SOTECHLAB
LABORATORIUM SPOŁECZNO-TECHNOLOGICZNE

Sztuczna inteligencja – czyli jaka? Teraźniejszość i przyszłość „inteligentnych” algorytmów

Czerwiec 2024

Katedra Studiów nad Społeczeństwem i Technologią WH AGH

SoTechLab | Czerwiec 2024

Niniejsza publikacja stanowi podsumowanie dyskusji pt. „Sztuczna inteligencja – czyli jaka? Teraźniejszość i przyszłość „inteligentnych" algorytmów", która odbyła się w formule zdalnej 25 czerwca 2024 r. w ramach Laboratorium Społeczno-Technologicznego (SoTechLab) przy Katedrze Studiów nad Społeczeństwem i Technologią Wydziału Humanistycznego AGH.

Osoby zaproszone do dyskusji:

- Rami Al Naib, Co-owner Numlabs, lider zespołu Data Science&AI
- dr Dota Szymborska, filozofka, etyczka i socjolożka, wykładowczyni akademicka w Uniwersytecie WSB Meritow w Warszawie, członkini GRAI przy Ministerstwie Cyfryzacji
- Martyna Wiącek, AI & Data Strategy Advisor w Aigorithmics, doktorantka na Wydziale Matematyki Stosowanej AGH
- Rafał Wieczerzak, prawnik, doktorant w dyscyplinie nauki prawne w Szkole Doktorskiej w Uniwersytecie Śląskim w Katowicach

Prowadzenie dyskusji

- Katarzyna Cieślak, Szkoła Doktorska AGH, KSSiT WH AGH, SoTechLab, kcieslak@agh.edu.pl
- Jakub Mirek, Szkoła Doktorska AGH, KSSiT WH AGH, SoTechLab, jamirek@agh.edu.pl

Działania Laboratorium w latach 2022-2024 są finansowane w ramach programu „Społeczna odpowiedzialność nauki – Popularyzacja nauki i promocja sportu" Ministerstwa Edukacji i Nauki (SONP/SP/548668/2022).

Więcej informacji o projekcie: https://sotechlab.agh.edu.pl/o-projekcie/

Katedra Studiów nad Społeczeństwem i Technologią WH AGH 2024

https://www.facebook.com/KSSTWHAGH

https://www.youtube.com/@kssitwhagh

https://www.instagram.com/kssitwhagh

Spis treści

Kluczowe pojęcia

Czarna skrzynka (ang. black box) - pojęcie to odnosi się do systemów informatycznych, w których znamy wprowadzane komendy lub dane oraz uzyskujemy wynik końcowy, ale proces przetwarzania wewnątrz systemu pozostaje nieznany czy też niezrozumiały dla przeciętnego użytkownika. Termin ten nie dotyczy wyłącznie systemów informatycznych, lecz także technologii takich jak np. smartfony czy sztuczna inteligencja. Czarno skrzynkowość tych rozwiązań oznacza brak zrozumienia ich działania, co może prowadzić do braku zaufania, szczególnie gdy technologia przestaje być kontrolowalna. Pojęcie to można powiązać z cytatem z pisarza science fiction, Artura C. Clarka, który stwierdził: „Każda wystarczająco zaawansowana technologia jest nierozróżnialna od magii."

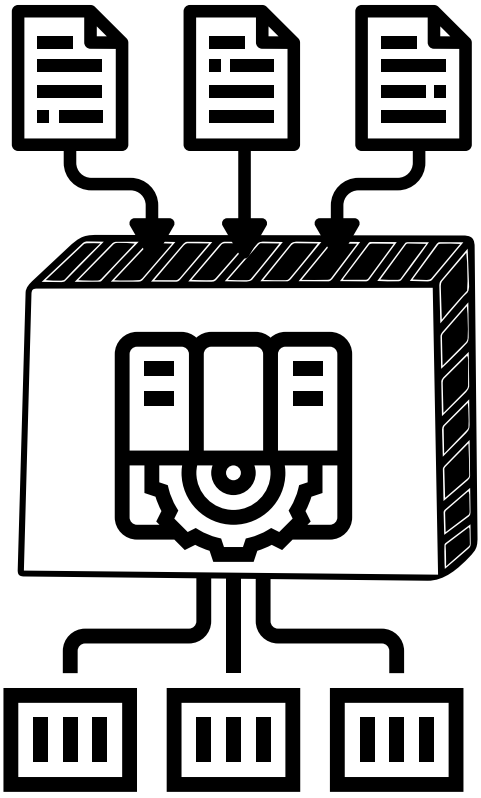
Wyjaśnialna AI (ang. Explainable AI, XAI) - termin ten odnosi się do modeli sztucznej inteligencji (ang. Artificial Intelligence, AI), których działanie można wyjaśnić i zinterpretować. Oznacza to, że rozumiemy, jak przetwarzają dane i generują wyniki, co zwiększa zaufanie do tych systemów. Na przykład, możemy zrozumieć, dlaczego samochód autonomiczny na widok pieszego skręca z drogi lub dlaczego model językowy interpretuje zapisy prawa w określony sposób. Specjaliści, tacy jak matematycy i informatycy orientują się w działaniu tych modeli dzięki znajomości funkcji matematycznych i procesów. W sytuacji gdy działanie AI staje się zbyt skomplikowane, korzystają zazwyczaj z frameworków i programów pomagających wyjaśnić sposób działania systemu. Mimo to, wiele zaawansowanych modeli AI pozostaje niewyjaśnialnych, działając jako "czarne skrzynki".

