Proposing the use of an "Advocatus Diaboli" as a pragmatic approach to improve transparency in qualitative data analysis and reporting

Judith Friedl-Knirsch*
University of Applied Sciences Upper Austria

ABSTRACT

Qualitative data analysis is widely adopted for user evaluation, not only in the Visualisation community but also related communities, such as Human-Computer Interaction and Augmented and Virtual Reality. However, the data analysis process is often not clearly described and the results are often simply listed in the form of interesting quotes from or summaries of quotes that were uttered by study participants. This position paper proposes an early concept for the use of a researcher as an "Advocatus Diaboli", or devil's advocate, to try to disprove the results of the data analysis by looking for quotes that contradict the findings or leading questions and task designs. Whatever this devil's advocate finds can then be used to reiterate on the findings and the analysis process to form more suitable theories. On the other hand, researchers are enabled to clarify why they did not include this in their theory. This process could increase transparency in the qualitative data analysis process and increase trust in these findings, while being mindful of the necessary resources.

Index Terms: Qualitative data analysis, evaluation methodology.

1 INTRODUCTION AND METHODOLOGICAL BACKGROUND

The advocatus diaboli (AD), while originally being an advocate against the canonization of a saint [5], is also understood as a rhetoric technique. It has for example been explored as a method to avoid group-thinking in focus groups by reflecting a different perspective to the group [11], in improving managerial decision making [12], or in finding and defining negative constraints for ontologies [2]. In this proposed methodology, the AD would be used to identify statements, behaviours or artefacts that (seemingly) contradict the results of the data analysis. The method presented in this position paper is not yet a fully fledged and tested process. It is an early hypothetical idea meant for discussion and further improvement before being adopted in studies.

There are two main directions for qualitative data analysis that are also used in the context of Human-Computer Interaction (HCI): the positivist approach of coding based on theory or a-priori codes and the interpretivist approach of emergent coding. Both of these approaches are based on philosophical directions that hold different views of the meaning of the data and its analysis. From the positivist perspective, there is an objective reality within the qualitative data that can be uncovered. Therefore, two researchers coding the data will reach very similar conclusions. This also allows for the use of inter-rater reliability. The interpretivist perspective, on the other hand, is based on the assumption that there are multiple ways of interpreting the data and results are always subjective to the researchers and their biases as well as the overall target of the research question [1]. The proposed method is based on the interpretivist approach to qualitative data analysis, with the base assumption being that there is no objective truth but rather an emerging subjective answer to a research question. To generate this result it is therefore necessary that the data analysis is conducted by a researcher with experience in qualitative data analysis (QDA) and knowledge about the domain. However, since the assumption is that all results are based on interpretation, the data can also be interpreted differently. The interpretation of the data and its context is therefore influenced by the biases of the researcher. The AD method now aims to reveal biases and enhance transparency of their thought process by contextualising statements that seemingly contradict the results of their data analysis, by bringing in an outside perspective. There is early psychological evidence that suggests that the process of "considering the opposite" is effective in reducing social bias [9, 7]. This also brings in a small part of the positivist idea, as this AD is an independent researcher who does not possess the contextual knowledge of the main researcher. Their role in the QDA process is to challenge the interpretation based on the same data.

In the past, there has been a discussion around the value of qualitative methods [6] and how researchers with a positivist perspective tend to dismiss qualitative approaches [8]. While quantitative methods are still prevalent, there also seems to be increasing recognition of qualitative methods. While a review of cross-virtuality analytics found barely any qualitative studies [4], a recent review of user evaluation in immersive analytics found that 57 of the included 231 studies were qualitative and further 25 used a mixed methods approach [3]. However, the respective data analysis processes were not analysed. QDA is an iterative process, yet the results rarely reflect this. By transparently reporting the process we could make it more clear what goes into QDA and it could make the results more comprehensible to researchers not as familiar with this process.

2 PROPOSED METHOD

After the qualitative data is collected, the main researcher conducts the qualitative data analysis. This person should be experienced with QDA, have the research question in mind and has enough background knowledge of the problem domain to produce meaningful results. This first data analysis follows the standard procedure of the chosen methodology, e.g. thematic coding. After the main researcher has reached conclusions, ideally underpinned with participant statements, another researcher is included. This second researcher takes on the role of the AD. They do not need to be experienced with QDA as their purpose is not to produce alternative results, but rather to find statements that (seemingly) do not fit the results of the main researcher. In contrast to the review process for submissions, the AD checks the validity of the qualitative results on the same data that these results are based on. The results of the AD process are then discussed with the main researcher. The main researcher decides for each instance of contradictory material whether this is truly incompatible with the results or whether there is an explanation for it in the context. Depending on this decision, they will then either adapt their results and include the finding of the AD in their theory building process or they contextualise a statement. For example, a participant in a user study might have a different opinion on a functionality of the tested system, based on misunderstanding or misinterpretation. This can be observed during the study and even revealed in the interview. Nevertheless, the first contradictory statement still exists. With this method it would be possible to transparently reflect the QDA process and contex-

*e-mail: judith.friedl-knirsch@fh-ooe.at

Table 1: Example of a table that could be used to analyse and later report the findings of the AD

AD #	Participant	Contradictory Statement	1st Round Result	Influence on Final Result

tualise the data, but the main researcher also has the opportunity to reflect on their biases to form stronger theories and more fitting results.

