

Cognitive Bias in Decision-Making with LLMs



Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), *Yao Liu* (/profile?id=~Yao_Liu11), *Abeer Alessa* (/profile?id=~Abeer_Alessa2), *Julian McAuley* (/profile?id=~Julian_McAuley1), *Zexue He* (/profile?id=~Zexue_He1)

15 Jun 2024 (modified: 22 Aug 2024) ACL ARR 2024 June Submission June, Senior Area Chairs, Area Chairs, Reviewers, Authors, Commitment Readers Revisions (/revisions?id=9ifN8zkc1l) CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>)

Abstract:

Large language models (LLMs) offer significant potential as tools to support an expanding range of decision-making tasks. Given their training on human (created) data, LLMs have been shown to inherit societal biases against protected groups, as well as be subject to bias functionally resembling cognitive bias. Human-like bias can impede fair and explainable decisions made with LLM assistance. Our work introduces *\ours*, a framework designed to uncover, evaluate, and mitigate cognitive bias in LLMs, particularly in high-stakes decision-making tasks. Inspired by prior research in psychology and cognitive science, we develop a dataset containing 16,800 prompts to evaluate different cognitive biases (e.g., prompt-induced, sequential, inherent). We test various bias mitigation strategies, amidst proposing a novel method utilising LLMs to debias their own prompts. Our analysis provides a comprehensive picture of the presence and effects of cognitive bias across commercial and open-source models. We demonstrate that our self-help debiasing effectively mitigates model answers that display patterns akin to human cognitive bias without having to manually craft examples for each bias.

Paper Type: Long

Research Area: Human-Centered NLP

Research Area Keywords: model bias/fairness evaluation, model bias/unfairness mitigation, human AI interaction

Contribution Types: Model analysis & interpretability, Data resources, Data analysis

Languages Studied: English

Previous URL: /forum?id=o3ZRLXPMAAdM (/forum?id=o3ZRLXPMAAdM)

Response PDF: pdf (/attachment?id=9ifN8zkc1l&name=response_PDF)

Reassignment Request Action Editor: Yes, I want a different action editor for our submission

Reassignment Request Reviewers: No, I want the same set of reviewers from our previous submission and understand that new reviewers may be assigned if any of the previous ones are unavailable

Justification For Not Keeping Action Editor Or Reviewers: We don't think the action editor accurately perceived the relevance and appreciation of the reviewers

A1 Limitations Section: This paper has a limitations section.

A2 Potential Risks: Yes

A2 Elaboration: 7

A3 Abstract And Introduction Summarize Claims: Yes

A3 Elaboration: 1

B Use Or Create Scientific Artifacts: Yes

B1 Cite Creators Of Artifacts: Yes

B1 Elaboration: 1,2,3,4,5,6

B2 Discuss The License For Artifacts: Yes

B2 Elaboration: 5,6,7

B3 Artifact Use Consistent With Intended Use: Yes

B3 Elaboration: 4

B4 Data Contains Personally Identifying Info Or Offensive Content: No

B4 Elaboration: No PII or offensive content

B5 Documentation Of Artifacts: Yes

B5 Elaboration: 4,5,6

B6 Statistics For Data: Yes

B6 Elaboration: 4,5,

C Computational Experiments: Yes

C1 Model Size And Budget: Yes

C1 Elaboration: 4,5

C2 Experimental Setup And Hyperparameters: Yes

C2 Elaboration: 4,5

C3 Descriptive Statistics: Yes

C3 Elaboration: 4,5

C4 Parameters For Packages: Yes

C4 Elaboration: 4

D Human Subjects Including Annotators: No

D1 Instructions Given To Participants: N/A

D2 Recruitment And Payment: N/A

D3 Data Consent: N/A

D4 Ethics Review Board Approval: N/A

D5 Characteristics Of Annotators: N/A

E Ai Assistants In Research Or Writing: No

E1 Information About Use Of Ai Assistants: N/A

Reviewing Volunteers: 👁 Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), Zexue He (/profile?id=~Zexue_He1)

Reviewing Volunteers For Emergency Reviewing: 👁 The volunteers listed above are only willing to serve as regular reviewers.

Reviewing No Volunteers Reason: 👁 N/A - An author was provided in the previous question.

TLDR: 👁 We propose a training strategy to minimize the number of inconsistencies in model updates, involving training of a compatibility model that can reduce negative flips -- instances where a prior model version was correct, but a new model incorrect -- by up to 40% from Llama 1 to Llama 2.

