



Lab Project Description

HCAI 2023/2024

Luís Macedo, Amílcar Cardoso

1 Introduction

The Human-Centered Artificial Intelligence (HCAI) course comprises: (i) a theoretic-practical component, evaluated in a written exam; (ii) a continuous assessment component, evaluated during the classes; and (iii) a practical work, consisting of the development of a Lab Project on a topic covered by the course. This document describes the latter, which will be carried out in two stages: Project Proposal, and Design and Development.

The main goal of this project is to **to tackle a challenge** related to, at least, two topics covered by the course.

2 Project Proposal

The deliverable of the first stage is a written document where the project to be developed is proposed. The document should have **no more than 3 pages in a 2-column format**¹, and should be written either in English or Portuguese.

¹Formatted according to the IJCAI guidelines, https://www.ijcai.org/authors_kit

The project must tackle a problem that is related to, **at least, two topics** covered in HCAI, namely *Human-AI Cooperation* (including Human in the loop, Human in command, Human out of the loop, and more specific aspects as Interactive Decision-making, Interactive Machine Learning, Active Learning, Learning from Demonstration/Apprenticeship Learning – Imitation Learning, Inverse Reinforcement Learning, Reinforcement Learning from Human Feedback–, Crowdsourcing, Text mining, opinion mining, and sentiment analysis), *Trustworthiness and Responsible AI* (including Explainability / Interpretability, Privacy, Ethics, etc.), *Recommender Systems, Intelligent Personal Assistants*.

This relation must be made explicit, either in the title of the project or in its documentation. Appendix A has a list of suggested topics.

The project proposal must be well-supported, by contextualizing and reviewing approaches to the problem with appropriate bibliographic references, while answering the following questions:

- **Problem and Motivation:** What is the problem? Why is it interesting? Why is it important?
- **Background:** What scientific literature provides context and background on the problem?
- **Goals:** What is the project’s aim? What would be considered a successful work? What are the expected contributions?
- **Approach:** What method(s) or algorithm(s) will be proposed? If there are existing implementations, are they going to be used? What is the plan to improve or modify such implementations? An exact answer is not necessary at this stage, but there should be a general sense of how the problem is going to be tackled.
- **Materials (including Data):** What data is going to be used? If new data is required, how is it going to be collected? Which frameworks and tools are going to be used for this?
- **Evaluation:** How will the results be evaluated? If qualitatively, what kind of results are expected (e.g., plots or figures)? If quantitatively, what type of analysis will be adopted for evaluating and/or comparing the results achieved (e.g., performance metrics or statistical tests)?

As long as their use is well-supported, any computational tools and data sources may be employed.

3 Design and Development

The proposed project will be developed throughout the semester and in the PL classes, where at least one professor will be available. Students are expected to perform several experiments, produce a piece of software, and write a scientific paper describing the outcomes of the project.

3.1 Paper

The paper should be written in English or Portuguese, and have **at most 7 pages in a 2-column format**², excluding the pages needed for bibliographic references. The following structure is suggested:

1. Introduction (including motivation, problem description, goal and approach; may reuse part of the Project Proposal);
2. Related Work (may include part of the Project Proposal);
3. Data & Approach (an overview of the data used and of the approach adopted for tackling the problem);
4. Experimentation (setup, quantitative and/or qualitative evaluation, and their discussion);
5. Conclusions (stressing the main achievements and main difficulties faced, and a reflection on the relation between the work to HCAI topics) and Future Work (what could still be done);
6. Bibliographic References.

3.2 Submission Instructions

The software developed must be made available on GitHub. The final submission must include two PDF files: (i) the paper, and (ii) a document with the URL of the repository where the software is available. Both documents must be submitted together, as a ZIP file through *Inforestudante*.

3.3 Evaluation

Overall, the project is worth **50%** of the final grade of the HCAI course (**10/20**), split into the following:

²Formatted according to the IJCAI guidelines, https://www.ijcai.org/authors_kit

- Project Proposal, **2/20**;
- Project (including paper, software, defence), **8/20**.

Evaluation will consider the following **criteria**:

- **Relevance** of the topic.
- **Significance** of the contributions.
- **Novelty/Originality** of the problem and approaches.
- **Quality/Technical Soundness** of the performed experiments, developed software, claims and conclusions.
- **Clarity** of the paper and of running instructions.

4 Important dates

The project has the following deadlines, with results to be submitted to the course area in *Inforestudante*:

- Project Proposal: **15th October, 2023**
- Final Project: **2nd, December, 2023**

Oral defences are scheduled to the **4th and 11th December, 2023**.

A Suggested Topics

A list of suggested topics is provided. For each topic, there is an associated professor, who will be the primary contact for further discussing the topic, providing basic bibliography and other insights. For any questions concerning each topic, do not hesitate to contact the associated professor by email, Slack, and/or in person during the PL classes.

- Ethics, law, and privacy in AI (Amílcar Cardoso)
- Human-AI Cooperation: Human in the Loop, Human in command/-Machine in the Loop, Human out of the loop (Luís Macedo)
 - Cooperative/Collaborative AI
 - Interactive Decision-making
 - Interactive Machine Learning
 - Active Learning
 - Learning from Demonstration/Apprenticeship Learning (Imitation Learning, Inverse Reinforcement Learning, Reinforcement Learning from Human Feedback)
 - Crowdsourcing
 - Text mining, opinion mining, sentiment analysis (Luís Macedo)
- Explainability, trustworthiness and transparency in AI (Amílcar Cardoso)
- Explainability in Generative Models (Amílcar Cardoso)
- Computational Co-creativity (Amílcar Cardoso)
- Learning Ethics, Moral and Values (Luís Macedo)
- Trustworthiness in Cooperative AI (Luís Macedo, Amílcar Cardoso)
- Evaluation metrics/measures for Explainable AI (Luís Macedo)
- Explainable Recommender Systems (Luís Macedo)