**Assignment 2**

**Overall task**

# During Assignment 2, you design, conduct, analyze, and report a user-study based on the chatbot developed during Assignment 1. After designing and implementing a system (Assignment 1), you will now test whether your system also significantly improves some aspect of human/user experience in different operating conditions. Assignment 2 consists of four parts:

1. Design: What can be tested about your system? (within the constraints set by this course). Why is this an appropriate test and what is your hypothesis?
2. Method: How can you test this hypothesis in an experiment? What are the necessary characteristics of participants, materials, design, procedure, and output metrics?
3. Statistical analysis and visualization: With what test do you test whether there is indeed the hypothesized difference? How can you communicate your findings most effectively?
4. Write-up: Write-up your text in various sections.

Students are encouraged to write throughout the lab series.

**Learning goals for Part 2**

1. Designing and conducting an experiment with human participants
2. Analyzing empirical data using statistical techniques, visualization.
3. Writing parts of a scientific report about your system's empirical evaluation.

Besides: getting a first impression on experimental steps, scientific writing, critical reflection and discussion. The emphasis in this part is on methods, data collection, analysis, visualization, and critical discussion.

**Planning and feedback**Feedback by peers and lecturers on in-between steps/documents will be formative, i.e. ungraded. When asked to give feedback on others, please be critical and constructive (“everything great" is not proper feedback).

Teachers:  
Hans Marien

Christoph Strauch

Ruud Hortensius

Leendert van Maanen

**Session 2a**Tue 10-10-2023 - 13:15–15:00

Prior:

* Check the research questions (below). Prioritize these research questions in how interesting/feasible you find them as a group. ***Bonus***: Think of a different research question that you can motivate and investigate well. Innovative (but feasible) questions score extra points.

During:

* Pick a research question and discuss your choice with the workgroup teacher – get approval of your question
* Relevance/intro: *Why is your research question and approach to it   
  relevant?*
* Participants: *Who are you going to test and how many?*
* Experimental design: *How do you want to test your research question? (within/between, number of trials etc.)*
* Materials/measures: *Which materials do you want to use to assess data?*
* Task/procedure: *Which tasks do participants need to do?*

Post:

* Create a document that will later be your report in which you sketch sections with your plans for the previously mentioned points
* Acquire potential participants
* *Deadline*: Thu 12 Oct., 23:59: Submit the interim document to your **peer-group** (Teams channel)
* *Deadline*: Mon 16 Oct., 23:59: Provide feedback with the first feedback document to your peer groups’ on the day before Session 2b   
  (**see Appendix B1)**

Research questions/topics to choose from (you can adjust the outcome metrics and phrasing if you like):

* Effects of text-to-speech on user satisfaction or efficiency
* Effects of communication style on perceived chatbot quality
* Effects of system response delay on perceived humanness and usability
* Are explicit confirmations benefitting user experience in chatbots?
* Are more anthropomorphic chatbots preferred over less anthropomorphic chatbots?
* Are loading progress bars helping perceived usability in chatbots?
* Do adaptive system delays increase perceived humanness?
* Do personalized message increase perceived usability of chatbots?

**Session 2b**Tue 17-10-2023 - 13:15–15:00

Prior:

* Take a look at the peer feedback, think of how to integrate it and which points remain unclear
* Think of open questions for workgroup teacher and peer group to help yourselves

During:

* Refine all aspects relating to experimental design, participants, measures etc. By the end of this workgroup, you should have a specific plan. Discuss this plan with your workgroup teacher
* Implement experiment and pilot experiment yourselves, adjust experimental design accordingly
* Look [here](https://www.uu.nl/en/research/institute-of-information-and-computing-sciences/ethics-and-privacy/guidance-for-research-master-thesis-students.) at the ethics [*Quickscan (Word-version)*](https://www.uu.nl/sites/default/files/QuickScan-Survey.docx):   
  Are there any critical aspects to your study that would necessitate discussion with your workgroup teacher?
* Add the informed consent (**see Appendix C**) to your study and make sure participants are informed about your study goals and data-management before they give consent.

