

## Popularnonaukowe streszczenie projektu

Jeśli Big Data stanowią ropę naftową naszych czasów, utratę prywatności można opisać jako nowy kryzys klimatyczny; w dobie zbierania i przetwarzania danych osobowych na bezprecedensową skalę osoby prywatne stają wobec rosnącej potęgi wielkich podmiotów gospodarczych oraz nieprzejrzystych systemów podejmowania decyzji z jakich podmioty te korzystają. Przedmiotowy projekt, powstały w oparciu o ideę siły przeciwważnej (*countervailing power*) sformułowaną przez Kena Galbraitha, ma za zadanie dowieść, że sztuczna inteligencja może zostać wykorzystana dla wzmocnienia pozycji współczesnego obywatela poprzez umożliwienie zautomatyzowanej oceny legalności tekstów dokumentów prawnych takich jak umowy użytkownika czy polityki prywatności, dokonywanej w sposób transparentny przez system wyjaśnialnej sztucznej inteligencji (XAI – *eXplainable Artificial Intelligence*). Elementem dokonywanej w ten sposób analizy byłaby również ocena przejrzystości tekstu, spełniając tym samym wymóg prawny transparentności tekstu jako niezbędnego warunku jego zgodności z prawem. Przedmiotowe badania stanowią istotny krok na drodze do stworzenia powszechnie dostępnej, darmowej technologii, która umożliwi szybką, łatwą i przejrzystą kontrolę przez użytkownika legalności postanowień umowy użytkownika lub polityki prywatności dostawcy usług, dostarczając zarazem prostych i niezawodnych wyjaśnień w zakresie wszelkich zidentyfikowanych problemów ze zgodnością postanowień z prawem.

Badania oparte będą na doświadczeniach zebranych przez Kierownika Projektu w trakcie projektów rozwijanych w latach 2018-2019 w ramach architektury Claudette (*portmanteau* nazwy [*unfair*] *Clause Detector*, Wykrywacz [nieodzwolonych] klauzul) w Europejskim Instytucie Uniwersyteckim. Zawdzięczający swe narodziny serii eksperymentów z uczeniem klasyfikatora uczenia maszynowego w zakresie rozpoznawania nieodzwolonych klauzul umownych, projekt Claudette rozwinął się w szeroką architekturę obejmującą analizę legalności postanowień umów użytkownika oraz polityk prywatności w świetle przepisów prawa (w tym Dyrektywy w sprawie nieuczciwych warunków w umowach konsumenckich oraz RODO). Badania te wykazały, że możliwym jest nie tylko przyuczenie systemu uczenia maszynowego do podawania wyjaśnień dla swoich ocen w zakresie identyfikacji nieodzwolonych klauzul; wykorzystanie sieci neuronowej czerpiącej z podłączonej bazy wiedzy zawierającej możliwe uzasadnienia prawne faktycznie zwiększa dokładność identyfikacji nieodzwolonych klauzul umownych. Dokonano również wyraźnych postępów w kierunku uczynienia systemu przystępnym dla użytkowników w całej Europie, dzięki eksperymentom z różnymi metodami umożliwienia wsparcia dla różnych języków, zarówno przy zachowaniu istniejących korpusów jak i przez połowicznie zautomatyzowane tworzenie nowych.

W oparciu o tą podstawę, planowane badania uczynią dwa istotne kroki naprzód. W pierwszej kolejności, dla wykazania że XAI może zostać wykorzystane dla dokonywania transparentnej analizy legalności tekstów dokumentów takich jak umowy użytkownika lub polityki prywatności, zbudowane zostaną bazy wiedzy w oparciu o ramy normatywne wyrosłe na gruncie Dyrektywy w sprawie nieuczciwych warunków umownych oraz RODO, z włączeniem wszelkich oficjalnych wykładni oraz orzecznictwa Trybunału Sprawiedliwości Unii Europejskiej. Bazy wiedzy zostaną następnie wykorzystane w celu analizy weryfikacji możliwości budowy systemu oceny legalności opartego na w pełni wyjaśnialnej sztucznej inteligencji. W drugiej kolejności, dla dokonania analizy i kwantyfikacji aspektów przejrzystości językowej, które zostaną ujęte w ogólnej ocenie legalności, projekt mieć będzie trzy cele: a) analizę pojęcia przejrzystości w oparciu o przepisy wspomnianej Dyrektywy oraz RODO, jak również opracowaniu na tej podstawie zestawu kryteriów dla dokonywania rozróżnienia pomiędzy przejrzystymi oraz nieprzejrzystymi elementami konstrukcji umów; b) identyfikację najlepiej dopasowanej metodologii dla lingwistycznej oceny czytelności w zakresie umów użytkownika oraz polityk prywatności, oraz próbę przełożenia jej na algorytm uczenia maszynowego; c) analizę empiryczną reprezentatywnej próbki klauzul z umów użytkownika oraz polityk prywatności.