Journal of Visualization and Interaction

Reviews for articles-2024-da-rutabaga

Review #1

Completed: 18-09-2024 00:32

Recommendation: Resubmit for Review

Conflict Declaration

I declare that I have no known conflicts of interest with the authors.

Review

The paper introduces RutaBAGA a visualization systems that enable university reviewers analyze their admission process and identify implicit biases. The authors present the results of a controlled experiment and a case study.

1. Advancement of Knowledge in InfoVis and HCI: This paper introduces "RutaBAGA," a visualization system aimed at helping university reviewers analyze the admission process and identify implicit biases. The paper contributes to the intersection of InfoVis and HCI by exploring how visualizations can enhance understanding of complex decision-making processes, particularly in institutional settings like university admissions. However, the manuscript needs a clearer explanation of the rationale behind the design choices. While it introduces the concept of using visual analytics to reveal biases, it does not sufficiently explain why visualization is the most appropriate solution for addressing the design goals. Strengthening this connection would clarify how the work contributes to advancing the fields of InfoVis and HCI, particularly by justifying the choice of visualization as a bias-detection mechanism in the admission process.
2. Credibility of the Manuscript's Claims: The paper claims that measuring the time spent reviewing applications can indicate bias. While this is an interesting hypothesis, the authors acknowledge that time is a noisy metric influenced by many factors, which undermines the credibility of this claim. The paper would benefit from a deeper analysis of confounding factors, such as the complexity or clarity of individual applications, to provide a stronger argument for why time spent reviewing is a valid indicator of bias. In the statistical analysis, there needs to be more clarity and methodology. For instance, while the method for Hypothesis 2 (H2) is indicated, the authors do not analyze whether the dataset meets the requirements to apply the selected test (linear regression). This omission reduces the reliability of the results. Moreover, using averages instead of medians in the statistical analysis is problematic since the data does not

appear to be normally distributed, as suggested by the results presented in Figure 2. Employing a median would provide a more accurate representation of central tendency in this context.

3. Clarity of the Evidence: The evidence presented is often unclear. For instance, there were 14 participants in the case study, but only 5 are included in Table 2. It is unclear why the authors chose to present data from only 5 participants, which creates confusion about the scope and completeness of the results. A more precise explanation of why these specific participants were selected or why the data from the remaining participants was excluded would help clarify the presentation of the evidence. Additionally, the study does not provide enough information on the participants' prior experience with the tool (except for P3, who had previous exposure). The inclusion of P3 introduces potential bias into the analysis. There is also a need for more detail about how the participants used the system and how long they received instructions, making it difficult to assess the robustness of the findings. The data collection methods need to be more clearly described. For instance, in Section 8, it is unclear how user interactions with the system were logged and what data was analyzed to support the results. The quotes from participants need to be clearly attributed, leaving ambiguity about whether they stem from the Zoom-based analysis sessions or individual evaluations of the applications. In general, the presentation of results could benefit from clearer explanations of both the procedure and the data analysis.
4. Expertise to Review: I am confident in reviewing most of this paper, particularly regarding the visualization design, the statistical methods used, and the overall experimental design. However, I may not have the specific expertise required to evaluate the nuances of the university admissions process and how bias is typically studied in this context. A reviewer with more expertise in educational policy or admissions practices could provide additional valuable insights on those aspects of the work. Overall Assessment: This paper presents an interesting approach to using visualization to identify biases in university admissions. However, several critical issues need to be addressed to strengthen the credibility and clarity of the work. The rationale for using visualization as the primary tool for bias detection needs to be better articulated, and the authors should address the limitations of using time spent as a bias indicator by analyzing confounding variables. Additionally, the statistical analysis requires clarification, particularly regarding whether the dataset fulfills the assumptions for the applied tests. The case study data also needs more transparency, especially regarding the selection and reporting of participants. Addressing these points will improve the overall quality and impact of the paper on the InfoVis and HCI communities.

