UKRI Centre for Doctoral Training in Safe & Trusted Artificial Intelligence

Group project, 2020-21

The aim of the group project is to identify how cutting-edge research in AI techniques could help to make some aspect of one of our partner's operations safer (in the sense of having some assurances over its behaviour) and/or more trustworthy.

Students this year will be working on a problem from Ernst and Young (EY). The people from EY leading the group project are Ariane Buescher and Luis Pizarro. There will be scheduled opportunities for you to get advice and feedback from Ariane, Luis and their colleagues (see *Timeline and scheduled activities* section below); outside of these sessions, any questions you have about the group project should be directed to Liz (elizabeth.black@kcl.ac.uk).

Project overview

EY has co-sponsored a new global survey, launched in Feb'20, assessing the current state of AI adoption by financial services (FS) organisations. Its findings are equally important for non-FS organisations:

- AI is expected to be an essential business driver with widespread adoption in the next two years.
- Most organisations have already implemented AI and expect to use the technology for new revenue generation.
- Increased adoption of AI technologies also comes with challenges, mainly trust and user adoption

You probably knew all that! However, beyond these encouraging lines EY know how difficult it is to implement an AI solution for a real industry problem successfully. A great portion of this success can be achieved by establishing an appropriate *AI Quality Assurance Process*, whose goal is to make all EY's developments comply with and monitor specific Trusted AI Dimensions: (i) robustness, (ii) fairness/bias, and (iii) explainability.

You will work in your group to address one of these dimensions (see group allocations below). Each group will need to do the following.

- Select two publicly available datasets to work with.
- For each dataset, identify a machine learning prediction problem you will explore (i.e., a question that you will seek to answer from the data using machine learning techniques).
- Implement at least 1 (ideally more) machine learning models to solve each problem (e.g., regression, classification, clustering).
- Identify or define a set of metrics that will be used to measure performance against your group's Trusted AI Dimension.
- For each model you develop, use existing libraries to test these along your group's Trusted AI Dimension.
- Produce a final report of your work as a Jupyter notebook.

EY will provide more details about this at the introductory session. You will also be provided with information about relevant libraries and metrics for monitoring the different Trusted AI Dimensions.

Group allocation

Group 1: robustness	Group 2: fairness/bias	Group 3: explainability
Mackenzie Jorgensen	Samuel Martin	Francis Ward
Sean Baccas	Anna Gausen	Mattia Villani
Fabrizio Russo	Munkhtulga Battogtokh	Lara Dal Molin
Alex Gaskell		Benjamin Batten

Final deliverables

Main report (one per group, must be written as a Jupyter notebook). This should:

- Describe the two datasets and machine learning prediction problems selected by your group.
- Describe how the machine learning models to address your problems were created, justifying the choices you have made.
- Define the metrics you are using to measure along your group's Trusted AI Dimension.
- Provide an evaluation of your models in terms of your group's Trusted AI Dimension.

Presentation (one per group, 20 minutes) of the key findings from your report. To be delivered to peers, EY colleagues, supervisors and CDT directors (if available).

Management report (one per group, <1 page) briefly describing how the group has ensured that everyone has had the opportunity to contribute to their full potential.

Peer feedback (one per student per other group member). For each member of their group, each student should provide brief feedback on (i) something they did well at, and (ii) something they could have improved on, to be shared with the individual. At the end of the Group Project, we will share a link to a Microsoft Form for you to provide this feedback.

Other deliverables

Preliminary report (one per group, approximately 4 – 5 pages, .pdf format). This should:

- Identify the two datasets and machine learning prediction problems selected by your group.
- Give an overview of the relevant libraries and how they are applicable to your problems.
- Summarise the results of the preliminary data analysis.

Timeline and scheduled activities

You are each expected to spend, on average, a day a week working on the Group Project. An overview of the Group Project timeline is shown in Figure 1.

The **final deliverables are due on 23 April,** with the exception of the presentation, which will be scheduled for some time after this deadline. The **preliminary report is due on 26 February**.

There are a number of scheduled sessions, to provide you with opportunities to get guidance, feedback and support.

- **Introductory session**: EY will present the problem for you to work on, and you will be able to ask any questions that you have.
- **Q&A sessions**: EY will be available to answer questions.
- **Preliminary report feedback sessions**: EY will provide feedback on your preliminary report.

These sessions will all take place on MS Teams, and will be recorded. Details of the dates and timings will be shared with you once finalised.

Intellectual property, confidentiality and contractual obligations

As is standard when working with other organisations, we have signed a formal collaboration agreement with EY that governs our engagement on the Group Project. This agreement means that we have a contractual obligation to ensure that you each abide by the following key points.

- All materials provided by EY (including the information about ML QA libraries, and the overview documents: robustness, fairness/bias, explainability) must only be used for the purpose of the Group Project, or for your own research. You cannot share these with anyone.
- The IP of all material produced during the Group Project will reside with EY. If you wish to publish anything resulting from the Group Project, this is permissible but EY must be provided with prior written notice of any such publication. Please contact the CDT Office as soon as possible if you think you may want to publish on anything resulting from the Group Project.
- You must not engage in any conduct with respect to any of EY's business, brand, reputation, products, services and/or materials which is illegal, deceptive, misleading or unethical including disparagement of EY's business, brand, reputation, products, services and/or materials.

UKRI Centre for Doctoral Training in Safe & Trusted Artificial Intelligence

Week commencing:	25-Jan	01-Feb	08-Feb	15-Feb	22-Feb	01-Mar	08-Mar	15-Mar	22-Mar	29-Mar	05-Apr	12-Apr	19-Apr	26-Apr
Methodology	■	■	■	■	■									
* Explore XAI libraries and examples	▨	▨	▨	▨	▨									
* Problem selection			▨	▨	▨									
* Exploratory data analysis (EDA)			▨	▨	▨									
Modelling						■	■	■	■	■	■	■	■	
* Select ML models						▨	▨	▨	▨					
* Identify XAI metrics						▨	▨	▨	▨					
* Build and evaluate ML models						▨	▨	▨	▨	▨	▨	▨	▨	
* Evaluation of XAI metrics						▨	▨	▨	▨	▨	▨	▨	▨	
Deliverables					**Preliminary report**								**Final report**	**Presentation**
Preliminary schedule of sessions (dates t.b.c.)	**Intro session**			**Q&A session**			**Preliminary report feedback sessions**			**Q&A session**				**Presentation**

Figure 1: Overview of Group Project timeline