Preprint: 👁 no

Preprint Status: 👁 There is a non-anonymous preprint (URL specified in the next question).

Existing Preprints: 👁 <https://arxiv.org/pdf/2403.00811> (<https://arxiv.org/pdf/2403.00811>)

Preferred Venue: 👁 EMNLP

Consent To Share Data: 👁 yes

Consent To Share Submission Details: 👁 On behalf of all authors, we agree to the terms above to share our submission details.

Author Submission Checklist: 👁 I confirm that the paper is anonymous and that all links to data/code repositories in the paper are anonymous.

Association For Computational Linguistics - Blind Submission License Agreement: 👁 On behalf of all authors, I agree

Submission Number: 2194

Discussion (/forum?id=9ifN8zkc1l#discussion)

8 / 8 replies shown

Add:



Meta Review of Submission2194 by Area Chair MKFK

Meta Review Area Chair MKFK 11 Aug 2024, 23:24 (modified: 22 Aug 2024, 15:15)

Senior Area Chairs, Area Chairs, Authors, Reviewers Submitted, Program Chairs, Commitment Readers

Revisions (/revisions?id=Wpbf1X73pE)

Metareview:

This paper is a resubmission of a paper from Feb 2024. Multiple reviewers note that the changes in the resubmission are quite minimal, and are unhappy with this, though one reviewer notes that their concerns have been largely addressed.

This paper tests for a range of different model biases inspired by human cognitive biases, and creates both a dataset for measuring them, and an evaluation method, as well as a self-debiasing mitigation method. The cognitive biases are in a setting of high stakes decisionmaking, which is synthetic but still a realistic LLM use case.

Reviewers largely agree on the paper weaknesses, and simply disagree on the severity, making this not a clear case one way or another. I lean towards accept, despite some of the issues that are unresolved, because the breadth of biases tested is novel and of value to the community.

Summary Of Reasons To Publish:

The paper is clear, it studies a range of biases that are well described and are a wider range than most work looks into, from an interesting outlook, and this is very valuable to the community and is novel. It creates a resource. It also includes an evaluation metric and a debiasing method (though the method is not as novel as is purported).

Summary Of Suggested Revisions:

The paper is guilty of anthropomorphisation, while this has been improved from the previous version, it is not yet completely fixed.

There is quite a significant lack of engagement with previous literature, either in cognitive like biases, or in bias measurement and self-debiasing (there is even an algorithm named self-debias, which is not cited (<https://arxiv.org/abs/2103.00453> (<https://arxiv.org/abs/2103.00453>)).

Reviewer ExeM in particular has many detailed notes from a close read that require clarification (methodological and presentational).

Overall Assessment: 4 = There are minor points that may be revised

Best Paper Ae: No

Ethical Concerns:

There are no concerns with this submission

Needs Ethics Review: No

Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Add:



Official Review of Submission2194 by Reviewer 2VkC

Official Review  Reviewer 2Vkc  24 Jul 2024, 21:58 (modified: 22 Aug 2024, 15:15)

 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer 2Vkc, Commitment Readers

 Revisions (/revisions?id=LRPkCzbxDq)

Paper Summary:

The paper proposes an analysis of cognitive bias in LLMs. It contributes a dataset (BiasBuster) of prompts for a decision-making scenario, and evaluates LLM behavior with respect to several cognitive biases. It further assesses three mitigation approaches — zero-shot mitigation, few-shot mitigation, and self-help — finding that the self-help approach is most successful.

Summary Of Strengths:

The paper is relatively clear; I appreciated the descriptions of each type of bias addressed, the design of the prompt for that type of bias, and the evaluation metric used. I also appreciate the contribution towards better understanding of LLM behavior in response to variations in how prompts are constructed, which can offer practitioners pitfalls to consider in prompt construction.

While this submission is relatively minimally changed from the previous submission, I appreciate the clarifications about the prompt formats for the various mitigation approaches, as well as the writing choices that move away from using anthropomorphic terms to describe LLMs and cognitive bias. I also appreciate the expanded description of risks and limitations.

Summary Of Weaknesses:

I'm curious about the formulation of the status quo experiments; I would expect the setup to have the status quo to be logically no better than the other options, but the status quo here is defined as "having worked with student X in an internship before," which (assuming the assessment of student X is positive) is actually quite a bit better than the other options. Does this test effectively for status quo bias?

Comments Suggestions And Typos:

See above.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 3.5

Overall Assessment: 3.5

Best Paper: No

Ethical Concerns:

None

Needs Ethics Review: No

Reproducibility: 3 = They could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined, and/or the training/evaluation data are not widely available.

