Współpraca międzynarodowa w oparciu o godną zaufania Sztuczną Inteligencję

Newsletter sekcji

11 grudnia 2024

Ministerstwo Cyfryzacji | Tu tworzymy przyszłość

GRAI GRUPA ROBOCZA DS. SZTUCZNEJ INTELIGENCJI

Od redakcji

Szanowni Państwo,
Drodzy Przyjaciele i Przyjaciółki Sekcji,

Z radością przekazujemy w Państwa ręce pierwszy numer newslettera naszej Sekcji ds. współpracy międzynarodowej w oparciu o godną zaufania sztuczną inteligencję. Dzięki wsparciu Ministerstwa Cyfryzacji oraz zaangażowaniu Grupy Roboczej ds. Sztucznej Inteligencji (GRAI) stworzyliśmy publikację, która ma na celu wspieranie wymiany wiedzy i budowanie międzynarodowego dialogu w tej dynamicznie rozwijającej się dziedzinie.

W GRAI mamy świadomość tego, że sztuczna inteligencja nie jest już wyłącznie obszarem technologii, bo niesie za sobą również wyzwania społeczne, etyczne i prawne. Chcemy wspólnie z Państwem rozwijać projekty i pomysły, które uczynią sztuczną inteligencję bardziej przejrzystą, inkluzywną i godną zaufania. Państwa opinie i sugestie będą dla nas nieocenionym źródłem inspiracji.

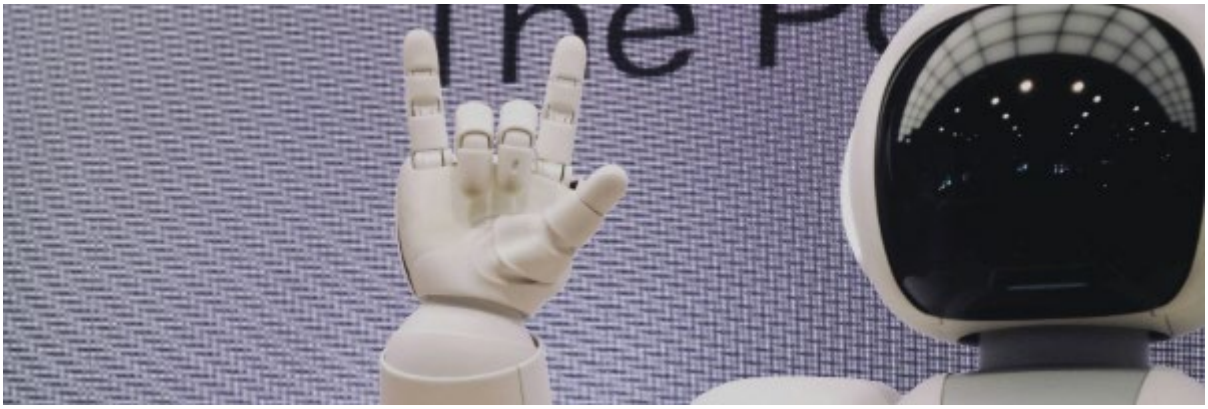
Serdecznie zapraszamy do lektury oraz współtworzenia przyszłych wydań. Dziękujemy za dołączenie do tego przedsięwzięcia – razem możemy kształtować przyszłość AI, która służy społeczeństwu i promuje międzynarodową współpracę[1].

Michał Jackowski
Zuzanna Karcz

[1] Newsletter przygotowali eksperci działający na zasadach pro publico bono w Grupie Roboczej ds. Sztucznej Inteligencji (GRAI) przy Ministerstwie Cyfryzacj w ramach sekcji Współpraca międzynarodowa w oparciu o godną zaufania Sztuczną Inteligencję.

