

[iTRUST] Interwencje przeciw zjawisku polaryzacji w społeczeństwie na rzecz mediów społecznościowych godnych zaufania: Od diagnozy do terapii

Katarzyna Budzyńska, Politechnika Warszawska, **Polska** (koordynator); Marie-Francine Moens, KU Leuven, **Belgia**; Andrea Rocci, Università della Svizzera italiana (Lugano), **Szwajcaria**; Carles Sierra, Artificial Intelligence Research Institute (IIIA), **Hiszpania**; Virginie Van Ingelgom, UCLouvain, **Belgia**

Cyfryzacja gwałtownie przeobraża nasze społeczeństwa, przekształca dynamikę naszych interakcji, przemienia kulturę naszych debat. Zaufanie odgrywa kluczową rolę w tworzeniu intelektualnej pokory i międzyludzkiej życzliwości w sporze i debacie: bez niego wiarygodność jest skazana na porażkę, reputacja jest zagrożona, współpraca narażona na szwank. Główne zagrożenia związane z cyfryzacją – **mowa nienawiści i fake newsy** – stanowią naruszenie podstawowego warunku dla zaufania i bycia godnym zaufania, które są kluczowe dla konstruktywnej, rozsądnej i odpowiedzialnej komunikacji, a także dla opartej na współpracy i etycznej organizacji społeczeństw. Zachowania te prowadzą ostatecznie do **efektu polaryzacji**, polegającego na tym, że użytkownicy mediów społecznościowych nieustannie wzajemnie się atakują w emocjonalny sposób, skupiając się na tym, co ludzi dzieli, a nie na tym, co ich łączy.

Koncentrując się na dwóch ważnych dziedzinach – równości płci i zdrowiu publicznym – iTRUST dostarczy:

- największego zbioru danych tekstowych, przeanalizowanych pod kątem strategii związanych z użyciem **etosu, patosu i technik ramowania**;
- nową metodologię wielkoskalowej analityki zaufania, służącej wykrywaniu **implicitnych wzorców i trendów** w mowie nienawiści i fake newsach;
- nowatorskiego **empirycznego ujęcia** tego, jak te wzorce wpływają na polaryzację komunikacji online i społeczeństwa jako całości; oraz
- zastosowań opartych na sztucznej inteligencji, które w oparciu o wyniki z poprzednich kroków dostarczą możliwości wykonywania **działań interwencyjnych** przeciwko mowie nienawiści, fake newsom i polaryzacji.

Ze względu na znaczenie dla **społeczeństwa opartego na wiedzy**, projekt kładzie duży nacisk na działania informacyjne i podnoszenie świadomości użytkowników, które zostaną podjęte we współpracy z mediami, muzeami i innymi partnerami.

Konsorcjum składa się z pięciu doświadczonych kierowników projektu, posiadających kompetencje w zakresie retoryki, politologii porównawczej, lingwistyki korpusowej, przetwarzania języka naturalnego, systemów wieloagentowych i obliczeniowej argumentacji. Grupa jest uzupełniona przez partnerów ze środowiska akademickiego oraz przemysłu, którzy zapewnią cenne uzupełnienie kompetencji konsorcjum w dziedzinach takich jak studia nad mediami czy technologie oparte na sztucznej inteligencji. Naszą długofalową ambicją jest ustanowienie europejskiej sieci i fundamentów dla **sztucznej inteligencji godnej zaufania** w odpowiedzi na priorytet Komisji Europejskiej *Europe fit for the Digital Age*.