



Krzysztof Boczek, 2022-05-05 11:50

Doktor AI coraz lepszy



Fot. Getty Images/iStockphoto

Przepracowanym lekarzom, pacjentom, którzy nie mogą czekać w kolejce do gabinetu, coraz częściej na świecie pomagają platformy ze sztuczną inteligencją. Miliony osób z tego korzysta. Pandemia drastycznie zwiększyła zapotrzebowanie na systemy. Już wiele z nich jest lepszych niż przeciętni lekarze pierwszego kontaktu.

Już w 2020 – pierwszym roku pandemii – Doctorlink zaliczył trzycyfrowy wzrost liczby użytkowników. Przybyło o 250% wykonujących konsultacje wideo i telefoniczne. Do tego potroiła się liczba przychodni lekarzy rodzinnych, które współpracują z tą platformą. O 200% wzrosła liczba ukończonych wywiadów medycznych, a prośby o przepisanie recepty online i liczba konsultacji wideo spuchły o – odpowiednio – 1470% i 905%! I to wszystko w ciągu 9 miesięcy!

W tym rozwoju pomogła decyzja brytyjskiego NHS. W marcu 2020 – czyli gdy rozpoczynała się pandemia – Doctorlink został wybrany platformą do zapewniania zarówno wstępnej selekcji pacjentów online, jak i konsultacji wideo podczas COVID-19.

Ale inne systemy podobnego typu też czas pandemii zaliczają do najbardziej udanych w swojej historii. Zwielokrotniają liczby swoich użytkowników i wykonanych konsultacji. ADA – system, którego głównie biuro

jest w Niemczech – w czasie już prawie 2 lat pandemii wykonała ok. 15 mln konsultacji. Inny światowy gracz w tej dziedzinie – Babylon – jest obecnie dostępny dla 24 mln ludzi, w samym 2021 r. dokonał 1,3 mln konsultacji. W Stanach współpracuje z nim już potentat ubezpieczeniowy Prudential. Babylon też trafił na giełdę nowojorską – w 2022 r. jego obroty mają wynieść między 900 mln dol. a okrągłym miliardem!

Z czego wynika tak dynamiczny rozwój?

W świecie pandemicznym i zdigitalizowanym takie platformy niosą bardzo wiele korzyści. Dla każdej ze stron. Wbrew pozorom – także dla lekarzy.

Platformy sprawdzające symptomy, i na tej podstawie stawiające diagnozy przedmedyczne, działają od co najmniej 6 lat. Wyposażone są w sztuczną inteligencję i chatboty – wirtualnych asystentów, którzy przeprowadzają wywiady z pacjentami. Czasem bardzo drobiazgowo – najlepsze stawiają pacjentom do 50 pytań. Na podstawie odpowiedzi systemy określają prawdopodobieństwo różnych diagnoz, podpowiadają, jakie poczynić dalsze kroki. Niektóre z nich kierują w razie potrzeby do konkretnych lekarzy/specjalistów, umożliwiają przeprowadzenie szybkiej konsultacji online – wideo czy audio. Inne nawet pozwalają już na wystawienie e-recepty.

Z tego zadowoleni są pacjenci – szybko, tanio lub bezpłatnie, zdalnie, precyzyjnie (który lekarz w gabinecie stawia diagnozę, wyliczając jej prawdopodobieństwo w procentach?). Szwedzi najczęściej konsultowali się w Babylonie przy relatywnie częstych i prostych problemach zdrowotnych: kobiety przy infekcji dróg moczowych, a mężczyźni – zakażeniu górnych dróg oddechowych. Z innych badań wynika, że aż 80% stosujących te systemy uznała, iż są one użyteczne, a z tego aż 90% deklarowała, że je użyje ponownie. Najlepsze platformy/aplikacje osiągają oceny bliskie 5 w sklepach online.

