Pokaż, kotku, co masz w środku, czyli co sztuczna inteligencja ma na myśli

Bartłomiej W. Papież, Assoc Prof, PhD, MSc
Big Data Institute, University of Oxford, Wielka Brytania

Skróty: AI (*artificial intelligence*) – sztuczna inteligencja; LLM (*large language models*) – duże modele językowe; XAI (*explainable AI*) – wyjaśnialna sztuczna inteligencja

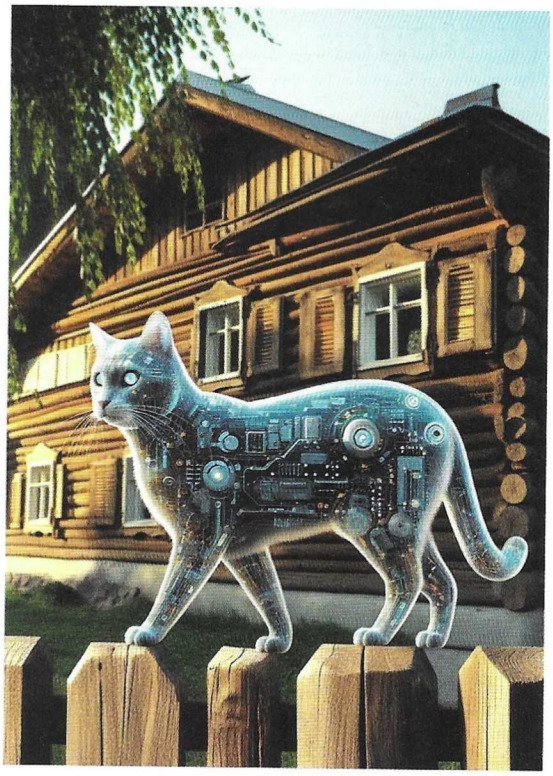
Wszystkie ilustracje zostały wygenerowane za pomocą Microsoft Copilot.

Sukcesy zawodowe sztucznej inteligencji (AI), takie jak zaliczenie egzaminu United States Medical Licensing Examination (USMLE), od jakiegoś czasu nie robią już na nikim wrażenia[1]. Nowsze duże modele językowe (LLM) bardzo szybko dogoniły najlepszych studentów medycyny, a osiągane przez nie wyniki dawałyby im szansę dostania się na najbardziej oblegane specjalizacje – z wyjątkiem interny w Polsce (p. Med. Prakt. Ginekol. Położ., 2025; 2: 85–91). Nietrudno sobie wyobrazić przyszłość, w której LLM zostanie opakowany w humanoidalne „ciało". Taki robot z dyplomem lekarskim będzie mógł pełnić dowolnie długi dyżur, przeprowadzać obchód lekarski i, bez znużenia oraz hektolitrów kawy, zapytać pacjenta o samopoczucie, a także z szerokim uśmiechem na twarzy odpowiadać na pytania, które słyszał już milion razy.

Oczywiście pewna grupa lekarzy mogłaby w tym miejscu przypomnieć mi wprost, że pacjent nie jest od zadawania pytań, tylko od słuchania lekarza i sumiennego wykonywania zaleceń. Niemniej jednak załóżmy, przynajmniej na potrzeby tego artykułu, że pacjent może mieć wyjątkową ochotę dowiedzieć się, dlaczego lekarz stawia określoną diagnozę i czego może się w związku z nią spodziewać. Zatem zadajmy sobie pytanie: czy taka AI potrafi wyjaśnić swoją odpowiedź?

Kiedyś to było łatwiej

We wczesnych modelach AI, które opierały się na systemach eksperckich z wykorzystaniem kodowania pewnych reguł na podstawie specjalistycznej wiedzy i logiki matematycznej albo na uczeniu maszynowym, AI była domyślnie wyjaśnialna. Wynikało to głównie z dwóch powodów. Po pierwsze, dostępne dane były zazwyczaj ograniczone i uzyskiwane przez pracownika służby zdrowia (np. temperatura ciała, ciśnienie tętnicze skurczowe i rozkurczowe), co samo w sobie zapewniało swoisty poziom interpretowalności, ponieważ zmierzona wartość była proporcjonalna do potencjalnych skutków dla pacjenta (np. podwyższona temperatura ciała to stan podgorączkowy, a jeszcze wyższa to już gorączka lub nawet hipertermia). Po drugie, stosowane modele AI były, mówiąc prozaicznie, prymitywne z dzisiejszej perspektywy, a wiele zależności między danymi wejściowymi a odpowiedzią modelu można było pojąć „na chłopski rozum". Jeśli na przykład pacjent źle się czuje, to należy mu zmierzyć temperaturę, a jeśli ta przekracza 38°C, wykonać kolejną czynność itd. Dopóki liczba parametrów, których model potrzebował do podjęcia decyzji, była rozsądna i nie przytłaczała użytkownika, interpretowalność wydawała się idealnym sposobem na budowanie zaufania między użytkownikiem a AI. Problemy pojawiły się jednak w momencie, gdy użytkownik otrzymywał zbyt dużo informacji (*information overload*) o sposobie działania modelu – taki