Openness

The study is pre-registered and provides sufficient details in the supplemental material for other researchers to reproduce the evaluations presented in the paper. This includes well-documented questionnaires, access to prototypes, and the source code for data analysis. Additionally, the manuscript ensures transparency by including a clear description of the data analysis procedure, making it easier for others to build on this work. If any artifacts were not included for ethical reasons, the paper adequately describes their characteristics

and justifies the trade-off between openness and ethical considerations.

Classification

Registered Report

Classification

Empirical Research - Quantitative

Classification

Emprirical Research - Qualitative

Classification

Systems or design research

Recommendation

Major Revisions

Revisions Requested

Related works: Revise the Related Work section to better distinguish your investigation from prior research using visualizations to promote bias awareness. Identify the scientific gap your paper addresses and outline how your approach is novel compared to existing visualization-based bias awareness tools. Explicitly state the limitations of previous approaches and explain how your work overcomes them or approaches the problem differently.

Requirements' justification: To address the limitations of relying on a convenience sample, the paper should either expand the number of participants or provide a clear rationale for why the current sample size is sufficient for the study's objectives. A more diverse pool of participants would improve the representativeness of the findings, especially given the potential variation in admissions processes across different institutions. The current sample may not fully capture the diversity of admissions practices, which could limit the generalizability of the tool's design. Involving a broader range of admissions chairs from various institutions would enhance the validity and relevance of the requirements, ensuring that the tool addresses a broader array of needs and contexts within university admissions.

Evaluation Scope: To strengthen the study's findings, it would be beneficial to broaden the scope of the case study or experiment by including more participants or involving a wider variety of users, such as admissions chairs from other institutions, faculty members, or external reviewers. This would provide a more comprehensive evaluation of the tool's effectiveness across different perspectives and decision-making roles. If expanding the participant pool is not feasible, the paper should explain why the current participants are representative enough to draw meaningful conclusions. This explanation should demonstrate how their expertise,

roles, and familiarity with the admissions process sufficiently reflect the intended user base, ensuring the results are still relevant and reliable despite the limited sample size.

Data sets: The comparison of time spent in Figure 3 needs revision, as each participant is reviewing a different set of applications, which are not directly comparable. The varying complexity and content of the applications could skew the analysis, leading to misleading conclusions about reviewer behavior or bias. To address this issue, ensure that participants review comparable sets of applications or implement normalization techniques to account for differences in complexity. This adjustment is critical to provide more accurate and reliable results in analysing time spent across different participants.

Reviewer Name

Leonel Merino

ORCID

0000-0002-5396-487X

Review #2

Completed: 25-09-2024 22:03

Recommendation: Revisions Required

Conflict Declaration

I declare that I have no known conflicts of interest with the authors.

Review

The paper presents an important and timely effort to mitigate implicit biases in university admissions through the use of visualizations. The authors introduce the rutaBAGA system, which provides visualizations of metrics such as time spent reviewing application components and ratings disaggregated by gender and race. These visualizations allow reviewers to reflect on their reviewing, helping them become more aware of potential implicit biases and adjust their ratings accordingly.

The topic is highly relevant, particularly in the current context of discussions surrounding bias in decision-making processes like admissions. The approach of using visualizations to increase awareness and potentially mitigate bias is promising, and the system addresses a critical gap in this area.

That said, I believe there are several areas where the paper could be improved or expanded:

1. Affirmative Action and its Effects: While the paper references the recent Supreme Court decision on affirmative action, it would be beneficial to include more detail on the effects of this decision, especially how it may influence future admissions practices. For instance, you could draw from recent discussions on

this topic, such as the articles in The New York Times and The Boston Globe (https://www.nytimes.com/2024/09/11/us/harvard-affirmative-action-diversity-admissions.html and https://www.bostonglobe.com/2024/09/25/metro/affirmative-action-colleges-black-enrollment/), to provide a more comprehensive context.

