PersLEARN : Research Training through the Lens of Perspective Cultivation

Yu-Zhe Shi[1 ⋆ ✉], Shiqian Li[1,2 ⋆], Xinyi Niu[1,3 ⋆], Qiao Xu[1,3 ⋆], Jiawen Liu[1,4 ⋆], Yifan Xu[1,5 ⋆],
Shiyu Gu[1], Bingru He[1,6], Xinyang Li[1,6], Xinyu Zhao[1,6], Zijian Zhao[1,6], Yidong Lyu[1,7], Zhen Li[1,8],
Sijia Liu[1,9], Lin Qiu[1,10], Jinhao Ji[1], Lecheng Ruan[1,11 ✉], Yuxi Ma[11], Wenjuan Han[12 ✉], Yixin Zhu[2 ✉]

[1]PersLab Research [2]Peking University [3]Huazhong University of Science and Technology [4]Tongji University [5]Kyushu University
[6]Tsinghua University [7]Chinese University of Hong Kong [8]University of Chinese Academy of Sciences [9]University of the Arts London
[10]University of Hong Kong [11]National Key Laboratory of General Artificial Intelligence, BIGAI [12]Beijing Jiaotong University

⋆Equal contributors ✉ syz@perslab.co, ruanlecheng@bigai.ai, wjhan@bjtu.edu.cn, yixin.zhu@pku.edu.cn

Abstract

Scientific research is inherently shaped by its authors' perspectives, influenced by various factors such as their personality, community, or society. Junior researchers often face challenges in identifying the perspectives reflected in the existing literature and struggle to develop their own viewpoints. In response to this issue, we introduce PersLEARN , a tool designed to facilitate the cultivation of scientific perspectives, starting from a basic seed idea and progressing to a well-articulated framework. By interacting with a prompt-based model, researchers can develop their perspectives **explicitly**. Our human study reveals that scientific perspectives developed by students using PersLEARN exhibit a superior level of logical coherence and depth compared to those that did not. Furthermore, our pipeline outperforms baseline approaches across multiple domains of literature from various perspectives. These results suggest that PersLEARN could help foster a greater appreciation of diversity in scientific perspectives as an essential component of research training. [1]

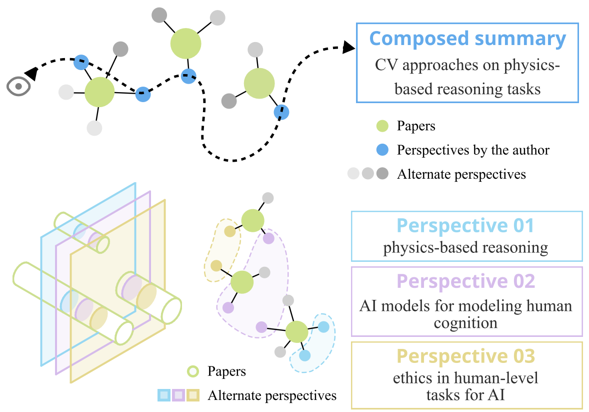


Figure 1: **Composed summaries *vs*. framed perspectives.** Composed summaries are subject to the authors' perspectives, whereas the perspective frames are directed by new ideas.

1 Introduction

The pursuit of science is driven by a desire to gain a deeper understanding of the natural world, not only through the collection of objective facts but also through interpreting those facts (Kuhn, 1970; Longino, 1990). As a result, scientific knowledge is shaped by a complex interplay of various factors that extend beyond the objective world. These factors include the personal characteristics of individual scientists (Heisenberg, 1958; Bybee, 2006), shared mindsets within scientific communities (Cetina, 1999), and broader societal contexts such as cultural and political influences (Latour and Woolgar, 1986; Latour, 1987; Lynch, 1993; Latour et al., 1999). Together, these factors contribute to forming perspectives regarding how best to interpret the natural world. Perspectives are essential to effectively process and communicate scientific knowledge with limited cognitive resources (Lewis et al., 2014; Griffiths et al., 2015; Gershman et al., 2015; Lieder and Griffiths, 2020).

However, junior researchers often face difficulties in developing their own scientific perspectives. They may struggle to identify the perspectives reflected in the existing literature and consequently struggle to develop and articulate their own viewpoints. This presents a significant obstacle to the progress of research training and deprives junior researchers of the opportunity to embrace the broader range of diverse perspectives that could contribute to their understanding of a particular topic (Duschl and Grandy, 2008). The challenge of developing scientific perspectives is particularly evident in one of the most significant research training approaches—writing literature reviews. In our pilot study, we asked students studying at the intersection of Artificial Intelligence (AI) and Cognitive Reasoning (CoRe) to write a review article from the perspective of "physics-based reasoning in Computer Vision (CV)" using a set of papers published on CV conferences. The assigned task aims to provide students with a multifaceted

[1]Website: https://perslearn.com/. Video: https://vimeo.com/802213150.

