.62 $\pm$ 1.15	13.02 $\pm$ 0.69	25.89 $\pm$ 0.64	25.80 $\pm$ 0.95	33.81 $\pm$ 1.03	12.68 $\pm$ 0.15	26.90 $\pm$ 0.88
LED-SB	31.07 $\pm$ 0.07	<b>49.67</b> $\pm$ 0.27	9.07 $\pm$ 0.3	32.17 $\pm$ 0.25	29.78 $\pm$ 0.4	<b>47.93</b> $\pm$ 0.42	10.41 $\pm$ 0.39	33.37 $\pm$ 0.41
LED-BSE	<b>32.24</b> $\pm$ 0.43	47.04 $\pm$ 2.38	12.85 $\pm$ 0.75	32.04 $\pm$ 0.87	<b>29.80</b> $\pm$ 0.55	46.39 $\pm$ 0.34	12.54 $\pm$ 0.08	33.20 $\pm$ 0.36
Per-LED	<b>32.24</b> $\pm$ 0.43	47.04 $\pm$ 2.38	<b>14.62</b> $\pm$ 0.32	<b>32.47</b> $\pm$ 0.69	<b>29.80</b> $\pm$ 0.55	46.39 $\pm$ 0.34	<b>14.77</b> $\pm$ 0.11	<b>33.44</b> $\pm$ 0.34



# Experiment

## \* Experimental Results & Analysis

\* Model's performance shows different levels in different answer types

Model	Dev				Test			
	Yes No	Generative	Extractive	Unanswerable	Yes No	Generative	Extractive	Unanswerable
T5-BSE	81.61 $\pm$ 1.63	25.67 $\pm$ 0.12	30.27 $\pm$ 0.84	13.04 $\pm$ 0.33	<b>83.61</b> $\pm$ 0.67	23.16 $\pm$ 0.11	29.58 $\pm$ 0.24	14.86 $\pm$ 0.30
Per-T5	81.61 $\pm$ 1.63	26.14 $\pm$ 0.30	30.38 $\pm$ 0.93	14.23 $\pm$ 0.22	<b>83.61</b> $\pm$ 0.67	23.39 $\pm$ 0.07	29.61 $\pm$ 0.21	<b>16.11</b> $\pm$ 0.03
LED-BSE	<b>82.18</b> $\pm$ 0.82	25.56 $\pm$ 0.81	34.78 $\pm$ 1.43	12.79 $\pm$ 0.57	78.69 $\pm$ 1.34	24.81 $\pm$ 0.36	33.54 $\pm$ 0.36	13.17 $\pm$ 0.29
Per-LED	<b>82.18</b> $\pm$ 0.82	<b>26.32</b> $\pm$ 0.79	<b>34.83</b> $\pm$ 1.44	<b>14.53</b> $\pm$ 0.11	78.69 $\pm$ 1.34	<b>25.66</b> $\pm$ 0.40	<b>33.57</b> $\pm$ 0.33	13.99 $\pm$ 0.23

\* Comparison of experimental results for various models on test set of SciMRC

\* SciMRC could pose challenges that cannot be directly addressed by GPT models.

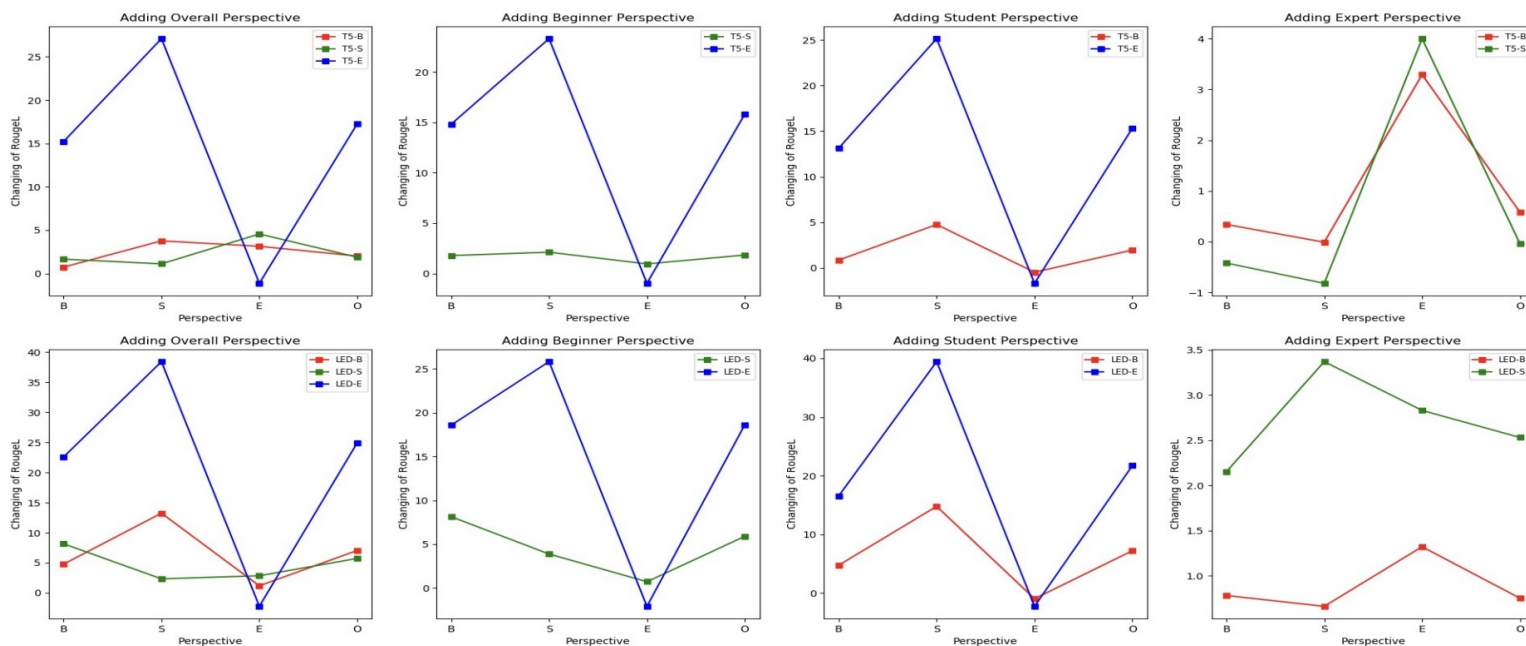
Model	Beginner	Student	Expert	Overall
Per-T5	26.34	45.28	<b>15.66</b>	31.66
Per-LED	<b>29.80</b>	<b>46.39</b>	14.77	<b>33.44</b>
GPT3.5-turbo	21.25	26.53	13.47	22.10



# Experiment

## ❄️ Interplay Among Perspectives

- ❄️ Based on the data of either perspective of S, B and E,
- ❄️ We further add other perspective data,
- ❄️ Then observe the change of Rouge-L scores on test set to quantity the influence of performance.



# Conclusion

## \* Conclusion

- \* We introduced SciMRC, a novel multi-perspective scientific machine reading comprehension dataset.
- \* SciMRC incorporates diverse perspectives of readers, including beginners, students, and experts.
- \* Our extensive experimental results highlight the inherent relationships and differences among these perspectives, underscoring the importance of perspective analysis in scientific machine reading comprehension.





# Thanks

---



# Thanks

