Plan Overview

A Data Management Plan created using DMPonline

Title: PhD research: Training healthcare providers on communication via a virtual agent simulation

Creator: Mohammad Al Owayed

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2021)

Project abstract:

This PhD project aims to examine how virtual agents can be utilised to train counsellors from the Dutch Children helpline. For that, we developed a conversational agent that simulates a child who is contacting the children's helpline through text. The project starts by examining the literature on virtual agent simulation, mainly for training. Then, we will look at how feedback can be added to the conversational agent to provide a learning tool. We are also planning to examine how different human factors (mainly, human values) can be integrated into such agents to provide a richer and more realistic learning experience. The primary purpose of such a system is staff education and training. Therefore, no sensitive medical data will be used.

ID: 96978

Start date: 17-01-2022

End date: 16-01-2026

Last modified: 08-01-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

PhD research: Training healthcare providers on communication via a virtual agent simulation

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

2. Date of consultation with support staff.

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Anonymised data of participants and their answers to evaluation questionnaires.	.csv	Through questionnaires	For conducting experiments.	SURF drive or TU Delft project storage	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman
Participant characteristics (age, gender)	.csv or .xlsx	Online questionnaire through Qualtrics	To describe our participants and/or determine their eligibility for the study and/or obtain a diverse sample	SURF drive or TU Delft project storage	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman
Program Code	Several (mostly; .py, .java, .html, .js)	By myself or from open-source libraries	To implement a program prototype used in evaluation or experiments	TU Delft GitLab or Azure	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman
Anonymised notes from focus groups (if needed)	.pdf	By myself	To understand how the design should look like	SURF drive or TU Delft project storage	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman
Informed consent forms	.pdf	online through Qualtrics	To get consent from participants to perform an experiment	SURF drive or TU Delft project storage	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman
Conversations scripts with the chatbot.	.txt	Will be automatically generated after the participant interacts with the chatbot	To know which user inputs are not recognised by the chatbot. also to know who wrote gibberish to exclude their data	SURF drive or TU Delft project storage	Mohammed Al Owayyed
Interaction behaviour (e.g., which button was clicked and when)	.csv or .xlsx	Will be generated by the system	To calculate the behaviour of interactions	SURF drive or TU Delft project storage	Mohammed Al Owayyed, Myrthe Tielman, Willem-Paul Brinkman

4. How much data storage will you require during the project lifetime?

- < 250 GB

II. Documentation and data quality

5. What documentation will accompany data?

- Methodology of data collection
- README file or other documentation explaining how data is organised
- Data will be deposited in a data repository at the end of the project (see section V) and data discoverability and re-usability will be ensured by adhering to the repository's metadata standards

The data accompanying publications (including the anonymised, non-tracable results, methodology and the analysis) will be uploaded to 4TU.

III. Storage and backup during research process

6. Where will the data (and code, if applicable) be stored and backed-up during the project lifetime?

- Git(lab)/subversion repository at TU Delft
- Project Storage at TU Delft
- SURFdrive

based on the project within the PhD, the data and informed consents will be stored in the project storage or SURFdrive.
The code will be stored on TU Delft Gitlab

IV. Legal and ethical requirements, codes of conduct

7. Does your research involve human subjects or 3rd party datasets collected from human participants?

- Yes

8A. Will you work with personal data? (information about an identified or identifiable natural person)

If you are not sure which option to select, ask your Faculty Data Steward for advice. You can also check with the privacy website or contact the privacy team: privacy-tud@tudelft.nl

- Yes

The age will be collected as age groups, and the gender will be collected as (male, female, non-binary, or other)
The other demographics (e.g., occupation) will also be collected as groups. These groups will be defined in each experiment's DMP.

8B. Will you work with any types of confidential or classified data or code as listed below? (tick all that apply)

If you are not sure which option to select, ask your Faculty Data Steward for advice.

- No, I will not work with any confidential or classified data/code

9. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your Faculty Contract Manager when answering this question. If this is not the case, you can use the example below.

The datasets underlying the published papers will be publicly released following the TU Delft Research Data Framework Policy. During the active phase of research, Mohammed will oversee the access rights to data (and other outputs), as well as any requests for access from external parties. They will be released publicly no later than at the time of publication of corresponding research papers.

10. Which personal data will you process? Tick all that apply

- Gender, date of birth and/or age
- Email addresses and/or other addresses for digital communication
- Signed consent forms
- Data collected in Informed Consent form (names and email addresses)

Prolific ID will be used to manage participants. No names will be linked.
The consent forms will be signed online through questionnaires (e.g., in Qualtrics)
The age will be collected as age groups, and the gender will be collected as (male, female, non-binary, other). The collected data will be used to analyse the demographic.

11. Please list the categories of data subjects

For the experiments, the participants will recruited in two ways:
1- General public through online platforms (e.g., Prolific, etc..)
2- Counsellors who have previous experience in their field (e.g., volunteer Counsellor)

12. Will you be sharing personal data with individuals/organisations outside of the EEA (European Economic Area)?

- No

15. What is the legal ground for personal data processing?

- Informed consent

16. Please describe the informed consent procedure you will follow:

All study participants will be asked for their consent to participate in the studies and for data processing before the experiments start. As this data management plan is for the whole 4-years period of the PhD program, we will attach the relevant consent form to each experiment's DMP.

17. Where will you store the signed consent forms?

- Same storage solutions as explained in question 6

18. Does the processing of the personal data result in a high risk to the data subjects?

If the processing of the personal data results in a high risk to the data subjects, it is required to perform a Data Protection Impact Assessment (DPIA). In order to determine if there is a high risk for the data subjects, please check if any of the options below that are applicable to the processing of the personal data during your research (check all

that apply).
If two or more of the options listed below apply, you will have to complete the DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to receive support with DPIA.
If only one of the options listed below applies, your project might need a DPIA. Please get in touch with the privacy team: privacy-tud@tudelft.nl to get advice as to whether DPIA is necessary.
If you have any additional comments, please add them in the box below.

- None of the above applies

22. What will happen with personal research data after the end of the research project?

- Anonymised or aggregated data will be shared with others

25. Will your study participants be asked for their consent for data sharing?

- Yes, in consent form - please explain below what you will do with data from participants who did not consent to data sharing

V. Data sharing and long-term preservation

27. Apart from personal data mentioned in question 22, will any other data be publicly shared?

- All other non-personal data (and code) underlying published articles / reports / theses

29. How will you share research data (and code), including the one mentioned in question 22?

- I will share my data and code via git(lab)/subversion and also create a snapshot in a repository
- All anonymised or aggregated data, and/or all other non-personal data will be uploaded to 4TU.ResearchData with public access

30. How much of your data will be shared in a research data repository?

- < 100 GB

31. When will the data (or code) be shared?

- As soon as corresponding results (papers, theses, reports) are published

32. Under what licence will be the data/code released?

- CC BY
- MIT License

CC BY for the data, and MIT for the code

VI. Data management responsibilities and resources

33. Is TU Delft the lead institution for this project?

- Yes, leading the collaboration - please provide details of the type of collaboration and the involved parties below

34. If you leave TU Delft (or are unavailable), who is going to be responsible for the data resulting from this project?

Myrthe Tielman (m.l.tielman@tudelft.nl)
Willem-Paul Brinkman (w.p.brinkman@tudelft.nl)

35. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

4TU.ResearchData is able to archive 1TB of data per researcher per year free of charge for all TU Delft researchers. We do not expect to exceed this and therefore there are no additional costs of long term preservation.