

NL4XAI: Interactive Natural Language Technology for Explainable Artificial Intelligence

According to Polanyi's paradox, humans know more than they can explain, mainly due to the huge amount of implicit knowledge they unconsciously acquire through culture, heritage, etc. The same applies for Artificial Intelligence (AI) systems mainly learnt automatically from data. However, in accordance with EU laws, humans have a right to explanation of decisions affecting them, no matter who (or what AI system) makes such decision.

In the NL4XAI project we will face the challenge of making AI self-explanatory and thus contributing to translate knowledge into products and services for economic and social benefit, with the support of Explainable AI (XAI) systems. Moreover, the focus of NL4XAI is in the automatic generation of interactive explanations in natural language (NL), as humans naturally do, and as a complement to visualization tools. As a result, the 11 Early Stage Researchers (ESRs) to be trained in the NL4XAI project are expected to leverage the usage of AI models and techniques even by non-expert users. Namely, all their developments will be validated by humans in specific use cases, and main outcomes publicly reported and integrated into a common open source software framework for XAI that will be accessible to all the European citizens. In addition, those results to be exploited commercially will be protected through licenses or patents.

It is worthy to note that we have selected some of the most prominent European researchers (from both academy and industry) in each of the related fundamental topics and created a joint high quality training program that can be seen as a pyramid with the main research objective (designing and building XAI models) on top. It will be achieved as result of jointly addressing the research objectives in the pyramid base (NL generation and processing for XAI; Argumentation Technology for XAI; Interactive Interfaces for XAI). ESRs are also to be trained in Ethical and Legal issues, as well as in transversal skills.

Interaktywne technologie języka naturalnego dla wyjaśnialnej sztucznej inteligencji

Zgodnie z paradoksem Polanyi'ego, 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-wy tłumaczalną 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, umożliwią wykorzystanie modeli i technik AI przez różnorodnych użytkowników, w tym użytkowników niebędących ekspertami.

Wybraliśmy do konsorcjum NL4XAI najwybitniejszych europejskich badaczy w każdym z podstawowych tematów projektu i stworzyliśmy wspólny, wysokiej jakości program szkoleniowy, który może być postrzegany jako piramida, której głównym celem badawczym jest projektowanie i

implementowanie modeli XAI. Młodzi badacze otrzymają również szkolenie w zakresie zagadnień etycznych i prawnych, jak również w zakresie umiejętności transferu wiedzy.