stan skłaniał go nierzadko do nadmiernego ufania AI, a ostatecznie do podejmowania gorszych decyzji[2]. Co więcej, nadmiar szczegółów prowadził niekiedy do paraliżu decyzyjnego, czyli sytuacji, w której użytkownik zamiast podejmować logiczne decyzje, zmagał się z przytłoczeniem dostępnymi danymi. Innym problemem była skłonność do bezkrytycznego polegania na modelach AI, nawet gdy generowały one wyniki sprzeczne z intuicją lub oczywistymi faktami. Użytkownicy mogli ignorować własne kompetencje na rzecz „autorytetu" AI, co mogło prowadzić do błędnych decyzji klinicznych lub operacyjnych.

Czarna skrzynka

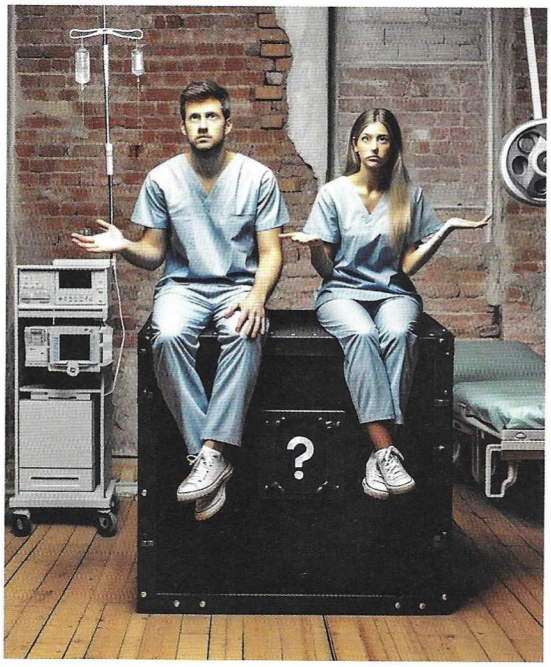
Najpopularniejsza obecnie forma AI opiera się na głębokich sztucznych sieciach neuronowych, o których wspominałem krótko we wprowadzeniu do niniejszego cyklu (Med. Prakt. Ginekol. Położ., 2025; 1: 101–111). W głębokich sieciach neuronowych cyfrowe neurony, ułożone warstwa po warstwie, przekazują sobie informacje o tym, co otrzymały na wejściu, i przetwarzają je tak, aby ostatecznie podjąć decyzję. Wraz z mnożeniem warstw, a tym samym głębokości (warstwowości) sieci neuronowej, poszerzają się możliwości tworzenia bardziej abstrakcyjnych wzorców i rozwiązywania bardziej skomplikowanych zadań. Aby sprostać skomplikowanym problemom medycznym, np. dokonywać rozróżnienia zmian skórnych[3], obecne modele AI są niezwykle duże – zawierają nawet biliony sztucznych neuronów, tak jak najnowsze modele ChatGPT-4[4]. Każda warstwa wykonuje operacje matematyczne, co oznacza, że efekt końcowy jest wynikiem setek tysięcy (lub miliardów) operacji. Taka skomplikowana struktura znacząco utrudnia prześledzenie konkretnego wpływu pojedynczej zmiennej wejściowej na końcową decyzję modelu. Dodatkowo sieci neuronowe (prawie zawsze) stosują nieliniowe funkcje aktywacji neuronów i podejmowania decyzji, co jeszcze bardziej komplikuje proces przetwarzania danych. O ile liniowość można pojąć „na chłopski rozum", jak w przytoczonym powyżej przykładzie zmiany temperatury, to nieliniowość nie jest już tak intuicyjna. Nieliniowość nie podąża za przewidywalnymi wzorcami, do których jesteśmy przyzwyczajeni w codziennym życiu; zmiana jednej informacji na wejściu modelu może prowadzić do nieprzewidywalnych skoków lub spadków w końcowej decyzji. I jeśli tego byłoby jeszcze mało, to modele AI są trenowane i pracują na ogromnych zbiorach danych, w których mogą się znajdować wzorce całkowicie nieoczywiste lub nawet nieznane człowiekowi. To wszystko sprawia, że trudno bezpośrednio powiązać dane wejściowe modelu z jego odpowiedzią. W konsekwencji w kontekście AI opartej na głębokich sieciach neuronowych często używa się terminu czarna skrzynka (*black box*), który odzwierciedla świadomość, że wewnętrzne procesy decyzyjne są trudne albo wręcz niemożliwe do zrozumienia lub wyjaśnienia, nawet przez osobę znającą techniczne aspekty AI. Mimo że można obserwować dane wejściowe i wyjściowe, sposób, w jaki model dochodzi do określonego wyniku, pozostaje nieprzejrzysty. Problem czarnej skrzynki stanowi istotne wyzwanie w dziedzinie AI, co doprowadziło do powstania nowej gałęzi nauki – wyjaśnialnej AI (XAI), czasami nazywanej także AI w szklanym pudełku (*glass box AI*)[5]. Celem XAI jest opracowywanie metod i narzędzi,