Kiedy Babylon miał już 60 tys. stałych pacjentów w Londynie i Birmingham, zbadano, kim oni są. To generalnie ludzie młodszy (94% poniżej 45 lat), względnie zdrowi (tylko 10% miało 1 z 5 głównych długoterminowych problemów zdrowotnych), zamożniejsi i – co istotne – lepiej wykształceni (ponad 80% ma min. ukończone studia). „Doświadczenia pacjentów były w dużej mierze bardzo pozytywne, nawet dla osób z przewlekłymi chorobami (choć mniej dla osób ze szczególnie złożonymi potrzebami), a pacjenci uznali, że Babylon Health oferuje opiekę lepszej jakości niż w ich poprzedniej praktyce” – opisywał w publikacji JAMA „Journal of American Medical Association”, zanim ta platforma została szerzej upowszechniona w USA. Tytuł zaznaczał, że wskaźnik rotacji pacjentów był wysoki – 1/4 odeszła – ale to było spowodowane zmieniającymi się potrzebami zdrowotnymi i chęcią bezpośredniego spotkania twarzą w twarz z lekarzem.

Korzyści dla systemów opieki zdrowia i lekarzy

Równie istotne jest zadowolenie instytucji narodowych zajmujących się opieką zdrowotną – to one przyczyniają się mocno do rozwoju tych systemów. Platforma Babylon, we współpracy z brytyjskim NHS, bazując na doświadczeniach ponad 100 tys. pacjentów Babylon GP at Hand w Londynie i Birmingham, wyliczyła, że dzięki takiej zdalnej weryfikacji symptomów przez AI, liczba wizyt w gabinetach lekarskich spadła o 28%. Wielu, zwłaszcza przy drobnych problemach zdrowotnych, wystarcza taka konsultacja oraz informacja o dalszych krokach, jakie powinien podjąć. Dlatego też te systemy czasem nazywane bywają „przedmedyczną pomocą”. Ci, którzy stale używają systemu i aplikacji, obniżyli szpitalne koszty NHS o 15-35%. Z tego tytułu platformy zyskują określenia „prewencyjnej medycyny”.

Systemy przynoszą nie tylko wymierne i spore oszczędności, ale też obniżają obciążenie systemu opieki zdrowotnej. Obecnie ok. 4 mld ludzi na Ziemi nie ma dostępu do ochrony zdrowia, lub jest on bardzo słaby. Dla nich taka platforma to szansa na jakąkolwiek konsultację lekarską. Według WHO w 2035 r. deficyt lekarzy i pielęgniarek wyniesie na świecie 12,9 mln. Nawet rozwinięte i bogate państwa zasysają medyków

w innych stron świata, bo rodzimych im już brakuje. NHS wyliczyło, że w latach 2028-29 w Wielkiej Brytanii będzie brakowało 11,5 tys. lekarzy pierwszego kontaktu. A i te prognozy mogą gwałtownie wzrosnąć z powodu pandemii – tylko w czerwcu 2021 r. odnotowano aż o 13% więcej konsultacji lekarskich w porównaniu z analogicznym okresem przed COVID-19. Wraz ze starzeniem się społeczeństw ta dostępność do medyków będzie spadać.

Korzyści dla personelu medycznego, zwłaszcza lekarzy? Jeśli trafia do nich pacjent skierowany przez platformę, już po weryfikacji w sprawdzaczu symptomów, doktor ma nie tylko gotowy wywiad, ale też na stole możliwe opcje interpretacji tych informacji. To nie tylko oszczędność czasu w gabinecie, ale także większa trafność stawianej diagnozy i zaleceń co do terapii.

Ruanda liderem

W Ruandzie system Babylon działa od 2016 r. Można z niego korzystać, dysponując samym tylko smartfonem. Tam ta platforma współpracuje też silnie z rządem – w 2021 firma podpisała umowę na 10 lat, doprowadzając do stworzenia pierwszego w Afryce cyfrowego systemu opieki zdrowotnej. – Ta ważna zmiana przybliży nas o krok do naszej misji oddania wysokiej jakości dostępnej i niedrogiej opieki zdrowotnej w ręce każdego człowieka na Ziemi – mówił wtedy dyrektor zarządzający Babyl – bo tak nazywany jest tam system – Shivon Byamukama.