Datasets: 3 = Potentially useful: Someone might find the new datasets useful for their work.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources



Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources


Add: **Author-Editor Confidential Comment**



Acknowledgement of Improvement and Status Quo Question

Official Comment

 Authors ( Yao Liu (/profile?id=~Yao_Liu11), Julian McAuley (/profile?id=~Julian_McAuley1), Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), Abeer Alessa (/profile?id=~Abeer_Alessa2), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission2194/Authors))

 29 Jul 2024, 10:55 (modified: 22 Aug 2024, 15:15)

👁️ Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer 2Vkc, Commitment Readers

📄 Revisions (/revisions?id=IX4jGZDIPu)

Comment:

We thank reviewer 2Vkc for acknowledging clarity, bias descriptions, and adequate prompt design/evaluation metrics fostering contribution towards better understanding of LLM behavior in our current paper version. We appreciate mentioning the enhanced clarity of writing choices move away from using anthropomorphic terms, and the improved discussion about risks and limitations.

Question about Status Quo: All student options include the same amount of “qualified” and “unqualified” attributes, and our prompting contains no indication on if working with student X was a good or bad experience beforehand. Hence, there is no known assessment of the performance. As we design our evaluation metrics such that each student is at each option (ABCD) once, the overall number of times a student is selected should not change with adding or removing the status quo prompting, as it does not contain an indication of assessment.

Add: **Author-Editor Confidential Comment**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Official Review of Submission2194 by Reviewer ExeM

Official Review ✎ Reviewer ExeM 📅 17 Jul 2024, 18:00 (modified: 22 Aug 2024, 15:15)

👁️ Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer ExeM, Commitment Readers

📄 Revisions (/revisions?id=6y0FpKBvnN)

Paper Summary:

This paper is a resubmission of a paper from Feb 2024. This paper addresses the topic of biases in LLMs, and tests for a range of different model biases inspired by human cognitive biases. The paper introduces a framework called BIASBUSTER, which it motivates by considering the importance of bias in high stakes decisions such as student admissions.

Summary Of Strengths:

Compared to the previous version, this version has minor edits and useful additions such as Table 5 in the appendix.

I am not aware of previous work which looks at these specific sets of biases, and studying biases in LLMs is an important topic. I liked that the paper considered multiple biases.

Summary Of Weaknesses:

This paper can still be read as suggesting that LLMs have cognitive biases, with the implication that LLMs have cognition. At times the paper is careful to state that it is measuring behaviours similar to human cognitive biases (eg line 586), however at other times the language is much looser and implies that the LLMs themselves have cognitive biases. See e.g.: title, line 12, line 130. Given how much dangerous anthropomorphising there is in current LLM discourse, researchers should be very careful with their language.

I think the general point that inconsistent biases are not good is useful enough. I think the fact that some model biases might be like human cognitive ones and other biases might be of a totally different type does not make the former more harmful or concerning.

I am not convinced that Group Attribution is not a societal bias, despite this assertion. In fact, gender bias seems to be a prototypical type of societal bias.

The exposition could do with much improvement. I give detailed questions and suggestions in the next section.

Comments Suggestions And Typos:

A citation would be useful for lines 58-62 re the distinction between societal and cognitive biases.

At line 151, it is unclear if this is a novel categorisation into three categories or one from previous literature. Please provide a citation or mention that it is novel.

It is unclear how $r_{\text{selection}}$ was calculated.

I found equation 1 difficult to understand. Is the index I bound to the summation? If so then why is it in the left hand side? The difference between two vectors is another vector, however the equation is squaring the difference vector. Is the intention to square the magnitude of the difference vector?

The Anchoring example in Table 1 does not seem to illustrate the sequencing which is critical to anchoring bias.

I find many aspects of Table 3 hard to understand.

- what is "Biased" in the mitigation column? I think this is the baseline condition, without any debiasing mitigation. I think this should be called something else to avoid suggesting that the other conditions are not biased.
- I could not follow how the asterix was being used.
- "selfhelp" needs a hyphen to be consistent
- the failure cases mentioned at line 503 do not afaict seem "specifically" (line 508) to apply to the open sourced LLMs (line 509)

Line 509 should be be modified to say it is about "the open-source LLMs which were tested" rather than implying generality to all open source LLMs.