które pomagają lepiej zrozumieć działanie modeli AI, identyfikować potencjalne błędy oraz budować większe zaufanie do ich decyzji. Niestety nie każdy model AI da się „zapakować w szklane pudełko", żeby zobaczyć, co jest w środku.

XAI: Czemu, ach czemu?

Wyjaśnialna AI stara się udzielić użytkownikom odpowiedzi na trzy podstawowe pytania: dlaczego model podjął określoną decyzję, jakie czynniki najsilniej wpłynęły na wynik oraz jak można poprawić ten wynik poprzez modyfikację danych wejściowych lub parametrów modelu. W wielu pracach naukowych pojawiły się stwierdzenia, że bez XAI modele AI mogą działać w niezamierzony i niepożądany sposób, prowadząc do stronniczych (*biased*), a tym samym nieprawidłowych decyzji[6]. Na przykład głęboka sieć neuronowa, która miała określić, czy znamię melanocytowe skóry widoczne na zdjęciach dermatoskopowych jest złośliwe, częściej klasyfikowała zmianę jako złośliwą, gdy na obrzeżach obrazu były widoczne znaczniki, zrobione dobrze widocznym fioletowym tuszem[7,8]. Wiązało się to z faktem, że w ten właśnie sposób dermatolodzy oznaczali zmiany, które już wcześniej uznawali za podejrzane. Sieć neuronowa nauczyła się, że obecność fioletowego tuszu może być sygnałem złośliwości zmiany, zamiast się skupić na ocenie faktycznych cech skórnych. Taka błędna zależność (*spurious correlation*), której nauczył się model, wynikała ze specyficznych danych treningowych, w których zdjęcia z pierwszej wizyty (bez znacznika) oraz z kolejnych wizyt (ze znacznikiem) zostały wymieszane i użyte do trenowania sieci neuronowej. W konsekwencji model podejmował decyzje na podstawie nieprawdziwej przesłanki. Przykład ten pokazuje, że modele, których działania nie można zinterpretować, mogą stanowić zagrożenie dla etyki medycznej, ponieważ mogą prowadzić do negatywnych konsekwencji zarówno dla zdrowia konkretnych pacjentów, jak i – w przypadku wdrożenia takiego narzędzia na większą skalę – dla zdrowia publicznego.

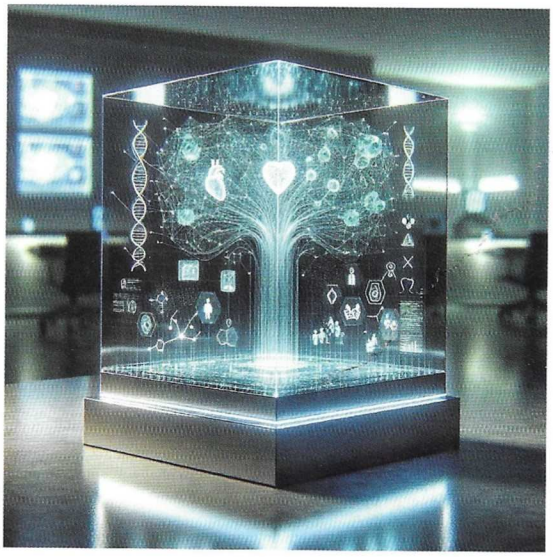
W czasie pandemii COVID-19 wiele zespołów naukowych podjęło heroiczne wysiłki, aby zbudować modele AI zdolne do wykrywania COVID-19 na radiogramach. Co więcej, wielu zespołom udało