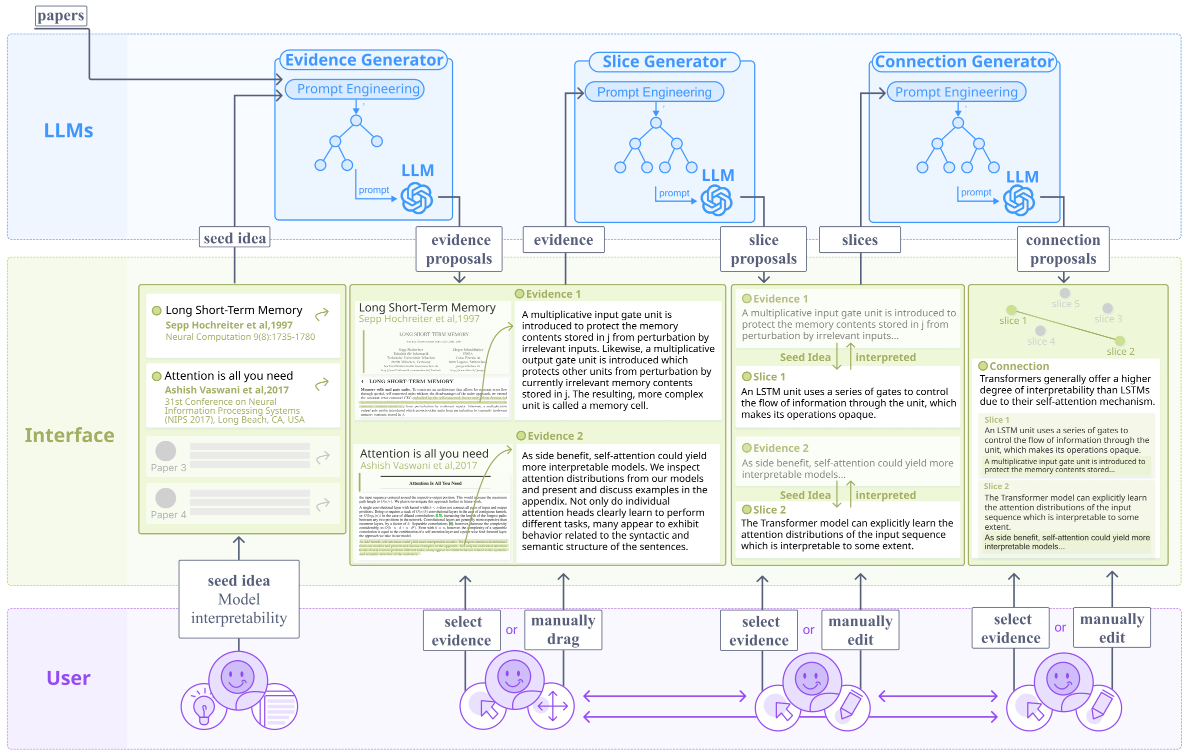


Figure 2: **The interactive workflow of** `PersLEARN` **.** This example showcases the scenario that a `PersLEARN` user intends to frame a rather novel perspective of "model interpretability" given the original papers on LSTM and Transformer. While the seed idea is not the major focus of both papers, evidence can be found to support that seed idea with the help of `PersLEARN` . The evidence includes the introduction of gates in LSTM to protect memory contents and the self-attention mechanism in Transformer, which can yield more interpretable models. Then the user re-interprets the evidence that LSTM operations are opaque due to the use of gates, while Transformer models have some level of interpretability through their attention distributions. It concludes that Transformers generally offer a higher degree of interpretability compared to LSTMs due to their self-attention mechanism. The process is assisted by prompt-engineered LLMs but is exactly determined by the user.

perspective on both computer vision and physics. Interestingly, most of the reviews the students composed do not have their own perspective; their reviews are titled "CV approaches on physics-based reasoning tasks" or have similar titles. This suggests that most students simply wrote summaries of every citing paper without considering an alternative perspective (see Fig. 1). To address this gap in research training, we propose `PersLEARN` , a tool that **explicitly** guides the process of cultivating scientific perspectives.

`PersLEARN` is grounded in classical theories drawn from the fields of cognitive and social sciences, particularly in the domain of scientific knowledge representation (Sec. 2.1). It provides an entire life-cycle of constructing a perspective frame that semi-automates researchers to start from a single seed idea and then iteratively interpret and structure relevant literature (Sec. 2.2). This process is facilitated through an interactive system that employs a hierarchical prompt-based approach to propose potential interpretations and structures based on a seed idea (Sec. 2.3). Experiments on both human evaluation (Sec. 3) and automatic evaluation of each module (Sec. 4) suggest that `PersLEARN` has the potential to enhance the quality of scientific research training significantly.

2 Design and Implementation

Designing `PersLEARN` is required to answer two questions: (i) What is the appropriate representation of perspective frames that makes researchers comfortable? (ii) How to informationize such representation for both user input and automated generation? In response to the questions, we highlight how `PersLEARN` is implemented from a theoretical framework to an interactive system step by step.

2.1 Theoretical Framework[2]

Following the principle of analogical education (Thagard, 1992; Aubusson et al., 2006), we create a system of analogies to ground the abstract concepts about perspectives. First, the scientific knowledge covered by the literature about a seed idea is in a

[2]View an abstract video illustration of the framework: https://vimeo.com/802213146.

