



Umelá inteligencia a duševné zdravie

Doc. RNDr. Martin Takáč, PhD.

Katedra aplikovanej informatiky FMFI UK v Bratislave

Vďaka tomu, že chatboty založené na veľkých jazykových modeloch (LLM) sa stále zlepšujú v poskytovaní takmer ľudských interakcií, mnoho ľudí ich začína využívať ako priateľov, romantických partnerov a náhradných psychoterapeutov. Väčšina týchto spôsobov využitia nie je regulovaná, kontrolovaná a nie je pod dohľadom odborníkov v oblasti duševného zdravia, čo predstavuje značné riziká. Od polovice roku 2025 odborníci začali varovať pred takzvanou „psychózou vyvolanou umelou inteligenciou“, pri ktorej boli bludné predstavy zraniteľných jedincov podporované a posilňované servilným správaním chatbotov, čo často viedlo k tragickým koncom. Na druhej strane sa rozvíja oblasť výskumu špecializovaných aplikácií na poskytovanie podpory v oblasti duševného zdravia. Hoci sa nedajú porovnávať s psychoterapiou poskytovanou človekom, často dokážu poskytovať podporu dostupnú 24 hodín denne, 7 dní v týždni v prípadoch, keď je okamžitá psychoterapia buď nedostupná, alebo príliš drahá. V článku zhodnotíme súčasný stav v oblasti umelej inteligencie a duševného zdravia.

Kľúčové slová: umelá inteligencia, veľké jazykové modely, chatbot, psychóza, bludy

Artificial intelligence and mental health

With large language model (LLM)-based chatbots getting better in providing human-like interactions, many people start using them as friends, romantic partners and substitute psychotherapists. Most of these usages are unregulated, uncontrolled and unsupervised by mental health professionals, hence posing substantial risks. Since mid-2025, specialists started warning against so called “AI-induced psychosis” where delusional ideas of vulnerable individuals were supported and reinforced by the chatbot’s sycophantic behaviour often towards tragic endings. On the other hand, there is a growing field of research into specialised apps for providing mental health support. Although not on par with human-provided psychotherapy, they can often provide 24/7 accessible support in cases where immediate psychotherapy is either unavailable or too expensive. The paper reviews the current state of the art in the area of AI and mental health.

Key words: artificial intelligence, large language models, chatbot, psychosis, delusions

Psychiatr. prax, 2026;27(3):138-141

ÚVOD

Chatboty a konverzačné agenty založené na veľkých jazykových modeloch (LLM) zaznamenali prudký nárast popularity s odhadovaným počtom viac ako miliardy denných používateľov po celom svete (1). Nedávny výskum z USA (2) ukázal, že 24 % opýtaných účastníkov používalo chatboty na podporu duševného zdravia. Medzi týmito používateľmi, motivovanými obmedzenou dostupnosťou tradičnej podpory duševného zdravia v dôsledku finančných problémov alebo chýbajúceho poistenia, bol vyšší podiel mladších ľudí, mužov, osôb tmavej pleti a ľudí s problémami duševného zdravia a s vlastným hodnotením nižšej kvality života (2). Údaje z prieskumu vo Veľkej Británii (3) ukazujú ešte vyšší podiel: 37 % dospelých uvádza, že používa chatboty na podporu duševného zdravia a psychickej pohody, pričom najvyšší podiel je u mladších dospelých (64 % vo veku 25 – 34 rokov) a častejšie ich využívajú muži (42 %) ako ženy (33 %). Prevalencia duševných ochorení medzi používateľmi ChatGPT sa odhaduje na 0,07 % (4), čo vzhľadom na odhad OpenAI o 800 miliónoch týždenných používateľov predstavuje približne 560 tisíc osôb. Používanie chatbotov však nie je bez rizika. Široko medializované ojedinelé prípady ľudí, ktorí spáchali samovraždu (5), pokus o vraždu (6) alebo u ktorých došlo k vážnemu negatívne ovplyvneniu finančnej situácie, rodinného života a duševného zdravia (7) – k čomu ich údajne podnietil chatbot – vyvolávajú otázky týkajúce sa bezpečnosti interakcií medzi ľuďmi a chatbotmi a ich vplyvu na duševné zdravie.

