



Dane to powietrze

Dane stanowią podstawę *evidence-based medicine*, dostęp do danych medycznych jest niezbędny do tworzenia prawie każdej nowej technologii w medycynie – te zwroty powtarzane są nieustannie w trakcie konferencji, znajdują się w dokumentach legislacyjnych i wszyscy zdają się z nimi zgadzać. Jak jednak zapewnić bezpieczny, transparentny dostęp do skompletowanych danych medycznych i jednocześnie zagwarantować poszanowanie praw i podmiotowości pacjenta? Podczas gdy Unia Europejska zastanawia się nad odpowiedzią na to pytanie, Polska być może znalazła rozwiązanie. I to z wykorzystaniem najnowszej technologii *blockchain*.

LIGIA KORNOWSKA

Podstawą działań diagnostyczno-leczniczych w dzisiejszych czasach jest medycyna oparta na faktach (*evidence-based medicine* – EBM), która korzysta w postępowaniu klinicznym z wiarygodnych dowodów naukowych dotyczących skuteczności i bezpieczeństwa terapii [1]. Wiarygodnych dowodów naukowych dostarczają badania, których podstawą są dane. Dotychczas ilość danych analizowanych w jednym badaniu była ograniczona dostępem do nich i możliwościami poznawczymi badacza. W dzisiejszych czasach szacuje się, że mamy 2314 exabajtów danych medycznych (1 exabajt to 1 miliard gigabajtów) [2], a komputery mogą prowadzić analizy na setkach tysięcy czy milionach rekordów w czasie rzeczywistym. Te dwa aspekty wykorzystane razem przenoszą medycynę na zupełnie inny poziom.

Dane medyczne przy wykorzystaniu odpowiedniej mocy obliczeniowej i dobrze zaprogramowanym kodzie informatycznym dają nam algorytmy sztucznej inte-

ligencji (*artificial intelligence* – AI), które są w stanie diagnozować i leczyć pacjenta tak samo bezpiecznie i skutecznie jak lekarz, a w niektórych przypadkach są znacznie lepsze niż zespół wykwalifikowanych lekarzy działających razem. Liczba dostępnych na świecie certyfikowanych algorytmów medycznych gwałtownie wzrasta – od 22 w 2015 r. do ponad 460 w pierwszym kwartale 2020 r. (ryc. 1) [3]. Z powodzeniem wykorzystywane są algorytmy, które samodzielnie stawiają diagnozę, jak w przypadku wykrywania retinopatii cukrzycowej w badaniu dna oka [4].

Duża część algorytmów została przebadana pod kątem większej czułości i swoistości w porównaniu z zespołem lekarzy, co wynika np. z głośnego badania opublikowanego w „Nature” dotyczącego analizy badań mammograficznych [5]. Algorytmy mogą także diagnozować w sposób, który dla lekarza – człowieka jest niedostępny. Sztuczna inteligencja może np. na

podstawie wykresu elektrokardiograficznego określić, czy pacjent ma... niedokrwistość [6].

Dane są paliwem nie tylko dla nowych technologii, takich jak AI. Podstawą naszej wiedzy jest EBM, dlatego dane są niezbędne przy odkrywaniu każdego nowego sposobu diagnostyki i leczenia. Są konieczne do tworzenia szczepionek, leków, wytycznych medycznych i technik diagnostycznych. Wykorzystanie ich na dużą skalę będzie prowadzić do rozwoju medycyny personalizowanej, która opiera się na zastosowaniu kryteriów molekularnych do doboru właściwej strategii terapeutycznej dla właściwego pacjenta we właściwym czasie i/lub na określeniu molekularnych predyspozycji do konkretnej choroby u konkretnej osoby, w celu podjęcia odpowiednich działań prewencyjnych w odpowiednim czasie [7]. Mówimy tu o zmianie paradygmatu leczenia jednym rodzajem terapii całej populacji na leczenie dostosowane do konkretnego pacjenta.

Wreszcie – dygitalizując szeroko rozumiane dane medyczne określonego pacjenta, będziemy w stanie stworzyć jego cyfrowego bliźniaka – *digital twin*, nad którym być może w niedalekiej przyszłości będzie mógł sprawować opiekę asystent medyczny AI i na którym będziemy mogli przeprowadzać cyfrowe testy, aby uzyskać odpowiedź, jaką terapię powinniśmy zastosować u prawdziwego pacjenta.