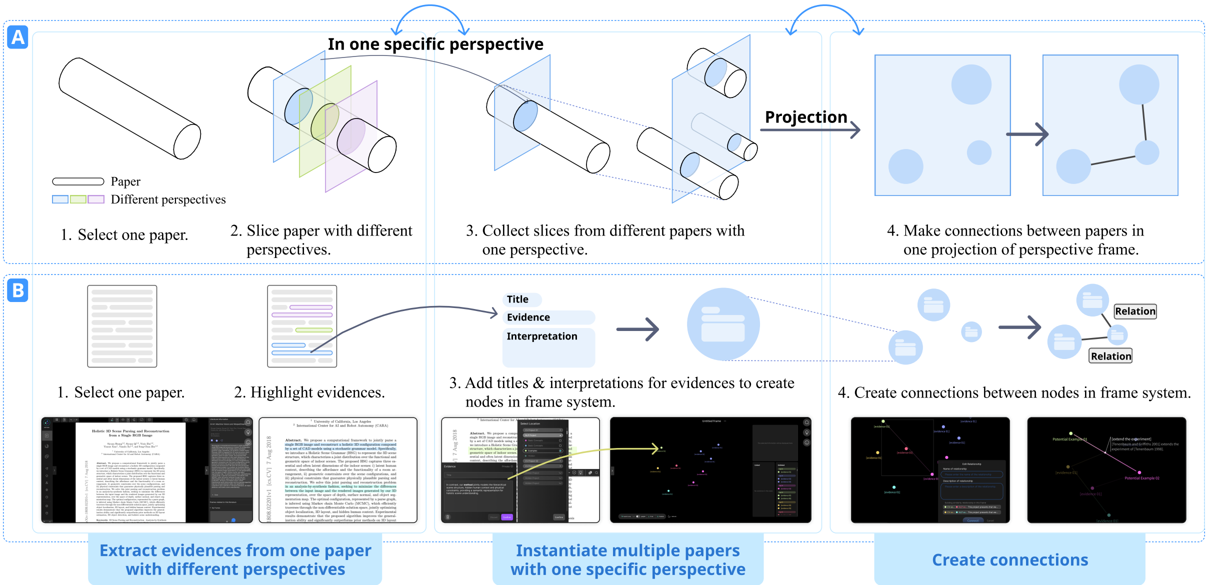


Figure 3: **Illustration of perspective cultivation.** (A) Visual analogy of the process: interpreting the evidence in the papers given the seed idea and structuring the papers with the relations between them. (B) User interfaces during the process.

higher-dimensional space than the perspective of a single paper (Duschl and Grandy, 2008). Here we set scientific knowledge of a seed idea as a 3D space and the specific perspective as a 2D plane for readability. For example, the seed idea "CV approaches on physics-based reasoning tasks" on the intersection of physics and CV can be framed as different specific perspectives, such as "physics-based reasoning" (Zhu et al., 2020), "AI models for modeling human cognition" (Lake et al., 2017), "evaluation metrics of new tasks in AI" (Duan et al., 2022), "ethics in human-level tasks for AI" (Jackson Jr, 2018), and "interpretability of physics-based reasoning AI models" (Edmonds et al., 2019). To not be trapped in a single perspective, we should pay attention to the ingredients of the papers rather than the ideas claimed by the authors. On this basis, framing another perspective is projecting the 3D space to another 2D plane by making *slices* from the papers, where each slice is a subset of ingredients. Such slices are articulated with others under the logic of the seed idea. Thus, a perspective frame is cultivated on the plane, growing from a seed idea with few slices to a graph with slices connected (see Fig. 3 for details).

Formally, the perspective frame is organized as a graph with information in nodes and edges on a 2D plane that instantiates the seed idea from the 3D space of scientific knowledge. The elements in a perspective frame can be described as follows:

- **Seed idea**: A rough textual description of the perspective, *e.g.*, "Physics-based reasoning using CV approaches," which serves as the starting point of the literature review and should be determined at the very beginning.
- **Evidence**: A piece of evidence comes from every paper in the selected set of literature, which contains the grounded information (a text span) supporting the given seed idea.
- **Slice**: A slice is the textual *interpretation* conditioned on the given seed idea based on a piece of evidence. A slice is a node in the graph.
- **Connection**: A connection between two slices is the textual interpretation conditioned on the perspective given the *relation* (*e.g.*, relations-in-common such as `inspire` and `parallel`; and relations-of-distinction such as `improve`, `alternate`, and `compete`) between two slices. A connection is an edge in the graph.

Fig. 2 shows the interactive workflow of `PersLEARN` . Suppose one concerns the *"model interpretability"* (seed idea) of LSTM and Transformer, which is not the major perspective of either original paper of the two models. Given the corresponding two papers *'Long short-term memory'* and *'Attention is all you need'*, the **evidence generator** finds the evidence to support the seed idea from the papers: *'A multiplicative input gate unit is introduced to protect the memory contents stored in j from perturbation by irrelevant inputs. Likewise, a multiplicative output gate unit is introduced which protects other units from perturbation by currently irrelevant memory contents stored in j. The resulting, more complex unit is called a memory cell.'*

and *'As side benefit, self-attention could yield more interpretable models. We inspect attention distributions from our models and present and discuss examples in the appendix. Not only do individual attention heads clearly learn to perform different tasks, but many also appear to exhibit behavior related to the syntactic and semantic structure of the sentences.'* The **slice generator** then generates the interpretations: *'An LSTM unit uses a series of gates to control the flow of information through the unit, which makes its operations opaque.'* and *'The Transformer model can explicitly learn the attention distributions of the input sequence which is interpretable to some extent.'* The **connection generator** finally provides the connection between these slices: *'Transformers generally offer a higher degree of interpretability than LSTMs due to their self-attention mechanism.'*. Such cultivation of a brand new perspective helps students *think outside the box*, which usually yields innovation in scientific research and should serve as one of the major parts in research training.