W Babylu zarejestrowanych jest 2,6 mln osób – 1/3 dorosłej populacji Ruandy. W żadnym kraju na świecie, tak duży procent mieszkańców nie korzysta z tej nowoczesnej, wygodnej i taniej formy kontroli stanu swojego zdrowia. Obecnie Babyl udziela średnio 5 tys. konsultacji dziennie. W 2021 r., po aktualizacji i uproszczeniu procedur rejestracji, między kwietniem a wrześniem, liczba rejestracji kobiet wzrosła o 64%, a mężczyzn o 55%. W pół roku!

Cyfrowe konsultacje kończą się tak jak w gabinetach lekarskich – cyfrową receptą i ew. skierowaniem na dalsze badania. Z Babyła korzystają też tamtejsze pielęgniarki w czasie triażu – system pomaga im zadawać właściwe pytania, podejmować odpowiednie decyzje. Informacje zebrane w czasie takich mikswanych konsultacji otrzymuje klinicysta, który już podejmuje decyzje odnośnie terapii.

Rozwój Babyłonu w Afryce wspiera Fundacja Billa i Melindy Gatesów. Oprócz Ruandy, system ten działa i jest dość popularny w Wielkiej Brytanii, Nowym Jorku, stanie Missouri, rozwija się też w 11 krajach Azji.

Inny z liderów tego rynku – ADA – działa od 2017 r. To wtedy ich aplikacja trafiła do sklepów Google Play. I jeszcze w tym samym roku uzyskała już pierwszy milion użytkowników. Została też uznana za najlepszą apkę medyczną zarówno w sklepie Google Play, jak i w App Store! W tym roku trafiła na niemiecki rynek, w następnym – jej chatboty gadały po hiszpańsku i portugalsku, a system doliczył się 5 mln konsultacji. Potem zostały dodane języki: francuski, suahili, rumuński... Obecnie platforma ma ponad 10 mln użytkowników na świecie, dokonała 25 mln konsultacji w 7 językach.

Za liderami podążyli inni. Gdy w 2018 r. opisywaliśmy te platformy do sprawdzania symptomów, to popularne były na świecie trzy: Babylon, ADA i Your.MD, choć działało na pewno więcej. Teraz już niektórzy badacze, do badań ich skuteczności czy bezpieczeństwa, biorą pod uwagę nawet kilkanaście systemów ze świata. Nawet pojedyncze, duże kliniki tworzą już własne.

AI vs człowiek

Wzrost zainteresowania to też efekt ciągłego rozwoju tych platform. Dzień po dniu dodawane są do ich baz danych nowe informacje, artykuły, wyniki kolejnych konsultacji. Systemy czerpią wiedzę z setek milionów faktów medycznych, z dziesiątek milionów konsultacji już udzielonych. Dla porównania – w I połowie 2018

r. Babylon korzystał z bazy, w której zapisano 530 mln faktów z wiedzy medycznej. Tymczasem przeciętny lekarz pierwszego kontaktu może korzystać z 5 do 10 mln takich informacji – kilkadziesiąt, a nawet kilkaset razy mniej. AI zdobywa wiedzę, jakiej człowiek nie jest w stanie osiągnąć. Czy więc dzięki temu i wypracowanym algorytmom platformy są lepsze niż człowiek w gabinecie?

Odpowiedź jest złożona. Choć można też ją ująć prosto – są kiepskie systemy, które osiągają bardzo niskie wyniki w stawianiu trafnych diagnoz, ale są też systemy, które przewyższają już umiejętności przeciętnego lekarza.

Przykładem niech będą dość rozbudowane badania przeprowadzone przez naukowców z Uniwersytetu Utah w USA w 2021 r. Opublikowali oni w „PloS ONE” wyniki, w których określili efektywność aż 12 systemów diagnozujących. By to osiągnąć, najpierw przetestowali na tychże platformach wystandaryzowany zestaw 50 winiety klinicznych. Dokładność systemu oceniano, mierząc procent przypadków, w których prawidłowa diagnoza była a) wymieniona jako pierwsza, b) wśród pięciu najczęstszych wymienionych diagnoz, c) wymieniona w ogóle.

