

## Ogłoszenie

Tytuł projektu: ARES: Wyjaśnienia odporne na ataki w kierunku bezpiecznej i godnej zaufania SI

Kierownik projektu: Przemysław Biecek  
E-mail: przemyslaw.biecek@gmail.com

### Opis projektu:

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.

### Warunki względem kandydata

- wykształcenie wyższe, preferowane kierunki: matematyka, statystyka, informatyka
- doświadczenie w pracy badawczej, preferowane opublikowane artykuły w obszarze wyjaśnialnej sztucznej inteligencji (XAI)
- biegła znajomość języka angielskiego.

Do konkursu może przystąpić osoba, która nie posiada stopnia naukowego doktora i nie jest uczestnikiem szkoły doktorskiej.

**Dyscyplina:** informatyka

**Limit miejsc:** 1

### Harmonogram rekrutacji

- rejestracja kandydatów w IRK: 5 maja – 20 maja 2022 r.
- postępowanie rekrutacyjne: 26 maja – 02 czerwca 2022 r.
- ogłoszenie listy rankingowej: do 09 czerwca 2022 r.
- przyjmowanie dokumentów od zakwalifikowanych kandydatów: 13 czerwca – 23 września 2022 r. do godz. 14.00
- ogłoszenie listy przyjętych do Szkoły Doktorskiej: do 30 września 2022 r.

### Opłata rekrutacyjna

200 zł

### Forma postępowania kwalifikacyjnego

W postępowaniu kwalifikacyjnym uwzględnia się ocenę:

1) aktywności naukowej kandydata na podstawie CV lub życiorysu udokumentowanej skanami materiałów załączonymi do wniosku o przyjęcie do Szkoły;

- 2) rozmowy kwalifikacyjnej z kandydatem / egzaminu kwalifikacyjnego;
- 3) innych osiągnięć.

### **Język postępowania kwalifikacyjnego, w tym rozmowy kwalifikacyjnej**

Rozmowa kwalifikacyjna odbywa się w języku polskim lub angielskim zgodnie z preferencjami kandydata zgłoszonymi w IRK. W przypadku wyboru języka polskiego, rozmowa kwalifikacyjna może zawierać część prowadzoną w języku angielskim.

### **Wymagane dokumenty**

Kandydat składa wyłącznie w IRK wnioszek o przyjęcie do Szkoły. Wniosek zawiera:

- 1) wskazanie wybranej dyscypliny, w której kandydat planuje podjąć kształcenie, numer PESEL lub numer paszportu, obywatelstwo, dane kontaktowe (adres zamieszkania, adres poczty elektronicznej, numer telefonu), informację, czy kandydat wyraża zgodę na doręczenie decyzji administracyjnych za pomocą środków komunikacji elektronicznej, zgodę na przetwarzanie danych osobowych na potrzeby postępowania rekrutacyjnego;
- 2) skan dyplomu ukończenia jednolitych studiów magisterskich bądź studiów drugiego stopnia lub równorzędny uzyskany na podstawie odrębnych przepisów, albo – w przypadku kandydatów realizujących kształcenie w ramach Europejskiego Obszaru Szkolnictwa Wyższego (European Higher Education Area) – oświadczenie, że dyplom lub zaświadczenie o uzyskaniu tytułu magistra zostanie dostarczone do dnia 23.09.2022 r. W przypadku dyplomu równorzędnego dyplomowi ukończenia jednolitych studiów magisterskich lub studiów drugiego stopnia, kandydat uzasadnia tę równorzędność. W przypadku gdy dyplom został wydany w języku innym niż polski lub angielski, kandydat załącza jego tłumaczenie przysięgłe;
- 3) życiorys lub CV zawierające informacje o aktywności naukowej, w tym zainteresowaniach i osiągnięciach naukowych kandydata w okresie pięciu lat kalendarzowych poprzedzających złożenie wniosku (w przypadku, gdy kandydat w tym okresie został rodzicem, co potwierdzi załączając we wniosku skan aktu urodzenia dziecka, termin ten podlega wydłużeniu o dwa lata na każde dziecko), w szczególności o:
  - publikacjach,
  - pracach badawczych i organizacyjnych w kołach naukowych,
  - udziale w konferencjach naukowych,
  - udziale w projektach badawczych,
  - nagrodach i wyróżnieniach,
  - stażach badawczych,
  - odbytych szkoleniach z zakresu umiejętności badawczych,
  - działalności popularyzującej naukę,
  - działalności w ciałach przedstawicielskich ruchu naukowego,
  - średniej ocen ze studiów,
  - karierze zawodowej,
  - znajomości języków obcych;
- 4) skany materiałów potwierdzających wskazaną w życiorysie lub CV aktywność naukową;
- 5) dokument potwierdzający znajomość języka angielskiego na poziomie co najmniej B2 lub oświadczenie o znajomości języka angielskiego w stopniu umożliwiającym kształcenie w szkole;
- 6) skan oświadczenia planowanego promotora o podjęciu się opieki promotorskiej oraz o liczbie doktorantów, dla których pozostaje wyznaczonym promotorem według wzoru, stanowiącego załącznik nr 4 do Uchwały nr 17 Senatu Uniwersytetu Warszawskiego z dnia 20 stycznia 2021 r. w sprawie zasad rekrutacji do szkół doktorskich Uniwersytetu Warszawskiego (Monitor UW z 2021, poz. 142), ponadto kandydat może dołączyć skan opinii planowanego promotora oraz opinie innych