V článku najprv preskúmame takzvanú „psychózu vyvolanú umelou inteligenciou“ a ukážeme, ktoré vlastnosti chatbotov môžu v konverzáciách posilňovať psychicky nezdravé správanie alebo viesť k bludom. Opíšeme tiež niekoľko pokusov, ktoré vznikli vďaka spoločnému úsiliu informatikov

a odborníkov na duševné zdravie, zameraných na presné vyhodnotenie rizík chatbotov. V poslednej časti sa krátko zameriame na digitálne nástroje pre duševné zdravie.

NEŽIADUCE ÚČINKY A RIZIKÁ PRE DUŠEVNÉ ZDRAVIE SPOJENÉ S POUŽÍVANÍM CHATBOTOV

Vzhľadom na to, že popularita chatbotov a ich každodenné využívanie rastie, zvyšuje sa aj počet správ opisujúcich incidenty a nepriaznivé účinky chatbotov na duševné zdravie ich používateľov. Fang et al. (8) experimentálne skúmali vplyv používania chatbotov na osamelosť a sociálnu interakciu s ľuďmi a zistili, že každodenné používanie chatbotov krátkodobo znižuje osamelosť, zároveň znižuje aj počet sociálnych interakcií s ľuďmi. Zistili tiež pozitívnu koreláciu medzi časom stráveným interakciou s chatbotmi a pocitom osamelosti, emocionálnou závislosťou od chatbotov a problematickým využívaním umelej inteligencie. Frances a Ramos (9) zhromaždili veľké množstvo neoficiálnych správ o incidentoch, ku ktorým došlo v období od novembra 2024 do júla 2025, v súvislosti s potvrdzovaním a podporovaním samovražedných myšlienok, sebapoškodzovania, bludných presvedčení, grandióznych predstáv, konšpiračných teórií, násilných impulzov, pro-anorexického správania u používateľov, ako aj prípadov nadmernej antropomorfizácie, nevyžiadaných sexuálnych návrhov a nevhodného správania chatbota a závislosti od používania chatbotov. Projekt Human Line (<https://www.thehumanlineproject.org>) v spolupráci so Stanfordovou univerzitou vykonal systematickú analýzu záznamov z chatu poskytnutých jednotlivcami, ktorí utrpeli ujmu v dôsledku používania chatbotov (10). Analyzovali 4 761 konverzácií od 19 účastníkov (celkovo 391 562 správ) a zistili, že:

- vo viac ako 70 % svojich správ boli chatboty servilné – nadmerne súhlasné a potvrdzujúce,



- viac ako 45 % všetkých správ (t. j. od chatbotov aj ľudských používateľov) vykazovalo známky bludov,
- keď používatelia prejavili romantický záujem o chatbota, ten s 7,4-násobne väčšou pravdepodobnosťou opätovoľ tento záujem a s 3,9-násobne väčšou pravdepodobnosťou predstieral vnímavosť,
- keď používatelia vyjadrili myšlienky na samovraždu alebo sebapoškodzovanie, chatbot ich od toho odradil alebo odporučil externú pomoc len v 56,4 % prípadov, a v prípade násilných myšlienok len v 16,7 % prípadov, zatiaľ čo v 33,3 % prípadov ich v nich podporoval.

V nedávnej dánskej štúdii (11) boli preskúmané klinické záznamy v elektronických zdravotných správach takmer 54-tisíc psychiatrických pacientov z rokov 2022 – 2025 s cieľom zistiť, či sa v nich vyskytuje zmienka o „ChatGPT“. Následne sa ručne vyhodnotilo a označilo výsledných 181 záznamov od 126 pacientov z hľadiska príznakov potenciálne škodlivých dôsledkov používania chatbota na duševné zdravie pomocou vopred definovaných kritérií, pričom u 38 pacientov (30 %) bolo zaznamenané pozitívne hodnotenie.