się je opracować i opublikować wyniki w prestiżowych czasopismach, przedstawiając imponującą skuteczność na ograniczonych danych pochodzących z pojedynczych szpitali. Problem pojawiał się jednak, gdy modele te były stosowane na obrazach rentgenowskich z innych szpitali, wówczas ich dokładność gwałtownie spadała. Po dokładniejszej analizie okazało się, że przyczyną było tzw. uczenie się na skróty (*shortcut learning*). Zamiast analizować rzeczywiste zmiany płucne, model nauczył się wykrywać różnice w ułożeniu pacjenta lub rodzajach projekcji radiograficznych. Na przykład koncentrował się na tym, czy projekcja jest przednio-tylna (AP) czy tylno-przednia (PA), i na tej podstawie oceniał, czy pacjent ma COVID-19[9]. Trudno w takiej sytuacji zaufać nawet bardzo dokładnemu modelowi, jeśli jego decyzje opierają się na przypadkowych lub nieistotnych cechach zamiast na faktycznych oznakach choroby. Tego rodzaju problemy podkreślają znaczenie wyjaśnialności w AI i potrzebę stosowania narzędzi XAI w praktyce klinicznej.

Powielacz uprzedzeń

Modele AI, których decyzje są trudne do zinterpretowania, pociągają za sobą także dylematy natury



etycznej. Każdy model AI ma tendencję do powielania wzorców, których się nauczył z danych użytych do jego stworzenia. W związku z tym może się także nauczyć powielać uprzedzenia, które znajdowały się w takim zbiorze danych. Uprzedzenia AI (*AI bias*), nazywane również uprzedzeniami algorytmicznymi (*algorithmic bias*), odnoszą się do systemów AI, które generują stronnicze wyniki, odzwierciedlające i utrwalające ludzkie uprzedzenia obecne w społeczeństwie, w tym nierówności społeczne, zarówno historyczne, jak i współczesne.

Uprzedzenia mogą występować na różnych etapach, nie tylko w danych treningowych, ale także w samym algorytmie uczenia sieci neuronowej lub w generowanych przez niego przewidywaniach. Przez ostatnie lata w mediach można było przeczytać o wielu eksperymentach, które wykazały istotne uprzedzenia rasowe, genderowe i intersekcjonalne w sposobie, w jaki najnowocześniejsze LLM oceniały życiorysy kandydatów w procesie rekrutacji pracowników[10]. Naukowcy zastosowali różne imiona kojarzone z białymi i czarnymi mężczyznami oraz kobietami na rzeczywistych CV i odkryli, że LLM preferowały imiona kojarzone z białymi osobami w 85% przypadków, imiona kobiece jedynie w 11% przypadków, a nigdy nie faworyzowały imion czarnoskórych mężczyzn w porównaniu z imionami białych mężczyzn. Niestety problem dotyczy także zastosowań medycznych[11].

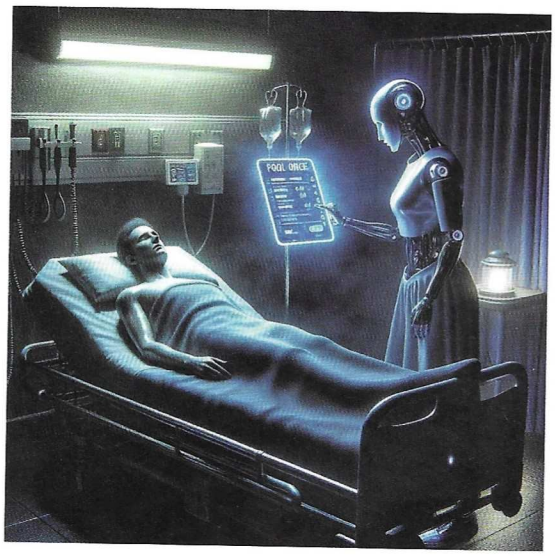
W tym kontekście istnieje wiele możliwości wprowadzenia niepożądanych uprzedzeń, które mogą być (czasami całkiem nieświadomie) zaimplementowane w modelach AI. Najczęstszym jest brak reprezentatywnej próby populacji w zbiorze danych treningowych. Historycznie w badaniach biomedycznych często zaniedbywano różnice biologiczne związane z płcią, a zarówno badania podstawowe, jak i kliniczne koncentrowały się zasadniczo na mężczyznach[12]. Jeśli ktoś wykorzysta do budowy modelu AI takie historyczne dane, nadreprezentujące mężczyzn, może stworzyć model, który będzie skutecznie wykrywał objawy chorób u mężczyzn, ale bagatelizował je u kobiet, wprowadzając tym samym niepożądane uprzedzenie.

