Competition notice

Project title: ARES: Attack-Resistant Explanations toward Secure and trustworthy AI

Research project manager: Przemysław Biecek

E-mail: przemyslaw.biecek@gmail.com

Project description

Machine learning explainability, fairness, robustness, and security are key elements of trustworthy Artificial Intelligence, an area of strategic importance. In this context, the main goals of the ARES project are:
G1. Develop adversarial attacks on state-of-the-art explanations to investigate vulnerabilities and limitations of the existing explainability and fairness approaches in machine learning.
G2. Introduce novel robust explanations that are stable against manipulation and intuitive to evaluate.

These are targeted at progressing explainable machine learning toward a secure and trustworthy adoption of AI solutions. Specifically, we aim to address the following research questions and hypotheses: Q1. What are the critical limitations of state-of-the-art explanations? Q2. How to detect the potential manipulation of explanations?
H1. State-of-the-art explanations for machine learning models trained on tabular data are not robust nor trustworthy in the context of adversarial attacks on explanations. Q3. What to improve toward developing robust explanations? Q4. How to evaluate explanations in the context of an adversary?
H2. Robust explanations are of novel quality and resist against the potential attacks. Answering Q1 and Q2 allows testing H1, which combined with answering Q3 and Q4 allows testing H2.

Requirement

- higher education, preferred majors: mathematics, statistics, computer science

- experience in research, preferably with some already published articles in the area of explainable artificial intelligence (XAI)

- proficiency in English.

The call is open to all those who are not PhD holders and are not students at the doctoral schools.

Discipline: Computer Science

Admission limit: 1

Recruitment schedule

- registration in the Internet Registration of Candidates, referred to as “IRK”, submitting an application to the IRK: 5 May – 20 May 2022
- qualification procedure: 26 May – 02 June 2022
- announcement of the ranking list: until 09 June 2022
- accepting documents from qualified candidates: 13 June – 23 September 2022
- announcement of the list of accepted candidates: until 30 September 2022

Recruitment fee

200 PLN

Form of the qualification proceedings

Qualification proceedings include the assessment of the following items:

1) the candidate's scientific activity, based on their CV or Resume, documented by scans of materials attached to the application for admission to the School;

2) an interview with the candidate;

3) other achievements.

Language of the selection process, including the interview

The interview shall be carried out in Polish or English – in accordance with the candidate's preferences presented in IRK. If the Polish language is selected, the interview may include parts in English.

Required documents

The candidate shall submit a School admission application only through the IRK. The application shall include the following:

1) indication of the selected discipline in which the candidate plans to pursue education, PESEL number or passport number, nationality, contact information (residence address, e-mail address, telephone number), information whether the candidate agrees to receive administrative decisions by means of electronic communication, consent for processing of personal data for the purposes of the admissions procedure;
2) a scan of the graduation diploma of uniform master's degree or postgraduate studies or an equivalent diploma obtained under separate regulations or in the case of candidates pursuing education within the European Higher Education Area – a certificate of obtaining a Master's degree or a declaration that the diploma or certificate of obtaining a Master's degree shall be provided by 23.09.2022 – declaration form. In the case of a diploma equivalent to a uniform master's degree or postgraduate studies graduation diploma, a candidate shall justify such equivalence. In case the diploma was issued in a language other than Polish or English, the candidate shall attach its certified translation;
3) a resume or CV outlining the candidate's scientific activity, including scholarly interests and achievements during the five calendar years preceding the application (if a candidate became a parent during this time, as evidenced by a scan of the child's birth certificate attached to the application, this period shall be extended by two years for each child), including, but not limited to:
    - publications,
    - research and organizational work at student research groups,
    - participation in scientific conferences,
    - participation in research projects,
    - awards and honorable mentions,
    - research internships,
    - research skills training programs completed,
    - activities promoting science,
    - activity in science movement representative bodies,
    - average of their university grades,
    - professional career,
    - level of proficiency in foreign languages;
4) scans of materials evidencing scientific activity mentioned in their CV and/or resume;
5) a document confirming at least B2 proficiency level in English or a declaration of the level of proficiency in English allowing education at the School;
6) the scan of a declaration by the planned supervisor, confirming their agreement to undertake the duties of a supervisor and of the number of doctoral students, for whom they perform the duties a designated supervisor, in accordance with the template constituting Appendix no.4 to the

Resolution no. 17 of the Senate of the University of Warsaw of 20th January 2021 on rules of admission to doctoral schools at the University of Warsaw (the University of Warsaw Monitor of 2021, item 142), the candidate may also attach a scan of their planned supervisor's opinion and opinions of other academics about the candidate and their scientific activity and/or proposed research project;

7) the photograph of a candidate's face that allows for their identification;
8) a declaration confirming whether the candidate was or is a doctoral student or a participant of doctoral studies or whether they have initiated a doctoral dissertation process or whether proceedings to award them a doctoral degree have been initiated – and if yes, the title of their doctoral dissertation or the research project prepared by a candidate, including the name and last name of the candidate's tutor or supervisor;
9) a declaration confirming that they have reviewed the Resolution no. 17 of the Senate of the University of Warsaw of 20th January 2021 on rules of admission to doctoral schools at the University of Warsaw (the University of Warsaw Monitor of 2021, item 142) and Articles 40 and 41 of the Code of Administrative Procedure;
10) scanned transcripts of records of the graduate and postgraduate studies or the uniform Master's degree studies, or equivalent documents (e.g. diploma supplement);
11) abstract of the master's thesis or master's project in English (up to 3,000 characters with spaces);

Evaluation criteria

The committee shall evaluate the candidates by awarding them points for their competencies to perform specific tasks in a research project and scientific achievements to date. On the basis thereof, the committee shall rank the candidates according to the following criteria:

a) competencies to perform specific tasks in a research project (70% of the final score)
    - 3 points - very good
    - 2 points – good
    - 1 point – poor
    - 0 points - no competencies
b) publication track record, including publications in renowned scientific papers / magazines (30% of the final score)
    - 4 points – prominent
    - 3 points - very good
    - 2 points – good
    - 1 point – poor
    - 0 points - no publication track record

Education program

The education lasts 4 years. It includes obligatory classes (no more than 300 hours in total during the whole period of education) and the implementation of an individual research program, carried out under the supervision of a supervisor. Beginning of education – October 1, 2022.

Scholarships

PRELUDIUM BIS doctoral scholarships shall amount to:

- PLN 4266.00 gross per month, until the month in which a PhD student's mid-term evaluation is performed at the doctoral school and
- PLN 5119.00 gross per month, after the month in which a PhD student's mid-term evaluation is performed at the doctoral school and

shall be awarded pursuant to the Act on Higher Education and Science of 20 July 2018.