Halucynacje AI - to zjawisko, w którym modele językowe AI generują odpowiedzi na zadane pytania lub zadania, które nie są oparte na realnych danych, ale na fikcyjnych informacjach stworzonych przez AI. Przykładem halucynacji mogą być biografie nieistniejących osób lub streszczenia książek, których AI nigdy nie analizowała. AI, aby spełnić żądanie użytkownika, tworzy odpowiedzi niezgodne z rzeczywistością, prezentując je z przekonaniem, jakby były prawdziwe. Zjawisko to wynika z faktu, że modele AI są zaprogramowane do generowania najbardziej prawdopodobnych odpowiedzi na podstawie dostępnych im danych, nawet jeśli te dane są niewystarczające lub nieistniejące.

Dane wejściowe

Przetwarzanie danych w “czarnej skrzynce”

Dane wyjściowe



Konteksty technologiczne i społeczne

Trudności definicyjne sztucznej inteligencji

Próby zdefiniowania sztucznej inteligencji za pomocą jednego wyjaśnienia łączącego różne perspektywy (nie tylko informatyczne, ale też społeczne) były podejmowane już w drugiej połowie XX wieku. Jak wskazała **Martyna Wiącek**:

Jeśli chodzi o samo sformułowanie czym ta sztuczna inteligencja jest, o tym, że nie jest to zadanie proste świadczy chociażby to, że kilkadziesiąt lat temu było powołane seminarium, po to, żeby właśnie ustalić definicję sztucznej inteligencji. Wtedy, kiedy jeszcze Alan Turing i jemu współcześni zajmowali się tą sztuczną inteligencją. Nie udało się tego zrobić w ciągu tych kilku dni.

Bardziej współczesne próby doprecyzowania terminu AI zostały podjęte podczas opracowywania przez Parlament Europejski dokumentu regulującego zastosowanie AI, czyli AI Act. Prace nad AI Act rozpoczęły się w lutym 2020 roku, kiedy Komisja Europejska opublikowała "Białą Księgę na temat sztucznej inteligencji – Europejskie podejście do doskonałości i zaufania". Natomiast wedle aktualnej ustalonej przez komisję definicji, system sztucznej inteligencji rozumiany jest jako:

System maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne.

Źródło: art. 3 pkt 1 Rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. (Ai Act)

Dwie cechy programu kwalifikujące go jako system oparty na AI:

- **Samodzielność (autonomia)**: pozwala systemom na dynamiczne reagowanie na zmieniające się warunki i otoczenie. Dzięki tej zdolności AI może nie tylko analizować nowe informacje, ale także dostosowywać swoje działanie i doskonalić wyniki w miarę zdobywania nowych danych. To sprawia, że systemy AI stają się bardziej precyzyjne, efektywne i zdolne do rozwiązywania coraz bardziej złożonych problemów bez konieczności bezpośredniej interwencji człowieka.

- **Zdolność do adaptacji i uczenia się na podstawie nowych danych**: Algorytmy AI stają się bardziej efektywne i dokładniejsze w odwzorowywaniu rzeczywistości, im więcej danych mają do analizy. W AI Act definiuje się AI jako system autonomiczny, który generuje prognozy, ale kluczowe jest to, że systemy te samodzielnie decydują, na jakich podstawach oprzeć swoje działania.