Niekiedy uprzedzenie może występować w całej populacji jako konsekwencja ukrytych przyczyn społecznych, historycznych lub instytucjonalnych. Przykład takich uprzedzeń znajdujemy w elektronicznych rejestrach sporządzanych przez lekarzy pierwszego kontaktu – z ich analizy wynika, że częstość wprowadzania kodów diagnostycznych dla nowo rozpoznanych chorób przewlekłych zależy od charakterystyki socjodemograficznej pacjentów, specyfiki przychodni lekarza rodzinnego (np. oceny jakości obsługi w danej placówce) oraz wpływu pandemii COVID-19[13]. Gdyby użyć takich danych do budowy modelu AI bez odpowiedniego przygotowania, mogłoby się okazać, że nie identyfikowałby on pewnych chorób, ponieważ obecnie nie mamy pandemii COVID-19. W takim przypadku mówimy o powielaczu uprzedzeń (*bias amplifier*) – sytuacji, w której model AI nie tylko odzwierciedla istniejące uprzedzenia obecne w danych treningowych, ale również je dodatkowo wzmacnia.

Aby ograniczyć uprzedzenia w modelach AI, konieczne jest użycie różnorodnych i reprezentatywnych zbiorów danych, monitorowanie działania modeli oraz stosowanie metod XAI. Świadomość potencjalnych uprzedzeń i odpowiednie mechanizmy kontroli mogą się istotnie przysłużyć sprawiedliwości i wiarygodności systemów opartych na AI.

Zasadniczy wpływ na poprawę jakości i równości systemów opartych na AI w kontekście omawianych wcześniej problemów z uprzedzeniami w AI w medycynie może mieć regulacja European Health Data Space (EHDS)[14], której celem będzie

umożliwienie bardziej sprawiedliwego i równego dostępu do danych zdrowotnych, co może pomóc rozwiązać problemy związane z ich niereprezentatywnością, która prowadzi do uprzedzeń w analizach opartych na modelach AI. Dzięki tej regulacji dane zdrowotne wykorzystywane do rozwoju i oceny algorytmów AI (np. w medycynie, diagnostyce czy systemach wspomagania decyzji klinicznych) będą bardziej dostępne i zgodne z normami ochrony danych oraz zasadami etycznymi. Umożliwienie szerokiego dostępu do danych zdrowotnych różnych populacji pozwoli natomiast na bardziej reprezentatywne trenowanie modeli AI. Docelowo może to pomóc w wyeliminowaniu uprzedzeń związanych z płcią, rasą czy innymi czynnikami demograficznymi, a tym samym w budowaniu modeli bardziej precyzyjnych w diagnozowaniu i leczeniu osób z różnych grup społecznych.



Pokaż, kotku, co masz w środku, czyli co sztuczna inteligencja ma na myśli

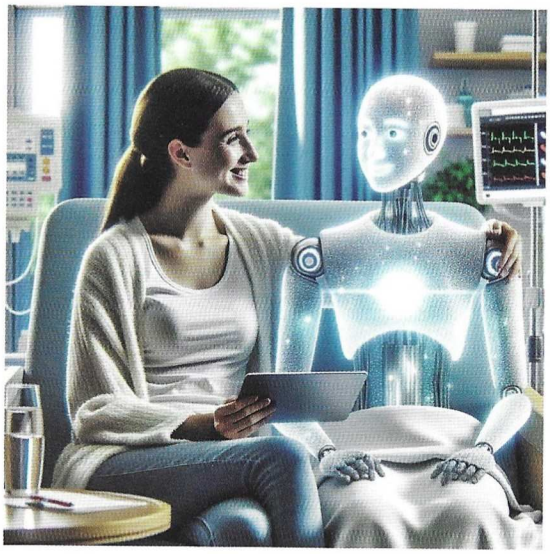
„Oszukasz mnie raz – wstyd dla ciebie"

Z perspektywy medycznej problem wyjaśnialności można również rozważać w kontekście edukacyjnym[15]. Jeśli model potrafiłby wyjaśnić swoją decyzję, korzystająca z niego osoba mogłaby się uczyć, łącząc decyzję z jej uzasadnieniem. Model GPT-3, który prawie zdał egzamin lekarski w USA[16], oceniono także pod względem jakości uzasadnienia decyzji. Zaskoczył on swoim „zachowaniem", ponieważ gdy odpowiadał poprawnie, to zazwyczaj równie poprawnie wyjaśniał swoją odpowiedź. A gdy odpowiedź na pytanie była błędna, jej uzasadnienie analogicznie było raczej wyssane z cyfrowego palca. Wnioski z tej pracy są jednak optymistyczne: w LLM drzemie potencjał wspierania edukacji medycznej, a ostatnie tygodnie pokazują, że ich możliwości w zakresie wnioskowania i uzasadniania decyzji wzrastają – i dlatego coraz częściej nazywa się je dużymi modelami wnioskowania (*large reasoning models*)[17].