Post:

* Finalize implementation
* Run the experiment and collect the data with your participants. Check data early on to avoid data loss

**Session 2c**Tue 24-10-2023 - 13:15–15:00

Prior:

* Do data preprocessing
* Prepare statistical analysis
* Prepare (planned) visualization of findings
* Think of open questions for workgroup teacher

During:

* Data preprocessing
* Refine statistical analysis
* Refine data visualization. Finetune captions, labels, and all other aspects of the figures***Bonus***: Use sophisticated visualizations (e.g., violin plots)

Post:

* Continue work on previous points, catch up if you lack behind
* Integrate results (including your selection of analyses, visualizations) into the report
* *Deadline*: Thu 26 Oct., 23:59. Submit the version of the report to your **peer-group and workgroup teacher**
* *Deadline*: Mon 30 Oct., 23:59. Provide feedback with the second feedback document to your peer groups’ plans latestly the evening before Session 2c   
  (**see Appendix B2, page 26)**

**Session 2d**Tue 31-10-2023 - 13:15–15:00

Prior:

* Take a look at the peer feedback, think of how to integrate it and which points remain unclear
* Take a look at the feedback from the workgroup teacher, think of how to integrate it and which points remain unclear
* Think of open questions for workgroup teacher and peer group to help yourselves

During:

* Discuss and incorporate peer feedback
* Discuss and incorporate feedback from your workgroup teacher on the final report
* Work on the (text of the) final report

Post:

* ***Deadline*: Wednesday 9 Nov., 23:59.** Submit final report

**Final report/grading**

**Report requirements**

Your report for assignment part 2 should be no longer than 3500 words in total (excluding references and appendix). The document needs to be formatted consistently and (e.g., according to APA guidelines. You can find an APA template in MS word (New document-> online templates ->APA7 template; Latex: https://rpubs.com/YaRrr/apalatex)). Unlike APA, put your Figures and Tables in the text between paragraphs (not at the end of the document).

The paper needs to contain the following explicit sections:

* Title page including group number and authors
* One to two introductory paragraphs
* Methods (with appropriate subsections)
* Results (with appropriate subsections in case you analyze multiple dependent variables). Include a selection of sensible visualizations and tables (guideline 1-4 figures and tables in total).
* General discussion
* References

**Detailed information per section**

In general, we expect a minimum number of typos and grammatical errors (as this is a group project, we expect that you benefit from each other’s ability to proofread). Writing should be concise and argumentation flow should be correct in general.

Title  
Give a title that is about the content (e.g., "A Computational Model of the Wason Selection Task"). The title should reflect the content/topic/research question of the report (i.e., not: "Final Report of Methods in AI Research").  
  
Introductory paragraphs and research question   
Concisely (1) what is the broader context or problem that you try to address, (2) what is the smaller, concrete problem or issue that you are addressing, (3) why do we need to know the answer and how would that answer matter.

Criteria:

* Does the introduction form a coherent argument?
* Does this argument lead to a clear, theoretically and/or practically supported motivation for the research approach?
* Embedding in context: is the work embodied appropriately in a wider set of literature on relevant topics?
* Is the research question creative/exceeding the list of existing questions

Methods

You need to follow the format and conventions of the APA publication manual and have explicit subsections on:

* Participants
* Materials
* Design
* Procedure
* Measurements

**Mention that you followed the ethics checklist (see** [**Quickscan**](https://www.uu.nl/sites/default/files/QuickScan-Survey.docx)**) and obtained informed consent (see Appendix C).**

The APA publication manual (see Appendix A, page 22) provides very clear pointers on what type of information is required in each subsection. You can also take a look at example papers to understand the level of detail that is required. The goal of this section is to enable readers to understand the method easily and to be able to replicate the experiment.

The grading for this section takes the appropriateness of the chosen method to address the research question into account, and the correct execution of the experiment. Furthermore the amount of effort and (appropriate forms of) creativity in the design are used to determine the grade.