Zaś bezpieczeństwo porad dotyczących postępowania po diagnozie oceniono, porównując je z krajowymi wytycznymi dla każdej winiety.

Generalnie – średnie wyniki dla wszystkich wyszły słabo. Przeciętna dokładność diagnostyczna systemów, tj. gdy prawidłowa diagnoza była obecna w pięciu najbardziej prawdopodobnych, wyszła na poziomie 51,0%. Rzut monetą. Ale – co trzeba podkreślić – różnice między systemami były ogromne – wyniki wahały się od totalnie nieakceptowalnego poziomu 22,2% skuteczności do aż 84,0% właściwych trafień. Dopiero co rozwijane systemy ciągnęły więc w dół ocenę całej branży – dość niesprawiedliwe działanie naukowców. Z bezpieczeństwem zaleceń było znacznie lepiej – w stanach nagłych w 71,8% systemy podawały trafne zalecenia, w przypadkach niepilnych – aż w 87,3%.

Badacze zauważyli to, co wydaje się dość oczywiste także w gabinecie – im więcej system zadawał pytań pacjentowi, im więcej czasu mu poświęcił, tym większa była prawidłowość postawionej diagnozy. Najbardziej efektywny system zadawał średnio aż 45,8 pytań, co przeciętnie zajmowało 217,4 sekund. Z kolei systemy, które uzyskały najgorsze wyniki zadawały średnio 9,5 pytania, co zajmowało im 37,7 sekundy. Krótka piłka.

„Niektóre systemy osiągają dobry poziom dokładności i bezpieczeństwa postępowania” – podkreślali badacze w podsumowaniu.

Które konkretnie?

Konkurencję bezapelacyjnie wygrała ADA – platforma, która swoją główną siedzibę ma w Niemczech. Aż w ponad 70% przypadków właściwa była pierwsza diagnoza, jaką ta platforma wymieniła. Doctorlink – miał 68% trafień za pierwszym strzałem. Wynik około 60% osiągnęły InforMedica oraz Your.Md. Kiepskie rezultaty – od 22 do 37% poziomu skuteczności – miały: CAIDR, Weg.MD oraz Isabel.

Właściwą diagnozę w pierwszych 5 postawionych miały: ponownie ADA (ponad 80%) i Doctorlink (86%). Z wyjątkiem systemu CAIDR, w tej dyscyplinie pozostałe osiągały skuteczność między 60 a 80% – czyli dość wysoką. Autorzy nie podają powodu, dla którego w tym zestawieniu nie został ujęty Babylon, w dalszej części artykułu już wspomniany.

Najbardziej właściwe dyspozycje po diagnozie wystawiał Doctorlink – wynik ponad 90%.

Ask NHS, Your.MD i ADA miały między 60 a 70% efektywności. Bezpieczeństwo tych zaleceń już zapewniały prawie wszystkie testowane – od 80% do prawie 100% (ADA i Doctorlink).

W innych badaniach publikowanych w „British Medical Journal” (BMJ), MJA i „PloS ONE”, także wygrała niemiecka ADA, osiągając skuteczność diagnoz na poziomie 77%, druga była Isabel (68%), trzeci – Symptomate (60%).

Jak ta efektywność ma się do tego, co robią lekarze?

Tylko w USA co roku doktorzy błędnie diagnozują aż... 12 mln pacjentów. Nic dziwnego więc, że w 2020 r. w opublikowanych badaniach w „Nature Communications” wykazano, że we właściwym diagnozowaniu system Babylon osiągnął wynik wyższy niż 72% lekarzy pierwszego kontaktu. Każdy z medyków, biorących udział w badaniach, ocenił średnio 159 przypadków klinicznych i stawiając diagnozę uzyskał średni wynik 71,40% skuteczności. Starsza sztuczna inteligencja Babylonu uzyskała skuteczność 72,52%, a nowa AI – 77,26%. O 6% lepsza od przeciętnego doktora.

