Podsumowanie debaty panelowej XVIII Kongresu ITS Polska 2026 r.

„Problem AI Alignment w obszarze ITS, czyli bezpieczne i odpowiedzialne działanie AI zgodne z intencjami ludzi”

(AI Alignment in ITS: Safe and Responsible AI Aligned with Human Intent)

Debata odbyła się podczas XVIII Polskiego Kongresu ITS i zgromadziła przedstawicieli środowiska naukowego, filozoficznego, technologicznego oraz praktyków wdrażających rozwiązania AI w obszarze transportu i systemów inteligentnych.

Moderator

- Paweł Góra – Fundacja Quantum AI

Uczestnicy panelu

- Cezary Dołęga – NEUROSOFT
- Ewa Agnieszka Lekka-Kowalik – Katolicki Uniwersytet Lubelski Jana Pawła II
- Katarzyna Marczuk – ALEET
- Adrian Mirowski – ONYX
- Andrzej Skowron – Instytut Badań Systemowych PAN

Debata ukazała, jak złożonym i wielowymiarowym zagadnieniem staje się dziś problem AI Alignment — rozumiany jako zgodność działania systemów sztucznej inteligencji z ludzkimi wartościami, intencjami, normami społecznymi oraz zasadami bezpieczeństwa.

Już na początku dyskusji zwrócono uwagę, że pojęcie AI Alignment wymyka się prostym definicjom. Paneliści podkreślali, że zagadnienie to obejmuje jednocześnie aspekty:

- technologiczne,
- matematyczne,
- filozoficzne,
- etyczne,
- społeczne,
- prawne,
- oraz operacyjne.

Ważnym wątkiem była refleksja nad ograniczeniami współczesnych modeli sztucznej inteligencji. Wskazywano, że rozwój technologiczny bardzo często wyprzedza nasze zdolności do pełnego rozumienia pojęć takich jak odpowiedzialność, autonomia, bezpieczeństwo czy etyczność działania systemów inteligentnych.

W części naukowo-filozoficznej debaty podkreślano, że pojęcia wykorzystywane w AI — takie jak „agent", „bezpieczeństwo", „wartości" czy „etyka" — są pojęciami nieostrymi, dynamicznymi i silnie zależnymi od kontekstu społecznego oraz rzeczywistości fizycznej, w której funkcjonują systemy sztucznej inteligencji.

Z kolei przedstawiciele środowiska praktyków i biznesu koncentrowali się na problemach implementacyjnych. Wielokrotnie podkreślano, że skuteczność systemów AI zależy przede wszystkim od:

- poprawnego zdefiniowania celu działania systemu,
- jakości danych,
- właściwego modelowania ograniczeń,
- oraz obecności kompetentnych ekspertów branżowych nadzorujących proces tworzenia i wdrażania rozwiązań.

Szczególną uwagę zwrócono na znaczenie danych treningowych. Paneliści wskazywali, że dane nigdy nie są całkowicie neutralne, ponieważ już sam wybór danych oraz sposobu ich interpretacji zawiera określone założenia i wartości. W tym kontekście podkreślano potrzebę współpracy interdyscyplinarnej obejmującej inżynierów, specjalistów transportu, filozofów, etyków, socjologów i ekspertów od organizacji procesów.

Istotnym elementem dyskusji były również:

- symulacje komputerowe,
- cyfrowe bliźniaki,
- systemy adaptacyjne,
- oraz możliwość walidacji decyzji AI w środowiskach symulacyjnych.

Paneliści zgodzili się, że narzędzia symulacyjne stają się dziś niezbędnym elementem nowoczesnych systemów ITS, choć jednocześnie zaznaczano, że żadna symulacja nie jest w stanie przewidzieć wszystkich możliwych zdarzeń i scenariuszy rzeczywistych.

W dyskusji pojawił się także problem odpowiedzialności za decyzje podejmowane przez systemy autonomiczne. Szczególnie interesującym wątkiem była odpowiedzialność rozproszona — sytuacja, w której odpowiedzialność za działanie systemu AI rozkłada się pomiędzy producenta, operatora, użytkownika oraz organizację wdrażającą rozwiązanie.

Jednym z najczęściej powracających wniosków debaty była konieczność utrzymania realnej obecności człowieka w procesie podejmowania decyzji („human-in-the-loop"), zwłaszcza w systemach mających wpływ na bezpieczeństwo ludzi i funkcjonowanie infrastruktury krytycznej.

Debata pokazała również, że rozwój AI w obszarze ITS nie jest wyłącznie problemem technologicznym. Jest to równocześnie wyzwanie cywilizacyjne wymagające dialogu pomiędzy nauką, techniką, biznesem oraz humanistyką.

Panel pozostawił uczestników z refleksją, że mimo ogromnego postępu technologicznego nadal znajdujemy się na etapie intensywnego poszukiwania metod odpowiedzialnego

projektowania systemów inteligentnych — systemów, które będą nie tylko skuteczne, ale również społecznie akceptowalne i zgodne z wartościami człowieka.

Opracowanie ma charakter syntetycznego podsumowania debaty panelowej i zostało przygotowane w celu dalszej refleksji nad problematyką AI Alignment w obszarze ITS.

Opracowanie: Barbara Wójtowicz