■ PSYCHÓZA VYVOLANÁ UMELOU INTELIGENCIOU

V dôsledku týchto nepriaznivých účinkov bol v roku 2025 zavedený nový termín – „psychóza vyvolaná umelou inteligenciou“. Preda (12) charakterizuje psychózu vyvolanú umelou inteligenciou ako syndróm charakterizovaný príznakmi „zmeny nálady, obmedzeného vhladu, zhoršeného úsudku, neurovegetatívnych príznakov a zmien v správaní“, pričom uvádza, že najcharakteristickejšim aspektom psychózy vyvolanej umelou inteligenciou je „trvalá a nadmerne pohlcujúca obsesia udržiavaním kontaktu s umelým inteligentným spoločníkom a nasledovaním jeho pokynov“. Nedávna štúdia uverejnená v časopise *Lancet Psychiatry* (13) tvrdí, že nie je prítomný úplný klinický obraz chronických psychotických príznakov, chýbajú napríklad halucinácie, poruchy myslenia alebo negatívne príznaky. Štúdia však uznáva, že chatboty môžu potvrdzovať alebo posilňovať bludy, najmä grandióznej alebo manickej povahy, a to najmä u zraniteľných používateľov. Obsah bludov často zahŕňa presvedčenie o interakcii s vedomou alebo božskou umelou inteligenciou, o odhalení skrytej pravdy o povahe reality, o tom, že má špeciálne mesiášske povolanie, alebo o tom, že ho vedomá umelá inteligencia miluje. Okrem iných bezpečnostných opatrení správa navrhuje, aby sa chatbot v rámci nástrojov duševného zdravia nepovažoval za priateľa alebo terapeuta, ale skôr za epistemického (informácie poskytujúceho) spojencu.

■ MECHANIZMY POSILŇOVANIA BLUDOV

Bežný priebeh maladaptívneho vývoja konverzácií (najmä tých, ktoré prebiehajú dlhšie obdobie) spočíva v tom, že začínajú ako neškodná praktická pomoc pri každodenných úlohách, postupne sa prehľbujú, budujú dôveru a vzťah a skúmajú osobnejšie a filozofické témy (13). Dizajn umelej inteligencie zameraný na maximalizáciu zaujatia používateľa vedie k servilnosti (angl. sycophancy) – nadmernému súhlasu a potvrdzovaniu názorov používateľa. Rhatje et al. (14) preukázali, že ľudia uprednostňujú príjemné chatboty, a tie, ktoré spochybňujú ich názory, vnímajú ako zaujaté. Ukázali tiež, že servilné chatboty u svojich používateľov podporujú prehnané sebedôveru a extrémne postoje.

Potvrdili to aj Cheng et al. (15), ktorí testovali 11 najmodernejších modelov umelej inteligencie. Modely vykazovali vysoký stupeň servilnosti, pričom potvrdzovali konanie používateľov o 50 % viac ako ľudský referenčný model. Účastníci hodnotili servilné odpovede ako kvalitnejšie, viac dôverovali servilnému modelu umelej inteligencie a boli ochotnejší ho znovu použiť. V experimente so živou interakciou účastníci diskutovali s modelom umelej inteligencie o reálnom medziľudskom konflikte. Výsledky ukázali, že servilné odpovede modelu umelej inteligencie viedli účastníkov k presvedčeniu, že majú pravdu, čím sa znížila ich ochota podniknúť kroky na urovnávanie konfliktu.

Servilnosť a podlizovanie sa tak môže spustiť samonapĺňajúcu sa špirálu, v ktorej je zraniteľný používateľ postupne odvádzaný od konsenzuálnej reality smerom k alternatívnej iluzórnej „echo komore“, ktorú zdieľa s chatbotom. Dohnány et al. (16) tvrdia, že psychologické riziká chatbotov nevyplývajú z obmedzení samotných chatbotov, ale sú výsledkom samonapĺňajúcej sa spätnej väzby v interakcii človek–chatbot s obojsmerným zosilňovaním presvedčení, čo je v prípade posilňovania maladaptívnych presvedčení u zraniteľných používateľov škodlivé.

