

Streszczenie popularnonaukowe w języku polskim

Projekt dotyczy uczenia maszynowego, gałęzi sztucznej inteligencji (Artificial Intelligence, AI) zajmującej się algorytmami autonomicznie pozyskującymi wiedzę z różnych źródeł danych. Główną siłą napędową współczesnego uczenia maszynowego jest głębokie uczenie (Deep Learning, DL): jego niedawny postęp spowodował, że wiele aplikacji AI osiągnęło swoje „punkty krytyczne”, przechodząc od fazy prototypów do użytecznych produktów i usług, z których obecnie korzystają miliardy użytkowników na całym świecie.

Jednak DL cierpi również z powodu fundamentalnych ograniczeń, między innymi (i) nienasyconego apetytu na dane i moc obliczeniową, co prowadzi do niebotycznego zużycia energii i skłania kluczowych graczy na rynku AI do rozważenia reaktorów jądrowych do zasilania swoich serwerów, oraz (ii) ograniczonej zdolności do spójnego, przejrzystego, interpretowalnego i niezawodnego rozumowania, co prowadzi do błędów, uprzedzeń (tzw. obciążen) i halucynacji, co ostatnio stało się szczególnie widoczne w dużych modelach językowych (Large Language Models, LLM). Wady te mają poważne konsekwencje w rzeczywistych zastosowaniach, stwarzając ryzyko m.in. podejmowania stronniczych decyzji i błędnych diagnoz medycznych. Dlatego też wyzwania podjęte w tym projekcie są perspektywnie istotne nie tylko dla społeczności naukowej, ale także dla wszystkich użytkowników rozwiązań opartych na sztucznej inteligencji.

Aby sprostać tym wyzwaniom, projekt ucieka się do systemów neurosymbolicznych, które są hybrydami konwencjonalnego DL z symbolicznym przetwarzaniem informacji, które odbywa się na wyższym poziomie abstrakcji, zbliżonym do języka naturalnego i języków programowania. Takie systemy łączą w sobie to, co najlepsze z obu światów, tj. znane zalety DL z transparentnym i uniwersalnym charakterem wnioskowania symbolicznego. W przeciwieństwie do rozwiązań opartych wyłącznie na DL, systemy neurosymboliczne są znacznie bardziej przejrzyste, interpretowalne i wymagają mniej danych i mocy obliczeniowej w uczeniu i użytkowaniu. Na przykład, w ramach przygotowań do tego projektu, zespół projektowy zaproponował pewną klasę systemów neurosymbolicznych, które nawet po nauczaniu na stosunkowo niewielu przykładach scen wizualnych (obiektów w polu widzenia kamery), były w stanie niezawodnie je interpretować i oferować podobną zdolność dla nowych scen, tj. dokonywać poprawnych predykcji w konfrontacji z wcześniej niewidocznymi scenami, czyli generalizować. Zespół zademonstrował te zalety na trudnym problemie diagnozowania medycznego opartego na mikroskopowym obrazowaniu tarczycy (skrawki histologiczne), budując dokładniejsze modele diagnostyczne niż konwencjonalne DL.

Projekt zaowocuje nowymi algorytmami sztucznej inteligencji, które są bardziej wydajne, wymagają mniej danych do uczenia i lepiej wyjaśniają swoje działania, a tym samym przyczyniają się zarówno do teorii, jak i praktyki sztucznej inteligencji. W odniesieniu do pierwszego, projekt poszerzy podstawy sztucznej inteligencji i zapewni społeczności badaczy i praktyków nowe algorytmy i nowe sposoby interakcji z ludźmi, w szczególności nowe sposoby objaśniania i uzasadniania podejmowanych decyzji. Projekt zweryfikuje też proponowane metody na rzeczywistych problemach, między innymi w (i) biomedycynie, która notorycznie cierpi z powodu niewystarczających danych i wysokich kosztów ich anotacji, wymagając jednocześnie wyjaśnień decyzji podejmowanych dla poszczególnych pacjentów, oraz w (ii) satelitarnym obrazowaniu Ziemi, gdzie szybki postęp jest niezbędny, aby poprawić nasze zrozumienie dynamicznych zmian, które obecnie wpływają na naszą planetę.

Wyniki projektu będą rozpowszechniane za pośrednictwem publikacji w renomowanych czasopismach naukowych, prezentacji na konferencjach i innych wydarzeniach naukowych, stale utrzymywanej strony internetowej poświęconej projektowi oraz mediów społecznościowych. Poza tym głównym namacalnym rezultatem projektu będzie biblioteka oprogramowania dostępna w ramach otwartej licencji (open-source) dla globalnej społeczności naukowej i wszystkich zainteresowanych.