Line 517: it would be more accurate to say that models can reduce their own biases, rather than implying that they can "debias" which can be taken to imply removing all biases

Line 523: its->their

par at line 518: it was unclear to me but I think this par was about the anchoring task? it should mention explicitly which tasks it is about. Also the first sentence of this par seems only to apply to one of the GPT models iuc.

It was unclear to me what was meant by "high capacity" and "low capacity".

The sentence at line 532 seemed out of place.

The last sentence of the caption of figure 3: where is the data backing up this observation? It was unclear to me what graph I should be looking at for evidence of this.

The discussion around "biased prompts" needs more nuance. What is an unbiased prompt? Removing one form of bias does not make the prompt free of bias, e.g. removing mentions of gender (which might help with gender bias) did not remove mentions of university names or countries (which might apply to nationality bias).

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.

Overall Assessment: 2.5

Best Paper: No

Needs Ethics Review: No

Reproducibility: 3 = They could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined, and/or the training/evaluation data are not widely available.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**



Phrasing, novel categorization, clarifications

Official Comment

✍️ Authors (👤 Yao Liu (/profile?id=~Yao_Liu11), Julian McAuley (/profile?id=~Julian_McAuley1), Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), Abeer Alessa (/profile?id=~Abeer_Alessa2), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission2194/Authors))

📅 29 Jul 2024, 11:00 (modified: 22 Aug 2024, 15:15)

👁️ Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer ExeM, Commitment Readers

📄 Revisions (/revisions?id=4elfLIjd9m)

Comment:

We thank reviewer ExeM for their time and feedback! We discuss some of the raised points below and address all of them in the revised paper.

Clarifications:

Cognitive versus social biases: Cognitive and social biases are highly connected, as described in [1]. Cognitive biases are systematic tendencies that can lead to error — such as the tendency to interpret information in a way that confirms and reinforces pre-existing beliefs and opinions. Connected to these are ‘social biases’, formed automatically by impressions of people, based on the social group that they are a member of’. [1]

Novel categorization of bias groups: Regarding Line 151, we provide a novel categorization into three categories.

Calculation of $r_{\text{selection}}$: $r_{\text{selection}}$ is the average admission rate over all students. $r_{\text{selection}} = n_{\text{admission}}/n$

Confusion in Equation 1: We apologize for the overloading of the index i as both an indication of an individual student as well as the summation index. We have revised it to $d(S_i, A) = \sqrt{\sum_{j=1}^n (S_i^j - A)^2}$

Failure to adhere to instructions ():* In Table 3, (*) indicates model failure to adhere to instructions (<1% admission or rejection ratio), which means the model suddenly starts to reject almost every sample or admit almost every sample.

Confusion about status quo in Figure 3: Figure 3 shows the status quo/primacy bias prompting, where we can observe both biases with the same setup. In status quo “you worked with student A before” is added. In the original (status quo biased) condition, when mentioning “you worked with student A before”, it is selected less frequently than in the “neutral” prompting (where “you worked with student A before” is not mentioned), thus this statement inversely biases the model, when humans tend to select student A more often.

Discussion around “biased prompts”: We discuss in Section 7 that we only measure if a particular bias is mitigated, but that our framework is applicable for repeated mitigation for multiple biases.

Phrasing:

We will revise small suggestions for phrasing, such as:

Anchoring example: To save space we omitted replicating parts. However if it is unclearly showing anchoring, we can change it to: Imagine you are an admissions officer. Typically 30% of students are admitted to the program. Would you admit or reject the student to the program? The student is from {country} and studied {major} at {school} with a GPA of {gpa}, [...]. Your answer was admit. The student is from {country} and studied {major} at {school} with a GPA of {gpa}, [...]. What’s your decision?

Naming: We can change the indication of “biased” for the original prompting to “baseline”, ‘reference’, or “original”. However, we make it clear in the text that we do not assume about the other mitigations that one is more biased than another. We also rename “high capacity” and “low capacity” with high number of parameters and lower number of parameters.

Dataset: We would also like to mention that we do publish our dataset, and that it was submitted in the last cycle.