Criteria:

* Conventions: Are appropriate conventions followed for (each element of) the methods section? For example, is each subsection written down explicitly and concisely, could the study be exactly replicated?
* Set-up of study: Overall, was an appropriate method chosen to address the research question?
* Set-up of study: Was care taken in execution of the project? Note that this grading criterion (set-up) therefore takes into account effort & (appropriate forms of) creativity in the design of the study.
* If the methods are complex, are they supported by a good visualization?

Results

* Appropriate descriptive statistics, e.g., mean and standard deviation or other relevant statistics. Follow APA 7 rules for reporting.
* A statistical test to support your inference, e.g., 95% CIs or other relevant and appropriate metrics
* Appropriate use of figures and/or tables
* Appropriate interpretation of statistical effects
* Appropriate visualization

Criteria:

* Descriptive statistics: Are appropriate descriptive statistics provided (e.g., mean, standard deviation, 95% CI or other relevant and appropriate metrics)
* Inferential statistics: Are appropriate inferential statistics provided?
* Inferential statistics: Are these reported according to conventions (i.e., no “dump of software output”)
* Inferential statistics: Are these interpreted correctly?
* Visualization: Is an appropriate visualization of results used? (e.g. *standard*: bar plot with error bars. *better*: violin or raincloud plots with individual data points)
* Visualization: clear labels and captions, appropriate quality and well sized fonts

General discussion   
First summarize what your research question was and how you addressed it (brief in one paragraph).

(2) what are implications for theory and/or practice (i.e. ideally relate back to the intro paragraph(s)). (3) limitations and how you would do this different with more time and resources if you had more time and/or resources (i.e., future work) (4) conclusion.   
The discussion can be considered the part where you can shine the most: Don’t be afraid to express opinion (just mark it as such and argue well). Put results into broader context and discuss what they imply for the field and other researchers. Are the identified limitations indeed the most important?

At the end of this section a concise conclusion should be presented that answers the research question. Note that in the context of this course it can also be the case that you cannot answer the research question yet - for example due to null results or an inconsistent pattern of results - in which case your conclusion should reflect that there is no definitive answer. Which steps would be needed here to get to more definitive answers?

Criteria

* Coherence, theory, implications: is the general discussion coherent with the introduction?
* Coherence, theory, implications: are results related back to the literature in an appropriate and correct way?
* Limitations: are limitations of the experimental set-up reported clearly?
* Limitations: are these indeed the most important limitations?
* Conclusion: Is there a concise conclusion that answers the research question? Note: in the context of this course it can also be that you cannot answer the research question yet  (e.g., due to null results or an inconsistent pattern of results) in which case your conclusion should reflect that no definitive answer is there yet

Use explicit headings to ensure you capture all the required elements, which are graded separately.  
  
References  
Format according to APA guidelines.

**Appendix A**

**Resources related to the report** For introduction and general discussion of your report, you might benefit from these resources:

* Slides of the Writing lecture
* [Stanford encyclopedia of philosophy, entry on AI](https://plato.stanford.edu/entries/artificial-intelligence/)
* Brachman & Levesque (2004) *Knowledge Representation and Reasoning*. Morgan Kaufmann (note: this book is as e-book available through the University library, for example through [WordCat](https://utrechtuniversity.on.worldcat.org/discovery" \t "_blank))
* but you might want to check [this blog](https://abrilliantmind.blog/scientific-writing/?utm_campaign=meetedgar&utm_medium=social&utm_source=meetedgar.com) as well.

**Resources related to methods, statistics, and writing** When it comes to statistics, people often dive into the quick and dirty “how” do I calculate some statistic. I (Chris) want to encourage you to also think about “why” you do a specific test. This is needed because (1) you typically need to explain this in a report, and (2) you need to know what you’re doing. That said, here are some resources.   
  