3 Reporting the AD

Since page limitations often inhibit researchers from stating every detail of their analysis, it could be helpful to include only the most relevant results of the AD method, which reflect the process of getting to the results. This should surely include the iterative nature of the data analysis that is sometimes lost in the reporting of data. This should reflect everything a reader needs to know to understand how the results are grounded in the data. However, while there is most likely not enough space to report every finding of the AD, this could be included as a table in the appendix or as supplemental material. While this might constitute more work when submitting, it is actually a table that will probably be generated in the analysis process anyways. This table should contain the contradictory statement, what part of the first round of results it contradicts, a participant code of the respective study participant and how it influenced the results or why it did not. This table could look similar to Table 1.

4 Exemplary Application

While this approach has not yet been tested and evaluated in practice, this section provides examples how the AD method could have been used in already published research. This is a hypothetical demonstration, intended to clarify the potential application of the AD method based on user studies that have already been conducted.

4.1 Vock et al. 2021 [13]

In this publication, a mainly qualitative approach is used to evaluate a prototype for interactive dashboards in AR. The prototype was designed to monitor the quality of the collected data in mobile intervention studies. The study was conducted with 15 domain experts who were presented with a real scenario from their research. They were then tasked to identify irregularities in the data, handle these errors and communicate this to their participants or team members.

In the analysis of their data, Vock et al. [13] use thematic analysis and even reference the specific approach they are referring to. The coding is then completed by one researcher to ensure consistency. The reporting of the qualitative results is then structured along their research objectives interlaced with quotes from their participants. At the end of each research objective, they provide a summary of implications for research and design.

An AD would read through these findings, with especially focusing on the implications and on summaries of findings. For example, there could be opposing opinions in the interview data to the statement *"Participants preferred the combination of head gaze and touch input, [...]"* [13]. This statement is most likely based on a quantitative observation of the log data, without a confirmatory quote. Some participants could, for instance, explicitly mention that they did not like this specific combination of interaction modalities or that they did not like the multimodality of interaction techniques at all. In the AD protocol they would then highlight this sentence. Afterwards the primary researcher would go through the AD protocol and could decide whether this statement that the AD found was enough to rethink this specific finding. If not, they could explain in the last column of the AD protocol why this statement was not deemed relevant enough to change the findings. For example, the same participant could have made a contradictory statement at a different time, or the statement was based on a misconception.

In this case, where the coding was performed by a single researcher, the AD method would actually add another step to the data analysis process. However, it could help the researchers to reflect on their biases in the analysis process, without having another researcher repeat the complete coding procedure.

4.2 Luo et al. 2023 [10]

A preliminary expert review is used to evaluate system for the analysis of human movement data in situ using mixed reality. The study utilised guided walkthrough tours to demonstrate the features of the prototype to the experts. This was then followed by an open ended discussion of two usage scenarios the system was based on, as well as a post-study interview.

For the data analysis, Luo et al. [10] again used thematic analysis and cited the specific approach they were referring to. The data for the thematic analysis was collected using observations, think-aloud comments and post-study interviews. The thematic analysis was then followed by a cross-validation of two authors. It is, however, unclear what Luo et al. [10] mean by "cross-validation". This leaves open the question whether the two authors performing this process actually looked at the data, if they looked at all the data, or whether they compared the findings of the thematic analysis to their own recollection or notes of the study. Having a specific process like the AD method could help structure such a process.

In the reporting of Luo et al. [10] the structure is then again defined by the themes. This analysis also includes confirmatory direct quotes from study participants. As opposed to the first example by Vock et al. [13], Luo et al. [10] do not provide a short summary with implications for each of the themes. However, the discussion section reiterates and summarises some of the results. Therefore, the AD would not only go through the results but also the discussion section to identify the most central findings. To look for contrasting views in the data, the AD could then go through the video recordings or the transcripts. They could also use the search function to find specific terms in the transcripts related to what they read in the results section. For example, in the work by Luo et al. [10] the use of lenses for filtering is discussed both in the findings and the discussion. The AD could then use the terms *filter* and *lens* to look for instances in the transcripts where this is talked about. In this type of research with only four experts involved in the analysis, it could also be interesting for the AD to check for statements where a specific expert was not mentioned. For example Luo et al. include the following statement about their filtering methods *"Both the actor-level (E1, E4) and data-point level filtering (E2-3) based on the Lenses were considered useful."* [10]. It could, for example, be interesting to check for what E2 and E3 statements on the actor-level filtering methods. Are there contrasting opinions about this feature or did it simply not come-up during the interview process. The primary researcher can then again chose to rethink their findings or to specify in the supplemental AD protocol, that this topic was not discussed with E2 and E3.