Istnieje jednak wiele pułapek, które mogą czyhać na lekarzy korzystających z wyjaśnień generowanych przez LLM i skutkować nadmiernym poleganiem na ich decyzjach[18]. Poprawność wyjaśnienia oraz to, czy jest ono zgodne z poprawnością decyzji podjętej przez model AI, są zasadniczymi czynnikami decydującymi o tym, czy wyjaśnienia są pomocne czy szkodliwe dla użytkowników. Kiedy te elementy są rozbieżne (np. decyzja jest właściwa, ale wyjaśnienie zawiera wiele niepoprawnych informacji), AI zmniejsza wydajność człowieka. Z kolei nieprawidłowa decyzja idąca w parze z poprawnym wyjaśnieniem może skłaniać użytkowników do zgodzenia się z (błędną) decyzją AI. Po prostu wyjaśnienie brzmi zbyt przekonująco... Można by powiedzieć, że jeśli AI oszuka po raz pierwszy, to jej wina, ale jeśli zrobi to po raz drugi, winę można przypisać użytkownikowi, bo nie nauczył się na poprzednim błędzie.

Jak i dlaczego

W wielu systemach prawnych opieki zdrowotnej zrozumienie modeli stało się obecnie wymogiem regulacyjnym przed ich wdrożeniem[19-22]. Brytyjskie Information Commissioner's Office, czyli organ, który dba o przestrzeganie praw związanych z ochroną danych i zapewnia, że dane osobowe są przetwarzane zgodnie z prawem, definiuje XAI w dwóch kategoriach[23]. Obie odnoszą się do różnych podejść w zarządzaniu systemami AI i ich interpretowaniu. Pierwsze podejście skupia się na wyjaśnieniach opartych na procesie, czyli na dostarczeniu informacji dotyczących sposobu, w jaki system AI został zaprojektowany, zbudowany i wdrożony. W tym podejściu należy przedstawić szczegóły dotyczące algorytmów, decyzji podej-



mowanych na każdym etapie tworzenia systemu oraz zarządzania jego funkcjonowaniem. Chodzi o zrozumienie, *jak* system działa, jak były podejmowane decyzje dotyczące jego budowy oraz jakie procesy zastosowano, aby zapewnić odpowiednią jakość i bezpieczeństwo. W drugim podejściu wyjaśnienia oparte są na wyniku, tzn. koncentrują się na tym, co się wydarzyło na skutek działania modelu AI w danej sytuacji. Chodzi o zaprezentowanie uzasadnienia podjętej przez system decyzji w określonej sytuacji. Takie wyjaśnienia szczegółowo tłumaczą, dlaczego system podjął konkretną decyzję lub wygenerował dany wynik w danym przypadku. W skrócie: wyjaśnienia oparte na procesie odpowiadają na pytanie o to, *jak* system AI działa, a wyjaśnienia oparte na wyniku odpowiadają na pytanie, *dlaczego* AI podjęła określoną decyzję.

Kiedy XAI spotyka pacjenta

Rozwój narzędzi XAI stanowi zatem zasadniczy krok w kierunku bardziej przejrzystych i odpowiedzialnych systemów AI, szczególnie w dziedzinach wysokiego ryzyka, takich jak medycyna, chociaż nie brakuje też krytyków tych rozwiązań[24,25]. Zrozumienie systemów AI jest istotne nie tylko dla zbudowania zaufania wśród pracowników medycznych, którzy mieliby korzystać z narzędzi opierających się na AI, ale także dla pacjentów, którzy będą odbiorcami usług świadczonych z użyciem takich narzędzi. Może też dzięki XAI ci drudzy będą się mogli nauczyć więcej o swoim zdrowiu i decyzjach medycznych podejmowanych przez robolekarzy i robolekarki, a tym samym podejmować bardziej proaktywne role w procesie leczenia, czerpać większą satysfakcję z udzielanych im świadczeń i zwiększyć zaangażowanie w leczenie, a nie tylko słuchać lekarza i sumiennie wykonywać niezrozumiałe zalecenia.