Także te uznane za lepsze systemy wywołują kontrowersje w kwestii efektywności stawianych diagnoz. W 2020 r. BBC opublikowała materiał, w którym kilku specjalistów podważało efektywność Babylonu. Pokazywali przykłady określonych objawów, które platforma niewłaściwie interpretowała, także przy poważnych problemach medycznych. Babylon argumentował, że platforma jest non stop udoskonalana, a jej algorytmy – poprawiane. Krytyczny materiał w BBC nie zaszkodził tak bardzo, skoro – jak donosił „British Medical Journal” – Babylon Health prowadzi rozmowy ze „znaczną liczbą” szpitali w Anglii, starając się rozszerzyć swoje usługi w NHS.

W marcu 2021, czyli po pierwszych dwóch falach COVID-19 i w czasie trzeciej, „British Medical Journal” pisał, iż systemy kontroli symptomów, wyposażone w AI, zawodziły – „konsekwentnie nie zidentyfikowały objawów ciężkiego COVID-19, bakteryjnego zapalenia płuc i sepsy, często doradzając tym przypadkom pozostanie w domu”. Naukowcy zwracają jednak uwagę, że ustalenie, którzy pacjenci z COVID-19 wymagają leczenia jest trudne, ponieważ infekcja może naśladować typowe stany, które rzadko wymagają pomocy medycznej, i ponieważ nie ma klinicznych objawów ani objawów, które rzetelnie przewidują, u kogo rozwinie się ciężka choroba.

Fińska lekcja

Fińscy naukowcy sprawdzili, czy kierujący ich placówkami medycznymi są chętni do stosowania platform sprawdzających symptomy. Ankiety przeprowadzono w 84 placówkach podstawowej opieki zdrowotnej. Z odpowiedzi wynika, iż:

- liderzy tych placówek w używaniu tych platform widzieli więcej wyzwań niż benefitów;
- wśród benefitów wymieniali: mniejszą liczbę telefonów i innych niepotrzebnych kontaktów, co dawało mniejsze obciążenie pracą. „Czas poświęcony na rozmowy telefoniczne można wykorzystać na bezpośrednią pracę z pacjentem” – argumentowali;
- w placówkach, które stosowały te platformy, usprawniono przepływ pracy, przyśpieszono proces segregacji pacjentów na podstawie wywiadów przeprowadzonych przez chatboty AI;
- w efekcie pacjent był szybciej umawiany na wizytę, więc ci mieli lepszy dostęp do opieki, także do dokumentacji medycznej;
- instrukcje, jakie system podpowiadał po wywiadzie, a wskazujące na konieczność samoopieki wzmacniały pacjentów w dbaniu o swoje zdrowie;
- platformy z AI szczególnie przydatne okazywały się, gdy pacjenci szukali opieki z powodu chorób przenoszonych drogą płciową;
- wyzwaniem było to, że algorytmy nie zawsze pytały o wszystko, co trzeba, więc personel medyczny musiał czasem dopytywać telefonicznie;
- problemem były też niewystarczające umiejętności internetowe pacjentów, zwłaszcza starszych osób – nie dawały sobie rady z przejściem przez wizytę online u chatbota.

Generalne wnioski? „Na podstawie odpowiedzi liderzy uznali, że korzystanie z SC (symptom checkers – sprawdzacze symptomów – red.) uczyniło ich organizację bardziej nowoczesną, co było szczególnie ważne w obliczu pandemii COVID-19, ponieważ SC umożliwiły organizacjom lepsze radzenie sobie z wyzwaniami związanymi z tą szybko zmieniającą się sytuacją”, piszą autorzy publikacji w „Health Informatic Journal”, w listopadzie 2021 r.

„Przedstawione wyniki wskazują, że twórcy rozwiązań SC powinni bardziej skoncentrować się na funkcjach, które zmniejszają i/lub upraszczają obciążenie pracowników służby zdrowia, a także na tym, jak rozwiązanie może przynieść korzyści pacjentom i organizacjom opieki zdrowotnej”, piszą autorzy w konkluzjach badań.

Źródło: „Służba Zdrowia”