AI dostaje ręce? Nowa funkcjonalność Anthropic Computer Use

Leszek Tasiemski



Zdjęcie autorstwa Possessed Photography on Unsplash

22 października 2024 firma Anthropic ogłosiła nową funkcjonalność swoich modeli Claude 3.5 Sonnet oraz Haiku: computer use. Pozwala ona **modelom AI na bezpośrednią obsługę komputera poprzez przejęcie kursora myszki oraz emulację klawiatury do wpisywania tekstów oraz komend.**

Możemy to uznać za przełomowy moment - tego typu funkcjonalność daje modelom AI "ręce" i sprawia, że mogą one nie ograniczać się do generowania treści albo kodu, który następnie użytkownik-człowiek wykorzysta, lecz bezpośrednio wykonywać akcje. Łatwo sobie wyobrazić, jakie konsekwencje mogą mieć takie działania w realnym świecie. Tego typu funkcjonalność ma oczywiste zalety, np. agent AI może za nas zamawiać usługi czy produkty w sieci, wysyłać wiadomości, wykonywać zadania związane z administracją firmy czy zarządzaniem infrastrukturą IT, bez potrzeby kosztownych skryptów oraz integracji API.

Musimy jednak pamiętać, że jest to potencjalnie bardzo potężne narzędzie w rękach przestępców, cyberprzestępczych grup APT. AI będzie mogło na masową skalę dokonywać ataków, które dotychczas wymagały obecności "człowieka przy klawiaturze". Wybiegając jeszcze dalej w przyszłość, możliwość bezpośredniej kontroli nad komputerem podłączonym do sieci, może stworzyć realny scenariusz “ucieczki” AI, poprzez ominięcie zabezpieczeń nałożonych na model przez człowieka. Na razie nie ma powodów do obaw - przy obecnym zaawansowaniu modeli AI, do takiego scenariusza jest nam jeszcze daleko.

Ogłoszenie Anthropic pokazuje kierunek rozwoju technologii AI, który sprawi, że będzie ona praktyczniejsza w użyciu i ograniczy konieczność tworzenia oraz utrzymywania kosztownych integracji. Funkcjonalność jest obecnie dostępna publicznie w wersji beta.

Więcej informacji w artykule pod tytułem Introducing computer use, a new Claude 3.5 Sonnet, and Claude 3.5 Haiku.

Narodowe podejście do sztucznej inteligencji

Maciej Majewski

Większość publikacji i wystąpień prezesów firm wspaniałej siódemki nasyconych się optymizmem i wiarą, że sztuczna inteligencja niesie ze sobą wspaniałe perspektywy rozwoju ludzkości. Takie hasła przyświecały też założycielom firmy OpenAI, która deklarowała, że bezpłatnie udostępniać nam będzie swoje narzędzia. Jednocześnie od lat osiemdziesiątych ubiegłego wieku jest oczywiste, że tak zwane wysokie technologie, na przykład użyte w programie Star Wars (Strategic Defence Initiative), są narzędziem uzyskiwania globalnej przewagi. Przypatrzmy się, jak rozwija się **nacjonalistyczne podejście do sztucznej inteligencji**.

W ostatnim czasie narodowi regulatorzy stworzyli szereg obowiązków i zakazów ograniczających globalny dostęp do technologii. Są tego liczne przykłady: Rosja i Chiny wymagają, aby zagraniczni dostawcy ujawniali kody źródłowe dostarczanych aplikacji

przetwarzających duże wolumeny danych. Kraje NATO generalnie zakazały sprzedaży do Chin i Rosji najnowszych modeli mikroprocesorów używanych do uczenia i przetwarzania modeli generatywnej sztucznej inteligencji. USA zakazały firmom sektora VC (kapitału wysokiego ryzyka) inwestycji w nowe przedsięwzięcia w Chinach. Liczne kraje, w tym Kanada i Chile, oferują ekspertom AI atrakcyjne warunki płacowe, mające zachęcić ich do emigracji do tych krajów. Te przykłady pokazują, że kraje konkurują o zasoby ludzkie, dzięki którym spodziewają się uzyskać przewagę w technologiach AI.