2. Clarification of Admissions Context: The paper mentions that the system was designed for graduate admissions, but it would help to specify whether this refers to a Master's or PhD program. These two levels often have distinct review processes, so understanding which was studied is important. It may also be useful to discuss any similarities or differences between the review processes for Master's and PhD programs, especially in the interviews that informed the system design.

3. Design Rationale: There appears to be a gap in explaining how the design requirements and goals were derived from the interviews. For instance, how was the decision made to use time spent reviewing applications as a proxy for bias? Was this suggestion driven by the interviewees, or was it based on other considerations? Providing more insight into the interviews would strengthen the connection between the formative research and the system's design.

4. Controlled Study Limitations: While the authors acknowledge two significant limitations of the controlled study (i.e., participants not being actual reviewers and the limited inclusion of only two racial categories), it would be beneficial to delve deeper into how these limitations may have impacted the study's outcomes. For instance, how might the results differ if actual admissions reviewers were recruited? Similarly, what would be the effect of including other race categories such as Hispanic and Asian applicants? Addressing these questions would provide a more nuanced understanding of the study's findings.

5. Demographic Concordance: I would also suggest considering whether concordance between the participants' demographic characteristics (e.g., race or gender) and those of the applicants influenced the ratings. This potential confounding variable should be controlled for in the analysis.

6. Hypothesis Directionality: The directionality of the hypotheses is unclear in some sections, such as on page 10 where the authors state that "the results contradict the directionality of our hypothesis." It would be helpful if the paper clearly stated the expected direction of each hypothesis upfront.

7. Results and Interpretation (H3): The results in H3, particularly in section 6.3.2, seem somewhat limited in their interpretability. For example, the authors provide only two participant examples when discussing changes made by participants. Expanding this section to include more participant behaviors or summarizing these findings with quantitative data would make this section more insightful.

8. Qualitative Analysis Counts: Including counts for the codes or qualitative analysis themes reported in Section 8 would help the reader understand the frequency of specific behaviors or insights.

9. Ethical Considerations: It raises some ethical concerns that the system was tested in a

real admissions context before being thoroughly validated in controlled settings. The case study suggests that the system influenced outcomes for some applicants, which is quite consequential. Given that the controlled study only partially supported the hypotheses, the authors should provide a stronger justification for deploying the system in such a high-stakes decision-making process. This could include further discussion on the ethical implications and how risks were mitigated.

Minor Point: 1. Since the paper primarily focuses on implicit biases, it would be more appropriate to remove mentions of cognitive biases in the abstract, as these are not addressed in the paper.

Overall, the paper presents an innovative and promising approach, but addressing these issues would significantly strengthen the work.

Openness

The controlled study was preregistered. Additionally, the supplementary materials provide details of the system implementation, the controlled study, and the case study. However, it should be noted that the paper does not include detailed documentation of the initial interviews conducted with admissions committee members.

Classification

Empirical Research - Quantitative

Classification

Emprirical Research - Qualitative

Classification

Systems or design research

Recommendation

Major Revisions

Revisions Requested

1. Expand the discussion on the effects of the affirmative action ruling, incorporating relevant sources.
2. Clarify whether the study was focused on Master's or PhD admissions, and discuss the similarities/differences in review processes.
3. Provide more detail on how design decisions (e.g., using time as a proxy for bias) were informed by interviews or other data.
4. Elaborate on how the controlled study's limitations (use of non-reviewer participants and limited racial categories) may have impacted the results.

5. Address potential effects of demographic concordance between participants and applicants on ratings.
6. Clearly state the directionality of each hypothesis, especially where results deviate from expectations.
7. Expand on results in H3, including more participant behaviors or a summary of findings with quantitative data.
8. Include counts for qualitative themes in Section 8 to provide clearer insights.
9. Justify the ethical implications of deploying the system in a real-world admissions process, given the partial support of the controlled study.
10. Revise the abstract to remove mentions of cognitive biases, focusing solely on implicit biases.

