

Szybki rozwój sztucznej inteligencji (SI) spowodował lawinowy przyrost publikacji na ten temat — artykułów naukowych, raportów technicznych, wpisów na blogach i innych. Większość prac naukowych na temat sztucznej inteligencji powstaje w języku angielskim, jednak znaczna ilość innych treści, takich jak raporty przemysłowe, posty na blogach i dokumenty programowe, jest opracowywana w różnych innych językach. Dlatego też krytyczne i aktualnie niezbędne jest opracowanie metod tłumaczenia maszynowego dla dokumentów z zakresu SI — nie tylko w celu przełamywania barier językowych, ale także w celu zapewnienia sprawiedliwego dostępu do najnowocześniejszych badań i postępu technologicznego. Jednak tradycyjne narzędzia tłumaczeniowe zazwyczaj nie są dostosowane do pracy z multimodalnymi danymi, które są charakterystyczne dla obszaru SI. Ponadto uzyskują one słabe wyniki dla języków o niskich zasobach z ograniczonymi danymi treningowymi. Stwarza to poważne przeszkody dla innowacji i globalnej współpracy w dziedzinie sztucznej inteligencji ponad granicami językowymi i kulturowymi.

Projekt LEXICAI ma na celu opracowanie nowej generacji narzędzi tłumaczeniowych dostosowanych do dokumentów AI. Nasz system zostanie zaprojektowany w sposób umożliwiający mu rozumienie kontekstu intermodalnego, włączanie informacji z tekstu, rysunków, równań, tabel i fragmentów (snippetów) kodu w celu poprawy dokładności tłumaczenia. System będzie również integrować informacje zwrotne od ekspertów dziedzinowych i lingwistów, aby zapewnić jasność tekstu, poprawność i spójność terminologiczną. Aby system był powszechnie dostępny, skupiamy się na opracowywaniu lekkich modeli, które mogą działać w środowiskach o ograniczonych zasobach. Nasze podejście obejmuje tłumaczenie zarówno dużych języków, jak i języków niedostatecznie reprezentowanych, aby zapewnić sprawiedliwy dostęp do wiedzy o sztucznej inteligencji i promować różnorodność językową w globalnych badaniach. Opracowana zostanie duża baza danych terminologicznych z zakresu SI w celu usprawnienia spójnego i niezawodnego tłumaczenia na różne języki. LEXICAI przełamie bariery językowe i techniczne w komunikacji SI, umożliwiając naukowcom, innowatorom i szerszej publiczności dostęp do najnowocześniejszych osiągnięć AI, ich zrozumienie i tworzenie, tym samym przyczyniając się do tworzenia lepiej połączonego i sprawiedliwego ekosystemu SI. Nasz projekt usprawni tłumaczenie maszynowe złożonych treści SI, zapewniając jednocześnie narzędzia tłumaczeniowe typu open source i bazę terminologiczną, dzięki którym AI będzie w przyszłości lepiej połączona, wielojęzyczna i inkluzywna.