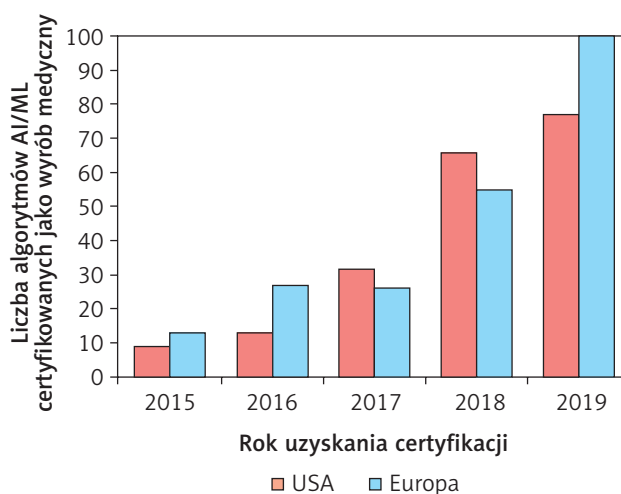
Nie ma wątpliwości, że dziś dane są absolutnie niezbędne do rozwoju każdej dziedziny naszego życia, nie tylko medycyny.

Ubóstwo danych medycznych

Produkujemy ogromne ilości danych medycznych, jednak dostęp do nich jest znacznie ograniczony. Sytuacja wygląda różnie w zależności od rodzaju danych i od kraju. Jeśli uda nam się zabezpieczyć dostęp do pewnej części danych, mamy spory problem z ich interoperacyjnością, czyli między innymi z możliwością skutecznego porównywania ich ze sobą. Nawet jeśli jest dostęp do danych, które potrafią „porozumiewać się” między sobą, cały czas jesteśmy narażeni na „stronniczość” (*bias*) rozwiązań tworzonych na podstawie danych ze względu na ich niereprezentatywność oraz niekompletność.

W ostatnim wydaniu „Lancet Digital Health” użyto określenia „ubóstwo danych medycznych”. Jego autorzy definiują je jako niezdolność do czerpania korzyści z innowacji i nowych osiągnięć przez jednostki, grupy lub populacje ze względu na brak dostatecznych danych je reprezentujących [8]. Aby zrozumieć, jak poważny jest to problem, wystarczy przytoczyć kilka przykładów, i to niekoniecznie z dziedziny nowych technologii.

Coraz więcej jest doniesień, z których wynika, że standardy i wytyczne medyczne, tworzone najczęściej na podstawie danych pochodzących od białych mężczyzn, nie są miarodajne u pewnych grup pacjentów. W „MIT Technology Review” opublikowano wyniki



Rycina 1. Liczba algorytmów AI/ML mających certyfikację FDA lub CE między 2015 r. a 2019 r.

Rok uzyskania certyfikacji CE jest porównywany z rokiem zatwierdzenia wyrobu medycznego przez FDA.

AI/ML – sztuczna inteligencja i uczenie maszynowe, CE – Conformité Européenne.

Opracowano na podstawie: Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. *Lancet Digit Health* 2021; 3(3): e195-e203.

badania dowodzące, że skala służąca do oceny zaawansowania choroby zwyrodnieniowej stawów (*Kellgren and Lawrence Grading Scale*) na podstawie badania RTG, za pomocą której określa się również możliwe dolegliwości bólowe u pacjenta, może nie być wiarygodna dla osób czarnoskórych [9]. Wśród pacjentów o ciemniejszej karnacji problemem jest także mniejsza wykrywalność czerniaka. Mimo większej częstości występowania tego raka u osób rasy białej, czas ich przeżycia był dłuższy niż wśród innych ras [10]. Ten błąd diagnostyczny może być powielany przez algorytmy służące do oceny znamion skórnych, ponieważ są one uczone przeważnie na zdjęciach osób o jasnej skórze.

Z kolei niedostateczna reprezentacja kobiet w badaniach naukowych ma ogromny wpływ na diagnostykę i leczenie tej połowy populacji [11]. Brak odpowiedniej reprezentacji pewnych grup społecznych w badaniach zmusił amerykańską Agencję ds. Żywności i Leków (*Food and Drug Administration – FDA*) do stworzenia wytycznych dotyczących zwiększania różnorodności populacji w badaniach klinicznych [12].