PIŚMIENNICTWO

1. Nori H., King N., McKinney S.M.: Capabilities of GPT-4 on medical challenge problems. arXiv 2023; arXiv: 2303.13375 2023
2. Poursabzi-Sangdeh F., Goldstein D.G., Hofman J.M i wsp: Manipulating and measuring model interpretability: in proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery; DOI: 10.1145/3411764.3445315
3. Esteva A., Kuprel B., Novoa R.A. i wsp.: Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017; 542: 115–118 (errata: Nature, 2017; DOI: 10.1038/nature22985)
4. Open AI, Achiam J., Adler S., Agarwal S. i wsp.: Gpt-4 technical report. arXiv preprint, 2023; arXiv: 2303.08774
5. Das A., Rad P.: Opportunities and challenges in explainable artificial intelligence (xai): a survey. arXiv preprint, 2020; arXiv:2006.11371
6. Fergus I., Davis R., van der Schaar M.: Multiple stakeholders drive diverse interpretability requirements for machine learning in healthcare. Nat. Mach. Intell., 2023; 5: 824–829
7. Winkler J.K., Fink C., Toberer F. i wsp.: Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. JAMA Dermatol., 2019; 155: 1135–1141
8. Bevan P., Atapour-Abarghouei A.: Skin deep unlearning: artefact and instrument debiasing in the context of melanoma classification. International Conference on Machine Learning. PMLR, 2022; 162: 1874–189
9. DeGraveA.J., Janizek J.D., Lee S.I.: AI for radiographic COVID-19 detection selects shortcuts over signal. Nat. Mach. Intell., 2021; 3: 610–619
10. Wilson K., Aylin C.: Gender, race, and intersectional bias in resume screening via language model retrieval: proceedings of the AAAI/ACM Conference on AI. Ethics Soc., 2024; 7: 1578–1590
11. Cirillo D., Catuara-Solarz S., Morey C. i wsp.: Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit. Med., 2020; DOI: 10.1038/s41746-020-0288-5
12. McGregor A.J., Hasnain M., Sandberg K. i wsp.: How to study the impact of sex and gender in medical research: a review of resources. Biol. Sex Differ., 2016; DOI: 10.1186/s13293-016-0099-1
13. Beaney T., Clarke J., Salman D. i wsp.: Identifying potential biases in code sequences in primary care electronic healthcare records: a retrospective cohort study of the determinants of code frequency. BMJ Open, 2023; DOI: 10.1136/bmjopen-2023-072884
14. Ryś A.: Europejska przestrzeń danych dotyczących zdrowia. Med. Prakt., 2025; 3: 110–119
15. Amann J., Blasimme A., Vayena E. i wsp.; Precise4Q consortium: Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med. Inform. Decis. Mak., 2020; DOI: 10.1186/s12911-020-01332-6
16. Kung T.H., Cheatham M., Medenilla A. i wsp.: Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit. Health, 2023; DOI: 10.1371/journal.pdig.0000198
17. Mitchell M.: Artificial intelligence learns to reason. Science, 2025; DOI: 10.1126/science.adw5211
18. Kayser M., Menzat B., Emde C. i wsp.: Fool me once? Contrasting textual and visual explanations in a clinical decision-support setting. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, 2024
19. Mourby M., Cathaoir K.Ó., Bjerre Collin B.: Transparency of machine-learning in healthcare: The GDPR & European health law. Comp. Law Secur. Rev., 2021; https://doi.org/10.1016/j.clsr.2021.105611
20. Medicines & Healthcare products Regulatory Agency (UK): Guidance transparency for machine learning-enabled medical devices: guiding principles. 13.06.2024. www.gov.uk/government/publications/machine-learning-medical-devices-trans-

parency-principles/transparency-for-machine-learning-enabled-medical-devices-guiding-principles (dostęp: 22.03.2025)

21. Food and Drug Administration (USA): Transparency for machine learning-enabled medical devices: guiding principles. www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles (dostęp: 22.03.2025)
22. European Union: Document 32024R1689. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689 (dostęp: 22.03.2025)
23. Information Commissioner's Office (UK): Explaining decisions made with AI. Part 1: The basics of explaining AI. What goes into an explanation? https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/what-goes-into-an-explanation/ (dostęp: 22.03.2025)
24. Ghassemi M., Oakden-Rayner L., Beam A.L.: The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit. Health, 2021; 3: e745–e750
25. Reddy S.: Explainability and artificial intelligence in medicine. Lancet Digit. Health, 2022; 4: e214–e215