Reviewer Name

anonymous

ORCID

N/A

Review #3

Completed: 16-10-2024 13:09

Recommendation: Revisions Required

Conflict Declaration

I declare that I have no known conflicts of interest with the authors.

Review

The paper presents a system / interface that supports the graduate admissions process. It tracks time spent on candidates and review scores. It then provides feedback to the reviewer on how much time she/he spends on different subpopulations of the candidate set. It then tests to see whether the explicit visualization of this information impacts the reviewer.

The work is well suited to the overall research area of 'responsible computing'. It gives empirical insight that confronting the reviewer with his/her biases is helpful and makes a difference. On that note, I encourage you to publish this manuscript.

However, I do also see shortcomings. My main concern is really with the design of a reviewing system.

The system design that supports the reviewing process is all but trivial. That being said, the current system design suffers from some more apparent troubles. Any introductory HCI course teaches parallel prototyping and iterative design approaches. However, from what I get from the paper, this iterative loop was cut short, and no parallel prototypes were designed:

"Based on our understanding of the program needs, we next sketched several solutions and built a preliminary prototype to ground the discussion in a second interview."

I know these kinds of things are labor intensive, but it raises the question of how important the system's design is for the underlying research questions. This should be discussed a bit more. I expect the results to change when the system is inconvenient. It seems that some committee members didn't use it at all.

That being said, I found the (qualitative) case study very insightful and enriching. Many of my concerns were addressed there. E.g., "Of course, we don't spend the exact same time on all applicants".

A second concern of mine is the little engagement with the literature on assessment that exists. See, e.g.,

Bias, Fairness, and Validity in Graduate-School Admissions: A Psychometric Perspective Sang Eun Woo et al. Perspectives on Psychological Science 2023, Vol. 18(1) 3–31

Another minor issue is that I would appreciate knowing which participants (e.g., in Figure 3) were male and female. Do you know if there is a correlation to be observed?

Would a percentage plot in addition to Fig 4 make sense?

Lastly, I would appreciate it if you were to go through the references. There are several issues with capitalization, and it would be nice to avoid "et al." and provide a complete list of authors.

Openness

The paper seemed to refer to supplemental material. However, I was not given access to such material. I hope I didn't overlook anything. In general, it would be nice if more details would be available, like - the tool itself - questionnaires and the detailed results - interview transcripts etc.

I appreciate the pre-registration for the study!

Classification

Empirical Research - Quantitative

Classification

Emprirical Research - Qualitative

Recommendation

Minor Revisions

Revisions Requested

As mentioned, I would love: - more thoughts on the design of an actual system or how it (the quality of the reviewing system) would impact your research. - a situation of this work in the context of graduate admissions research - some more additional supp material - a minor copy-edit pass.

Reviewer Name

anonymous

ORCID

N/A

Metareview

Completed: 2024-12-15 09:00

Recommendation: Accept

Conflict Declaration

I declare that I have no known conflicts of interest with the authors.

Review

The authors have responded thoroughly and constructively to the reviewers' comments in both their response letter and in their manuscript. In particular, the authors have addressed the comment on missing prior work and details about their methodology and the participants. They have also significantly improved their discussion to contextualize their work and its limitation better as well as other important factor regarding the system and its use, e.g., potential ethical consideration of using such a system in real-world applications. Since none of the concerns raised initially were critical, it was decided by the Associate Editor that the manuscript could be accepted as is after the revision round. Due to an error in the submission system, this meta-review and answer to the authors has unfortunately disappeared. As such, the acceptance meta-review was written by the JoVI responsible organizer, Lonni Besançon, and approved by the original Associate Editor, Michael Sedlmair.

Reviewer Name

Michael Sedlmair

ORCID

0000-0001-7048-9292