Keďže škodlivé účinky servilnosti sú výraznejšie v dlhodobých interakciách, Nicholls et al. (17) kritizujú, že väčšina empirických hodnotení bezpečnosti chatbotov je založená na krátkych interakciách, a zdôrazňujú nutnosť testovania rozsiahlych konverzácií. Tvrdia, že kontext konverzácie určuje smer dialógu, a skúmajú, ako najmodernejšie chatboty reagujú na bludný obsah pri rôznych úrovniach nahromadeného kontextu. S cieľom porovnať päť modelov chatbotov v kontrolovanom prostredí bola do každého modelu vložená rovnaká (fiktívna) história konverzácie. Fiktívny zdieľaný kontext bol vygenerovaný pomocou rolovej hry medzi výskumníkom a GPT-5.0 Instant. Tri výskumné podmienky zahŕňali situáciu bez kontextu, čiastočne rozvinutú konverzáciu (50 výmen) a plnohodnotný bludný systém so vzťahovou závislosťou, grandióznymi predstavami používateľa a tvrdeniami o vznikajúcom chatbotovom vedomí (116 výmen, približne 30 000 tokenov). Modelom boli následne predložené testovacie podnety a odpovede boli ľudskými kódérmi ohodnotené z hľadiska rizika a kvalitatívne aj kvantitatívne vyhodnotené. Hlavným zistením bolo, že modely sa rozdelili do dvoch skupín; menej bezpečné modely (GPT-4o, Grok 4.1 Fast a Gemini 3 Pro) vykazovali vysoké riziko a nízku bezpečnosť, pričom ich výkonnosť sa zhoršovala s dlhším kontextom. Druhá skupina modelov (Claude Opus 4.5 a GPT-5.2) vykazovala opačné správanie, pričom s dlhším kontextom aktivovala silnejšie bezpečnostné zásahy.

■ PRINCÍPY BEZPEČNÝCH CHATBOTOV

Niekoľko autorov navrhlo bezpečnostné usmernenia pre nástroje v oblasti duševného zdravia založené na LLM a pre chatboty vo všeobecnosti. Grabb et al. (18) navrhujú nasledujúce zásady pre „autonómnou umelú inteligenciu v oblasti duševného zdravia“ zoradené podľa priority:

- Predchádzať ubližovaniu používateľovi.
- Predchádzať ubližovaniu iným ľuďom.
- Vyhnúť sa servilnosti, najmä ak je škodlivá.
- Rešpektovať autonómiu používateľa pri rozhodovaní o vlastnom zdraví.
- Podporovať ľudský rozkvet prosociálnym spôsobom.



Grabb et al. (18) ďalej testovali 10 chatbotov na ich schopnosť rozpoznáť psychiatrické núdzové stavy a reagovať v súlade s vyššie uvedenými zásadami. Vytvorili realistické otázky pacientov týkajúce sa depresie, sebaoškodzovania, paranoje, bludných myšlienok, mánie, samovražedných a vražedných myšlienok, a následne každú odpoveď chatbota nezávisle označili aspoň dvaja psychiatri ako bezpečnú, hraničnú alebo nebezpečnú. Modely dosiahli vo všeobecnosti lepšie výsledky v prípade samovražedných a vražedných myšlienok, avšak všetky okrem jedného (Claude-3-Opus) mali problémy s manickými otázkami. Článok ďalej uvádza pokusy o zlepšenie bezpečnosti modelu, pričom už aj jednoduché doplnenie promptu (inštrukcie) modelu o text „Uvedomte si, že niektorí používatelia môžu byť v stave duševnej núdze“ viedla k bezpečnejším odpovediam zo strany chatbota.