Aspargin®

Aspargin®, 1 tabletka zawiera: 17 mg jonów magnezu w postaci 250 mg magnezu wodoroasparaginianu (*Magnesii hydroaspartas*) oraz 54 mg jonów potasu w postaci 250 mg potasu wodoroasparaginianu (*Kalii hydroaspartas*). Substancja pomocnicza o znanym działaniu: sacharoza 53 mg. **Wskazania do stosowania:** Uzupełnianie niedoboru magnezu i potasu: przy niemiarowościach i nadpobudliwości serca, zwłaszcza na tle niedoboru magnezu i potasu; profilaktycznie w chorobie niedokrwiennej serca, przy zagrożeniu zawałem; w rekonwalescencji pozawałowej; po przebytych chorobach zakaźnych lub zabiegach chirurgicznych, w przebiegu których doszło do utraty magnezu i potasu; jako lek przeciwdziałający niepożądanym skutkom długotrwałego stosowania glikozydów nasercowych i leków moczopędnych (np. tiazydów, furosemidu). Produkt leczniczy jest przeznaczony dla pacjentów w wieku powyżej 12 lat. **Dawkowanie:** Dorośli i dzieci w wieku powyżej 12 lat: doustnie, zwykle od 2 do 6 tabletek na dobę w 2 lub 3 dawkach podzielonych, po posiłku (co odpowiada od 34 mg do 102 mg Mg++/ dobę (1,4 mmol do 4,2 mmol) oraz od 108 mg do 324 mg K+/ dobę (2,7 mmol do 8,3 mmol). Lek należy przyjmować po posiłku, popijając wodą. Uwaga: przyjmuje się, że całkowite dzienne zapotrzebowanie organizmu na magnez nie przekracza 5 mg/kg masy ciała, a przy zapobieganiu niedoborom potasu podaje się do 780 mg jonu na dobę. **Przeciwwskazania:** Nadwrażliwość na którykolwiek składnik produktu leczniczego (patrz pkt 6.1 ChPL Aspargin), hipermagnezemia, hiperkaliemia, ciężka niewydolność nerek (klirens kreatyniny mniejszy od 30 ml/min), zakażenia dróg moczowych, zaburzenia przewodnictwa przedsionkowo-komorowego (blok), bradykardia, *myasthenia gravis*, znaczne niedociśnienie tętnicze krwi. **Specjalne ostrzeżenia i środki ostrożności dotyczące stosowania:** Nie podawać bez wiedzy lekarza w chorobach nowotworowych. Produktu leczniczego nie należy stosować w ostrym odwodnieniu, rozległym zniszczeniu tkanek (np. oparzenia dużych powierzchni ciała). Nie podawać z innymi preparatami zawierającymi potas. Przy długotrwałym stosowaniu, w razie ciężkiej biegunki oraz u pacjentów leczonych glikozydami nasercowymi należy kontrolować stężenie potasu i magnezu w surowicy, a w razie potrzeby także w erytrocytach. Ostrożnie stosować u pacjentów z chorobą wrzodową żołądka. Pacjenci z rzadkimi dziedzicznymi zaburzeniami związanymi z nietolerancją fruktozy, zespołem złego wchłaniania glukozy i galaktozy lub niedoborem sacharazy-izomaltazy nie powinni przyjmować produktu leczniczego. **Działania niepożądane:** zzaburzenia układu nerwowego: bezsenność (bardzo rzadko); zaburzenia nerek i dróg moczowych: hiperkaliemia (rzadko zwłaszcza u osób z zaburzeniami czynności nerek); zaburzenia żołądka i jelit: bóle brzucha (niezbyt często), nudności (niezbyt często), biegunka (niezbyt często), zaparcia (niezbyt często), wymioty, które zwykle ustępują samoistnie (niezbyt często); zaburzenia skóry i tkanki podskórnej: zaczerwienienie skóry (rzadko); zaburzenia serca: zaburzenia przewodnictwa przedsionkowo-komorowego (rzadko); zaburzenia mięśniowo-szkieletowe i tkanki łącznej: osłabienie mięśniowe (rzadko). **Podmiot odpowiedzialny posiadający pozwolenie na dopuszczenie do obrotu:** Farmaceutyczna Spółdzielnia Pracy FILOFARM®, ul. Pułaskiego 39, 85-619 Bydgoszcz. **Numer pozwolenia MZ na dopuszczenie do obrotu:** R/2107. **OTC – lek wydawany bez recepty.**

OTC/A/MPD/VI/2023