Therefore, the AD process could help to formalise the vague process of "cross-validation" in this publication. It could also raise potentially interesting questions about the findings and it could make the analysis process more transparent by including specifications on the content of different interviews.

5 Discussion

This approach is primarily intended as an actual method in qualitative data analysis. It is intended to bring two schools of thought in epistemology, the interpretivist and the positivist side, a little bit closer together or at least make their viewpoints more transparent and comprehensible. One (interpretivism) sees data analysis as always subjective. There is no objective truth because all researchers

and participants are inherently subjective. The other (positivism) sees data analysis as trying to find the objective truth in the subjective data.

Nevertheless, there is also a very pragmatic component to the method of the AD. A classic analysis variant would be, for example, for two researchers to carry out the iterative data analysis process independent of each other, e.g. using thematic coding. The results are then compared and the inter-rater reliability can be specified. Therefore, it takes two researchers who are familiar with the research topic and QDA and who carry out the complete analysis and then discuss the specific result together. Realistically, even having access to two such researchers might be difficult and the dual analysis takes up a lot of time for these researchers. At the very least this QDA process is currently often not reflected in the reporting in qualitative results.

In the AD method, the first data analysis would be carried out by a researcher who is experienced with qualitative methods. The second researcher, who then takes on the role of AD, does not have to be able to do this, but takes the result from the main researcher and tries to find contrasting examples in the data. They do not face the pressure of creating a conceptually sound data analysis but instead act as a detective that tries to uncover plot holes in a story. These examples can then either be explained by the context, i.e. why they do not contradict the results of the main researcher, or it enables the realisation that the results do not quite fit the data yet. In this case this finding of the AD is integrated into results.

This also shares similarities with a review process, as there the intention is also to retrace the results when judging how appropriate the methods of a submission are and how strong the contribution is. However, in the current review process there is mainly not enough time for the reviewers to this, even if the full data is and can be published. Additionally, the AD method is part of producing the results instead of a review process. This is also the reason why the AD should be actually interested in finding contradictory statements. By doing so the results can be more meaningful, the reasoning of the results can be more transparent, and the publication can be more valid and understandable.

Overall, the benefits of this method could be:

- The formalisation of a pragmatic approach for QDA that is easy to follow and enhances the transparency in reporting the methodology of QDA.
- The process takes up less resources than double coding, as the AD process is less intensive than a second coding process and can be done by a researcher who is not proficient in QDA.
- Reflecting on biases in the QDA process that might even go beyond inter-rater reliability as two researchers working at the same problem with a similar background might be prone to think alike anyways.

6 Limitations

As for every method, there are limitations to the proposed AD method. This method is mainly suitable for smaller amounts of qualitative data and in instances where resources in the form of researchers experienced in QDA is sparse. However, when there is a more complex and in-depth qualitative study with several different types of data such as interview transcripts, videos of observations and other artifacts, the AD method might not save any time. Moreover, trying to go through all this evidence may be difficult for researchers who are inexperienced with QDA.

Additionally, while this method of looking for contradictory statements and artifacts may increase the transparency and make reasonings more comprehensible, there is no way of knowing whether the AD followed the procedure thoroughly. There is only indication to this in what researchers report in the publication and possibly a table of all findings in the appendix or supplemental material.

Having an AD that is not as familiar with the study and the collected data could also lead to them overestimating or underestimating the relevance of specific contradictory statements they identified. Therefore, it is first necessary to make sure that the AD understands the findings and for the AD to document all seemingly contradictory statements they find and leave the judgement on their relevance to the experienced primary researcher. However, having an inexperienced researcher that is not perfectly familiar with the context can also help uncover otherwise hidden insights, as they add a different perspective that is not yet influenced by the immersion into the problem domain.

Finally, as mentioned in the context of reporting, page limitations restrict how much details on the QDA process can be included in a publication. There are workarounds by submitting an appendix or supplemental material, however, this also comes with the additional workload for the researchers of preparing these materials.

7 Future Work

Since this is an initial concept for using an AD in QDA, the first and most relevant future work is the application of the proposed methodology in a real research scenario. This could be done in the form of a case study, to demonstrate how the method can be described and reported in practice and to describe the benefit and drawbacks when compared to traditional methods. Furthermore, a study including several groups of qualitative researchers can be used to compare the results of QDA with different types of methodologies, including the proposed AD method.