Czarne skrzynki - od funkcji matematycznych do „magii" przetwarzania danych

W trakcie dyskusji Martyna Wiącek odniosła się do koncepcji „czarnej skrzynki" na przykładzie sztucznych sieci neuronowych, które składają się z więcej niż trzech warstw. Wskazała, że choć takie sieci, jak na przykład Chat GPT 3.5, mogą mieć aż 175 miliardów parametrów, to zasadniczo są one jedynie złożonymi funkcjami matematycznymi. Jak podkreśliła Martyna:

> *Niektórzy są zawiedzeni czasem, ale tam nie ma żadnej magii. Tam są liczby na wejściu, po prostu zestaw liczb, który później nam te modele przekładają na ludzki język, bo tak zostały zaprogramowane.*

Jednak analiza działania tych funkcji, choć teoretycznie możliwa, jest rzadko realizowana ze względu na ich złożoność. Jedną z metod wyjaśnialności jest zastosowanie frameworków do interpretowania działania wyników modeli językowych (opartych na metodach uczenia maszynowego). Analiza z ich udziałem pozwala na zrozumienie, które cechy wejściowe wpływają na decyzje podejmowane przez model. Jest to istotne, zwłaszcza w kontekście oceny, czy model podejmuje właściwe decyzje, czy np. nie opiera się na błędnych lub nieetycznych założeniach.

Jednym z przykładowych frameworków, przywołanych w trakcie dyskusji jest LIME (Local Interpretable Model-agnostic Explanations).

Inne platformy o podobnych działaniu wspierające metody wyjaśnialności systemów AI:

- **SHAP (SHapley Additive exPlanations)**: Bazuje na wartości Shapleya z teorii gier i umożliwia wyjaśnianie predykcji modeli poprzez ocenę wkładu każdej cechy wejściowej.

- **Anchors**: Metoda ta tworzy tak zwane „kotwice" (anchors), czyli wyjaśnienia w postaci jeśli-wtedy. „Kotwice" określają warunki, w ramach których model predykcyjny zachowa się w określony sposób.

- **DeepLIFT** (Deep Learning Important FeaTures): Technika ta analizuje różnice w aktywacjach neuronów w głębokich sieciach neuronowych, aby przypisać wartości wejściom, które prowadzą do danej decyzji modelu.

- **Counterfactual Explanations**: Koncepcja ta polega na przedstawieniu alternatywnych scenariuszy, które mogłyby zmienić wynik modelu, aby wyjaśnić decyzje modelu. Na przykład, można zmienić pewne cechy wejściowe i zobaczyć, jak wpłynie to na predykcję.

Wykorzystanie AI w praktyce

Codzienne zastosowania modeli AI opartych na interfejsie konwersacyjnych są różnorodne i obejmują zarówno sfery zawodowe, jak i osobiste. O zastosowaniach AI w mikroskali, do realizacji codziennych zadań, mówiły na swoich przykładach zaproszone do SoTechLabu osoby.

Rami Al Naib korzysta z AI do sprawdzania prezentacji, ofert i maili, co pozwala mu poprawić jakość swojej pracy i zaoszczędzić czas. Dodatkowo, AI generuje dla niego grafiki, które może np. wykorzystać w prezentacjach.

Dota Szymborska wykorzystuje AI w swojej pracy naukowej do przeprowadzania burzy mózgów i tworzenia streszczeń artykułów. Dzięki temu może szybciej się rozwijać i efektywniej przyswajać wiedzę.

Rafał Wieczerzak używa AI jako notatnika i pamiętnika do zapisywania pomysłów naukowych i biznesowych, a także do tworzenia streszczeń publikacji. AI pomaga mu w organizacji i szybszym przetwarzaniu informacji.

Martyna Wiącek oprócz wykorzystania AI w pracy naukowej, używa AI do inspirowania się, organizowania materiałów do prezentacji i szkoleń, co ułatwia jej codzienną pracę i pozwala na lepsze przygotowanie wystąpień.

Przede wszystkim jednak, oprócz codziennych zastosowań, AI wykorzystywane jest również w skali makro, w biznesie (na coraz większą skalę) czy sektorze publicznym (jeszcze na niewielką skalę).

Predykcja i prognoza metodami AI

Martyna Wiącek zwracała uwagę na rosnącą rolę AI w prognozowaniu, które jest istotne w wielu dziedzinach, jak biznes i energetyka. Wskazywała, że tradycyjne metody matematyczne, takie jak analiza danych z przeszłości, były i są używane do przewidywania przyszłych trendów, na przykład cen akcji czy zapotrzebowania na energię.