Jednocześnie tworzone są reguły gry, ograniczające rozwój tych dziedzin u globalnych konkurentów. Ograniczenia mogą dotyczyć wszystkich rodzajów zasobów zasilających łańcuch wartości SI, a więc kapitału, danych, umiejętności i zdolności przetwarzania. Z uwagi na ograniczoność tych zasobów, preferencyjne warunki dla krajowych firm muszą negatywnie odbić się na konkurentach z innych krajów. Stąd wzięło się pojęcie narodowej sztucznej inteligencji. Używa się również pojęcia suwerennej SI zdefiniowanej jako: „oparty o narodową strategię rozwój SI, którego celem jest ochrona interesów narodowych, bezpieczeństwa, konkurencyjności gospodarki oraz dobrostanu społeczeństwa".

Wiele z tych krajów, które wprowadzają bariery współpracy w dziedzinie AI, jest jednocześnie sygnatariuszami różnych międzynarodowych porozumień promujących zasady ładu korporacyjnego, odpowiedzialnego rozwoju itd. Jak kraje te chcą zachować wiarygodność, pozostaje pytaniem bez odpowiedzi.

Więcej na ten temat w artykule aut. Susan Ariel Aaronson: The Age of AI Nationalism and Its Effects, wyd. Center for International Governance Innovation, #306, 2024 [dostęp 11.12.2024]

Tajlandia: Google inwestuje 1 mld USD w rozwój tajskiego AI

Agata Konieczna

Google ogłosiło inwestycję o wartości 36 miliardów batów (ok. 1 miliarda dolarów) w nowe centrum danych w prowincji Chonburi w Tajlandii. To posunięcie jest częścią strategii Google w regionie Azji, mającej na celu wzmocnienie infrastruktury chmurowej i ekspansję w obszarze sztucznej inteligencji.

Inwestycja wpisuje się w globalny wyścig technologiczny, zwłaszcza w obliczu rosnącej konkurencji ze strony Microsoftu i OpenAI. Google stawia na rozwój swoich modeli

transformacyjnych, będących fundamentem współczesnych narzędzi AI, takich jak generatywne modele językowe (jak np. ChatGPT, a w przypadku Google - Gemini). Waga inwestycji w Tajlandii jest o tyle znacząca, że pokazuje, jak,istotna staje się Azja w globalnej strategii Google, szczególnie w obszarach ogólnie pojętych technologii AI i usług chmurowych.

Wzmocnienie infrastruktury technologicznej w Azji nie tylko umożliwi Google skuteczniejsze dostarczanie usług na regionalnym rynku, ale również przyczyni się do rozwoju lokalnych ekosystemów technologicznych. Jest to istotna inicjatywa - warto będzie się przyglądać jej z perspektywy globalnych trendów w rozwoju technologii cyfrowych, a zwłaszcza z perspektywy Polski, w której także Alphabet Inc. zainwestowało greenfieldowo znaczne kwoty.

Więcej informacji dostępne jest m. in. w artykule Business Insidera .

EMA przyjmuje dokument refleksyjny na temat AI/ML w cyklu życia produktów leczniczych

Jagoda Miszewska

Europejska Agencja Leków (EMA) 30 września 2024 opublikowała zaktualizowany dokument Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle, w którym przedstawiono aktualne podejście odnośnie wykorzystania sztucznej inteligencji w celu wspierania bezpiecznego i skutecznego opracowywania, regulowania i stosowania leków zarówno tych przeznaczonych dla ludzi, jak i leków weterynaryjnych. Celem publikacji jest omówienie ogólnych wytycznych dla stosowania AI i uczenia maszynowego na każdym etapie cyklu życia produktów leczniczych.

Po rocznym okresie konsultacji, uwzględniono opinie 66 interesariuszy, w tym organów regulacyjnych i konsorcjów publicznych. Rewizja została oparta na 1342 otrzymanych komentarzach oraz niedawno przyjętym unijnym AI Act. EMA wyjaśniła, że niektóre sugestie, wykraczające poza zakres wstępnych rozważań, zostaną wzięte pod uwagę przy opracowywaniu przyszłych formalnych wytycznych naukowych w dziedzinie AI. W ramach aktualizacji zharmonizowano terminologię techniczną, rozszerzono słowniczek pojęć oraz doprecyzowano zakres dokumentu, obejmując również technologie tradycyjnego uczenia maszynowego.