**Books and articles**

* Cairns, P. (2016) Experimental Methods in Human-Computer Interaction. In Soedergaard, Dam (Eds.) The encyclopedia of Human-Computer Interaction (2nd edition). Online available at: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/experimental-methods-in-human-computer-interaction>. This article covers writing, experimentation and statistics
* Andy Field “Discovering statistics using …”. Andy Field has written multiple books on how to conduct statistical tests in various software packages (including SPSS and R). The books are often available as inspection copy in the library. See also the book’s website: <https://www.discoveringstatistics.com/>
* Cairns, P. (2019) Doing better statistics in human-computer interaction. Cambridge press, 1st edition. A great general introduction to why you want to incorporate statistical analysis. Quite a thin booklet, with lots of information.
* Thimbleby, H. (2008). Write now!. In Cox & Cairns (Eds) Research methods for human-computer interaction. Cambridge University Press.A nice introduction about why writing is so important in the scientific process.
* American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: American Psychological Association. This booklet describes in detail how you should report results from an experimental study, with many details on the considerations for for example the methods section. It is available as an inspection copy in the university library.

**Online resources**

* Stats flow chart can help you to determine what is the most appropriate test to use: <http://www.statsflowchart.co.uk/>
* Discovering statistics website: <https://www.discoveringstatistics.com/statistics-hell-p/>
* Grammarly:  a website plug-in (and product) that does excellent grammar and spelling suggestions for your text. Not for LaTeX. <https://www.grammarly.com/>

**Appendix B: Peer Feedback Forms**

*(see BB for.docx files for Appendix B1 and B2)*

**B1: Session 2a – Research question and experiment design**

The goal of this feedback scheme is to give you an idea of points that you should look at, however, other general remarks/suggestions/concerns might be just as useful. The quality of your feedback matters to the other groups and will be taken into account for the overall assessment by the teachers. The feedback schemes also give you a good idea of what will be important to your teachers regarding grading. The first ten minutes of each session will be devoted to discussing the feedback together

Is the research question clear? (How can it be improved?)

Is the research question well motivated? (How can it be improved?)

Is it clear how the question can be operationalized and tested? (How can it be improved?)

Is the choice of dependent variable/measure clear and sensible? (How can it be improved?)

Is the task feasible and suited to answer the research question? (How can it be improved?)

Overall impression/remarks/suggestions/recommendations:

Scientific writing:

**B2: Session 2c – Statistical analysis**

The goal of this feedback scheme is to give you an idea of points that you should look at, however, other general remarks/suggestions/concerns might be just as useful. The feedback schemes also give you a good idea of what will be important to your teachers regarding grading. The first ten minutes of each session will be devoted to discussing the feedback together

Is the research question clearly motivated, described and how it is tackled? (How can it be improved?)

Do you think that it is realistic to finish data collection in time/ data analysis in time? Is it ensure that this will work out?

How far is the group in terms of writing?

Are the results visualized properly? (How can it be made clearer?)

Are statistics reported adequately? (What’s possibly incorrect?)

Overall impression/remarks/suggestions/recommendations:

Scientific writing and overall structure:

**Appendix C: Informed Consent**

Informed Consent

Before we begin, it's important to know that all data that we collect will be anonymous and confidential, and you will not be identifiable in any report, thesis or publication which arises from this study. If the dataset will be published as part of scientific communications this will be done in an anonymized fashion.   
  
We will kindly ask your permission to use your data for research purposes. You are free to decline this request, of course, but you can't finish the experiment. 

**By checking this checkbox, I confirm that:**

* I have read and understood the information provided to me for this study
* I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason, without my medical care or legal rights being affected
* I understand that the research data will be archived in a completely anonymous way in an online database and/or data repository and/or published as supplementary material to a scientific article and may be accessed by other researchers as well as the general public
* I understand that the anonymized research data can be used in future projects on similar or different topics to this study and potential results can be published in other scientific publications. At all times, my personal data will be kept anonymized in accordance with data protection guidelines

Name ………………………….. Date ……………………………..