Powyższe przykłady to wierzchołek góry lodowej. Niekompletne i niereprezentatywne dane już teraz są dużym problemem w praktyce klinicznej, ale staną się problemem ogromnym, gdy posłużą do tworzenia nowych technologii informatycznych wspierających diagnostykę i leczenie. Tak zaawansowane osiągnięcia jak AI mogą wynieść medycynę na wyższy poziom, oferując szybszą, lepszą, skuteczniejszą opiekę, ale jednocześnie mogą powiększyć nierówności w dostępie do tej opieki oraz wpłynąć negatywnie na diagnostykę i leczenie pewnych grup pacjentów. To tylko jeden

„Nie ma wątpliwości, że dziś dane są absolutnie niezbędne do rozwoju każdej dziedziny naszego życia, nie tylko medycyny”

z powodów, dlaczego Polski nie stać na nierozwijanie własnych algorytmów i własnych technologii medycznych na podstawie danych polskich pacjentów.

Polska i Europa

Świadoma tego problemu jest Unia Europejska, która podejmuje szereg działań zmierzających do stworzenia europejskiej składnicy danych. W 2020 r. opublikowano dokument „A European strategy for data” [13], który swoim zakresem obejmuje dużo więcej sektorów niż tylko ochrona zdrowia. Ma on być podstawą dla *European Health Data Space*, której celem jest maksymalne wykorzystanie potencjału digitalizacji zdrowia dla zapewnienia wysokiej jakości opieki zdrowotnej i zmniejszenia nierówności. Inicjatywa ta powinna promować dostęp do danych medycznych na potrzeby badań, strategii profilaktycznych i poprawy wyników diagnostyczno-leczniczych. Jeszcze nie wiadomo, jaki ostatecznie będzie miała kształt – ogromny wpływ mają na to poszczególne regulacje państw członkowskich w zakresie dostępu i dzielenia się danymi medycznymi, a także dojrzałość w zakresie interoperacyjności systemów IT.

Polska zdaje się zauważać ten problem i stara się nadążyć za gwałtownym rozwojem nowych technologii. Ostatnio opublikowana „Polityka AI” ma co prawda ogólnosektorowy charakter, ale wskazuje bardzo wyraźnie na potrzebę zajęcia się problemem dostępu do polskich danych medycznych. Centrum e-zdrowia natomiast wraz z zespołem ekspertów pracuje nad strategią e-zdrowia dla Polski na lata 2021–2026, której jednym z rozdziałów jest interoperacyjność, a drugim AI i *big data*. W rozdziale na temat AI zespół poświęcił dużo uwagi tematowi danych medycznych.

Obecnie dostęp do danych anonimowych i osobowych w medycynie jest problematyczny. Teoretycznie dostęp do danych anonimowych powinien być nieograniczony, bo nie jest objęty przepisami RODO, a skuteczna anonimizacja wyklucza przyporządkowanie danych do konkretnego pacjenta. Jednocześnie w obliczu wątpliwego w świetle przepisów RODO art. 26 ust. 4 Ustawy o prawach pacjenta, który reguluje dostęp do dokumentacji w celach naukowych, podmioty lecznicze mają uzasadnione wątpliwości, czy mogą przekazywać anonimowe dane instytucjom innym niż uczelnie wyższe. Problemem jest też korzy-

stanie z anonimowych danych zdrowotnych zgromadzonych w rejestrach centralnych. Rada Ministrów w marcu 2021 r. przyjęła „Program otwierania danych na lata 2021–2027”, który mimo uwag zgłoszonych w trybie konsultacji przez Polską Federację Szpitali i Koalicję AI w Zdrowiu nie uwzględniła w należyty sposób sektora ochrony zdrowia, chociaż podkreślono w nim kluczową wartość danych medycznych. Nadal także nie istnieją wytyczne w zakresie skutecznej anonimizacji danych medycznych, które mogłyby istotnie zwiększyć bezpieczeństwo i zmniejszyć strach przed dzieleniem się anonimowymi danymi. Mamy więc sytuację, w której większość interesariuszy się ze sobą zgadza, a jednocześnie problem wciąż nie jest rozwiązany.