Jednak dzięki AI, prognozowanie staje się coraz bardziej zaawansowane. AI, wykorzystując techniki takie jak sieci neuronowe, może automatycznie analizować dane i dostarczać prognozy. Przykładowo, w Pythonie można łatwo zaimplementować te techniki za pomocą kilku linijek kodu.

Martyna opisała dwa sposoby, w jakie AI może współpracować z tradycyjnymi metodami:

- **Kalibracja modeli matematycznych przy pomocy AI**: AI może pomóc dostosować parametry matematycznych modeli prognozowania na podstawie rzeczywistych danych. To oznacza, że możemy poprawić dokładność prognoz, precyzyjniej dostosowując model do rzeczywistości.

- **Wykorzystanie wyników matematycznych w SI**: Wyniki z tradycyjnych metod matematycznych mogą być używane jako dodatkowe dane dla algorytmów SI, co może poprawić dokładność prognoz.

Martyna podkreślała, że prognozowanie szeregów czasowych w rzeczywistych zastosowaniach stanowi wyzwanie, wymagające starannego doboru odpowiednich metod. Na przykład, w projektach, które mają prognozować zapotrzebowanie na produkty farmaceutyczne, gdzie mnogość zmiennych znacząco wpływa na trafność przewidywań.

Wykorzystanie AI w przedsiębiorstwie

Metody AI leżące u podstaw modeli językowych zdaniem **Ramiego Al Naib** dobrze radzą sobie z zadaniami o ogólnym charakterze, jednak napotykają trudności w rozwiązywaniu specyficznych, specjalistycznych problemów wymagających pogłębionej wiedzy domenowej. Firmy często unikają inwestycji w zaawansowane rozwiązania AI, ze względu na wysokie koszty i skomplikowany proces wdrażania.

W momencie jak zaczynamy mówić o wiedzy domenowej, o wiedzy niespopularyzowanej i zaczynamy rozwiązywać zadania które są rzeczywiście bardzo specyficzne i wymagają ekspertyzy, to koszt opracowania rozwiązania jest na tyle duży że wiele firm się po prostu na to nie decyduje. **- powiedział Rami Al Naib**

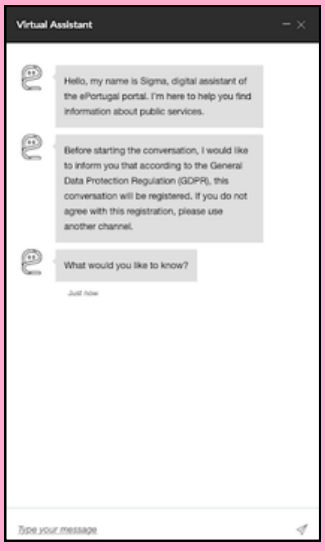
Nasz rozmówca podał przykład projektu "Mecenas AI," który ma na celu zautomatyzować weryfikację dokumentów prawnych. Zastosowanie metod AI w tym przypadku pozwoliłoby na odciążenie z rutynowej pracy zespoły prawnicze. Choć rozwój takich technologii skrócił się z miesięcy do tygodni, a nawet dni, wciąż widoczny jest sufit możliwości obecnych narzędzi.

Wykorzystanie AI w systemie prawnym

Prawo to dziedzina szczególnie "wrażliwa" na wdrożenia AI do codziennego użytku ze względu na ryzyko stronniczości (osób lub podejmowania decyzji w oparciu o obciążone nią dane tzw. AI bias). W trakcie spotkania podjęliśmy dyskusję na temat zastosowań AI w obszarze sądownictwa. **Rafał Wieczerzak** wyróżnił kilka obszarów obecnych i możliwych wdrożeń:

- Organizacja działania sądów

Automatyzacja w zakresie ewidencjonowania i organizacji dokumentacji wpływającej do sądu. Wspieranie komunikacji z osobami pracującymi w sądzie i zwracających się z prośbą o udzielenie informacji za pomocą interfejsów konwersacyjnych (np. czatbot może podać informacje o statusie sprawy na podstawie podanej sygnatury).



Dzięki zastosowaniu czatbotów osoby zainteresowane statusem sprawy mogą szybko uzyskać informacjeo aktualnym stanie sprawy.