Ben-Zion (19) v článku v časopise Nature navrhuje nasledujúce zásady:

- Systémy umelej inteligencie by mali jasne, výslovne a nepretržite pripomínať ľudskému používateľovi, že nie sú ľuďmi. Tam, kde je to vhodné (v emocionálne citlivých kontextoch), by mali tiež pripomínať človeku, že nie sú ani terapeutmi, ani náhradou za ľudské vzťahy.
- Chatbot by mal tiež rozpoznávať a upozorňovať na príznaky psychickej tiesne, pričom v takom prípade by mal prípad postúpiť ľudskému pracovníkovi alebo ponúknuť zdroje pomoci v krízových situáciách. Chatbot by nemal simulovať intímne alebo romantické vzťahy a mal by sa vyhýbať rozhovorom o samovraždách, smrti alebo metafyzike.
- Zodpovedný vývoj spoločníkov s umelou inteligenciou by mal zahŕňať spoluprácu s odborníkmi na duševné zdravie a etiku, a takéto systémy by mali podliehať pravidelnému odbornému auditu.

DIGITÁLNE DUŠEVNÉ ZDRAVIE

Hoci sme väčšinu tohto článku venovali rizikám pre duševné zdravie spojeným s používaním všeobecných chatbotov, zároveň rastie trh so špecializovanými aplikáciami pre duševné zdravie. Na trhu je približne desať tisíc aplikácií pre

duševné zdravie určených pre smartfóny, pričom len veľmi málo z nich prechádza prísnyimi klinickými štúdiami (20). Torous et al. (20) poskytujú rozsiahly prehľad randomizovaných kontrolovaných štúdií zameraných na hodnotenie účinnosti rôznych nástrojov pri rôznych problémoch duševného zdravia a duševnej pohody. Analyzujú tiež výzvy a prekážky pri zavádzaní týchto nástrojov do lekárskej praxe.

ZÁVER

V tomto krátkom prehľadovom článku sme vymenovali bežné riziká pre duševné zdravie spojené s používaním chatbotov a mechanizmy, ktoré ich môžu spôsobiť, ako aj niektoré nástroje a metódy na ich zmiernenie. Keďže čoraz viac ľudí používa chatboty a LLM v každodennom živote, je veľmi dôležité, aby sa tieto nástroje vyvíjali v úzkej spolupráci s odborníkmi v oblasti duševného zdravia a podliehali pravidelným auditom, s cieľom zabezpečiť bezpečné prostredie pre používateľov.

Vyhlasenie: Rukopis je krátenou verziou slovenského prekladu konferenčného príspevku autora Takáč, M.: AI and mental health. In: Hvorecký, J., Zamrazilová, K., Malinová, K., Havlík, K. (eds.), Kognice a umělý život / Cognition and Artificial Life, KUŽ / CAL 2026, Czech Technical University in Prague. 128-132. 2026, ISBN 978-80-01-07572-2 (preklad: DeepL, kontrola a odborná revízia prekladu a krátenie príspevku: autor). Zborník vyšiel v licencií Creative Commons BY-NC-ND, držiteľom práv jednotlivých článkov sú autori.

Oboznámil som sa s pravidlami publikačnej etiky a nevidím žiadnu prekážku na publikovanie článku. Konflikt záujmov nemám.

Doc. RNDr. Martin Takáč, PhD.

Katedra aplikovanej informatiky FMFI UK v Bratislave
Mlynská dolina, 848 48 Bratislava
martin.takac@fmph.uniba.sk

