

Antares Gryczan
Dyrektor Departamentu Utrzymania i Rozwoju Systemów
Polski Fundusz Rozwoju Operacje

„Czy aktualnie GenAI jest gotowa do służby w sektorze publicznym?”

Generatywna sztuczna inteligencja otacza nas ze wszystkich stron, atakuje z nagłówków gazet, portali społecznościowych, reklam, czy wreszcie, z coraz większej liczby artykułów naukowych - trzecia wiosna AI trwa w najlepsze; nie ma dnia, żeby nie pojawiały się ekscytujące zastosowania, nowe modele i teorie. Także w obszarze usług sektora publicznego potencjalne zastosowania wydają się całkiem oczywiste. Na przykład zwiększona efektywność, większa dostępność usług w kanałach samoobsługowych dla obywateli oraz ich personalizacja. Wdrożenie AI w tak specyficznym środowisku jak sektor publiczny wymaga jednak zastanowienia się nad kilkoma zagadnieniami. Ich omówieniu poświęcona będzie moja prezentacja, należą do nich następujące kwestie:

1. Technologia GenAI jest jeszcze na bardzo wczesnym etapie rozwoju, co wiąże się z konsekwencjami, takimi jak: halucynacje, rozliczalność i powtarzalność rezultatów, czy wsparcie modeli AI dla języka polskiego.
2. Istnieją obiektywne przeszkody w implementacji w sektorze publicznym związane z wyzwaniem, takimi jak: integracja z systemami dziedzicznymi, luka kompetencyjna, czy wyzwania, jakie stawia przetwarzanie modeli AI w chmurze obliczeniowej.
3. Oprócz klasycznych zagrożeń związanych z bezpieczeństwem danych, AI niesie ze sobą nowe:
 - Kradzież Modelu (Model Stealing)
 - Wycieki danych (Data Leakage)
 - Ataki poprzez gradienty (Gradient Leakage)
 - Ataki poprzez manipulacje danymi wejściowymi (Adversarial Attacks)
 - Wstrzyknięcie danych (Data Poisoning)
 - Ataki poprzez usuwanie cech (Feature Removal Attacks)
 - Ataki poprzez modelowane wycieki (Model Inversion Attacks)
4. Implementacja musi uwzględnić aktualny krajobraz regulacyjny:
 - AI ACT
 - NIS2/UKSC
 - RODO

Na koniec w swojej prezentacji postaram się przestawić propozycje praktycznego podejścia do zagadnienia implementacji AI w domenie publicznej w kontekście ww. zagadnień.