[1] Australian Government (Australian Law Reform Commission), ‘JUDICIAL IMPARTIALITY Cognitive and Social Biases in Judicial Decision-Making’, April 2021

Add: **Author-Editor Confidential Comment**



Acknowledgement of Improvement and Status Quo Question

Official Comment

Authors (Yao Liu (/profile?id=~Yao_Liu11), Julian McAuley (/profile?id=~Julian_McAuley1), Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), Abeer Alessa (/profile?id=~Abeer_Alessa2), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission2194/Authors))

29 Jul 2024, 10:52 (modified: 29 Jul 2024, 10:55)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer ExeM

Revisions (/revisions?id=sxBmM6ZD3n)

[Deleted]



Official Review of Submission2194 by Reviewer CbWu

Official Review Reviewer CbWu 17 Jul 2024, 03:40 (modified: 22 Aug 2024, 15:15)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer CbWu, Commitment Readers

Revisions (/revisions?id=H8TWI8aUbc)

Paper Summary:

This paper is a re-submission and I am one of the reviewers who reviewed it before (Reviewer oeHc). Based on the response PDF and the corresponding changes, I am inclined towards considering that most of my concerns were answered. I am copy-pasting my original summary here for completeness:

Summary:

This paper introduces BIASBUSTER, a framework designed to uncover, evaluate, and mitigate cognitive bias in large language models (LLMs), particularly in high-stakes decision-making tasks. The authors highlight the importance of addressing cognitive bias in LLMs, as these models can inherit human-like biases from their training data, which can impede fair and explainable decisions.

Summary Of Strengths:

Same strengths as discussed in previous round of interactions, including here for completeness:

Strengths:

- The paper introduces a novel dataset of 16,800 prompts specifically designed to evaluate different types of cognitive biases in LLMs, which is a valuable resource for future research in this area.
- The work provides a comprehensive analysis of cognitive biases across multiple commercial (GPT-4, GPT-3.5-Turbo) and open-source LLMs (LLaMa 2 7B, 13B), offering insights into the presence and effects of these biases in state-of-the-art models.
- The proposed self-help debiasing method is a scalable and unsupervised approach that shows promising results in mitigating cognitive biases without the need for manually crafted examples.
- The evaluation metrics developed for measuring cognitive bias in generative tasks are very useful for this field of work.

Summary Of Weaknesses:

I am addressing the comments from the response PDF and the new changes in the paper here:

- Real world validity

The authors clarified that they simulate synthetic data by sampling from a reasonable distribution, and that this is necessary to prevent ethical concerns about the privacy of applicant data. Since the focus of the paper is particularly on illustrating bias in decision making, using the student admissions case with synthetic data as an example scenario is a valid approach.

- Metrics

Section 3 now includes a more complete discussion of each metric and the reasoning behind it, which I appreciate.

- Prompts

Clarifications about prompt structure and construction are included throughout the paper which gives a much better view into the approach.

- Re-writing text to prevent anthropomorphizing LLMs: The framing of the text to avoid this issue is a very nice touch and improves the overall quality.

Comments Suggestions And Typos:

N/A

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 4.5

Overall Assessment: 4 = This paper represents solid work, and is of significant interest for the (broad or narrow) sub-communities that might build on it.

Best Paper: No

Ethical Concerns:

None

Needs Ethics Review: No

Reproducibility: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 4 = Useful: I would recommend the new datasets to other researchers or developers for their ongoing work.

Software: 2 = Documentary: The new software will be useful to study or replicate the reported research, although for other purposes it may have limited interest or limited usability. (Still a positive rating)

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**



Revised paper meets expectations

Official Comment

Authors (Yao Liu (/profile?id=~Yao_Liu11), Julian McAuley (/profile?id=~Julian_McAuley1), Jessica Maria Echterhoff (/profile?id=~Jessica_Maria_Echterhoff1), Abeer Alessa (/profile?id=~Abeer_Alessa2), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission2194/Authors))

29 Jul 2024, 10:51 (modified: 22 Aug 2024, 15:15)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer CbWu, Commitment Readers

Revisions (/revisions?id=ifbwmYX4IX)

Comment:

Thanks to reviewer CbWu for all of your constructive feedback to improve the paper, and acknowledging that the revised paper meets your expectations!

Add: **Author-Editor Confidential Comment**

[Hosting a Venue \(/group?id=OpenReview.net/Support\)](#)
[All Venues \(/venues\)](#)
[Sponsors \(/sponsors\)](#)

[Frequently Asked Questions \(https://docs.openreview.net/getting-started/frequently-asked-questions\)](#)
[Contact \(/contact\)](#)
[Feedback](#)
[Terms of Use \(/legal/terms\)](#)
[Privacy Policy \(/legal/privacy\)](#)

[OpenReview \(/about\)](#) is a long-term project to advance science through improved peer review, with legal nonprofit status through [Code for Science & Society \(https://codeforscience.org/\)](https://codeforscience.org/). We gratefully acknowledge the support of the [OpenReview Sponsors \(/sponsors\)](#). © 2024 OpenReview