Notably, elements such as evidence, slices, and connections are not determined at once but may be revised in multiple iterations. As the perspective frame grows, the researcher's understanding of the seed idea goes deeper, and the contents of slices and connections are sharpened accordingly. Hence, instead of answering a *chicken-or-the-egg* problem between slices and connections, our users generate them iteratively. Varied by the seed ideas, a perspective can be a well-organized collection of information (*e.g.*, "performance comparison between backbone models on physical-reasoning tasks" (Duan et al., 2022)), a statement (*e.g.*, "intuitive physics may explain people's ability of physical reasoning" (Kubricht et al., 2017)), or a problem (*e.g.*, "physical reasoning by CV approaches" (Zhu et al., 2020)). Though coming with different levels of abstraction, they all bring information gain, more or less (Abend, 2008).

PersLEARN well echoes the established theories, suggesting our design's integrity. In a perspective frame, elements are contextualized in the entire frame by connecting with each other (Grenander, 2012; Shi et al., 2023); no element's meaning is determined solely by itself. Moreover, any revision of an element influences the larger structure. Such representation has been shown as an innate knowledge representation of humans—*theory theory* (Gopnik, 1994; Gopnik and Meltzoff, 1997;

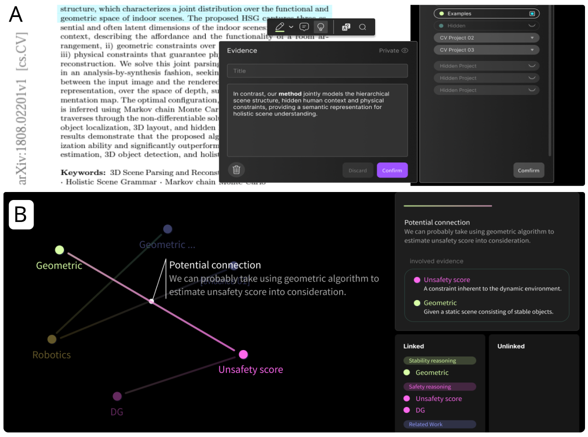


Figure 4: **UI showcases.** (A) A selected piece of evidence and its interpretation. (B) A generated perspective frame.

Carey, 1985, 2009). Furthermore, Carey (1986, 2000) have shown that such a framework can be captured and gradually revised by young students in terms of science education. To the best of our knowledge, current tools for literature review composing (*e.g.*, ResearchRabbit, Connected Paper, Inciteful, and Litmaps) all focus on visualizing literature relationships based on similarity and citation relationships without explicitly considering the framing of diverse perspectives.

2.2 Implementing User Interface (UI)

A researcher may develop a seed idea when reading a few papers, even if it is far from a mature perspective. The user first locates the evidence in a paper by dragging the mouse to select the text span through the `PDFViewer` and adds the selected span into `Evidence Hub`. Next, the user could generate a slice by writing a textual interpretation of the paper based on the evidence; this would trigger the initialization of a new perspective frame, and the first slice can be dragged into the canvas (implemented by `D3.js` library (Bostock et al., 2011)). The user can get back to the papers for more pieces of evidence and back to revising interpretations by clicking on the slices and editing the information at the right bar. With more than one slice in the canvas, the user can connect two slices by dragging the mouse around them and then write a textual interpretation of the relation between them. Likely to edit the slices, the user can also edit the connections by clicking on them and editing the information at the right bar. The perspective frame is cultivated by repeating these steps, buliding up the mindset for perspective framing in the *learning by doing* principle (Schank et al., 1999). Please refer to Fig. 4 for an exemplar perspective frame.

In the user-centered design of UI (Zaina et al.,

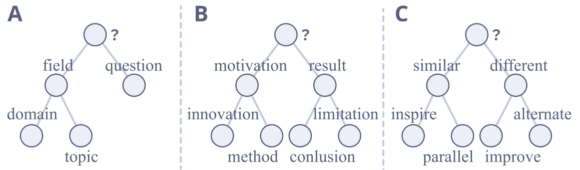


Figure 5: **Prompt Engineering.** (A) Evidence location. (B) Slice generation. (C) Connection generation.

2021), we follow the established theories in design for education (Hu et al., 1999; Miraz et al., 2016), such as color classification (Wen, 2021) and hierarchical information display (Jinxian, 2020). These support the integrity of our UI design.

2.3 Semi-automating the Procedure

We employ a hierarchical prompt-based approach to semi-automate the slice generation and connection generation. `PersLEARN` automatically generates some candidate proposals of slices and connections, and users can choose to accept, delete or modify these proposals.