W Portugalii Ministerstwo Sprawiedliwości prowadzi badania nad asystentem AI, do wsparcia osób w komunikacji z sądem. W szczególności skupiono się na obszarze prawa rodzinnego. Program ma na celu ułatwiać kontakt z sądem w sprawach rodzinnych pomagając np. w kompletowaniu dokumentów potrzebnych do ustalenia alimentów. Obecnie na stronie Ministerstwa można rozmawiać z chatbotem Sigma.

- Analiza danych

Metody AI są wykorzystywane do przetwarzania dużych zbiorów danych, a w sądownictwie systemy AI mogą identyfikować i generować linie orzecznicze, np. w sprawach kredytów frankowych, gdzie umowy są często bardzo podobne. Systemy AI mogą być stosowane w prostych sprawach rejestrowych, takich jak zmiana adresu spółki w Krajowym Rejestrze Sądowym Choć takie operacje nie są skomplikowane, nadal wymagają, aby takiej zmiany dokonał referendarz sądowy W tym przypadku nie jest potrzebna złożona analiza prawna, więc AI mógłyby zautomatyzować rutynowe czynności orzecznicze.

- Wspieranie decyzji orzeczniczych

AI może sugerować sędziom decyzje na podstawie analizy i wydanych orzeczeniach w innych sprawach, co może przyczynić się do większej spójności i przewidywalności wyroków. Choć jak podkreślał w trakcie dyskusji Rafał Wieczerzak „(...) nie wyobrażam sobie stosowania systemów sztucznej inteligencji w takich postępowaniach jak rodzine, np. w zakresie przyznania praw rodzicielskich, a także w postępowaniach karnych. W tym przypadku mamy aprzykład amerykańskiego systemu COMPAS , który służy do oceny ryzyka recydywy. No i tutaj były dosyć spore problemy związne ze stronniczością rasową i nieprzejrzystością tego rozwiązania.

Wyzwania

Praca ludzi a rozwój AI

AI to narzędzie, które przez wiele osób postrzegane jest jako przyczynek do kolejnej rewolucji na rynku pracy. Obecnie, jak zauważył Rafał Wieczerzak, AI jest przede wszystkim używana do prostych zadań, ale nie jest jeszcze gotowa do podejmowania poważnych decyzji, na przykład w sprawach sądowych.

Mimo tego, AI może wpływać na rynek pracy, szczególnie dla osób na początku swojej kariery, takich jak stażyści czy juniorzy. Rami Al Naib zauważał, że AI może wpłynąć na pracowników wykonujących mniej skomplikowane zadania:

> „Rzeczywiście, tymi rzeczami zajmują się osoby, które jeszcze tego doświadczenia nie mają. I tak, jest istotne ryzyko, że w nadchodzących latach będziemy zwalniać ludzi w działach customer supportu, czy w działach administracji".

Rami podkreślał potrzebę analizy, czy rzeczywiście takie zmiany będą prowadzić wyłącznie do zaniku stanowisk, czy raczej do ich transformacji. Porównując obecną sytuację z rewolucją przemysłową, można zauważyć pewne podobieństwa.

Wprowadzenie nowych technologii w przeszłości również prowadziło do obaw o utratę miejsc pracy, ale ostatecznie przyniosło zmiany w strukturze rynku pracy, a efektywność wzrosła.

Jak zauważył Rami:

> „Proces produkcji stał się wielokrotnie szybszy i tańszy, ale rynek się sam wyregulował. W długim terminie to wcale nie doprowadziło do negatywnych skutków, wręcz przeciwnie, do pozytywnych."

Podobnie, obecne zmiany związane z AI mogą prowadzić do przekształcenia ról w pracy, gdzie osoby będą zajmować się nadzorowaniem procesów, które wcześniej były wykonywane ręcznie. Jak zauważyła Katarzyna Cieślak: „AI być może nie zabierze nam pracy, ale zabierze nam taski".

Współczesne podejście do AI w pracy często zakłada model hybrydowy, gdzie AI (jak w sądownictwie) wspiera ludzką pracę, a nie całkowicie ją zastępuje. To podejście może pomóc w adaptacji do zmian na rynku pracy i w efektywnym wykorzystaniu nowych technologii.

Jakość i braki danych a rzetelność i zaufanie do AI

Sztuczna inteligencja boryka się z wyzwaniami związanymi z jakością i wiarygodnością danych, na których się uczy (wkład/dane, który poza samym procesem przetwarzania, determinuje uzyskany wynik).

W trakcie dyskusji Martyna Wiącek podkreślała, że AI ma swoje ograniczenia i może być podatne na błędy, szczególnie gdy przyjmuje za pewne informacje przekazane przez użytkowników, nawet jeśli są one błędne.