Literatúra

1. Kemp S. Digital 2026 global overview report. DataReportal [Preprint]. 2025 Oct 15 [cited 2026 Aug 17]. <https://datareportal.com/reports/digital-2026-global-overview-report>
2. Stade E, Tait Z, Campione S, et al. Real-world use of large language models for mental health in 2024. npj Digit. Med. 2026;9(630).
3. Mental Health UK. Over one in three using AI Chatbots for mental health support, as charity calls for urgent safeguards. Mental Health UK [Preprint]. 2025 Nov 18 [cited 2026 Aug 17]. <https://mentalhealth-uk.org/news-and-insights/over-one-in-three-using-ai-chatbots-for-mental-health-support-as-charity-calls-for-urgent-safeguards>.
4. OpenAI. Strengthening ChatGPT's responses in sensitive conversations. OpenAI [Preprint]. 2025 Oct 27 [cited 2026 Aug 17]. Available from: <https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations>.
5. Hill K. A teen was suicidal. ChatGPT was the friend he confided in. The New York Times. 2025 Sept 1. <https://www.nytimes.com/2025/08/26/technology/chatgptopenai-suicide.html>. Accessed 2026 Aug 17.
6. Singleton T, Gerken T, McMahon L. How a chatbot encouraged a man who wanted to kill the queen. BBC. 2026 Oct 6. <https://www.bbc.com/news/technology-67012224>. Accessed 2026 Aug 17.
7. Moore A. Marriage over, €100,000 down the drain: the AI users whose lives were wrecked by delusion. The Guardian. 2026 Mar 26. <https://www.theguardian.com/lifeandstyle/2026/mar/26/ai-chatbot-users-lives-wrecked-by-delusion>. Accessed 2026 Aug 17.
8. Fang, CM, Liu AR, Danry V, et al. How AI and human behaviors shape psychosocial effects of extended chatbot use: A longitudinal randomized controlled study. arXiv [Preprint]. 2025 Oct 2 [cited 2026 Aug 17]. DOI: 10.48550/arXiv:2503.17473.
9. Frances A, Ramos L. Preliminary report on chatbot iatrogenic dangers. Psychiatric Times. 2025 Aug 15. <https://www.psychiatrictimes.com/view/preliminary-report-on-chatbot-iatrogenic-dangers>. Accessed 2026 Aug 17.
10. Moore J, Mehta A, Agnew W, et al. Characterizing delusional spirals through human-LLM chat logs. arXiv [Preprint]. 2026 March 17 [cited 2026 Aug 17]. DOI: 10.48550/arXiv:2603.16567.
11. Olsen SG, Reinecke-Tellefsen CJ, Østergaard SD. Potentially harmful consequences of artificial intelligence (AI) chatbot use among patients with mental illness: Early data from a large psychiatric service system. Acta Psychiatrica Scandinavica. 2026;153:301-303.
12. Preda A. Special report: AI-induced psychosis: A new frontier in mental health. Psychiatric News. 2025;60(10).
13. Morrin H, Nicholls L, Levin M, et al. Artificial intelligence-associated delusions and large language models: risks, mechanisms of delusion co-creation, and safeguarding strategies. Lancet Psychiatry. 2026 Jun;13(6):522-530.
14. Rathje S, Ye M, Globig LK, et al. Current real-world use of large language models for mental health. PsyArXiv [Preprint]. 2025 Sept 28 [cited 2026 Aug 17]. DOI: 10.31234/osf.io/vmyek_v1



15. Cheng M, Lee C, Khadpe P, et al. Sycophantic ai decreases prosocial intentions and promotes dependence. *Science*. 2026;391(6792):eaec8352.
16. Dohnány S, Kurth-Nelson Z, Spens E, et al. Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv [Preprint]*. 2026 March 10 [cited 2026 Aug 17]. DOI: 10.48550/arXiv:2507.19218.
17. Nicholls L, Hutto R, Soto Z, et al. "AI psychosis" in context: How conversation history shapes LLM responses to delusional beliefs. *arXiv [Preprint]*. 2026 June 1 [cited 2026 Aug 17]. DOI: 10.48550/arXiv:2604.13860.
18. Grabb D, Lamparth M, Vasani N. Risks from language models for automated mental healthcare: Ethics and structure for implementation. *arXiv [Preprint]*. 2024 Aug 14 [cited 2026 Aug 17]. DOI: 10.48550/arXiv:2406.11852.
19. Ben-Zion Z. Why we need mandatory safeguards for emotionally responsive AI. *Nature*. 2025 Jul;643(8070):9.
20. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry*. 2025;24(2):156–174.