Generate proposals of slices Scientific papers generally have similar and main-streaming structures (Doumont, 2014). Humans read scientific papers effectively while considering this prior structure rather than browsing aimlessly. We leverage this intuition by proposing the hierarchical prompt-based approach. This approach takes the seed idea and partial texts of the paper (*i.e.*, `Abstract`, `Introduction`, `Discussion`, and `Conclusion`) sections as input, and outputs the proposals of slices. We designed a hierarchical prompt-based approach (see Appx. A.1). First, we parse the seed idea to identify the specific field and domain of interest and fit the parsed terms into the prompting schema. Next, the prompted Large Language Model (LLM) extracts sentences from papers as evidence proposals. The LLM generates slice proposals conditioned on the evidence.

Specifically, it consists of two prompting stages: prompt generation and answer extraction. In the first stage, we first prompt an LLM with a generated prompt. After the LLM generates a response, we extract the information as the answer. Next, we traverse the hierarchical prompting schema from the top down to adopt a prompt template. Finally, we concatenate it with the texts of the paper as the prefix to generate the response. In the second stage, we post-process the response by removing repeated words and punctuation marks such as extra spaces.

Generate proposals of connections Similar to slice generation, generating proposals of connections follows a prompt-based approach. The LLM takes two slices as input and outputs the relation between these two slices. A connection shows the relation between two slices (*e.g.*, relations-in-common such as `inspire` and `parallel`; and relations-of-distinction such as `improve`, `alternate`, and `compete`). Hence, we design the prompt as a multiple-choice question.

Our approach avoids uncontrollable and time-consuming manual designing while achieving comparable performance compared to existing fully-manual methods. Since we use the zero-shot setting, labor-consuming labeling is not required.

3 Human Evaluation

To validate `PersLEARN` for research training, we conducted a human study following the standard protocols of digital device auxiliary scenarios in higher education (Van den Akker, 1999; Neuman, 2014). This study is approved by the Institutional Review Board (IRB) of Peking University.

3.1 Method

Materials We created a scenario that simulates the training on writing literature reviews. The literature used in our simulation is five papers published at computer vision conferences. These papers have different topics varying from 3D scene parsing and reconstruction to learning object properties and using tools. However, they can be integrated together by interpreting from a physics-based perspective.

Participants We recruited 24 participants from the Peking University participant pool (11 female; mean age = 22.63). Every participant was paid a wage of $14.6/h. We evenly divided participants into the control and experimental groups.

Procedures All participants were required to read the five papers and compose a short paragraph of literature review given the perspective "Physics-based reasoning." Only the abstract, introduction, and conclusion/discussion were mandatory to read to reduce workload. The experiments lasted for 1 hour. The control group followed the standard procedure of writing reviews without `PersLEARN` as researchers usually do in their studies: reading the raw papers and writing the review. The experimental group utilized `PersLEARN` to create the review: locating evidence, interpreting, illustrating relations, and synthesizing the review. All participants were free to use the internet for extra help, such as searching for new concepts and unfamiliar words.

3.2 Result

We evaluate PersLEARN both quantitatively and qualitatively to verify whether it helps students compose more logical and pertinent reviews.

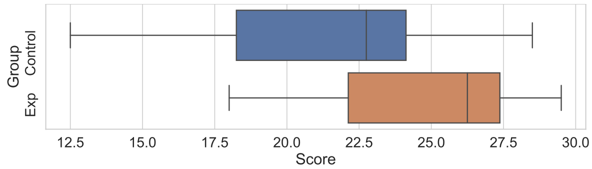
Quantitative evaluation The reviews from the control and experimental groups were shuffled and sent to experts to grade. The grading metrics include logicality and pertinence. Specifically, we asked 3 experts to grade on consistency (Farkas, 1985), rationality (Kallinikos and Cooper, 1996), organization (Kallinikos and Cooper, 1996), topic relevance (Hayes, 2012), opinion clarity (Williams, 1990), and concreteness (Sadoski et al., 2000); each ranks from 1 to 5. All of the experts hold Ph.D. degrees in related fields of AI, have been working on AI for at least eight years, and have no conflict of interest with the authors of this paper.

The average scores of the control and experimental groups are 21.25 and 25.08, respectively. Fisher's exact test on the two variables (*i.e.*, whether PersLEARN was used and the score) reveals that the experimental group significantly outperforms the control group in both logicality and pertinence ($P = 0.0361$; see Fig. 6a), suggesting participants exploit the interpretations from a particular perspective and organize them by inducing their relations. Such a paradigm equips them with improved research training. Detailed scores on 6 evaluation metrics are shown in Fig. 6b. The results demonstrate a noticeable improvement in academic review writing in terms of logicality and pertinence for the experimental group; the experimental group's performance shows a clear shift towards higher scores.

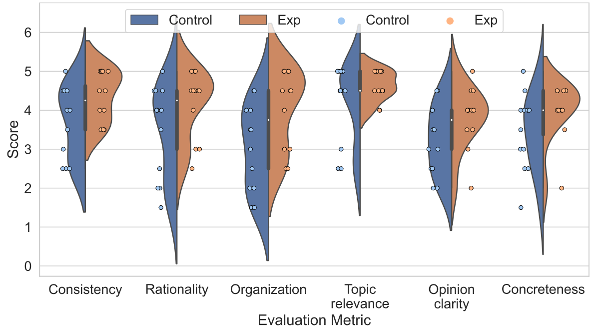
Qualitative evaluation We further conducted an interview to record qualitative comments after the participants in the control group finished their experiments. We interviewed them on how PersLEARN contributed to reading papers and composing reviews; see Appx. B.3 for interview questions. Most participants stated that PersLEARN helped them better understand the content of articles, think more clearly, and organize their writing expediently. For future work, they hoped to embed intelligent agents to provide protocols for each procedure.