Jak zauważyła Martyna Wiącek: „Można mu [czatowi] wmówić, że coś co powiedział dobrze, jest nieprawdą i on wtedy też się poprawi, bo on zakłada, że ten jego rozmówca jest najmądrzejszy na świecie". To pokazuje, jak kluczowe są dane i sposób, w jaki są one interpretowane przez AI.

Problem ten wynika z faktu, że AI jest uczona na podstawie danych, które mogą być niepełne lub błędne. Modele AI, takie jak Chat GPT, mogą tworzyć i prezentować z pełnym przekonaniem nieistniejące artykuły naukowe czy książki, gdy brakuje im wiarygodnych informacji. Takie „halucynacje" mogą prowadzić do błędnych decyzji i twierdzeń, co wpływa na spadek zaufania do tych systemów.

W miarę jak AI staje się bardziej zaawansowane, pojawia się również obawa, że przyszłe generacje modeli mogą uczyć się na danych generowanych przez inne AI, co może prowadzić do degeneracji wiedzy. Jak donoszą naukowcy w czasopiśmie Nature, istnieje ryzyko, że takie „samouczące się" systemy przestają być efektywne, gdyż będą oparte na coraz mniej rzetelnych źródłach danych.

Martyna Wiącek wskazywała na znaczenie przygotowania danych szkoleniowych dla AI: „Odpowiednie przygotowanie danych jest tutaj kluczowe, ponieważ stanowią one fundament wszelkich decyzji podejmowanych przez systemy sztucznej inteligencji." Dlatego konieczne jest zapewnienie AI rzetelnych informacji oraz systematyczne udzielanie jej informacji zwrotnych, aby poprawić dokładność i wiarygodność jej odpowiedzi.

Niemniej jednak, wyzwaniem pozostaje „czarnoskrzynkowość" AI, czyli brak transparentności w działaniu modeli. Jak zauważyła Martyna: „Mają dużo parametrów, natomiast jesteśmy w stanie

rozwijać metody, które pomogą zrozumieć, w oparciu o co te metody działają, jak reguły zostały sformułowane." Zrozumienie, jakie cechy wpływają na działanie modelu, może pomóc w wykrywaniu błędów i poprawie jakości odpowiedzi.

Kwestie etyczne w AI

Jak zauważyła **Dota Szymborska**, AI, nie jest również wolne od problemów związanych z moralnością i decyzjami, które muszą być podejmowane w sytuacjach kryzysowych. Badaczka zwracała uwagę na to, że SI może powielać ludzkie błędy i uprzedzenia. Jak mówiła: „Technologia nie jest wolna od problemów związanych z równością i sprawiedliwością, które są obecne w społeczeństwie.

Dota wskazywała na potrzebę humanizowania AI oraz wyzwań związanych z rozumieniem i implementowaniem etyki przez te systemy. Porównywanie generatywnej sztucznej inteligencji do ludzi, bywa jednak mylące. Z jednej strony lubimy mówić, że Chat GPT 3 to kilkulatek, a Chat GPT to już student, ale takie porównania nie odpowiadają rzeczywistości, bo tak naprawdę czym innym jest ludzka inteligencja a czym innym ta sztuczna. O czym wiedział już Turing, proponując sławny test, sprawdzający czy sztuczna inteligencja osiąga już poziom ludzkiej.

Dylemat wagonika, znany również jako problem etyczny w kontekście autonomicznych pojazdów, jest przykładem trudnych decyzji, które AI musi podejmować. Jak wspominała Szymborska, „W przypadku prób unikania wypadków samochodowych, AI może stanąć przed wyborem między śmiercią kierowcy a uratowaniem pieszego." To stawia pytanie, jak AI powinna podejmować decyzje w sytuacjach, które mają poważne konsekwencje moralne.

AI Act, który wyznacza standardy prawne i etyczne dla sztucznej inteligencji, jest krokiem w kierunku rozwiązania tych problemów. Jak podkreślała Szymborska, ten akt reguluje kwestie etyczne, ale nie rozwiązuje wszystkich dylematów związanych z moralnością w kontekście AI.

Prywatność danych wykorzystywanych do szkolenia AI

Każdy autorski model AI potrzebuje danych, aby możliwy był proces uczenia maszynowego. Dane te są zbierane z różnych źródeł, w tym mediów społecznościowych, które dostarczają ogromnych ilości informacji o zachowaniach, preferencjach i interakcjach użytkowników.