Dostęp do danych osobowych wygląda zgoła inaczej. Dane osobowe podlegają przepisom RODO, a więc ich wykorzystanie powinno być poprzedzone zdobyciem szczegółowej zgody pacjenta określającej, jaki zakres danych, przez jaki czas, w jakim celu i przez kogo będzie przetwarzany. Pacjenci w większości chętnie takich zgód udzielają. Proces ten ma obecnie miejsce głównie w konkretnych szpitalach, często jednostkach badawczych, które prowadzą specyficzne badania nad wybranym zagadnieniem i w tym celu potrzebują wyselekcjonowanych danych określonych pacjentów. W praktyce jest to np. zgoda udzielona chwilę przed pobraniem krwi czy zgoda pobierana przez asystenta profesora, który opiekuje się pacjentami na danym oddziale. Jest to doskonały przykład altruistycznego dzielenia się swoimi danymi na potrzeby nauki, jednak ma on słabe punkty. Po pierwsze grupa badana może nie być reprezentatywna. Jest zawężona tylko do specyficznej populacji (np. osoby mieszkające w dużym mieście, mające wyższe wykształcenie), a pozyskane w ten sposób dane pochodzą tylko z jednego źródła (z podmiotu leczniczego, w którym udzielana jest zgoda) i z dużym prawdopodobieństwem nie są kompletne. Po drugie pobieramy zgodę od pacjenta w sytuacji dla niego niepewnej, kiedy jest zależny od personelu medycznego. Istnieją uzasadnione wątpliwości co do dobrowolności udzielenia takiej zgody.

Idea dawstwa danych

Stosunkowo nową koncepcją jest *data altruism*, co można przetłumaczyć wprost jako altruizm danych. Ta idea opiera się na tezie, że pacjent jako właściciel swoich danych może je świadomie podarować dla rozwoju medycyny. Koncepcja ta jest promowana w różnych dokumentach Komisji Europejskiej, między innymi w „Assessment of the EU Member States’ rules on health data in the light of GDPR”. Znajduje się tam informacja, że do umożliwienia dawstwa danych szykują się Niemcy i Dania, jednakże na razie brakuje gotowych rozwiązań w tym zakresie [14]. W Estonii są elementy dawstwa danych – pacjent może wyrazić zgodę

na wykorzystanie swoich danych dla celów badawczych [15]. Cały czas jednak toczą się dyskusje, w jaki sposób umożliwić pacjentowi bycie altruistą danych i jak zrobić to w całkowicie bezpieczny dla niego sposób.

Polska może być w tym zakresie liderem. Konsorcjum kilku podmiotów – z Polską Federacją Szpitali na czele – tworzy obecnie fundację „Podaruj Dane”, która będzie stanowić zaufaną trzecią stronę dla danych medycznych. Pacjent, tak samo jak daruje swoją krew czy szpik, za pomocą kilku kliknięć na stronie internetowej będzie mógł złożyć oświadczenie woli, którym upoważni fundację „Podaruj Dane” do dostępu do swoich danych medycznych. Dzięki temu fundacja, będąca quasi-repozytorium danych, będzie mogła pobierać je z różnych źródeł, a następnie kompletować i anonimizować. Zaufanie do fundacji będzie oparte na dwóch najważniejszych filarach – zaangażowaniu wielu organizacji zaufania publicznego do nadzorowania działania fundacji oraz zastosowania najbezpieczniejszej technologii *blockchain*.

Fundacja zrealizuje ideę dawstwa danych, opierając się na technologii rozproszonych rejestrów DLT (*blockchain*), poprzez zapisywanie, przechowywanie i zarządzanie elektronicznymi oświadczeniami woli pacjentów oraz zapisywanie i przechowywanie rejestru operacji na danych medycznych, co zapewni pełną transparentność całego procesu. Instytucje zaufania publicznego, czyli organizacje pacjenckie, uniwersytety i instytucje publiczne będą mogły zostać kuratorami wierzchołków *blockchain*, przez co zapewnią bieżący nadzór nad prawidłowością przepływu danych.

W prostych słowach – każda transakcja przeprowadzona na danych pacjenta (tj. złożenie oświadczenia woli przez pacjenta, przepływ danych, ich skompletowanie i anonimizacja) będzie zapisana w rozproszonej sieci, a tym samym będzie nieusuwalna i niezmiennalna. Pacjent w każdym momencie będzie mógł zalogować się na swoje konto i sprawdzić historię procesów zachodzących na jego danych. I będzie miał pewność, że ta historia nie jest zmanipulowana. W przypadku wątpliwości pacjent będzie mógł poprosić o audyt operacji przeprowadzonych na jego danych jedną z organizacji będących kuratorami wierzchołka *blockchain*. Pacjent będzie mógł także w każdej chwili wycofać zgodę lub ją zmodyfikować.

