

Interactive Intelligence

Checklist for Review of Dataset

(Version 2)

We recommend that students or employees wishing to publish on their data and results for a given research project in the form of a dataset asks a fellow student or colleague to review this dataset with regard to the points in this checklist. The purpose of the checklist is to ensure that all data that can be made available is made available, that all analyses were conducted conscientiously by the researchers, that all results are reported accurately, and that all methods are transparent and sufficiently clear to be reproducible.

If you choose to have your code reviewed according to this checklist, we advise you to upload this document together with your dataset to the research data repository of your choice (e.g. 4TU Research Data) upon publication of your work.

I. Basic Data

Paper title:	German and Dutch Translations of the Artificial-Social-Agent Questionnaire Instrument for Evaluating Human-Agent Interactions
Name(s) of researcher(s):	Anonymous
Name of the reviewer:	Anonymous
Data repository platform (e.g. 4TU Centre for Research Data):	4TU Centre for Research Data

II. Checklist

Statement	Yes	No
1. The dataset contains a README file that fulfils the requirements of the data repository platform that the researcher wishes to use. If no such requirements can be found, the dataset nonetheless contains a README file that clearly explains the contents of the dataset?	yes	
2. Either within the README file or within an extra, easily findable file, the researchers have explained their data. This means that, for example, for every column of a tabular dataset, all column names and possible cell values are explained.	yes, explanation for the csv columns are given in xlsx files named analogously.	
3. data is in readily readable file formats. If this should not be the case, the README (or similar) clearly explains the file format and which software can be used to access the contents.	Yes, csv file and additionally needed, linked data in xlsx	
4. All data has been anonymized in accordance to promises made in the Data Management Plan.	yes	
5. The analysis file or files contain a header with meta-data (name of author, date of writing, required input files and generated output files).	yes	
6. All required input files for the analysis are available in the dataset.	yes	

Statement	Yes	No
7. There is an output file that is generated by the analysis script that neatly combines code and commentary (e.g. markdown output file). This output file is in a readily readable file format (e.g. pdf).	yes	
8. The analysis script is clean and comprehensible in the sense that: <ul style="list-style-type: none"> • There is sufficient, useful, and clearly written commentary • Irrelevant code (such as old analyses) has been removed • The details of analyses that are not reported in the paper (e.g. assumption checks) are proportional to those that are reported in the paper. This means that unreported analyses should not clutter up the script, making it long and unreadable. 	Yes, just very few unneeded codeblocks are left, which will be removed	
9. The analysis script can be run successfully.	Yes, only MCMC did not finish suvcessful	
10. All preprocessing steps are clearly described and traceable, especially when preprocessing code cannot be executed because raw data is not available.	yes	
11. The analyses and results reported in the manuscript can be found back in the analysis script with labels according to where they appear in the manuscript.	yes	
12. All results reported in the manuscript accurately correspond to the output produced by the analysis script.	yes	
13. Results reported in the manuscript can be reproduced via alternative means of computation.	Yes, I reproduced means and SDs of data using a newly created script and doing the data composing again by hand. Difference values could not be reproduced since the Bayesian values are not directly the differences of the primary data	

III. Additional comments by reviewer

Please state any additional things you noticed in reviewing the dataset or possible points of improvement for the reviewer.

Comments for review round 1:

Should we maybe remove the building description for the docker image from the readme, since we actually want them to use your prebuild image, since otherwise version can once again be different. We can nevertheless provide the Dockerfile as is for reference. I however, was unable to build it. Before getting this far I also had to change the \"into ' in line 40. But this is not an issue if everyone should use the provided image. -> This is strange as I can build the image without problems from scratch. So I would like to keep the building instructions for those who would like to try it. But I will add a comment that in case this fails, people should just pull the Docker image from Dockerhub.

* In "Data Transformation Summative first half"

* lines 109 to 111 are not needed, and should be remove (same is true for the Dutch first half transformation) -> Fixed!

* In "Transformation-Culture_from_raw_English"

* line 63/232/373 should be "AU1: AU3" instead of "AU1:AU2,AU3" for consistency (no different result as

far as I observed) [this also appears e.g. in -> The current format is flexible to there being other columns between AU2 and AU3. I checked and there do not seem to be other columns between AU2 and AU3 in the three data frames that we load, so I fixed this as well as you suggested.

* In GermanSummative"

* the comment in line 89ff should be removed as it is "identical" as lines 94ff -> Fixed!

* lines 70 and 82 are superfluous as the select in line 122 can be done directly from data01 and data02, as d1 still contains the attention checks for one language, which is confusing -> Done, both for GermanSummative and DutchSummative.

* In "cultural_comparison"

* the "Running MCMC with 4 parallel chains" always finishes unsuccessfully, saying " cannot open the connection" Can you double-check that this works with the docker image (I run the docker image which I pulled) -> I also get this when I knit this file in RStudio, so I have to knit it via the terminal instead. I will add this as a note to the readme.

* In the "Readme.pdf"

* In section "Types of analyses and content" the numbered enumeration 1),2),3) is not shown correctly, but for all numbers apart from 1) appended to one of the other points -> Fixed!

IV. Review log

Round	Date	Paper Status	Checklist Items	Signature Reviewer	Signature Researcher
1	15.03.2024	writing	all		
2	19.03.2024	writing	13		