W odpowiedzi na uwagi interesariuszy **EMA zastąpiła określenie "wysokiego ryzyka" terminami "wysokiego wpływu regulacyjnego" i "wysokiego ryzyka dla pacjenta"**, aby dostosować się do specyficznego języka zawartego w załączniku 1 do AI Act. Agencja podkreśliła, że w przypadku wykorzystania modeli AI lub usług stron trzecich o wysokim wpływie regulacyjnym lub wysokim ryzyku dla pacjenta, oczekuje się od twórców oprogramowania zastosowania procesu kwalifikacji metodologii. EMA zaznaczyła, że choć technologia AI ma potencjał do usprawnienia wielu aspektów cyklu życia produktów leczniczych, kluczowe znaczenie ma zachowanie wiarygodności dla regulatorów, płatników i pacjentów przy wprowadzaniu nowych technologii.

Polska jako członek UE, jest zobowiązana do przestrzegania wytycznych EMA, co może wymagać dostosowania krajowych przepisów oraz procedur, szczególnie w obszarze rejestracji produktów leczniczych (organem odpowiedzialnym jest Urząd Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych). Polskie firmy farmaceutyczne i biotechnologiczne będą musiały uwzględnić omawiane rekomendacje, aby uniknąć ryzyka komplikacji przy wnioskowaniu o pozwolenie na dopuszczenie do obrotu dla swoich produktów.

Zdecydowaną zaletą jest zwiększenie bezpieczeństwa pacjentów poprzez zalecane kontrole systemów AI stosowanych przy produkcji leków. Jednak wprowadzenie nowych standardów wiąże się również z potrzebą zwiększonych inwestycji w infrastrukturę IT i bezpieczeństwo danych w przemyśle farmaceutycznym oraz koniecznością podnoszenia kompetencji pracowników tej branży.

Więcej informacji w artykule EMA adopts reflection paper on AI/ML in drug development.

Sekcja ds. współpracy międzynarodowej w oparciu o godną zaufania sztuczną inteligencję

Grupa Robocza ds. Sztucznej Inteligencji

Sekcja działa w ramach Grupy Roboczej ds. Sztucznej Inteligencji (GRAI) przy Ministerstwie Cyfryzacji. Jej głównym celem jest wspieranie międzynarodowej współpracy na rzecz odpowiedzialnego rozwoju AI. Dążymy do budowania pomostów pomiędzy administracją publiczną, biznesem i środowiskiem naukowym, aby tworzyć warunki dla rozwoju sztucznej inteligencji zgodnej z zasadami etyki, cyberbezpieczeństwa oraz praw człowieka.

Pośród naszych kluczowych działań znalazły się:

a) **Konsultacje prezentowanych przez Ministerstwo Cyfryzacji polityk** – bierzemy aktywny udział w opiniowaniu dokumentów strategicznych oraz tworzenie spójnych stanowisk w imieniu grupy.

b) **Katalog ekspercki** to baza specjalistów gotowych wspierać rozwój AI w Polsce, w tym w obszarze współpracy z MŚP i administracją.

c) **Newsletter specjalistyczny** to regularne publikacje prezentujące analizy, komentarze i wydarzenia związane z AI na poziomie lokalnym, europejskim i globalnym.

Sekcja organizuje także cykliczne spotkania, które umożliwiają wymianę doświadczeń i integrację środowiska. Dzięki wspólnemu zaangażowaniu tworzymy przestrzeń dla dyskusji, które kształtują przyszłość AI jako technologii transparentnej, godnej zaufania i dostępnej dla wszystkich.

Członkowie i członkinie naszej sekcji działają na zasadach pro publico bono, przyczyniając się do dynamicznego rozwoju tej kluczowej dziedziny technologicznej na arenie międzynarodowej. Dołącz do nas i współtwórz przyszłość sztucznej inteligencji!

Osoby zainteresowane dołączeniem do GRAI i naszej sekcji, współtworzenia opracowań eksperckich na zasadach pro bono prosimy o przesłanie zgłoszenia na adres mailowy: grai@cyfra.gov.pl.

Dowiedz się więcej na stronie GRAI.