pracowników naukowych na temat kandydata oraz jego aktywności naukowej lub propozycji projektu badawczego;

- 7) fotografię przedstawiającą twarz kandydata, umożliwiającą jego identyfikację;
- 8) oświadczenie, czy jest lub był doktorantem albo uczestnikiem studiów doktoranckich lub posiada bądź posiadał wszczyć przewód doktorski albo postępowanie w sprawie nadania stopnia doktora, a jeżeli tak – tytuł rozprawy doktorskiej albo projektu badawczego przygotowywanego przez kandydata, z podaniem imienia i nazwiska jego opiekuna naukowego lub promotora;
- 9) oświadczenie o zapoznaniu się z treścią uchwały nr 17 Senatu Uniwersytetu Warszawskiego z dnia 20 stycznia 2021 r. w sprawie zasad rekrutacji do szkół doktorskich Uniwersytetu Warszawskiego (Monitor UW z 2021, poz. 142) oraz treścią art. 40 i art. 41 Kodeksu postępowania administracyjnego;
- 10) skany kart przebiegu studiów I i II stopnia lub jednolitych studiów magisterskich, lub dokumentów równoważnych (np. suplement do dyplomu);
- 11) streszczenie pracy lub projektu magisterskiego w języku angielskim (do 3000 znaków ze spacjami).

### **Kryteria oceny**

Komisja ocenia kandydatów w skali punktowej biorąc pod uwagę kompetencje do realizacji projektu badawczego oraz dotychczasowy dorobek naukowy kandydata, opracowując na tej podstawie ranking kandydatów wg następujących kryteriów:

a) kompetencje do realizacji określonych zadań w projekcie badawczym (70% oceny końcowej)

- 3 pkt - bardzo dobre
- 2 pkt – dobre
- 1 pkt - słabe
- 0 pkt - brak kompetencji

b) dorobek naukowy kandydata, w tym publikacje w renomowanych wydawnictwach / czasopismach naukowych (30% oceny końcowej)

- 4 pkt – wyróżniający
- 3 pkt - bardzo dobry
- 2 pkt – dobry
- 1 pkt – słaby
- 0 pkt - brak dorobku naukowego

### **Program kształcenia**

Kształcenie trwa 4 lata. Obejmuje zajęcia obowiązkowe (nie więcej niż 300 godz. łącznie przez cały okres kształcenia) oraz realizację indywidualnego programu badawczego, realizowanego pod kierunkiem promotora. Rozpoczęcie kształcenia – 1 października 2022.

Przygotowanie rozprawy doktorskiej w ramach programu nie może trwać dłużej niż 4 lata.

### **Stypendium**

Stypendium doktoranckie PRELUDIUM BIS w miesięcznej wysokości:

- 4266,00 zł brutto/mies. do miesiąca, w którym została przeprowadzona ocena śródkresowa doktoranta w szkole doktorskiej,
- 5119,00 zł brutto/mies. po miesiącu, w którym została przeprowadzona ocena śródkresowa doktoranta w szkole doktorskiej

otrzymywane jest na zasadach określonych przepisami ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce.