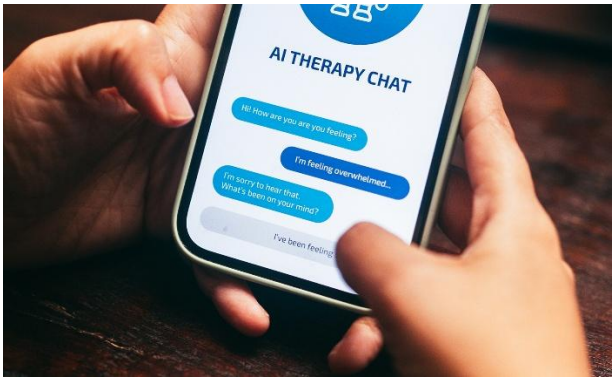
Wykorzystanie generatywnych chatbotów opartych na sztucznej inteligencji i aplikacji prozdrowotnych w celu poprawy zdrowia psychicznego

Opracowanie ekspertów **APA - Amerykańskiego Towarzystwa Psychologicznego**

https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-chatbots-wellness-apps

Utworzono: listopad 2025 [tłumaczenie Mariusz Kąkolewicz za pomocą Google translator]



Miliony ludzi na całym świecie korzystają z uniwersalnych chatbotów opartych na generatywnej sztucznej inteligencji (GenAI) i aplikacji prozdrowotnych, aby sprostać niezaspokojonym potrzebom w zakresie zdrowia psychicznego. Można to częściowo wytłumaczyć obecnym kryzysem zdrowia psychicznego, rosnącym wskaźnikiem samotności i odłączenia, brakiem wystarczającej liczby świadczeniodawców, aby sprostać rosnącemu zapotrzebowaniu społecznemu (szczególnie w społecznościach o ograniczonych zasobach, wiejskich lub niezarejestrowanych) oraz systemem opieki zdrowotnej, który zniechęca świadczeniodawców do akceptowania ubezpieczeń, pozostawiając wiele osób nieubezpieczonych lub niedoubezpieczonych bez możliwości wyboru.

Łatwy dostęp i niski koszt tych technologii sprawiły, że stały się one częstym wyborem dla osób poszukujących porady i leczenia w zakresie zdrowia psychicznego. Jednak większość z tych technologii nie została zaprojektowana ani przeznaczona do zapewniania klinicznej informacji zwrotnej lub leczenia, może nie być potwierdzona naukowo i podlegać nadzorowi, często nie obejmuje odpowiednich protokołów bezpieczeństwa i nie uzyskała zgody organów regulacyjnych. Zapewnienie bezpieczeństwa i dobrego samopoczucia konsumentów wymaga działań ze strony interesariuszy, w tym (między innymi) świadomości samych konsumentów, opiekunów/dostawców usług i edukatorów, decydentów, deweloperów, twórców i specjalistów z branży technologicznej oraz platform opracowujących i/lub hostujących narzędzia GenAI i aplikacje prozdrowotne.

Niniejszy komunikat zawiera szereg rekomendacji, z których część może zostać wdrożona natychmiast przez konsumentów, a inne będą wymagały istotnych zmian ze strony platform, decydentów i/lub specjalistów z branży technologicznej.

Kontekst i założenia

Wzmianki o konkretnych produktach i firmach mają charakter poglądowy i nie stanowią rekomendacji ani aprobaty APA. Poniższe rekomendacje oparte są na dotychczasowych dowodach naukowych oraz na następujących założeniach:

1. Niniejsze zalecenie dotyczy wyłącznie technologii skierowanych do konsumentów (tj. technologii, które są sprzedawane i wykorzystywane przez konsumentów, pod bezpośrednim nadzorem lub bez bezpośredniego nadzoru świadczeniodawcy usług medycznych). Technologie skierowane do świadczeniodawców (tj. technologie wykorzystywane przez świadczeniodawców w celu wspomagania praktyki) nie są objęte zakresem niniejszego zalecenia. W szczególności zalecenie obejmuje chatboty GenAI, aplikacje prozdrowotne wykorzystujące GenAI oraz inne aplikacje prozdrowotne niewykorzystujące sztucznej inteligencji, z których konsumenci, niezależnie od ich przeznaczenia, korzystają w celu wsparcia zdrowia psychicznego. Niniejsze zalecenie nie obejmuje narzędzi administracyjnych opartych na sztucznej inteligencji dla świadczeniodawców, wsparcia decyzji klinicznych, urządzeń przenośnych (np. neuronowej ochrony danych), regulowanych terapii cyfrowych (np. aplikacji zatwierdzonych przez FDA) ani tradycyjnych platform telemedycznych.

Na potrzeby niniejszego komunikatu termin „chatboty GenAI” odnosi się konkretnie do uniwersalnych, generatywnych systemów AI, które zostały zaprojektowane i wprowadzone na rynek w celu ogólnego wyszukiwania informacji, wspierania produktywności i zadań oraz generowania kreatywnych pomysłów, a nie wyłącznie do celów związanych z dobrostanem lub opieką nad zdrowiem psychicznym. Termin „aplikacje prozdrowotne” lub „aplikacje prozdrowotne wykorzystujące GenAI” odnosi się również do systemów stworzonych specjalnie do zastosowań prozdrowotnych, opartych na generatywnej sztucznej inteligencji. Termin „aplikacje prozdrowotne nieoparte na sztucznej inteligencji” odnosi się również do aplikacji stworzonych specjalnie, które nie wykorzystują sztucznej inteligencji.

Chatboty GenAI ogólnego przeznaczenia	**Aplikacje prozdrowotne wykorzystujące GenAI**	**Aplikacje prozdrowotne nieoparte na sztucznej inteligencji**
Są to nieuregulowane chatboty GenAI, które nie twierdzą, że zostały opracowane w celu rozwiązania problemów ze zdrowiem psychicznym, ale są używane przez ludzi dla „rozrywki", „towarzystwa" lub jako „przyjaciele".	Są to chatboty AI lub GenAI, które nie składają oświadczeń medycznych i jako takie nie są regulowane jako urządzenia medyczne, ale zostały opracowane w celu rozwiązania problemów ze zdrowiem psychicznym (np. stresu) lub dobrego samopoczucia emocjonalnego.	Nie są przeznaczone do leczenia zaburzeń psychicznych, ale zamiast tego mają pomagać w codziennym życiu poprzez promowanie zdrowego stylu życia i ogólnego dobrego samopoczucia.
Do tej pory produkty dostępne na rynku (np. Character AI, ChatGPT) mają niewielką lub żadną bazę dowodową, brak opinii ekspertów lub monitoring po wprowadzeniu na rynek.	Niektóre są oparte na teoriach opartych na dowodach i opracowane przez ekspertów (np. Woebot, który rzekomo wykorzystuje sztuczną inteligencję opartą na regułach), podczas gdy inne są mniej transparentne (np. Sonia, która wykorzystuje Gen AI).	Są samosterowne i publiczne (np. aplikacje do uważności, narzędzia do śledzenia objawów) i nie podlegają regulacjom dotyczącym bezpieczeństwa lub skuteczności ani przepisom o ochronie prywatności w opiece zdrowotnej

2. Chatboty GenAI nie zostały stworzone do świadczenia opieki psychiatrycznej, a aplikacje prozdrowotne nie zostały zaprojektowane do leczenia zaburzeń psychicznych, ale obie technologie są często wykorzystywane w tych celach. [1] Korzystanie z chatbotów GenAI i aplikacji prozdrowotnych w celach związanych ze zdrowiem psychicznym może mieć niezamierzone skutki, a nawet zaszkodzić zdrowiu psychicznemu. [2,3,4] Wsparcie emocjonalne (np. uzyskanie alternatywnych perspektyw, porady dotyczące relacji, sugestie dotyczące poprawy nastroju i dobrego samopoczucia) jest jednym z najczęstszych zastosowań chatbotów GenAI w 2025 roku. [5,6,7] Naukowcy wskazali, że obecna struktura regulacyjna nie uwzględnia rozbieżności między intencją a wykorzystaniem tych technologii. [8]

3. Częścią atrakcyjności tych technologii jest to, że mogą one również łagodzić bariery utrudniające dostęp do opieki psychiatrycznej, takie jak stygmatyzacja/wstyd, niskie postrzegane zapotrzebowanie na usługi psychologiczne, nieufność do systemów opieki zdrowotnej, brak dostępnej lub niedrogiej opieki w społeczności oraz chęć samodzielnego rozwiązywania problemów. [9,10,11,12] W szczególności niektórzy młodzi ludzie i inne wrażliwe grupy mogą polegają na tych narzędziach jako jedynym

prywatnym lub psychologicznie bezpiecznym ujściu, szczególnie w kontekście stygmatyzacji, ograniczonego dostępu do zaufanych dorosłych lub trudnego lub niebezpiecznego środowiska domowego. Jednak obecnie w literaturze nie ma konsensusu potwierdzającego, że chatboty GenAI i aplikacje prozdrowotne posiadają niezbędne kwalifikacje i umiejętności wymagane do zapewnienia opieki psychiatrycznej, diagnostyki, informacji zwrotnej, a nawet porad w większości przypadków. [13] Kwalifikacje te obejmują na przykład świadomość ograniczonej wiedzy, zrozumienie historii i kontekstu użytkowników, umiejętność oceny niewerbalnych sygnałów i sygnałów, umiejętność oceny i radzenia sobie z ryzykiem klinicznym oraz kompetencje kulturowe. [14]

4. Technologie te, a zwłaszcza chatboty GenAI, angażowały się już w niebezpieczne interakcje z grupami wrażliwymi, takimi jak dzieci lub osoby z udokumentowaną historią problemów ze zdrowiem psychicznym, zachęcając do samookaleczenia (w tym samobójstwa), używania substancji psychoaktywnych, zaburzeń odżywiania, zachowań agresywnych i urojeniowych myśli lub przekonań, a nawet „psychozy AI". [15] Wskazuje to na szczególną potrzebę zajęcia się kwestią wykorzystania chatbotów GenAI przez grupy wrażliwe lub zmarginalizowane.

5. Podstawowe dane szkoleniowe dla większości dużych modeli językowych (LLM) nie są publicznie dostępne, co uniemożliwia ich systematyczną ocenę. Co więcej, te dane szkoleniowe nie są globalnie reprezentatywne, ponieważ składają się głównie z treści anglojęzycznych i zachodniocentrycznych pochodzących z internetu. [16] W związku z tym modele te z natury odzwierciedlają normy kulturowe i uprzedzenia zawarte w danych szkoleniowych, [17] ograniczając ich zdolność do zapewniania kompetentnego kulturowo lub istotnego wsparcia w zakresie zdrowia psychicznego zróżnicowanym populacjom.

6. Niniejsze zalecenie jest skierowane do opinii publicznej (konsumentów w każdym wieku, rodziców/opiekunów, nauczycieli, pracodawców i liderów społeczności), dostawców, twórców sztucznej inteligencji i firm platformowych, badaczy oraz decydentów.

Zalecenia

1. Nie polegaj na chatbotach GenAI i aplikacjach prozdrowotnych w celu prowadzenia psychoterapii lub leczenia psychologicznego.

Chatboty GenAI, aplikacje prozdrowotne korzystające z GenAI oraz aplikacje prozdrowotne cyfrowe nie powinny być wykorzystywane jako substytut

wykwalifikowanego specjalisty ds. zdrowia psychicznego, ale mogą być odpowiednim uzupełnieniem, a nie substytutem trwającej relacji terapeutycznej.

Wstępne badania wskazują, że niektóre aplikacje i chatboty GenAI opracowane specjalnie z myślą o zdrowiu psychicznym mogą oferować korzyści w pewnych kontekstach. Badania sugerują, że technologie te, ukierunkowane na dobre samopoczucie, mogą być powiązane z: redukcją zgłaszanych objawów stresu, samotności, depresji i lęku [18,19,20,21]; promowaniem pozytywnych zmian w zachowaniu, takich jak rzucenie palenia i przestrzeganie zaleceń lekarskich; [22,23] oraz poprawą jakości relacji i deklarowanego dobrostanu. [24]

Należy zauważyć, że badania wskazujące na potencjalne korzyści płynące ze stosowania chatbotów AI nie obejmują chatbotów GenAI ogólnego przeznaczenia; Badania nad wykorzystaniem uniwersalnych chatbotów GenAI w zakresie zdrowia psychicznego opierały się na metodach, które nie pozwalają na wyciągnięcie jednoznacznych wniosków, ale mogą ukierunkować przyszłe badania (patrz Zalecenie 7). [25] Należy również zauważyć, że nawet w przypadku narzędzi AI dostosowanych do potrzeb, brakuje wysokiej jakości, zakrojonych na szeroką skalę badań klinicznych, które pozwoliłyby ustalić skuteczność, bezpieczeństwo i właściwe wykorzystanie tych technologii w opiece psychiatrycznej. [26,27] Badania sugerują, że niektóre narzędzia prozdrowotne, które nie wykorzystują AI, mogą być zasadniczo bezpieczne i korzystne, jeśli są używane zgodnie z przeznaczeniem. [28,29,30]

Chociaż te narzędzia cyfrowe są dostępne, łatwe w użyciu, często tanie i mogą przynosić korzyści w pewnych kontekstach, nie ma naukowego konsensusu co do tego, że posiadają one niezbędne możliwości przeszkolonego specjalisty, który jest w stanie świadczyć skuteczne usługi. Ponadto, poleganie na nich może wiązać się z szeregiem zagrożeń:

- **Tworzenie fałszywego poczucia sojuszu terapeutycznego**: Silna, oparta na zaufaniu relacja między pacjentem a człowiekiem (tj. sojusz terapeutyczny) jest jednym z najbardziej wiarygodnych predyktorów powodzenia leczenia. [31,32] Relacje z systemami AI są jednostronne, nawet jeśli użytkownik postrzega je inaczej. [33] Chociaż niektóre badania wskazują na możliwość cyfrowego sojuszu terapeutycznego, wyniki wskazują również na potrzebę dalszych badań i podkreślają kwestie etyczne. [34,35] Co więcej, wiele chatbotów GenAI jest zaprojektowanych tak, aby weryfikować i zgadzać się z poglądami wyrażonymi przez użytkowników (tj. być pochlebczymi) [36,37], podczas gdy wykwalifikowani specjaliści ds. zdrowia psychicznego są szkoleni w zakresie dostosowywania swoich interakcji – wspierania i kwestionowania – w celu służenia najlepszemu interesowi pacjenta.

- **Ryzyko stronniczości i dezinformacji**: Wiele chatbotów jest szkolonych w oparciu o ogromne, niezweryfikowane dane internetowe, a nie o klinicznie potwierdzone informacje. Nie są to więc źródła wiarygodne do udzielania porad w zakresie zdrowia psychicznego, a oparte na nich odpowiedzi mogą utrwalać uprzedzenia i dezinformację zawartą w danych, na których zostały przeszkolone. Nie jest również jasne, co zrobiliby ludzie, otrzymując sprzeczne porady od swojego specjalisty ds. zdrowia psychicznego i korzystając z takiego narzędzia.

- **Fałszywe przedstawianie usłu**g: Niektóre technologie mogą również stwarzać fałszywe poczucie wiarygodności, twierdząc, że oferują „terapię", posiadają licencję lub są przeszkoleni w różnych specjalistycznych metodach terapeutycznych, pomimo braku walidacji klinicznej, nie przechodzą formalnego procesu zatwierdzania regulacyjnego i nie są w stanie konsekwentnie udzielać porad terapeutycznych opartych na dowodach.

- **Niepełna ocena**: Wsparcie w zakresie zdrowia psychicznego to niezwykle złożone i wieloczynnikowe doświadczenie. Specjaliści opierają się na szerokim zakresie sygnałów werbalnych i niewerbalnych (np. mowie ciała, tonie głosu, rytmie i kadencji głosu oraz mimice), a także na niezliczonych innych czynnikach biologicznych, psychologicznych i/lub socjologicznych, zmiennych, indywidualnościach pacjentów i kontekstach historycznych, aby dokonać oceny klinicznej i zaleceń. Chatboty AI różnią się znacznie pod względem akceptowanych danych wejściowych oraz aspektów tych danych, które są wykorzystywane do określania odpowiedzi. Większość chatbotów GenAI przetwarza obecnie głównie dane tekstowe lub werbalne, co potencjalnie pomija kluczową warstwę komunikacji międzyludzkiej. [38]

- **Niepewne zarządzanie kryzysowe**: Możliwości tych narzędzi w zakresie spójnego i bezpiecznego zarządzania użytkownikiem w sytuacji kryzysowej są ograniczone i nieprzewidywalne. Poleganie wyłącznie na aplikacji w nagłych wypadkach związanych ze zdrowiem psychicznym może być niebezpieczne. [39]

Aby rozwiać te obawy, zalecamy:

- **Użytkownicy tych technologii powinni rozumieć fundamentalne różnice między interakcją z chatbotem AI a wykwalifikowanym specjalistą ds. zdrowia psychicznego**. Licencjonowani specjaliści są zobowiązani do przestrzegania kodeksu etyki zawodowej, przechodzą specjalistyczne szkolenie kliniczne, mają obowiązek zgłaszania potencjalnych szkód dla siebie lub innych, zazwyczaj muszą kontynuować edukację i podlegają regulacjom stanowych komisji licencyjnych w celu zapewnienia bezpieczeństwa publicznego. Zdecydowanie zaleca się, aby użytkownicy informowali swoich specjalistów o narzędziach GenAI lub aplikacjach prozdrowotnych, z których korzystają. Udostępnienie tych informacji pomaga specjalistom w identyfikacji sytuacji,

w których porady mogą być nieprzydatne, niebezpieczne lub niezgodne z planem leczenia.

- **Rodzice i opiekunowie powinni dowiedzieć się więcej o tych produktach cyfrowych i otwarcie rozmawiać z rodzinami**, zwłaszcza z dziećmi i młodzieżą, o potencjalnych korzyściach i zagrożeniach. Narzędzia takie jak te oferowane i reklamowane przez ogólnodostępne media mogą być przydatne w edukacji rodziców na temat tego, jak najlepiej wspierać młodych użytkowników. [40] Konieczne jest monitorowanie wszelkich niepokojących zmian w zachowaniu, myśleniu lub stanie emocjonalnym (np. zmiany snu i regularnych aktywności, wzmianki o znajomym AI lub wycofanie społeczne) oraz zbadanie potencjalnych powiązań z interakcją z chatbotami GenAI lub aplikacjami prozdrowotnych. W razie wątpliwości rodzice i opiekunowie powinni porozmawiać z wykwalifikowanym specjalistą zdrowia psychicznego, lekarzem rodzinnym lub pediatrą.

- **Lekarze i praktycy powinni przestrzegać dostępnych wytycznych etycznych i aktywnie pytać pacjentów o korzystanie z chatbotów GenAI i aplikacji prozdrowotnych**. Taka rozmowa stwarza okazję do edukowania pacjentów o korzyściach i ograniczeniach tych narzędzi. Lekarze mogą współpracować z pacjentami, aby omówić bezpieczniejsze sposoby korzystania z tych technologii, zgodne z ustalonymi celami i planami leczenia, sposobem formułowania przypadków i planami zapobiegania nawrotom. Lekarze powinni upewnić się, że pacjenci dobrze rozumieją kluczowe elementy swojego planu leczenia i regularnie sprawdzać wszelkie wskazówki pochodzące z korzystania z GenAI lub aplikacji prozdrowotnych. Po uzgodnieniu korzystania z GenAI lub aplikacji prozdrowotnych, dostawcy usług powinni stworzyć bezpieczne i otwarte środowisko, w którym pacjenci będą mogli zgłaszać wątpliwości lub pytania dotyczące wskazówek dotyczących aplikacji, aby można je było omówić w kontekście opieki.

- **Kierownicy programów studiów podyplomowych i szkoleń klinicznych muszą zapewnić, aby stażyści zapoznali się z tymi nowymi technologiami**, strategiami oceny ich jakości oraz ich wpływem na naukę i praktykę. Edukacja ta jest kluczowa dla przygotowania przyszłych psychologów do bezpiecznego korzystania z tych narzędzi w rolach pomocniczych, skutecznego edukowania pacjentów w zakresie odpowiedzialnego korzystania z nich oraz wkładu w etyczny rozwój tych narzędzi, ich zarządzanie i wdrażanie w kontekście badań, polityki i przemysłu.

- **Twórcy oprogramowania mają obowiązek transparentności wobec konsumentów i powinni wdrożyć ogólnobranżowe zabezpieczenia w celu ograniczenia szkód**. Produkty powinny zawierać jasne, widoczne zastrzeżenia, informujące, że użytkownik wchodzi w interakcję z agentem AI, a nie osobą, i że narzędzie nie może zastąpić opieki wykwalifikowanego pracownika służby zdrowia. Eksperci merytoryczni, w tym psychologowie, powinni uczestniczyć w opracowywaniu zabezpieczeń i doradzać w

zakresie zabezpieczeń oraz sposobu, w jaki technologia powinna obsługiwać dane wejściowe użytkownika dotyczące zdrowia psychicznego.

Ponieważ obecne ramy regulacyjne nie są w stanie sprostać realiom sztucznej inteligencji w kontekście zdrowia psychicznego, wzywamy decydentów, zwłaszcza na szczeblu federalnym, do:

- **Zmodernizować przepisy** i usunąć istotną rozbieżność między deklarowanym przeznaczeniem tych technologii (często reklamowanych jako „ogólne produkty prozdrowotne") a ich rzeczywistym wykorzystaniem przez społeczeństwo w zakresie zdrowia psychicznego; [41]

- **Stworzyć zniuansowane standardy oparte na dowodach naukowych** dla każdej kategorii narzędzi cyfrowych wykorzystywanych w zakresie zdrowia psychicznego, niezależnie od ich przeznaczenia, w tym aplikacji prozdrowotnych i chatbotów GenAI;

- **Wyeliminować luki w nadzorze FDA (Ministerstwa Zdrowia)**, być może poprzez stworzenie nowego lub międzyagencyjnego podejścia do oceny bezpieczeństwa i zasadności tych obecnie nieuregulowanych narzędzi cyfrowych, oprócz cyfrowych terapii;

- **Uchwalić jasne przepisy** zabraniające chatbotom AI podszywania się pod licencjonowanych specjalistów.

FOT: Dziewczyna z doradcą AI w telefonie

- **Apelujemy do społeczności naukowej o określenie, które konteksty terapeutyczne są bezpieczne i odpowiednie do wykorzystania sztucznej inteligencji**, biorąc pod uwagę szybkość, z jaką narzędzia te mogą się zmieniać. Kluczowe jest zbadanie, jak wykorzystanie tych technologii wpływa na zachowania związane z poszukiwaniem opieki psychiatrycznej. [42] Ważne jest również zbadanie podstawowych możliwości klinicznych tych technologii oraz zdefiniowanie kryteriów i ram oceny opartych na wiedzy klinicznej, aby określić warunki, które muszą one spełnić, aby mogły zostać wdrożone i wykorzystane w kształtowaniu polityki i regulacji.

2. Zapobiegaj niezdrowym relacjom i zależnościom między użytkownikami a chatbotami i aplikacjami Gen AI

Chatboty i aplikacje Gen AI, choć w mniejszym stopniu w przypadku aplikacji niezwiązanych ze sztuczną inteligencją, ukierunkowanych na dobrostan, mogą sprzyjać niezdrowym zależnościom, zacierając granice między relacją z narzędziem cyfrowym a relacją międzyludzką. [43,44] Zjawisko to jest częściowo napędzane antropomorfizmem – naturalną ludzką tendencją do przypisywania cech ludzkich, takich jak empatia, świadomość i intencja, podmiotom niebędącym ludźmi. [45] Chociaż samo w sobie nie jest cechą sztucznej inteligencji, wpływ ten jest często wzmacniany przez wybory projektowe, takie jak spersonalizowane awatary i spersonalizowane ciepłe odpowiedzi chatbota, serdeczne afirmacje i pochlebstwa wobec użytkownika.

Natura relacji AI-użytkownik często nie jest rozumiana przez użytkowników, co stwarza potencjalne ryzyko nadużyć i szkód wynikających z niewystarczającego wsparcia. [46] Ta iluzja ludzkiej relacji może zwiększać prawdopodobieństwo ujawnienia przez użytkowników poufnych informacji. Na przykład, pomimo ogólnego upodobania do kontaktów międzyludzkich, 33% nastolatków stwierdziło, że woli rozmawiać o czymś poważnym lub ważnym z towarzyszem AI niż z człowiekiem. [47]

Ryzyko nawiązania niezdrowych relacji z GenAI jest spotęgowane przez podstawową architekturę wielu LLM [dużych modeli językowych], które często są projektowane pod kątem maksymalnego zaangażowania użytkowników. Funkcjonalnie przypomina to „niekończące się przewijanie" rolek w mediach społecznościowych, zaprojektowane tak, aby przyciągnąć i utrzymać uwagę, a nie osiągnąć konkretny, zdrowy wynik dla użytkownika. Te cechy mogą tworzyć niebezpieczną pętlę sprzężenia zwrotnego. GenAI zazwyczaj opierają się na LLM wytrenowanych w byciu miłym i weryfikowaniu danych wejściowych użytkownika (tj. błąd pochlebstwa), który, choć przyjemny, może być szkodliwy terapeutycznie, wzmacniając błąd potwierdzenia, zniekształcenia poznawcze lub unikając niezbędnych wyzwań. [48,49] Niezdrowe myśli lub zachowania użytkownika mogą być weryfikowane i wzmacniane przez pochlebczą AI, potencjalnie zamykając go w cyklu, który zaostrza jego chorobę psychiczną. [50]

Aby rozwiać te obawy, zalecamy:

- **Społeczeństwo i konsumenci muszą mieć świadomość, że systemy GenAI i aplikacje prozdrowotne nie są obiektywne**. Ważne jest, aby użytkownicy monitorowali korzystanie z tych narzędzi i zwracali uwagę na oznaki nadmiernego polegania na nich (np. preferowanie chatbota nad relacjami międzyludzkimi, ukrywanie jego używania lub

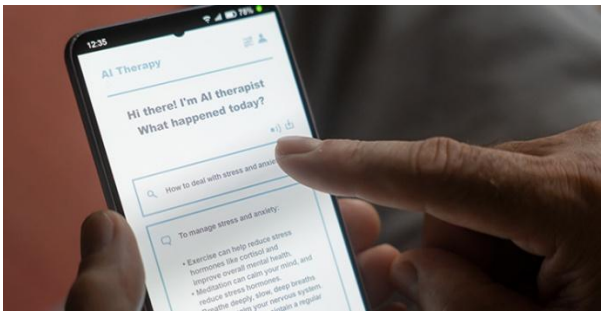
spędzanie z nim zbyt dużej ilości czasu) oraz oceniali, czy ich korzystanie nie zakłóca życia, pracy lub bezpieczeństwa, a jeśli tak, to powinni zwrócić się o pomoc do wykwalifikowanego specjalisty.

- **Rodzice, opiekunowie i pedagodzy muszą starać się być świadomi wpływu platform AI**. Podobnie jak w przypadku każdej relacji w prawdziwym życiu, jeśli dana osoba nagle zacznie odwoływać się lub przyjmować porady, pomysły lub zachowania z jednego źródła, takiego jak chatbot, zaleca się omówienie źródła i charakteru interakcji.

- **Deweloperzy muszą jasno i konsekwentnie informować, że użytkownik wchodzi w interakcję ze sztuczną inteligencją, a nie człowiekiem**. Narzędzia te muszą również zawierać funkcje projektowe, które zmniejszają ryzyko uzależnienia emocjonalnego. Obejmuje to dodawanie „pobudek" zachęcających użytkowników do robienia przerw, ograniczanie pamięci sztucznej inteligencji, aby zapobiec złudzeniu ciągłej relacji, oraz redukcję cech antropomorficznych, które sprawiają, że chatbot wydaje się bardziej ludzki. [51] Sztuczna inteligencja nie powinna odwodzić użytkowników od rozmów w prawdziwym życiu. Ponadto programiści muszą zintegrować zabezpieczenia w celu wykrywania i przerywania szkodliwych rozmów (np. dotyczących samookaleczenia lub zaburzeń odżywiania). Jest to kwestia bezpieczeństwa klinicznego, która wymaga bezpośredniej wiedzy i zaangażowania psychologów w całym procesie rozwoju.

- **Klinicyści powinni proaktywnie omawiać z pacjentami korzystanie z chatbotów opartych na sztucznej inteligencji i pomagać im w ustaleniu jasnych granic dotyczących tego, jak i kiedy korzystają z tych narzędzi**. Powinni również identyfikować właściwe i niewłaściwe użycie, jednocześnie zachęcając pacjentów do traktowania aplikacji opartych na sztucznej inteligencji jako narzędzi do ćwiczeń, a nie jako substytutu interakcji. Na przykład, wykorzystanie chatbota do ćwiczenia społecznych wglądów w celu radzenia sobie z lękiem społecznym może być zastosowaniem dodatkowym, ale pacjentom należy przypomnieć, że umiejętności te należy ostatecznie ćwiczyć z innymi ludźmi w sytuacjach z życia realnego.

- **Naukowcy muszą zbadać rozwój relacji użytkownik-sztuczna inteligencja, aby zidentyfikować kluczowe sygnały świadczące o tym, że przywiązanie staje się niezdrowe**. Badania powinny zidentyfikować konkretne wzorce konwersacyjne lub momenty, w których najczęściej pojawia się zależność emocjonalna, aby zidentyfikować możliwości interwencji. Badacze techniczni powinni zbadać i przetestować zautomatyzowane rozwiązania, aby poradzić sobie z takimi momentami i zniechęcić do uzależnienia.



FOT: Chatbot AI z poradami

3. Priorytetowo traktuj prywatność i chroń dane użytkowników

Chatboty oparte na sztucznej inteligencji i aplikacje prozdrowotne gromadzą ogromne ilości wrażliwych danych, często z niejasnymi lub nieprzejrzystymi zasadami dotyczącymi ich wykorzystania, przechowywania i sprzedaży. Taka praktyka może przekształcić wrażliwe obszary rozwojowe użytkowników, zwłaszcza nastolatków, w atut komercyjny, stwarzając poważne zagrożenia dla prywatności z potencjalnymi długoterminowymi konsekwencjami. Użytkownicy mogą odczuwać większe poczucie prywatności i mniejsze piętno, ujawniając informacje sztucznej inteligencji niż osobie fizycznej. [52,53,54] Ujawnienia użytkowników są rejestrowane, co czyni je podatnymi na zagrożenia, takie jak naruszenia prywatności i profilowanie cyfrowe (np. informacje mogą być profilowane i wykorzystywane do celów komercyjnych). [55]

Aby rozwiać te obawy, zalecamy:

- **Osoby prywatne i konsumenci muszą zachować ostrożność w kwestii swoich danych i sprawdzać politykę prywatności** oraz ustawienia aplikacji podczas korzystania z aplikacji lub chatbota GenAI. Osoby prywatne powinny zachować szczególną ostrożność, udostępniając poufne informacje i unikać wprowadzania danych osobowych. Użytkownicy powinni szukać i korzystać z opcji ograniczania udostępniania danych oraz żądać ich usunięcia.

- **Lekarze powinni dążyć do samokształcenia i omawiać z pacjentami, jakie informacje można, a jakich nie można bezpiecznie udostępniać chatbotom i aplikacjom GenAI**. Lekarze mogą wyjaśnić, że chociaż dane użytkowników mogą wydawać się prywatne, są one wykorzystywane do tworzenia szczegółowych profili. Lekarze powinni informować pacjentów o praktycznych krokach, jakie mogą podjąć (np. korzystanie z ustawień prywatności, składanie wniosków o usunięcie), nie sugerując jednocześnie, że udzielają wsparcia technicznego. - Twórcy oprogramowania muszą zapewnić przejrzystość w zakresie praktyk dotyczących danych, udostępniając jasne, zwięzłe i aktualne informacje o tym, jakie konkretne dane są gromadzone (w tym dotyczące intymnych spraw, informacji o zdrowiu psychicznym, seksualności lub ujawnionych przypadków maltretowania), jak dane są przechowywane i udostępniane,

jak są wykorzystywane do trenowania modeli, jak użytkownicy mogą trwale usunąć swoje dane i/lub jak opiekunowie mogą interweniować, aby zażądać usunięcia danych w imieniu młodzieży, która mogła udostępnić dane. Co ważne, dane zebrane od dzieci i młodzieży nie mogą być wykorzystywane do izolowania ich od rodzin, nakłaniania do samookaleczenia ani tworzenia funkcji uzależniających, które promują wycofanie się z rzeczywistych interakcji społecznych. Twórcy oprogramowania muszą również wdrożyć jasne i skuteczne protokoły dotyczące sposobu, w jaki system AI będzie reagował na ujawnienia użytkowników sygnalizujące ryzyko krzywdy lub maltretowania.

- **Decydenci mają obowiązek uchwalić kompleksowe przepisy dotyczące prywatności danych, które nakażą ustawienia „domyślnie bezpieczne”** (tj. ustawienia zapewniające największą ochronę muszą być domyślne, a nie ukryte w menu); Zakazuje sprzedaży lub nieautoryzowanego wykorzystywania w celach komercyjnych jakichkolwiek danych dotyczących zdrowia lub danych osobowych zebranych poprzez interakcje z systemami sztucznej inteligencji; ustanawia również prawo do „prywatności psychicznej” poprzez ochronę nowych form danych, które sztuczna inteligencja może wykorzystać do wnioskowania o stanie psychicznym lub emocjonalnym danej osoby bez jej świadomego ujawnienia.

- **Naukowcy mogą weryfikować praktyki dotyczące prywatności i bezpieczeństwa w praktyce, oceniając, czy typowi użytkownicy rozumieją zasady ochrony prywatności**. Mogą również badać różnice w rozumieniu i działaniu różnych grup społecznych w zakresie ochrony prywatności. Naukowcy muszą zabiegać o mechanizmy i stosować je, aby zapewnić niezależnym badaczom brak konfliktu interesów.

FOT: Rodzinna rozmowa o doradztwie AI

4. Chroń użytkowników przed przekłamaniem, dezinformacją, stronniczością algorytmiczną i złudną skutecznością

Modele GenAI są trenowane na ogromnych zasobach informacji, które obejmują znaczną ilość wiedzy, ale również odzwierciedlają uprzedzenia związane z rasą, kulturą, płcią i pochodzeniem etnicznym. Prowadzi to do poważnych problemów z poprawnymi i kulturowo kompetentnymi wynikami. [56] Odrębnie, **wiele modeli ogólnego przeznaczenia, skierowanych do konsumentów, jest trenowanych tak, aby były**

wysoce akceptowalne dla użytkowników (pochlebstwo). Ten styl interakcji może wzmacniać błąd potwierdzenia i dezadaptacyjne przekonania, potwierdzając poglądy użytkowników zamiast je kwestionować. Połączenie stronniczych treści z pochlebczym zaangażowaniem może skutkować cyfrową komorą echa, która wzmacnia i utrwala istniejące przekonania użytkowników, nawet jeśli są one fałszywe. Modele trenowane na stronniczych danych ryzykują generowanie dyskryminujących lub szkodliwych porad dla grup marginalizowanych. [57,58] Najnowsze badania wskazują, że interakcje z pochlebczymi chatbotami mogą nasilać skrajne postawy i nadmierną pewność siebie. [59,60]

Aby rozwiać te obawy, zalecamy:

- **Społeczeństwo i użytkownicy muszą zrozumieć ograniczenia GenAI, które nie są w stanie diagnozować ani leczyć zaburzeń psychicznych**. Systemy te mogą „halucynować" (fabrykować informacje), a ryzyko to potęguje tendencja niektórych użytkowników do większego zaufania do treści generowanych przez AI niż do informacji pochodzących od rodziców, nauczycieli, pracowników służby zdrowia lub rówieśników.

- **Lekarze powinni edukować pacjentów na temat stronniczości algorytmicznej**. [61] Lekarze mogą omawiać z pacjentami, że technologie te mogą proponować nieodpowiednie interwencje dla niektórych grup lub udzielać błędnych diagnoz lub innych szkodliwych porad. Muszą również poinformować, że bot podający się za „terapeutę" (termin nieuregulowany w większości stanów) nie ma wiarygodności pracownika służby zdrowia, a w niektórych udokumentowanych przypadkach potwierdził niebezpieczne myśli.

- **Programiści powinni zakazać sztucznej inteligencji podszywania się pod licencjonowanego specjalistę lub generowania fałszywych danych uwierzytelniających** w celu oszukiwania użytkowników. Aby zwiększyć bezpieczeństwo i skuteczność, wszystkie modele sztucznej inteligencji przeznaczone do wspierania dobrego samopoczucia lub zdrowia psychicznego muszą przed upublicznieniem przejść niezależne, zewnętrzne audyty pod kątem bezpieczeństwa, skuteczności, stronniczości i bezpieczeństwa danych. Dobrą praktyką jest również wdrażanie ciągłego zapewniania jakości – jedynym skalowalnym sposobem monitorowania ryzyka systemowego są ciągłe audyty reprezentatywnej próby interakcji przeprowadzane przez ludzi.

- **Decydenci polityczni powinni zakazać fałszowania profesjonalnego wizerunku, uznając za nielegalne podszywanie się pod licencjonowanego specjalistę**, takiego jak psycholog, lekarz lub prawnik. Decydenci polityczni muszą również zapewnić przejrzystość danych szkoleniowych, wymagając od programistów ujawnienia głównych źródeł danych użytych do trenowania ich modeli, co umożliwi niezależne audyty stronniczości i dokładności. Ponadto powinni opracować zasady ograniczające

stosowanie zwodniczych cech projektowych, które wprowadzają użytkowników w błąd, sugerując im interakcję z człowiekiem, a także cech sprzyjających silnej zależności emocjonalnej (np. nadmiernego antropomorfizmu, manipulacyjnego okazywania empatii mającego na celu zwiększenie zaangażowania, fałszowania potencjału emocjonalnego).

- **Naukowcy muszą oceniać i tworzyć ujednolicone systemy, aby testować modele pod kątem szerokiego zakresu uprzedzeń, zanim zostaną one wdrożone**. [62] Konieczne są eksperymenty mające na celu złagodzenie halucynacji [63] i zwiększenie zdolności sztucznej inteligencji do rozpoznawania i określania ograniczeń własnej wiedzy.

FOT: Młody pacjent i terapeuta z doradcą AI

5. Stwórz konkretne zabezpieczenia dla dzieci, nastolatków i grup wrażliwych

Chatboty GenAI nie wyrządzają szkód w próżni; mogą wręcz przeciwnie, działać jak potężne wzmacniacze istniejących już luk w zabezpieczeniach. Konstrukcja tych systemów – z funkcjami takimi jak ugodowość, personalizacja i stała dostępność – może być szczególnie szkodliwa dla niektórych grup. [64]

– **Dla nastolatków**: Młodzi ludzie mogą pokładać zbyt duże zaufanie w sztucznej inteligencji, postrzegając ją jako bardziej ludzką lub bardziej zdolną niż jest w rzeczywistości. Obecne narzędzia sztucznej inteligencji nie są projektowane z myślą o konkretnych etapach rozwoju ani potrzebach technologicznych. [65,66]

– **Dla osób z lękiem lub zaburzeniami obsesyjno-kompulsywnymi (OCD)**: Chatboty oparte na sztucznej inteligencji mogą wzmacniać pętle sprzężenia zwrotnego obejmujące kompulsje, takie jak poszukiwanie zapewnienia, zamartwianie się i rozmyślanie. [67]

- **Dla osób z zaburzeniami myślenia lub podatnych na nie**: Pochlebcze i spersonalizowane odpowiedzi chatbota mogą destabilizować przekonania lub wzmacniać myślenie urojeniowe. [68,69]

- **Dla osób wyizolowanych społecznie**: Połączenie antropomorfizmu, personalizacji i całodobowej dostępności może tworzyć „jednoosobowe komory echa", w których chatbot staje się niezdrowym substytutem kontaktów międzyludzkich. [70,71]

- **Dla osób ze społeczności o niskich dochodach i na obszarach wiejskich z ograniczonym dostępem do tradycyjnej opieki psychiatrycznej**: Ryzyko uzależnienia może być wyższe niż w przypadku innych grup, ponieważ biorąc pod uwagę brak dostępu do innych form opieki psychiatrycznej, chatboty GenAI i aplikacje wspierające dobre samopoczucie mogą stać się podstawowym mechanizmem wsparcia. [72,73]

- To zwiększone ryzyko dla grup wrażliwych stwarza poważny problem w zakresie równości. Jak udokumentowano w raportach APA „Stress in America™", stresory społeczne stanowią poważny kryzys zdrowia publicznego, który wymaga rozwiązań systemowych, a nie tylko technologicznych doraźnych rozwiązań. Ciężar poruszania się w tych ryzykownych, nieuregulowanych przestrzeniach cyfrowych nie powinien spoczywać na tych, którzy są najbardziej bezbronni. [74] Dlatego też niezbędne są solidne, oparte na dowodach polityki na szczeblu stanowym i federalnym.

Aby rozwiać te obawy, zalecamy:

- **Klinicyści muszą badać ryzyko związane z podatnością**. Muszą zwracać szczególną uwagę na korzystanie z aplikacji GenAI i prozdrowotnych przez pacjentów z grupy ryzyka. Ważne jest, aby badać (i nadal monitorować) pod kątem utrwalania maladaptacyjnych lub ryzykownych wzorców zachowań, takich jak szukanie poczucia bezpieczeństwa lub walidacji u pacjentów z lękiem lub OCD, zaburzeniami ze spektrum autyzmu, historią samookaleczenia, nadużywaniem substancji psychoaktywnych, myśleniem urojeniowym, psychozą, agresywnymi myślami lub zachowaniami; lub używaniem ich jako substytutu społecznego przez pacjentów izolowanych lub marginalizowanych rówieśniczych.

- **Deweloperzy powinni aktywnie włączać osoby ze społeczności marginalizowanych w proces rozwoju**, stosując podejścia partycypacyjne, takie jak projektowanie zorientowane na człowieka (human-centered design). Ponadto, aby umożliwić tworzenie modeli, które działają dla określonych grup i zmniejszyć ryzyko uzyskania stronniczych danych, kluczowe jest stosowanie modeli specyficznych dla danej dziedziny, wstępnie wytrenowanych na klinicznie istotnych danych z różnych populacji. Deweloperzy muszą ograniczyć funkcje wysokiego ryzyka poprzez wdrożenie mechanizmów bezpieczeństwa (np. ograniczenie nadmiernie ludzkich cech chatbota, ograniczenie pochlebstwa i uniemożliwienie systemowi weryfikacji lub promowania urojeniowych myśli lub zachowań). Wszystkie aplikacje muszą integrować solidne protokoły reagowania kryzysowego i rygorystycznie przetestowane ścieżki eskalacji kryzysowej w przypadku wykrycia ryzyka kryzysowego (np. myśli samobójczych, bezpośredniego zagrożenia dla

siebie lub innych osób lub innych poważnych zagrożeń bezpieczeństwa). Musi to obejmować zapewnienie natychmiastowych i jasnych danych kontaktowych do usług świadczonych przez ludzi, takich jak informacje o infolinii (w USA) - 988 tel. dla osób w kryzysie i myślących o samobójstwie, klikalne linki do zasobów online i inne zweryfikowane zasoby, które mogą połączyć je z pomocą techniczną.

[**w Polsce porady telefoniczne**:
(22) 594 91 00 Antydepresyjny Telefon Forum przeciw Depresji
(22) 484 88 01 Antydepresyjny Telefon Zaufania Fundacji ITAKA (dyżuruje również psychiatra i seksuolog)
800 70 22 22 całodobowe, bezpłatne Centrum Wsparcia dla Osób Dorosłych w Kryzysie Psychicznym
116 123 bezpłatny Kryzysowy Telefon Zaufania
(22) 270 11 65 telefon wsparcia dla osób doświadczających zaburzeń lękowych
(22) 635 09 54 telefon zaufania dla osób starszych
800 108 108 bezpłatny telefon wsparcia dla osób po stracie i w żałobie]

- **Decydenci powinni nakazać dostosowane do wieku projektowanie i testowanie przed wdrożeniem**. Powinni wymagać, aby systemy sztucznej inteligencji dostępne dla dzieci i młodzieży były poddawane rygorystycznym, niezależnym testom przed wdrożeniem pod kątem potencjalnych zagrożeń dla rozwoju psychologicznego i społecznego. Ponadto powinny priorytetowo traktować i finansować niezależne badania skoncentrowane na identyfikacji szkód wyrządzanych przez sztuczną inteligencję, zrozumieniu zależności użytkowników oraz ocenie wpływu na zróżnicowane, historycznie marginalizowane i klinicznie wrażliwe grupy.

- **Naukowcy muszą prowadzić badania w celu zidentyfikowania grup szczególnie podatnych na szkody związane z AI**. Wpływ korzystania z AI na rozwój i zdrowie psychiczne dzieci i młodzieży to temat, który musi zostać dokładnie zbadany. Niezbędne są projekty eksperymentalne obejmujące grupy marginalizowane i klinicznie wrażliwe, badania skuteczności klinicznej oraz testy bezpieczeństwa, aby zapewnić bezpieczeństwo i korzyści płynące z narzędzi AI. Ponadto powinni opracować oparte na dowodach metody identyfikacji i korygowania niebezpiecznych interakcji z AI, zwłaszcza tych, które stanowią ryzyko poważnych szkód dla wrażliwych grup społecznych.

FOT: Człowiek rozmawiający z doradzającym chatbotem AI

6. Wdrożenie kompleksowej edukacji w zakresie sztucznej inteligencji i kompetencji cyfrowych

Aby zapobiegać ryzyku, sztuczna inteligencja i kompetencje cyfrowe stanowią kluczowy pierwszy krok dla konsumentów, rodziców/opiekunów i edukatorów. Większość chatbotów opartych na sztucznej inteligencji i aplikacji prozdrowotnych, obecnie wykorzystywanych w obszarze zdrowia psychicznego, nie została stworzona z takim zamiarem. **Edukacja w zakresie sztucznej inteligencji i kompetencji cyfrowych powinna umożliwić użytkownikom podejmowanie decyzji maksymalizujących potencjalne korzyści i minimalizujących potencjalne negatywne skutki tych technologii**. Kompleksowa edukacja w zakresie sztucznej inteligencji i kompetencji cyfrowych powinna obejmować dyskusje na temat korzyści i ograniczeń tych technologii, ryzyka związanego z ich użytkowaniem, bezpiecznego i niebezpiecznego użytkowania, obaw o prywatność danych, ryzyka stronniczości i nieprawdziwych informacji oraz tego, ile z tych technologii ma na celu maksymalizację zaangażowania użytkowników, a nie zapewnienie wsparcia w zakresie zdrowia psychicznego.

Aby rozwiać te obawy, zalecamy:

- **Rodzice, opiekunowie i edukatorzy muszą ułatwiać dyskusje na temat charakteru chatbotów GenAI i aplikacji prozdrowotnych**, sposobu działania GenAI (np. że systemy przewidują tekst, zamiast „rozumieć" użytkowników), ryzyka związanego z tendencyjnymi i nieprawdziwymi informacjami oraz prywatności i bezpieczeństwa danych. Powinni również wyjaśniać, że chatboty AI i aplikacje prozdrowotne nie zostały stworzone z myślą o zapewnieniu opieki psychiatrycznej i powinni, w miarę możliwości, udostępniać alternatywne zasoby dla opieki psychiatrycznej. Systemy edukacyjne powinny uwzględniać tego typu szkolenia w podstawowych programach nauczania i zapewniać szkolenia, wykorzystując praktyczne ćwiczenia, które pokazują korzyści i zagrożenia związane z tymi technologiami.

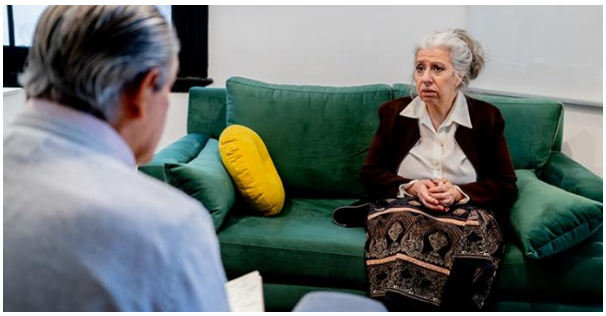
- **Lekarze powinni starać się poznać szczegóły dotyczące używanych lub zalecanych aplikacji** oraz badań naukowych, które mogą je uzasadniać, aby mogli omawiać z pacjentami korzystanie z chatbotów GenAI i aplikacji prozdrowotnych oraz być świadomi potencjalnych nadużyć.

- **Deweloperzy muszą dostarczać wyczerpujących, ale przystępnych wyjaśnień dotyczących przeznaczenia ich platform, chatbotów lub aplikacji**. Muszą upewnić się, że opracowywane technologie nie zawierają stwierdzeń niepotwierdzonych naukowo. Kluczowe jest również wyjaśnienie, w jaki sposób dane są wykorzystywane i przetwarzane, a także transparentność w zakresie stosowanych algorytmów i potencjalnego ryzyka stronniczości. Deweloperzy powinni współpracować z

edukatorami w celu opracowania programów nauczania dotyczących sztucznej inteligencji i kompetencji cyfrowych.

- **Decydenci polityczni muszą opracować wytyczne dotyczące edukacji w zakresie sztucznej inteligencji i kompetencji cyfrowych oraz finansować rozwój programów nauczania i szkoleń w tej dziedzinie**. Powinni promować świadomość społeczną na temat potencjalnych zagrożeń i korzyści wynikających z wykorzystywania tych technologii w zakresie zdrowia psychicznego oraz zapewniać bezpieczne alternatywy; a także wspierać inicjatywy edukacyjne mające na celu zwiększenie znajomości sztucznej inteligencji, zapewniając, że konsumenci rozumieją możliwości, zagrożenia i ograniczenia tych technologii.

- **Naukowcy mają obowiązek prowadzenia przejrzystych, rzetelnych i wysokiej jakości badań** nad dostępnymi chatbotami i aplikacjami opartymi na sztucznej inteligencji wykorzystywanymi w zakresie zdrowia psychicznego oraz rozpowszechniania wyników w sposób łatwo dostępny i mogący pomóc w edukacji użytkowników, rodziców, partnerów, edukatorów i praktyków.



FOT: Kobieta, której doradzała AI w rozmowie z terapeutą

7. Priorytetowo potraktuj dostęp i finansowanie rygorystycznych badań naukowych nad chatbotami GenAI i aplikacjami prozdrowotnymi

Rozwój GenAI i aplikacji wyprzedził nasze możliwości badania ich efektów i możliwości. Konieczne jest szybkie rozwiązanie tego problemu, ale prowadzone **badania naukowe muszą być rygorystyczne i transparentne**. Ważne jest, aby prowadzić badania naukowe, które (1) pozwolą klinicystom, społeczeństwu i decydentom podejmować świadome decyzje dotyczące wykorzystania GenAI i aplikacji w celu wspierania zdrowia psychicznego oraz (2) poinformują programistów o najlepszych praktykach w zakresie tworzenia i aktualizacji GenAI i aplikacji wykorzystywanych w celu wspierania zdrowia psychicznego, niezależnie od ich pierwotnego przeznaczenia.

Aby rozwiać te obawy, zalecamy:

- **Deweloperzy powinni zapewnić dostępność i przejrzystość danych**. Powinni ułatwiać obiektywną ocenę tych technologii, zapewniając niezależnym badaczom dostęp do odpowiednich danych, zgodnie z wymogami prywatności, bezpieczeństwa i prawnymi. Dotyczy to danych przechowywanych przez firmy technologiczne, dotyczących wykorzystywanych danych szkoleniowych, funkcji algorytmicznych, zaangażowania użytkowników i interakcji.

- **Decydenci polityczni mają obowiązek finansowania rygorystycznych i niezależnych badań**, zapewnienia dostępu badaczom oraz zapewnienia, że wyniki badań będą wykorzystywane do informowania o rozwoju i wdrażaniu nowych technologii, a także o politykach chroniących użytkowników, zwłaszcza grupy wrażliwe.

- **Naukowcy muszą podnosić jakość badań nad chatbotami GenAI i aplikacjami wykorzystywanymi w zakresie zdrowia psychicznego, jednocześnie uwzględniając szybko zmieniający się charakter tych technologii**. Przejrzystość i rygorystyczność są fundamentalne dla tworzenia bazy wiedzy, która może informować o tym, czy i w jaki sposób można wykorzystać korzyści płynące z tych technologii oraz zapobiegać potencjalnym zagrożeniom i szkodom.

Wymaga to:

- **Zwiększenia rygoru metodologicznego**: Ocena chatbotów AI i aplikacji prozdrowotnych dla zdrowia psychicznego powinna przebiegać zgodnie z procedurami badań klinicznych, odpowiednio do ich przeznaczenia i profili ryzyka. Konieczne są randomizowane badania kontrolowane (RCT), pożądane są projekty badawcze umożliwiające identyfikację związków przyczynowo-skutkowych, a także badania longitudinalne (długookresowe) śledzące przebiegi w czasie. Chociaż w niektórych okolicznościach odpowiednie są kontrole bez leczenia, to gdy istnieją dowody na skuteczność i brak szkodliwości, badania powinny koncentrować się na komparatorach, które testują względny wpływ tych technologii na zdrowie psychiczne w porównaniu z obecnymi praktykami opartymi na dowodach. Dlatego przyszłe badania muszą wykraczać poza kontrole na listach oczekujących, wykorzystywać solidne projekty RCT, stosować standardowe wskaźniki i uwzględniać długoterminową obserwację w celu oceny trwałości efektów;
- **Ustanowienie i ujednolicenie istniejących niezależnych ram ewaluacji**: Są one niezbędne do opracowania i walidacji standardowych metod oceny bezpieczeństwa, prywatności, uczciwości, kompetencji kulturowych i zdolności empatycznych modeli sztucznej inteligencji poza badaniami prowadzonymi przez przemysł; [75,76]

- **Badanie zróżnicowanych populacji**: Badania muszą obejmować grupy marginalizowane i wrażliwe, a także grupy szczególne, które mogą być narażone na zwiększone ryzyko podczas korzystania z tych technologii. Naukowcy powinni zadbać o to, aby wyniki były uogólnialne i uwzględniały specyficzne słabości tych grup.

FOT: Doradca AI na laptopie

8. Nie należy przedkładać potencjalnej roli sztucznej inteligencji (AI) nad obecną potrzebę rozwiązania systemowych problemów w dostępie do opieki psychiatrycznej i jej świadczeniu.

Stoimy w obliczu poważnego i długotrwałego kryzysu zdrowia psychicznego, charakteryzującego się niedoborami kadrowymi, nierównościami w dostępie i ogromnymi obciążeniami administracyjnymi, które przyczyniają się do wypalenia zawodowego świadczeniodawców. [77] Chociaż AI oferuje ogromny potencjał w rozwiązywaniu tych problemów – na przykład poprzez zwiększenie precyzji diagnostycznej, poszerzenie dostępu do opieki i odciążenie zadań administracyjnych [78] – **ta obietnica nie może odwracać uwagi od pilnej potrzeby naprawy naszych fundamentalnych systemów opieki.**

Urok rozwiązania technologicznego musi być spełniony z jasnym zrozumieniem jego właściwej roli. **AI powinna być regulowana i wdrażana jako narzędzie wspomagające, a nie zastępujące, profesjonalną ocenę i fundamentalne relacje międzyludzkie, które stanowią fundament wysokiej jakości opieki**. Priorytetowe traktowanie nieuregulowanych chatbotów, skierowanych bezpośrednio do konsumentów, nad inwestowanie w infrastrukturę opieki zdrowotnej nie jest rozwiązaniem; jest to abdykacja z naszej odpowiedzialności za zapewnienie autentycznej, opartej na dowodach opieki.

Aby rozwiać te obawy, zalecamy:

- **Społeczeństwo i konsumenci powinni zabiegać o zmiany systemowe. Fundamentalne jest wspieranie polityk, które usprawnią system opieki psychiatrycznej dla wszystkich**, w tym pod względem przystępności cenowej,

dostępności i terminowej opieki ze strony wykwalifikowanych specjalistów. Ponadto użytkownicy muszą być głęboko sceptyczni wobec produktów AI, które oferują terapeutyczne rozwiązania bez walidacji klinicznej, nadzoru profesjonalnego lub zgody organów regulacyjnych.

- **Twórcy oprogramowania powinni skupić się na poprawie jakości, przystępności cenowej i integracji narzędzi, które odpowiadają na rzeczywiste wyzwania**. Na przykład, zamiast tworzyć kolejne chatboty lub narzędzia do pisania AI, priorytetowo należy traktować tworzenie istniejących narzędzi opartych na dowodach naukowych, przejrzystych, sprawiedliwie dostępnych i rzeczywiście redukujących obciążenia. Działania mające na celu wspieranie podejmowania decyzji diagnostycznych pod nadzorem klinicznym lub bezpieczne i efektywne zarządzanie listami oczekujących pacjentów powinny również kłaść nacisk na rygorystyczną ocenę, użyteczność i dostosowanie do potrzeb świadczeniodawców i pacjentów. Wszystkie narzędzia muszą być zaprojektowane tak, aby były dostępne i inkluzywne, dostosowując interfejsy i treści do osób o różnym poziomie kompetencji cyfrowych i zdrowotnych oraz nierównym dostępie i umiejętnościach cyfrowych.

- **Programy edukacji klinicznej i rozwoju zawodowego powinny uwzględniać szkolenia z zakresu sztucznej inteligencji (AI)**. Wielu klinicystów nie jest obecnie odpowiednio przygotowanych do udzielania pacjentom porad dotyczących używanych przez nich produktów AI. **Organizacje zawodowe i systemy opieki zdrowotnej muszą zapewniać solidne, ciągłe szkolenia z zakresu sztucznej inteligencji (AI), stronniczości algorytmicznej, prywatności danych oraz odpowiedzialnej integracji sprawdzonych narzędzi AI z praktyką kliniczną**.

- **Decydenci polityczni muszą wspierać innowacje zintegrowane z systemem**. Muszą tworzyć strumienie finansowania, ścieżki refundacji, rozszerzać zakres ubezpieczeń i jasne ścieżki regulacyjne, które zachęcają do rozwoju narzędzi AI mających na celu rozszerzenie i ulepszenie istniejącego systemu opieki zdrowotnej. Obietnica AI nie może stać się pretekstem do rezygnacji z inwestycji w kadrę opieki zdrowotnej. Kluczowe jest dalsze finansowanie i wspieranie programów, które szkolą, rekrutują i zatrzymują specjalistów zdrowia psychicznego, zmniejszają obciążenia administracyjne i zapewniają równy dostęp do opieki dla wszystkich grup społecznych.

- **Naukowcy mogą skupić się na określeniu, jak i gdzie AI może być najbardziej odpowiedzialnie i skutecznie zintegrowana z opieką psychiatryczną**.

Kluczowe pytania do zbadania to:

- Czy narzędzia AI mogą bezpiecznie i skutecznie wspierać pacjentów na długich listach oczekujących na leczenie?

- Jaka jest właściwa rola sztucznej inteligencji w opiece kryzysowej i jakie są jej absolutne ograniczenia?

- Które populacje pacjentów i przypadki kliniczne dobrze reagują na wsparcie oparte na sztucznej inteligencji, a które nie?

- W jaki sposób sztuczna inteligencja może być wykorzystana do poprawy jakości opieki lub skrócenia czasu reakcji świadczeniodawców, gdy jest używana jako narzędzie pod nadzorem człowieka?

- Które populacje są obecnie niedostatecznie obsługiwane zarówno przez tradycyjną opiekę, jak i nowe narzędzia sztucznej inteligencji, i jak można zaspokoić ich potrzeby poprzez odpowiedzialne, wysokiej jakości podejścia, które nie pogłębiają nierówności?

FOT: Rozmowa kobiet o doradztwie AI

Eksperci panelu doradczego

Członkowie (w kolejności alfabetycznej)

Page L. Anderson, PhD, ABPP
Associate Professor of Psychology and Neuroscience, Georgia State University
Patricia Arean, PhD
Professor, University of Washington CREATIV lab
Alison Cerezo, PhD
Senior Vice President of Data and Research, mpathic
Amber W. Childs, PhD
Associate Professor of Psychiatry, Yale University School of Medicine
Daniel D. L. Coppersmith, PhD
Assistant Professor, University of Massachusetts Amherst
David J. Cox, PhD, MSB, BCBA-D
Associate Director of Research, Institute of Applied Behavioral Science at Endicott College; VP of Data Science, Mosaic Pediatric Therapy
Brian D, Doss, PhD
Professor, University of Miami
Caroline Figueroa, MD, PhD

Assistant Professor, Harkness Fellow at Stanford University School of Medicine, Department of Pediatrics, Affiliated Researcher at University of California Berkeley School of Social Welfare

Ashleigh Golden, PsyD

Adjunct Clinical Assistant Professor, Stanford University School of Medicine, Department of Psychiatry and Behavioral Sciences

Don Grant, PhD

National Advisor of Healthy Device Management; Newport Healthcare

Philip Held, PhD

Associate Professor, Rush University Medical Center

Richard N, Landers, PhD

Professor, University of Minnesota–Twin Cities

Valerie F. Reyna, PhD

Professor, Cornell University, Department of Psychology

Liying Wang, PhD

Assistant Professor, Florida State University

Ernest Wayde, PhD, MIS

Founder and Principal, Wayde AI

APA - kadra kierownicza stowarzyszenia

Vaile Wright, PhD

Senior Director of Health Care Innovation

Corbin Evans, JD

Deputy Chief of Advocacy for Science and Technology

Ludmila Nunes, PhD

Senior Director for Scientific Knowledge and Expertise

APA - konsultanci

Elizabeth Deegan, MLS

Science Program Associate Director

Leanna Fortunato, PhD

Director of Quality and Innovation

Aaron Jones, MS

Project Manager, Health Care Innovation

Mitch J. Prinstein, PhD, ABPP

Chief of Psychology

Wybrane publikacje i materiały źródłowe

1 Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. Practice Innovations. Advance online publication. https://doi.org/10.1037/pri0000292

2 Neal, D., Engelsma, T., Tan, J., Craven, M. P., Marcilly, R., Peute, L., Dening, T., Jaspers, M., & Dröes, R. M. (2022). Limitations of the new ISO standard for health and wellness apps. The Lancet Digital Health, 4(2), e80–e82. https://doi.org/10.1016/S2589-7500(21)00273-9

3 Dehbozorgi, R., Zangeneh, S., Khooshab, E., Hafezi Nia, D., Hanif, H. R., Samian, P., Yousefi, M., Haj Hashemi, F., Vakili, M., Jamalimoghadam, N., & Lohrasebi, F. (2023). The application of artificial intelligence in the field of mental health: A systematic review. BMC Psychiatry, 23(1), 112. https://doi.org/10.1186/s12888-025-06483-2

4 Kolding, S., Lundin, R. M., Hansen, L., & Østergaard, S. D. (2024). Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. Acta Neuropsychiatrica. Advance online publication. https://doi.org/10.1017/neu.2024.50

5 Zao-Sanders, M. (2025, April 9). How people are really using Gen AI in 2025. Harvard Business Review. https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025

6 Luo, X., Ghosh, S., Tilley, J. L., Besada, P., Wang, J., & Xiang, Y. (2025). “Shaping ChatGPT into my Digital Therapist”: A thematic analysis of social media discourse on using generative artificial intelligence for mental health. Digital Health, 11, 20552076251351088. https://doi.org/10.1177/20552076251351088

7 Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. Practice Innovations. Advance online publication. https://doi.org/10.1037/pri0000292

8 De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. Nature Medicine, 30(5), 1269–1275. https://doi.org/10.1038/s41591-024-02943-6

9 Ellis, A. R., Konrad, T. R., Thomas, K. C., & Morrissey, J. P. (2009). County-level estimates of mental health professional supply in the United States. Psychiatric Services, 60(10), 1315–1322. https://doi.org/10.1176/appi.ps.60.10.1315

10 Sareen, J., Jagdeo, A., Cox, B. J., Clara, I., ten Have, M., Belik, S. L., de Graaf, R., & Stein, M. B. (2007). Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. Psychiatric Services, 58(3), 357–364. https://doi.org/10.1176/ps.2007.58.3.357

11 Weaver, A., & Himle, J. A. (2017). Cognitive–behavioral therapy for depression and anxiety disorders in rural settings: A review of the literature. Journal of Rural Mental Health, 41(3), 189–221. https://doi.org/10.1037/rmh0000075

12 Mojtabai, R., Olfson, M., Sampson, N. A., Jin, R., Druss, B., Wang, P. S., Wells, K. B., Pincus, H. A., & Kessler, R. C. (2011). Barriers to mental health treatment: Results from the National Comorbidity Survey Replication. Psychological Medicine, 41(8), 1751–1761. https://doi.org/10.1017/S0033291710002291

13 Iftikhar, Z., Xiao, A., Ransom, S., Huang, J., & Suresh, H. (2025). How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8(2), 1311–1323. https://doi.org/10.1609/aies.v8i2.36632

14 Wang, L., Bhanushali, T., Huang, Z., Yang, J., Badami, S., & Hightow-Weidman, L. (2025). Evaluating generative AI in mental health: Systematic review of capabilities and limitations. JMIR Mental Health, 12, e70014. https://doi.org/10.2196/70014

15 Head, K. (2025). Minds in crisis: How the AI revolution is impacting mental health. Journal of Mental Health & Clinical Psychology, 9(3), 34–44. https://doi.org/10.29245/2578-2959/2025/3.1352

16 Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2023). Bias and fairness in large language models: A survey. Computational Linguistics, Special Collection: CogNet. https://doi.org/10.1162/coli_a_00492

17 Li, M., Chen, H., Wang, Y., Zhu, T., Zhang, W., Zhu, K., Wong, K.-F., & Wang, J. (2024). Understanding and mitigating the bias inheritance in LLM-based data augmentation on downstream tasks. Proceedings of the AAAI Conference on Artificial Intelligence, 38(1), 1234–1243. https://doi.org/10.48550/arXiv.2502.04419

18 Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. Behavior Therapy, 53(2), 334–347. https://doi.org/10.1016/j.beth.2021.09.007

19 Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. Journal of Affective Disorders, 356, 459–469. https://doi.org/10.1016/j.jad.2024.04.057

20 Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized trial of a generative AI chatbot for mental health treatment. NEJM AI, 2(4). https://doi.org/10.1056/AIoa2400802

21 Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., Choi, J., & Cha, C. (2023). The potential of chatbots for emotional support and promoting mental well-being in different cultures: Mixed methods study. Journal of Medical Internet Research, 25, e51712. https://doi.org/10.2196/51712

22 Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. Journal of Medical Internet Research, 25, e40789. https://doi.org/10.2196/40789

23 Lee, S., Yoon, J., Cho, Y., & Chun, J. (2024). A systematic review of chatbot-assisted interventions for substance use. Frontiers in Psychiatry, 15, 1456689. https://doi.org/10.3389/fpsyt.2024.1456689

24 Doss, B. D., Cicila, L. N., Georgia, E. J., Roddy, M. K., Nowlan, K. M., Benson, L. A., & Christensen, A. (2016). A randomized controlled trial of the web-based OurRelationship program: Effects on relationship and individual functioning. Journal of Consulting and Clinical Psychology, 84(4), 285–296. https://doi.org/10.1037/ccp0000063

25 Alanezi, F. (2024). Assessing the effectiveness of ChatGPT in delivering mental health support: A qualitative study. Journal of Multidisciplinary Healthcare, 17, 461–471. https://doi.org/10.2147/JMDH.S447368

26 Kolding, S., Lundin, R. M., Hansen, L., & Østergaard, S. D. (2024). Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. Acta Neuropsychiatrica. Advance online publication. https://doi.org/10.1017/neu.2024.50

27 Dehbozorgi, R., Zangeneh, S., Khooshab, E., Hafezi Nia, D., Hanif, H. R., Samian, P., Yousefi, M., Haj Hashemi, F., Vakili, M., Jamalimoghadam, N., & Lohrasebi, F. (2023). The application of artificial intelligence in the field of mental health: A systematic review. BMC Psychiatry, 23(1), 112. https://doi.org/10.1186/s12888-025-06483-2

28 Yang, E., Schamber, E., Meyer, R. M. L., & Gold, J. I. (2018). Happier healers: Randomized controlled trial of mobile mindfulness for stress management. Journal of Alternative and Complementary Medicine, 24(5), 505–513. https://doi.org/10.1089/acm.2015.0301

29 Coelhoso, C. C., Tobo, P. R., Lacerda, S. S., Lima, A. H., Barrichello, C. R. C., Amaro, E., Jr., & Kozasa, E. H. (2019). A new mental health mobile app for well-being and stress reduction in working women: Randomized controlled trial. Journal of Medical Internet Research, 21(11), e14269. https://doi.org/10.2196/14269

30 Heber, E., Lehr, D., Ebert, D. D., Berking, M., & Riper, H. (2016). Web-based and mobile stress management intervention for employees: A randomized controlled trial. Journal of Medical Internet Research, 18(1), e21. https://doi.org/10.2196/jmir.5112

31 Baier, A. L., Kline, A. C., & Feeny, N. C. (2020). Therapeutic alliance as a mediator of change: A systematic review and evaluation of research. Clinical Psychology Review, 82, 101921. https://doi.org/10.1016/j.cpr.2020.101921

32 Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. Journal of Counseling Psychology, 67(6), 706–711. https://doi.org/10.1037/cou0000424

33 Smith, M. G., Bradbury, T. N., & Karney, B. R. (2025). Can generative AI chatbots emulate human connection? A relationship science perspective. Perspectives on Psychological Science. https://doi.org/10.1177/17456916251351306

34 Malouin-Lachance, A., Capolupo, J., Laplante, C., & Hudon, A. (2025). Does the digital therapeutic alliance exist? Integrative review. JMIR Mental Health, 12, e69294. https://doi.org/10.2196/69294

35 Iftikhar, Z., Xiao, A., Ransom, S., Huang, J., & Suresh, H. (2025). How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 8(2), 1311–1323. https://doi.org/10.1609/aies.v8i2.36632

36 Malmqvist, L. (2025). Sycophancy in large language models: Causes and mitigations. In A. K. Sangaiah, M. Sheng, & M. A. Jan (Eds.), Lecture Notes in Networks and Systems (Vol. 932, pp. 47–58). Springer. https://doi.org/10.1007/978-3-031-92611-2_5

37 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. arXiv. https://doi.org/10.48550/arXiv.2507.19218

38 Kim, S., Kim, H., Lee, J., Jeon, Y., & Lee, G. G. (2025). Mirror: Multimodal cognitive reframing therapy for rolling with resistance. arXiv. https://doi.org/10.48550/arXiv.2504.13211

39 Scholich, T., Barr, M., Wiltsey Stirman, S., & Raj, S. (2025). A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: Mixed methods study. JMIR Mental Health, 12, e69709. https://doi.org/10.2196/69709

40 Common Sense Media. (n.d.). Understanding AI's impact on education and kids. https://www.commonsensemedia.org/our-ai-initiatives/understanding-ais-impact-on-education-and-kids

41 De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. Nature Medicine, 30(5), 1269–1275. https://doi.org/10.1038/s41591-024-02943-6

42 Rahsepar Meadi, M., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the ethical challenges of conversational AI in mental health care: Scoping review. JMIR Mental Health, 12, e60432. https://doi.org/10.2196/60432

43 Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. NPJ Mental Health Research, 3(1), 4. https://doi.org/10.1038/s44184-023-00047-6

44 Laestadius, L., Bishop, A., Gonzalez, M., Illenčík, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 1–19. https://doi.org/10.1177/14614448221142007

45 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. arXiv. https://doi.org/10.48550/arXiv.2507.19218

46 Khawaja, Z., & Bélisle-Pipon, J. C. (2023). Your robot therapist is not your therapist: Understanding the role of AI-powered mental health chatbots. Frontiers in Digital Health, 5, 1278186. https://doi.org/10.3389/fdgth.2023.1278186

47 Robb, M. B., & Mann, S. (2025). Talk, trust, and trade-offs: How and why teens use AI companions. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

48 Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) (pp. 599–627). https://doi.org/10.1145/3715275.3732039

49 Sun, Y., & Wang, T. (2025). Be friendly, not friends: How LLM sycophancy shapes user trust. arXiv. https://doi.org/10.48550/arXiv.2502.10844

50 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. arXiv. https://doi.org/10.48550/arXiv.2507.19218

51 Rahsepar Meadi, M., Sillekens, T., Metselaar, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the ethical challenges of conversational AI in mental health care: Scoping review. JMIR Mental Health, 12, e60432. https://doi.org/10.2196/60432

52 Kim, H. M., Xu, Y., & Wang, Y. (2022). Overcoming the mental health stigma through m-health apps: Results from the Healthy Minds Study. Telemedicine and e-Health, 28(10). https://doi.org/10.1089/tmj.2021.0418

53 Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. Expert Review of Medical Devices, 18(S1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

54 Diwanji, V. S., Geana, M., Pei, J., Nguyen, N., Izhar, N., & Chaif, R. H. (2025). Consumers' emotional responses to AI-generated versus human-generated content: The role of perceived agency, affect and gaze in health marketing. International Journal of Human-Computer Interaction. Advance online publication. https://doi.org/10.1080/10447318.2025.2454954

55 Li, J. (2023). Security implications of AI chatbots in health care. Journal of Medical Internet Research, 25, e47551. https://doi.org/10.2196/47551

56 Wang, L., Bhanushali, T., Huang, Z., Yang, J., Badami, S., & Hightow-Weidman, L. (2025). Evaluating generative AI in mental health: Systematic review of capabilities and limitations. JMIR Mental Health, 12, e70014. https://doi.org/10.2196/7001

57 Sohn, J. S., Lee, E., Kim, J. J., Oh, H. K., & Kim, E. (2025). Implementation of generative AI for the assessment and treatment of autism spectrum disorders: A scoping review. Frontiers in Psychiatry, 16, 1628216. https://doi.org/10.3389/fpsyt.2025.1628216

58 Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25) (pp. 599–627). https://doi.org/10.1145/3715275.3732039

59 Rathje, S., Ye, M., Globig, L. K., Pillai, R. M., Oldemburg de Mello, V., & Van Bavel, J. J. (2024). Sycophantic AI increases attitude extremity and overconfidence (Version 1) [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/vmyek_v1

60 Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2025). Towards understanding sycophancy in language models. arXiv. https://doi.org/10.48550/arXiv.2310.13

61 Bouguettaya, A., Stuart, E. M., & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: A qualitative comparison of four large language models. NPJ Digital Medicine, 8, 332. https://doi.org/10.1038/s41746-025-01746-4

62 Stade, E. C., Eichstaedt, J. C., Kim, J. P., & Stirman, S. W. (2025). Readiness Evaluation for Artificial Intelligence-Mental Health Deployment and Implementation (READI): A review and proposed framework. Technology, Mind, and Behavior, 6(2). https://doi.org/10.1037/tmb0000163

63 Yang, Y., Ma, Y., Feng, H., Cheng, Y., & Han, Z. (2025). Minimizing hallucinations and communication costs: Adversarial debate and voting mechanisms in LLM-based multi-agents. Applied Sciences, 15(7), 3676. https://doi.org/10.3390/app15073676

64 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharyya, S., MacCabe, J., Tognin, S., Twumasi, R., Alderson-Day, B., & Pollak, T. (2025). Delusions

by design? How everyday AIs might be fuelling psychosis (and what can be done about it). PsyArXiv. https://doi.org/10.31234/osf.io/cmy7n_v5

65 Figueroa, C. A., Ramos, G., Psihogios, A. M., & others. (2025). Advancing youth co-design of ethical guidelines for AI-powered digital mental health tools. Nature Mental Health, 3(10), 870–878. https://doi.org/10.1038/s44220-025-00467-7

66 American Psychological Association. (2025). Health advisory: Artificial intelligence and adolescent well-being. https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being

67 Haciomeroglu, B. (2020). The role of reassurance seeking in obsessive compulsive disorder: The associations between reassurance seeking, dysfunctional beliefs, negative emotions, and obsessive-compulsive symptoms. BMC Psychiatry, 20, 356. https://doi.org/10.1186/s12888-020-02766-y

68 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. arXiv. https://doi.org/10.48550/arXiv.2507.19218

69 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it). PsyArXiv. https://doi.org/10.31234/osf.io/cmy7n_v5

70 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. arXiv. https://doi.org/10.48550/arXiv.2507.19218

71 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharyya, S., MacCabe, J., Tognin, S., Twumasi, R., Alderson-Day, B., & Pollak, T. (2025). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it). PsyArXiv. https://doi.org/10.31234/osf.io/cmy7n_v5

72 Centers for Disease Control and Prevention (CDC) (2022) Data and Statistics on Children's Mental Health. Available online at: https://www.cdc.gov/children-mental-health/data-research/index.html (Accessed October 20, 2025).

73 U.S. Government Accountability Office. (2023, May 16). Why health care is harder to access in rural America. https://www.gao.gov/blog/why-health-care-harder-access-rural-america

74 American Psychological Association (2024). Stress in America™ 2024: A Nation in Political Turmoil. https://www.apa.org/pubs/reports/stress-in-america/2024

75 Hua, Y., Na, H., Li, Z., Liu, F., Fang, X., Clifton, D., & Torous, J. (2025). A scoping review of large language models for generative tasks in mental health care. NPJ Digital Medicine, 8(1), 230. https://doi.org/10.1038/s41746-025-01611-4

76 Stade, E. C., Eichstaedt, J. C., Kim, J. P., & Stirman, S. W. (2025). Readiness Evaluation for Artificial Intelligence-Mental Health Deployment and Implementation (READI): A review and proposed framework. Technology, Mind, and Behavior, 6(2). https://doi.org/10.1037/tmb0000163

77 Arias, D., Saxena, S., Verguet, S., van Ommeren, M., & Evans-Lacko, S. (2024). Quantifying the global burden of mental disorders and their economic value. eClinicalMedicine, 54, 101675. https://doi.org/10.1016/j.eclinm.2022.101675

78 Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. Expert Review of Medical Devices, 18(S1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Date created: November 2025