Korporacja Meta, znana z platform takich jak Facebook i Instagram, zamierza używać postów, zdjęć i ich podpisów opublikowanych na tych platformach, do szkolenia swojej AI. Chociaż Meta teoretycznie umożliwia użytkownikom wyrażenie sprzeciwu wobec takich działań, proces ten jest skomplikowany.

Z kolei Elon Musk, właściciel portalu X (dawniej Twitter), wprowadził politykę, zgodnie z którą użytkownicy są domyślnie zobowiązani do udostępniania swoich danych (co jednak ważne - i tak ogólnodostępnych dla użytkowników portalu), co ma na celu wspieranie rozwoju sztucznej inteligencji oraz innych technologii.

W obu przypadkach pojawiają się pytania o etykę i transparentność w zakresie wykorzystania danych użytkowników. Rodzi to również wyzwania dotyczące prawa autorskiego. Czy AI ma prawo korzystać i przetwarzać cudze dzieła?

Dostęp do rozwiązań AI

Obecnie, korzystając z AI, "płacimy" za to głównie poprzez dostarczanie danych i ocen, które pomagają w szkoleniu modeli (np. w Chat GPT oceniamy przyciskiem "łapki" w górę lub dół czy odpowiedź nas zadowala). W przyszłości jednak, AI mogą stać się narzędziami, za które płaci się bezpośrednio. Przykładem tego są już istniejące wersje Chat GPT – zarówno darmowa, jak i płatna w abonamencie.

W kontekście pracy z AI, kluczowe stają się kompetencje związane z obsługą tych technologii. Należy do nich umiejętność efektywnego promptowania, czyli wydawania precyzyjnych komend do AI. Rafał Wieczerzak podkreślał, że skuteczne tworzenie promptów wymaga odpowiednich umiejętności; ogólne prompty dają słabe wyniki, dlatego istotne jest precyzyjne formułowanie zapytań i edytowanie wyników generowanych przez AI.

Z drugiej strony Dota Szymborska zwracała uwagę, że AI może być narzędziem, które umożliwia osobom mniej zaawansowanym osiągnięcie wyższego poziomu umiejętności. W ten sposób, AI nie tylko automatyzuje istniejące procesy, ale także zmienia sposób, w jaki zdobywamy i rozwijamy kompetencje.

Predykcje i rekomendacje

Na zakończenie dyskusji uczestnicy zostali poproszeni o prognozy dotyczące przyszłości sztucznej inteligencji.

Dota Szymborska podkreślała, że jednym z ważnych czynników będzie dostępność energii, ponieważ rozwój AI wymaga ogromnych zasobów energetycznych, co może wpłynąć na tempo jej adopcji.

Rami Al Naib zwrócił uwagę, że obecne technologie umożliwiają automatyzację prostych, powtarzalnych zadań, co może prowadzić do eliminacji niektórych zawodów. Prognozował, że w ciągu najbliższych pięciu lat nastąpi wzrost zastosowań AI, ale tempo zmian będzie dostosowane do tempa zmian społecznych.

Rafał Wieczerzak zauważył, że sektor publiczny, w tym sądownictwo, pozostaje w tyle za sektorem prywatnym w zakresie wdrażania AI, ale wskazał na istniejące już systemy wspomagające pracę sądów w innych krajach, takich jak Francja, Brazylia i Chiny. Podkreślił, że na poziomie krajowym konieczne jest najpierw przeprowadzenie cyfryzacji i digitalizacji, aby móc wprowadzać bardziej zaawansowane technologie.

Martyna Wiącek wskazała, że technologię AI najszybciej adaptują pojedyncze osoby oraz biznesy, które dążą do zwiększenia efektywności i zysków. Podkreśliła jednak, że wdrożenie rozwiązań AI w biznesie nie jest proste i wymaga dużych nakładów pracy. Wskazała także na brak gotowości regulacyjnej oraz na fakt, że rozwój AI będzie szybki, ale nie należy oczekiwać, że w ciągu kilku lat zdominuje ona wszystkie sektory.

Na podstawie analizy (wspomaganej Chatem GPT 4.0) poczynionych w trakcie dyskusji obserwacji, rekomendujemy poniższe działania związane z rozwojem AI.