Dzięki powstaniu społeczności dawców danych, a także wykorzystaniu najnowszych technologii umożliwiających bieżącą interakcję między dawcami a odbiorcami danych stworzymy przestrzeń do wynagradzania pacjenta za podarowanie danych. Tak jak w przypadku dawstwa krwi pacjent jest wynagradzany niefinansowo, tak w przypadku dawstwa danych otrzyma korzyść niematerialną – token. Dodatkowo posiadanie tokena da pacjentowi realny wpływ na kierunek rozwoju systemu dawstwa danych. Pacjent dysponujący tokenem będzie się znajdować w centrum systemu dawstwa danych.

Główną funkcją tokena będzie obsługa operacji w systemie umożliwiająca realizację racjonalnych interakcji ekonomicznych między wszystkimi jego uczestnikami. Pozwoli to na utrzymanie i rozwój systemu, jednocześnie wynagradzając rzeczywistego bohatera – dawcę danych. W ramach tworzonego przez nas systemu chcemy przekazywać tokeny za uczestniczenie w systemie dawstwa danych. Będą one wymienne na szereg korzyści dla pacjentów. Zadbamy także o to, by jednostki dostarczające i utrzymujące bazy danych medycznych miały środki na ich odpowiednie zabezpieczenie. W ten sposób przełamamy bariery związane z wymianą danych, jednocześnie tworząc sprawiedliwy obieg korzyści, z uwzględnieniem dobra pacjenta na pierwszym miejscu.

Rozwiązanie to jest innowacją na skalę globalną i jeśli się powiedzie, może wskazać drogę dla całej Unii Europejskiej. Takie podejście rozwiązuje problem niekompletności danych (przez możliwość ich pobierania z różnych źródeł), a także dostępu do danych (przez anonimizację na bazie zgody pacjenta). Oczywiście wciąż otwarta pozostaje kwestia interoperacyjności, a także możliwego wykluczenia pewnych grup społecznych z tego rozwiązania (np. osób starszych, które nie korzystają z komputera). Są to jednak wyzwania, którymi twórcy fundacji zamierzają się zająć na dalszych etapach projektu.

Brakujące dane mają znaczenie. Wykorzystując niepełne dane na masową skalę do tworzenia nowych technologii medycznych, ryzykujemy, że braki te mogą znacznie pogłębiać nierówności w opiece zdrowotnej, powodować dyskryminację i zaburzyć równowagę w społeczeństwie. Dziś jest ostatni moment, aby zapewnić warunki do stabilnego i dobrego rozwoju zdrowia cyfrowego.

Ligia Kornowska,
dyrektorka zarządzająca Polskiej Federacji Szpitali

Piśmiennictwo

1. https://pl.wikipedia.org/wiki/Evidence-based_medicine
2. EMC Digital Universe with Research and Analysis by IDC. The digital universe of opportunities: rich data and the increasing value of the internet of things. April, 2014. <https://www.emc.com/leadership/digital-universe/2014iview/index.htm>
3. [https://doi.org/10.1016/S2589-7500\(20\)30292-2](https://doi.org/10.1016/S2589-7500(20)30292-2)
4. <https://www.eyeworld.org/first-artificial-intelligence-system-approved-fda-detect-diabetic-retinopathy>
5. <https://www.nature.com/articles/s41586-019-1799-6>
6. <https://www.nature.com/articles/s41586-019-1799-6>
7. <http://pkmp.org.pl/strona/czym-jest-medycyna-personalizowana>
8. [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30317-4/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30317-4/fulltext)
9. <https://www.technologyreview.com/2021/01/22/1016577/ai-fairer-healthcare-patient-outcomes/>
10. [https://www.jaad.org/article/S0190-9622\(16\)30380-2/abstract](https://www.jaad.org/article/S0190-9622(16)30380-2/abstract)
11. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1761670/>
12. <https://www.fda.gov/media/127712/download>
13. https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
14. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health_data_en.pdf
15. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5741780/>