Zaufanie do sztucznej inteligencji

Sztuczna inteligencja to nie tylko nowa technologia. To potężna siła, która przekształca nasze codzienne praktyki, osobiste oraz zawodowe interakcje, a także środowisko społeczne, w którym żyjemy. Daje nam niezrównane możliwości, ale także pociąga za sobą nowe, czasem nieznane wyzwania. Głównym wyzwaniem związanym z rozwojem sztucznej inteligencji jest ustanowienie harmonijnych relacji człowiek-sztuczna inteligencja, niezbędnych dla dobrego wykorzystania jej potencjału. Aby sprostać temu wyzwaniu, obecny projekt koncentruje się na zaufaniu – kluczowym składniku każdego społeczeństwa. Projekt ma na celu rozpoznanie różnic między zaufaniem do ludzi i agentów sztucznej inteligencji oraz zidentyfikowanie kluczowych czynników psychologicznych, które determinują rozwój zaufania do sztucznej inteligencji. Do tej pory niewiele wiadomo na temat zaufania w relacjach człowiek-sztuczna inteligencja, a prawie nic na temat zaangażowanych mechanizmów psychologicznych. Projekt przyczyni się do lepszego zrozumienia, w jaki sposób budowane jest zaufanie między ludźmi a agentami sztucznej inteligencji, co sprawia, że agenci sztucznej inteligencji są postrzegani jako godni zaufania, oraz jak ocenia się ich moralność.

Celem tego projektu jest udzielenie odpowiedzi na następujące pytania badawcze: Jakie są podobieństwa i różnice między zaufaniem interpersonalnym a zaufaniem do sztucznej inteligencji? Jakie czynniki psychologiczne wpływają na gotowość ludzi do zaufania sztucznej inteligencji? Jaka jest rola moralności i kompetencji przy ocenie wiarygodności sztucznej inteligencji? Jak ważne jest to, co zrobiła sztuczna inteligencja oraz to, jakie były konsekwencje jej działań, przy dokonywaniu osądów moralnych?

Projekt składa się z siedmiu badań podzielonych na trzy odrębne linie badawcze. Pierwsza linia związana jest z czynnikami psychologicznymi wpływającymi na zaufanie do sztucznej inteligencji, takimi jak optymizm, lęk lub odporność na zmiany. W tych badaniach zidentyfikuję jakie cechy sprawiają, że ludzie chętnie lub niechętnie ufają sztucznej inteligencji. Druga linia badawcza skoncentruje się na dwóch uniwersalnych wymiarach, na których ludzie oceniają innych – jednym związanym z moralnością, oraz drugim związanym z kompetencjami. Zbadam, w jaki sposób informacje o moralności i kompetencjach agenta sztucznej inteligencji wpływają na gotowość do zaufania im. Trzecia linia dotyczy sądów moralnych na temat sztucznej inteligencji. W tych badaniach sprawdzę, czy ludzie przywiązują większą wagę do tego, aby działania sztucznej inteligencji były etycznie słuszne (np. mówienie prawdy, przestrzeganie zasad) czy na ich konsekwencje, jako że dobry uczynek może doprowadzić do katastrofy, podczas gdy działanie niezgodne z prawem może nastąpić w imię większego dobra. Badania będą prowadzone zarówno w laboratorium, jak i przy użyciu internetowych paneli badawczych, wykorzystując zarówno istniejące metody narzędzia badawcze, jak i takie, które zostaną stworzone specjalnie na potrzeby tego projektu.