

Nr umowy: 860621 – NL4XAI; Kierownik projektu: dr hab. Katarzyna Budzyńska, prof. uczelni

Interactive Natural Language Technology for Explainable Artificial Intelligence (Interaktywne technologie języka naturalnego dla wyjaśnialnej sztucznej inteligencji)

Zgodnie z paradoksem Polanyi, ludzie wiedzą więcej niż mogą wyjaśnić, głównie dzięki ogromnej ilości ukrytej wiedzy, którą zdobywają nieświadomie poprzez kulturę, dziedzictwo, itp. To samo odnosi się do systemów sztucznej inteligencji (*Artificial Intelligence*, AI), które głównie uczą się automatycznie na podstawie danych. Jednak zgodnie z prawem UE ludzie mają prawo do uzyskania wyjaśnienia decyzji, które ich dotyczą, niezależnie od tego kto (lub jaki system AI) podejmuje taką decyzję. W projekcie NL4XAI stajemy przed wyzwaniem uczynienia sztucznej inteligencji samo-wytłumaczalną i tym samym przyczynienia się do przełożenia wiedzy na produkty i usługi, przynoszące korzyści gospodarcze i społeczne, przy wsparciu systemów wytłumaczalnej inteligencji (*Explainable AI*, XAI). Ponadto, NL4XAI skupia się na automatycznym generowaniu interaktywnych wyjaśnień w języku naturalnym (*natural language*, NL), tak jak robią to ludzie w naturalnych kontekstach. W rezultacie oczekuje się, że jedenastu młodych naukowców (*Early Stage Researchers*, ESR), którzy otrzymają szkolenie w ramach projektu NL4XAI, przyczynią się do wykorzystania modeli i technik AI przez różnych użytkowników, w tym użytkowników niebędących ekspertami.

Kierownikiem projektu w PW jest dr hab. Katarzyna Budzyńska, prof. uczelni (<https://www.ans.pw.edu.pl/Pracownicy/Nauczyciele-akademicy-WAiNS/Budzynska-Katarzyna>), a doktorantem jest Martijn Demollin (<https://nl4xai.eu/people/esr8/>), zajmujący się tematem „Customized interactive argumentation schemes for XAI”. NL4XAI realizowany jest we współpracy z następującymi uniwersytetami i przedsiębiorstwami: University of Santiago de Compostela (Hiszpania, koordynator projektu), University of Aberdeen (Wielka Brytania), Agencia Estatal del Consejo Superior de Investigaciones Científicas (Hiszpania), Technische University of Delft (Holandia), Centre National de la Recherche Scientifique (Francja), Università ta’Malta (Malta), Universiteit Utrecht (Holandia), Universitat Autònoma de Barcelona (Hiszpania), Universiteit Twente (Holandia), University of Lorraine (Francja), University of Dundee (Wielka Brytania), Indra Soluciones Tecnologías de la Información SL (Hiszpania), Arria NLG Limited (Wielka Brytania), Wizenoze (Holandia), Orange (Francja), Info Support (Holandia) oraz Accenture (Hiszpania). Projekt realizowany jest na WAiNS w okresie od września 2020 do sierpnia 2023 r.