3.3 Discussion

We present a case study to show how the experimental group composes better reviews than the control group; see representative paragraphs in



(a) The total scores with 30.0 at most.



(b) The scores of the 6 metrics with 5.0 for each. The presented scores in the plot may surpass 5.0 due to smoothing.

Figure 6: **The scores of the control and experimental groups.** The scores of each review are averaged across experts.

Appx. C.1. We conclude from our human study that PersLEARN can boost literature reading and review writing by providing a perspective-guided thinking framework of evidence locating, interpretation deriving, and relation inducing.

4 Automatic Evaluation

To automatically evaluate PersLEARN at scale, we introduce a *perspective reconstruction* task with three sub-tasks (slice generation, connection generation, and diversity evaluation), requiring the system to recover an established perspective frame given the same seed idea.

4.1 Benchmark Construction

Dataset Collection We carefully construct a testing set with reputation-established narrative reviews, expert reviews, and opinion articles to obtain a high-quality ground truth of perspectives. Systematic reviews and articles of information collection are removed from the set because such papers do not provide a sharp and unique perspective; we ensure that all articles are developed around a concrete and coherent perspective. Moreover, we ensure that every title is the epitome of the perspective held by the article; we treat the titles as seed ideas.

36 review articles are collected from diverse domains standing at the intersection of AI and CoRe, including CV, Natural Language Processing (NLP), Intuitive Physics (Phy), Causality (Cau), Abstract Reasoning (AbsRe), Mirroring and Imitation (MrIm), Tool Use (Tool), Non-verbal Communication (NvComm), Intentionality (Int), Theory

Table 1: **Result of our pipeline.** w/o and w/ are with and without prompt engineering, respectively. The performance of slice generation, connection generation, and perspective diversity indicate the efficacy of our prompt engineering.

Metric	w/o	w/
Slice BLEURT	0.238	**0.795**
Connection CR	0.450	**0.550**
Perspective VMR	0.028	**0.006**

of Mind (ToM), and Utility (U). This generates a literature set with 333 papers cited by at least one of the articles. Among these, 24 papers are cited by more than one article. Some of the papers are directly obtained from S2ORC (Lo et al., 2020), while others are parsed from raw PDF.

Evaluation Metrics of slice generation For a cited paper in the original review, we treat the coherent sentences around the citation mark as the ground truth for the corresponding slice, following the same protocol as in Li et al. (2022). Because the semantic meaning is critical (rather than the wording and phrasing), we employ BLEURT (Sellam et al., 2020) rather than word-wise evaluation metrics like ROUGE and BLEU (Lin, 2004; Papineni et al., 2002). BLEURT score indicates the similarity between two statements; larger scores mean better performance.

Evaluation Metrics of connection generation Since the connection between two papers under the same perspective is only conditioned on the slices, we focus on the logical consistency between the generated connection and the two input slices. Following the setting of Natural Language Inference (NLI), we calculate the Consistent Rate (CR), the proportion of entailment prediction in all predictions. Higher CR indicates better performance. We employ the state-of-the-art model, DeBertaV3 (He et al., 2021b,a), as the NLI model for evaluation.

Evaluation Metrics of Diversity This is an extended case study based on slice generation. We specially study how different perspectives drive the interpretations from the same set of papers. We calculate the normalized Variance-to-Mean Ratio (VMR) over the BLEURT scores on all established perspectives of a set of papers for each approach. Lower VMR indicates that an approach generates slices conditioned on different perspectives well.

4.2 Experiments of Slice Generation

Setup We use **InstructGPT** as the backbone LLM model (Ouyang et al., 2022) for our prompt-based approach. The input and output are the same as in Sec. 2.3. The baseline approach directly prompts the LLM with the target output without the proposed hierarchical prompting schema.

Results The BLEURT results in Tab. 1 show that the generation with prompt engineering outperforms that without by a large margin (233%). This result validates our pipeline in abstract understanding and perspective-based interpreting; see representative slices in Appx. C.2.

4.3 Experiments of Connection Generation

Setup We use **InstructGPT** as the backbone LLM model. The input and output of this evaluation are the same as the connection proposed in Sec. 2.3. The baseline approach directly prompts the LLM with the target output.

Results As shown in Tab. 1, our connection generation module surpasses the baseline approach in CR by a large margin (22%). It means more logical connections are generated by our approach and thus contribute to more entailment predictions. See representative connections in Appx. C.3.

4.4 Experiments on Diverse Perspectives

Setup We use **InstructGPT** as the backbone LLM model for both the slice and the connection generation modules. The baseline approach adopts the slice and connection generation modules without the proposed schema.

Results The VMR results in Tab. 1 show that `PersLEARN` generates slices of richly diverse perspectives, surpassing the baseline by a large margin (79%). We present some examples of the interpretations of different perspectives; see representative slices in Appx. C.4.

5 Discussion

We present `PersLEARN` to facilitate scientific research training by explicitly cultivating perspectives. Human study shows that `PersLEARN` significantly helps junior researchers set up the mindset for jumping out of perspective given by the literature and framing their own ones. Extensive benchmarking shows that our system has the potential to mine perspectives out of diverse domains of literature without much human effort. These experiments suggest that `PersLEARN` has the potential to support scientific research training in general—from explicating one's own perspective to embracing the diverse perspectives of others. Readers can refer to the "Broader Impact" and "Limitation" sections (Sec. 5) for further discussions.

Ethics Statement

The human study presented in this work has been approved by the IRB of Peking University. We have been committed to upholding the highest ethical standards in conducting this study and ensuring the protection of the rights and welfare of all participants. Considering that the workload of the procedure for participants is relatively high among all human studies, we paid the participants a wage of $14.6/h, which is significantly higher than the standard wage (about $8.5/h). Every expert was paid $240 for grading the 24 review paragraphs composed by the participants.

We have obtained informed consent from all participants, including clear and comprehensive information about the purpose of the study, the procedures involved, the risks and benefits, and the right to withdraw at any time without penalty. Participants were also assured of the confidentiality of their information. Any personal data collected (including name, age, and gender) was handled in accordance with applicable laws and regulations.

Broader Impact

The underlying impact of the mindset brought by `PersLEARN` goes beyond research training toward science education in general. Specifically, `PersLEARN` provides the infrastructure for further investigation in two aspects: (1) embracing the diverse perspectives of the same scientific topic to construct a stereoscopic understanding of the topic; (2) facilitating the communication between junior researchers with different mindsets.

The broader impact is analogous to the classic fable *Blind men and an elephant*, where each man interpreted the elephant differently because they were standing on different perspectives. Though this has been a metaphor complaining that science is limited by observation (Heisenberg, 1958), it highlights the virtue of scientific research—focused, and every young researcher understands and interprets science from a focused perspective. Hence, to gain a more comprehensive view of the elephant, the blind men may put their understandings of it together and then try to synthesize it based on their perspectives. In contrast, a sighted person may view the elephant from a distance and capture a holistic view at first—she ends up with a superficial understanding of the elephant if not selecting a perspective and going close to the elephant, like the blind. Thus, by embracing diverse perspectives (*i.e.*, visualizing the perspective frames in a hub), one gets a stereoscopic view and, more importantly, a deeper understanding of the scientific topic. Moreover, when the metaphorical *blind men* in the fable attempt to articulate their distinct perspectives, they may be hindered by the gap between mindsets. To exemplify, individual might struggle to comprehend the concept of a "fan", which in their perception, the elephant appears to resemble. This suggests that the communication of science should be executed in a listener-aware way and that the speaker's perspective should be transformed (*i.e.*, by changing the terms used in slices and connections) to its analogical equivalent in the listener's mindset. Thus, science can be communicated easily, facilitating its transparency, reliability, and the chances of cross-domain collaboration. In summary, our framework of scientific perspective may bring science education to a future with better student-centered considerations (Leshner, 2018).

Limitations

As a preliminary work, the design and evaluation of `PersLEARN` come with limitations, leading to further investigations:

- Can we construct a larger scale dataset of explicit perspective frames of the literature for more fields in the sciences, such as biology, sociology, *etc.*?
- Can we fine-tune LLMs on the larger dataset to obtain better performance on slice and connection generation?
- Can we carry out a human study at a larger temporal scale, say during one semester, to track the progress of students using `PersLEARN` ?

With many questions unanswered, we hope to facilitate research training and science education in a broader way.

Acknowledgements

The authors would like to thank Dr. Yujia Peng, Yuxi Wang, and Zili Li for their kind help in setting up the human study. This work is supported by the Institute for Artificial Intelligence at Peking University. W.H. is supported by the Talent Fund of Beijing Jiaotong University (2023XKRC006). All authors declare no conflict of interest.

References

Gabriel Abend. 2008. The meaning of 'theory'. *Sociological Theory*, 26(2):173–199.

Peter J Aubusson, Allan G Harrison, and Stephen M Ritchie. 2006. *Metaphor and analogy: Serious thought in science education*. Springer.

Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graph*, 17(12):2301–2309.

Rodger W Bybee. 2006. Scientific inquiry and science teaching. *Scientific inquiry and nature of science: Implications for teaching, learning, and teacher education*, pages 1–14.

Susan Carey. 1985. *Conceptual change in childhood*. MIT Press.

Susan Carey. 1986. Cognitive science and science education. *American Psychologist*, 41(10):1123.

Susan Carey. 2000. Science education as conceptual change. *Journal of Applied Developmental Psychology*, 21(1):13–19.

Susan Carey. 2009. *The Origin of Concepts*. Oxford University Press.

Karin Knorr Cetina. 1999. *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Jean-luc Doumont. 2014. Structuring Your Scientific Paper. In *English Communication for Scientists*. Nature.

Jiafei Duan, Arijit Dasgupta, Jason Fischer, and Cheston Tan. 2022. A survey on machine learning approaches for modelling intuitive physics. In *International Joint Conference on Artificial Intelligence*.

Richard A Duschl and Richard E Grandy. 2008. *Teaching scientific inquiry: Recommendations for research and implementation*. BRILL.

Mark Edmonds, Feng Gao, Hangxin Liu, Xu Xie, Siyuan Qi, Brandon Rothrock, Yixin Zhu, Ying Nian Wu, Hongjing Lu, and Song-Chun Zhu. 2019. A tale of two explanations: Enhancing human trust by explaining robot behavior. *Science Robotics*, 4(37).

David K Farkas. 1985. The concept of consistency in writing and editing. *Journal of Technical Writing and Communication*, 15(4):353–364.

Samuel J Gershman, Eric J Horvitz, and Joshua B Tenenbaum. 2015. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278.

Alison Gopnik. 1994. The theory theory. In *Mapping the mind: Domain specificity in cognition and culture*, pages 257–293. Cambridge University Press.

Alison Gopnik and Andrew N Meltzoff. 1997. *Words, thoughts, and theories*. MIT Press.

Ulf Grenander. 2012. *A calculus of ideas: a mathematical study of human thought*. World Scientific.

Thomas L Griffiths, Falk Lieder, and Noah D Goodman. 2015. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2):217–229.

John Hayes. 2012. Modeling and remodeling writing. *Written communication*, 29:369–388.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021a. Debertav3: Improving deberta using electra-style pretraining with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021b. Deberta: Decoding-enhanced bert with disentangled attention. In *International Conference on Learning Representations*.

Werner Heisenberg. 1958. Physics and philosophy: The revolution in modern science. *Physics Today*, 11(9):36.

Paul Jen-Hwa Hu, Pai-Chun Ma, and Patrick YK Chau. 1999. Evaluation of user interface designs for information retrieval systems: a computer-based experiment. *Decision Support Systems*, 27(1-2):125–143.

Philip C Jackson Jr. 2018. Toward beneficial human-level ai... and beyond. In *AAAI Spring Symposia*.

Lin Jinxian. 2020. Application of guiding design in ui design of mobile terminal. In *2020 IEEE International Conference on Innovation Design and Digital Technology (ICIDDT)*.

Jannis Kallinikos and Robert Cooper. 1996. Writing, rationality and organization: An introduction. *Scandinavian Journal of Management*, 12:1–6.

James R Kubricht, Keith J Holyoak, and Hongjing Lu. 2017. Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, 21(10):749–759.

Thomas S Kuhn. 1970. *The structure of scientific revolutions*. University of Chicago Press: Chicago.

Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. 2017. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253.

Bruno Latour. 1987. *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Bruno Latour and Steve Woolgar. 1986. *Laboratory life: The construction of scientific facts*. Princeton University Press.

Bruno Latour et al. 1999. *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.

Alan I. Leshner. 2018. Student-centered, modernized graduate stem education. *Science*, 360(6392):969–970.

Richard L Lewis, Andrew Howes, and Satinder Singh. 2014. Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, 6(2):279–311.

Xiangci Li, Biswadip Mandal, and Jessica Ouyang. 2022. Corwa: A citation-oriented related work annotation dataset. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Falk Lieder and Thomas L Griffiths. 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*.

Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel S Weld. 2020. S2orc: The semantic scholar open research corpus. In *Annual Meeting of the Association for Computational Linguistics*.

Helen E Longino. 1990. *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Michael Lynch. 1993. *Scientific practice and ordinary action: Ethnomethodology and social studies of science*. Cambridge University Press.

Mahdi H Miraz, Peter S Excell, and Maaruf Ali. 2016. User interface (ui) design issues for multilingual users: a case study. *Universal Access in the Information Society*, 15:431–444.

Delia Neuman. 2014. Qualitative research in educational communications and technology: A brief introduction to principles and procedures. *Journal of Computing in Higher Education*, 26:69–86.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, et al. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*.

Mark Sadoski, Ernest T Goetz, and Maximo Rodriguez. 2000. Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text types. *Journal of Educational Psychology*, 92:85.

Roger C Schank, Tamara R Berman, and Kimberli A Macpherson. 1999. Learning by doing. *Instructional-design theories and models: A new paradigm of instructional theory*, 2(2):161–181.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. Bleurt: Learning robust metrics for text generation. In *Annual Meeting of the Association for Computational Linguistics*.

Yu-Zhe Shi, Manjie Xu, John E. Hopcroft, Kun He, Joshua B Tenenbaum, Song-Chun Zhu, Ying Nian Wu, Wenjuan Han, and Yixin Zhu. 2023. On the complexity of bayesian generalization. In *International Conference on Machine Learning*.

Paul Thagard. 1992. Analogy, explanation, and education. *Journal of Research in Science Teaching*, 29(6):537–544.

Jan Van den Akker. 1999. Principles and methods of development research. *Design Approaches and Tools in Education and Training*, pages 1–14.

Guoying Wen. 2021. Research on color design principles of ui interface of mobile applications based on vision. In *2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*.

Joseph Williams. 1990. Toward clarity and grace. *Chicago: The University of Chicago*.

Luciana AM Zaina, Helen Sharp, and Leonor Barroca. 2021. Ux information in the daily work of an agile team: A distributed cognition analysis. *International Journal of Human-Computer Studies*, 147.

Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. 2020. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345.