

DMP C14/21/072

Project Name Mixed-initiative explanation methods: towards the next generation of interactive machine learning steered with rich feedback of non-expert users - DMP C14/21/072

Grant Title C14/21/072

Principal Investigator / Researcher Katrien Verbert

Project Data Contact Robin De Croon

Description While eXplainable Artificial Intelligence (XAI) is only recently gaining widespread visibility, the Machine Learning (ML), Artificial Intelligence (AI) and Recommender Systems literature contain a long history of work on explanations. Recent work in this area focuses not only on explaining outcomes of models, but also on interactive, mixed-initiative approaches that enable end-users to improve the accuracy of models. While such approaches have been shown to have huge potential, most of this work is targeted to ML experts. Little work has been done to support such human-in-the-loop approaches for non-expert users, i.e. users with little or no ML knowledge. In addition, existing approaches for non-expert users mostly focus on rather simple ML models that are easy to explain. The objectives of this project are threefold: first, we will research how existing model-agnostic explanation methods can be extended with interaction capabilities to capture feedback of non-expert users. Second, we will research how feedback elicited with these mixed-initiative explanation can be incorporated in ML models. Third, we will research which mixed-initiative explanation methods work better for particular user groups and how they can be personalised. By researching the effectiveness of different mixed-initiative explanation methods from a Human-Computer Interaction perspective, and using model-agnostic explanation methods to capture user feedback, we aim to provide a generic framework that can be used to incorporate user feedback in a range of ML models, including more complex deep learning models.

Institution KU Leuven

1. General Information

Name of the project lead (PI)

Katrien Verbert

Internal Funds Project number & title

C14/21/072 - Mixed-initiative explanation methods: towards the next generation of interactive machine learning steered with rich feedback of non-expert users

2. Data description

2.1. Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

Type of data	Format	Volume	How created
User-centred design data	.docx	< 1GB	Focus groups and formative evaluations will be organised. This data will be transcribed and stored in .docx files
Source code of explanation methods and framework	.js, .py	< 1GB	Source code.
Observational data	.json	< 1 GB	Interaction logs and eyetracking data of participants.
Questionnaires	.xlsx	< 1 GB	Pre- and poststudy questionnaire data
User interface designs	.tiff, .svg	1 - 10GB	User-centered design process and evaluations will result in intermediate screen designs

Reuse: public recommender datasets could be used for testing purposes, e.g., MovieLens. Applicable licenses will be complied to, e.g., <https://files.grouplens.org/datasets/movielens/ml-25m-README.html>

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

Short description of the kind of personal data that will be used:

- Focus groups and interviews will be conducted and recorded. Afterwards the recordings will be transcribed and removed.
- Personality questionnaires will be conducted. The objective is to adapt explanations according to different variables, including the user profile and the context. We will research how these variables can be detected and how explanation interfaces can be adapted on-the-fly.
- Contact details for the informed consent and (optional) study updates.

Privacy Registry Reference: We are applying for a PRET application (including ethics evaluation by SMEC). Personal data will only be collected after approval of PRET.

3.2. Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s).

No

3.3. Does your research possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

Whenever opportunities for valorisation present themselves, it will be discussed with all project members. The conclusions will be appended to this data management plan.

3.4. Do existing 3rd party agreements restrict dissemination or exploitation of the

3.4. Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions regarding reuse and sharing are in place?

No

4. Documentation and metadata

4.1. What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?

Research methods and practices will be fully documented as docx-files. Details on the setting of the interviews, the selection of interview subjects and the instructions given to interviewers will be documented in a docx-file. The data from interview questionnaires will be added to the documentation, as well as an overview of all steps taken to remove direct identifiers in the data. Finally, we require that each algorithm/code is accompanied by instructions on how to store, open and read it: all data fields should have meaningful names; all files should be accompanied by a README file that describes the goal of the study, the data format and the meaning of all stored quantities; all files should be accompanied by a package file (e.g., yarn or pip) that describes the data format as well as the (version of) the software needed to run the code.

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

All steps in the research will be documented as explained in Section 4.1 to ensure findability and reusability of the data. Furthermore, the use of git has become a defacto standard for codesharing and documentation.

5. Data storage and backup during the project

5.1. Where will the data be stored?

1. The time-stamped master copy of the data will be kept on our research unit central storage facility using Nextcloud. Copies can be made and kept on personal devices.
2. All personal data will be pseudo-anonymised as quickly as possible, the coding table will not be stored locally and only in an access controlled environment using Nextcloud. The data will be stored in the university's secure environment using Nextcloud managed by the system group of the Department of Computer Science.
3. Since we will collaborate with researchers from other research units and groups, we will use Nextcloud for active use of the data during the project.

5.2. How will the data be backed up?

The Computer Science department guarantees that our nextcloud file servers are backedup correctly. Once each researcher automatically syncs with the centrally controlled system, backup is automatic.

5.3. Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The computer science department ensures sufficient storage for all server backups. Nextcloud ensures consistent sync between the researchers' computers and the data server.

5.4. What are the expected costs for data storage and backup during the project? How will these costs be covered?

Data storage is already available. The HCI division in the Department of Computer Science has recently expanded its storage capacity using NetApp technology. This additional storage is accessible through Nextcloud (managed by the System Group of the Department of Computer Science).

5.5. Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

We rely on the Department of Computer Science System Group for security. The nextcloud setup

enables us to limit access control to data. It can only be accessed when specific clearance is granted.

6. Data preservation after the end of the project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

All research data will be stored for 10 years - according to KU Leuven's research data management policy.

6.2. Where will these data be archived (= stored for the long term)?

Data will be archived using the KU Leuven Tivoli system for archiving.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Data preservation costs are already included in the yearly Computer Science department contribution.

7. Data sharing and re-use

7.1. Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions or because of IP potential)?

No.

7.2. Which data will be made available after the end of the project?

Third parties will have access to our outcome data through our publications and after setting up a new collaboration. For industrial users, the conditions will be determined in collaboration with LRD.

7.3. Where/how will the data be made available for reuse?

- Upon request by mail
- Other (specify):

The source code will be released on GitHub. Aggregated, anonymized data will be available on request after signing a data sharing agreement. The procedure for requesting access to data is available on the project website.

7.4. When will the data be made available?

- Upon publication of the research results

7.5. Who will be able to access the data and under what conditions?

- The source code will be released on GitHub.
- Aggregated, anonymized data will be available on request after signing a data sharing agreement. The procedure for requesting access to data is available on the project website.

7.6. What are the expected costs for data sharing? How will these costs be covered?

The nextcloud setup provides us with the ability to grant temporary read access to authorized persons. No additional costs are expected.

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

A data management guide has been distributed in the research team to inform all researchers on how to manage the data they generate.

8.2. Who will be responsible for data storage & back up during the project?

Data backup is done automatically using nextcloud and the KU Leuven Tivoli backup system.

8.3. Who will be responsible for ensuring data preservation and sharing?

Katrien Verbert and Robin De Croon

8.4. Who bears the end responsibility for updating & implementing this DMP?

The end responsibility for updating and implementing the DMP is with the supervisor (promotor).