Zwiększenie przejrzystości i wyjaśnialności systemów AI

- Implementacja metod wyjaśnialności: Warto kontynuować rozwijanie i wdrażanie frameworków takich jak LIME, SHAP, Anchors oraz DeepLIFT, aby umożliwić lepsze zrozumienie i interpretację wyników generowanych przez modele AI. Umożliwi to użytkownikom lepsze zrozumienie podstaw decyzji podejmowanych przez AI.
- Edukacja na temat wyjaśnialności: Warto promować edukację w zakresie wyjaśnialności AI wśród społeczeństwa, aby zwiększyć zdolność do rozumienia i interpretowania działania systemów AI, a także zwiększać zaufanie do nich.

Poprawa jakości i wiarygodności danych

- Weryfikacja danych: Warto inwestować w technologie i procesy umożliwiające dokładniejszą weryfikację danych wejściowych, aby zminimalizować ryzyko błędnych informacji i „halucynacji" AI.
- Standardy Jakości danych: Warto tworzyć standardy jakości danych, które będą stosowane podczas trenowania modeli AI. Wprowadzenie takich standardów pomoże w zapewnieniu rzetelności wyników i zwiększy zaufanie do systemów AI.

Rozwój i zastosowanie etycznych ram dla AI

- Przeciwdziałanie uprzedzeniom: Warto opracować strategie i technologie do identyfikowania i minimalizowania uprzedzeń w danych oraz algorytmach. Należy regularnie audytować i testować modele AI pod kątem etycznym.
- Kodeks etyczny: Warto wprowadzić kodeks etyczny dla rozwoju i stosowania AI, który będzie obejmować zasady odpowiedzialności, przejrzystości i sprawiedliwości w działaniu algorytmów (lub rozwijać już istniejące).

Analiza wpływu AI na rynek pracy

- Badanie wpływu na zawody: Warto prowadzić regularne badania wpływu AI na rynek pracy, zwłaszcza w kontekście młodszych pracowników i osób na początku kariery zawodowej, aby lepiej zrozumieć, jakie zmiany mogą zajść i jak najlepiej przygotować się na nadchodzące wyzwania.
- Wsparcie w przekształceniu ról: Warto myśleć już teraz o strategiach wsparcia dla pracowników, których zadania zostaną zautomatyzowane, koncentrując się na przekształceniu ich ról oraz rozwoju nowych umiejętności potrzebnych w zmieniającym się środowisku pracy.

Promowanie interdyscyplinarnej współpracy

- Współpraca międzybranżowa: Warto wspierać współpracę między specjalistami z różnych dziedzin (np. inżynierami AI, etykami, prawnikami, socjologami) w celu tworzenia bardziej zrównoważonych i odpornych na błędy systemów AI.
- Integracja społeczna i technologiczna: Warto rozwijać projekty i inicjatywy, które integrują aspekty społeczne i technologiczne, aby zapewnić, że rozwój AI uwzględnia zarówno techniczne, jak i społeczne perspektywy.

Warto przeczytać

- European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council on harmonized rules on Artificial Intelligence* (AI Act), https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=OJ:L_202401689.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). *People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and machines*, 28, 689-707.
- Heaven, D. (2020). *Why asking an AI to explain itself can make things worse*. MIT Technology Review. https://www.technologyreview.com/2020/01/29/304857/why-asking-an-ai-to-explain-itself-can-make-things-worse/.
- Morkisz, P., Wiącek, M., & Wochlik, I. (2023). *Wykorzystanie metod obliczeniowych i sztucznej inteligencji w bezpieczeństwie energetycznym*. 'Wiedza Obronna', 283(2), http://wiedzaobronna.edu.pl. https://doi.org/10.34752/2023-e283.
- Wamba, S. F., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). *Are we preparing for a good AI society? A bibliometric review and research agenda*. 'Technological Forecasting and Social Change', 164, 120482.

Nagrania z dyskusji w ramach Laboratorium społeczno-technologicznego (SoTechLab) są dostępne na Kanale YouTube Katedry Studiów nad Społeczeństwem i Technologią WH AGH: youtube.com/@kssitwhagh

Działania Laboratorium społeczno-technologicznego (SoTechLab) w latach 2022-2024 są finansowane w ramach programu „Społeczna odpowiedzialność nauki – Popularyzacja nauki i promocja sportu" Ministerstwa Edukacji i Nauki (SONP/SP/548668/2022).

https://sotechlab.agh.edu.pl/

Katedra Studiów nad Społeczeństwem i Technologią WH AGH 2024

https://www.facebook.com/KSSTWHAGH

https://www.youtube.com/@kssitwhagh

https://www.instagram.com/kssitwhagh