Another aspect of future work that can be explored after exploring the usefulness of the method itself, is the improvement of the procedure. It could be possible to train a large language model (LLM) to look for contradictory statements in transcriptions of interview and observation data. However, this has to be done very carefully as each LLM is biased by its training data. However, it could be an interesting approach, as this does not require complete trust in an LLM. The analysis itself would still be performed by a skilled researcher, then the LLM would try to identify counterarguments, and afterwards the researcher checks the results of the LLM again. So this would only be used as a tool in a human lead and supervised process. Additionally, this process needs to specifically take care of data protection and compliance with its respective regulations, like the General Data Protection Regulation in the European Union.

8 Conclusion

In conclusion, this position paper presents an early concept of how a data analysis procedure including an "Advocatus Diaboli" could look like. Such a process could improve the transparency in reporting the qualitative data analysis process and make it easier to comprehend how these results relate to the collected data. It could also serve as a pragmatic process for data analysis where it is difficult to employ two researchers who are experienced with qualitative data analysis methods.

Acknowledgments

This publication is a part of the X-PRO project. The project X-PRO is financed by research subsidies granted by the government of Upper Austria.

References

[1] A. Blandford, D. Furniss, and S. Makri. *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers, 2016. 1

[2] S. Ferré and S. Rudolph. Advocatus Diaboli – Exploratory Enrichment of Ontologies with Negative Constraints. In D. Hutchison, T. Kanade,

J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, A. Ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez, eds., *Knowledge Engineering and Knowledge Management*, vol. 7603, pp. 42–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Lecture Notes in Computer Science. doi: 10.1007/978-3-642-33876-2_7 1

[3] J. Friedl-Knirsch, F. Pointecker, S. Pfistermüller, C. Stach, C. Anthes, and D. Roth. A Systematic Literature Review of User Evaluation in Immersive Analytics. *Computer Graphics Forum*, 43(3), 2024. 1

[4] B. Fröhler, C. Anthes, F. Pointecker, J. Friedl, D. Schwajda, A. Riegler, S. Tripathi, C. Holzmann, M. Brunner, H. Jodlbauer, H. Jetter, and C. Heinzl. A Survey on Cross-Virtuality Analytics. *Computer Graphics Forum*, 41(1):465–494, Feb. 2022. doi: 10.1111/cgf.14447 1

[5] J. Gray. *The Evolution of the Promoter of the Faith in Causes of Beatification and Canonization: A Study of the Law of 1917 and 1983*. PhD thesis, Pontifica Universitas Lateranensis Facoltas Iuris Canonici, Rome, 2015. 1

[6] S. Greenberg and B. Buxton. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 111–120. ACM, Florence Italy, Apr. 2008. doi: 10.1145/1357054.1357074 1

[7] A. G. Greenwald. The open-mindedness of the counterattitudinal role player. *Journal of Experimental Social Psychology*, 5(4):375–388, Oct. 1969. doi: 10.1016/0022-1031(69)90031-6 1

[8] T. Isenberg, P. Isenberg, Jian Chen, M. Sedlmair, and T. Moller. A Systematic Review on the Practice of Evaluating Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2818–2827, Dec. 2013. doi: 10.1109/TVCG.2013.126 1

[9] C. G. Lord, M. R. Lepper, and E. Preston. Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6):1231–1243, 1984. Place: US Publisher: American Psychological Association. doi: 10.1037/0022-3514.47.6.1231 1

[10] W. Luo, Z. Yu, R. Rzayev, M. Satkowski, S. Gumhold, M. McGinity, and R. Dachselt. Pearl: Physical Environment based Augmented Reality Lenses for In-Situ Human Movement Analysis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15. ACM, Hamburg Germany, Apr. 2023. doi: 10.1145/3544548.3580715 2

[11] C. MacDougall and F. Baum. The Devil's Advocate: A Strategy to Avoid Groupthink and Stimulate Discussion in Focus Groups. *Qualitative Health Research*, 7(4):532–541, Nov. 1997. Publisher: SAGE Publications Inc. doi: 10.1177/104973239700700407 1

[12] C. R. Schwenk. Devil's Advocacy in Managerial Decision-Making. *Journal of Management Studies*, 21(2):153–168, Apr. 1984. doi: 10.1111/j.1467-6486.1984.tb00229.x 1

[13] K. Vock, S. Hubenschmid, J. Zagermann, S. Butscher, and H. Reiterer. IDIAR: Augmented Reality Dashboards to Supervise Mobile Intervention Studies. In *Mensch und Computer 2021*, pp. 248–259. ACM, Ingolstadt Germany, Sept. 2021. doi: 10.1145/3473856.3473876 2