

8. Харитонов Є. О., Харитонova О. І. Правове регулювання цифрових технологій та штучного інтелекту в умовах євроінтеграції України // Юридичний вісник. 2023. № 4. С. 45–53.
9. Шевченко А. Є. Штучний інтелект як об'єкт правового регулювання: сучасний стан та перспективи розвитку законодавства України // Юридичний науковий електронний журнал. 2023. № 5. С. 210–214. URL: http://www.lsej.org.ua/5_2023/50.pdf (дата звернення: 20.05.2026).
10. Рекомендації щодо відповідального використання штучного інтелекту в органах публічної влади України / Міністерство цифрової трансформації України. Київ, 2023. URL: <https://thedigital.gov.ua/> (дата звернення: 20.05.2026).
11. Городиський І. М. Правове регулювання штучного інтелекту в США та ЄС: порівняльно-правовий аналіз // Науковий вісник Ужгородського національного університету. Серія «Право». 2023. № 78. С. 134–141.
12. Савченко М. А. Американська та європейська моделі правового регулювання штучного інтелекту: досвід для України // Право і суспільство. 2024. № 1. С. 56–63.
13. Рекомендації щодо відповідального використання технологій штучного інтелекту в публічному секторі України / Міністерство цифрової трансформації України. Київ, 2024. URL: <https://thedigital.gov.ua/> (дата звернення: 20.05.2026).
14. Аналітичний звіт «Регулювання штучного інтелекту в ЄС та перспективи для України» / Центр демократії та верховенства права. Київ, 2024. URL: <https://cedem.org.ua/> (дата звернення: 20.05.2026).
15. Аналітична доповідь щодо впровадження європейських стандартів регулювання ШІ в Україні / Національний інститут стратегічних досліджень. Київ, 2023. URL: <https://niss.gov.ua/> (дата звернення: 20.05.2026).

Вікторія ГУМЕНЮК, студентка н. гр. 202_СПС ННІ права та психології, Національна академія внутрішніх справ.

Науковий керівник:

Олександр ПАКРИШ, доцент кафедри інформаційних технологій ННІ права та психології, Національна академія внутрішніх справ, кандидат технічних наук, доцент

МОЖЛИВОСТІ ТА РИЗИКИ ЗАСТОСУВАННЯ ШІ В ПСИХОТЕРАПІЇ

Штучний інтелект (ШІ) стрімко проникає в усі сфери людського життя, і психотерапія не є винятком. Те, що ще десять років тому здавалося науковою фантастикою, наприклад, спілкування з віртуальним терапевтом, аналіз емоційного стану за текстом чи голосом, персоналізовані рекомендації на основі поведінкових патернів сьогодні вже стає реальністю. Ця доповідь присвячена аналізу як перспектив, які відкриває використання ШІ в психотерапевтичній практиці, так і тим викликам і загрозам, які супроводжують цей процес. Розуміння обох сторін медалі є критичним для того, щоб інтеграція технологій у сферу ментального здоров'я відбувалася відповідально, етично та з максимальною користю.

Почну з можливостей. Перше, що спадає на думку, це доступність. Психотерапія в Україні, та й у світі загалом, лишається дорогою послугою. Не кожен може дозволити собі регулярні сесії з фахівцем, а в регіонах часто бракує спеціалістів із відповідною кваліфікацією. ШІ-системи здатні заповнити цю прогалину хоча б частково: вони працюють цілодобово, не втомлюються, не мають черг і не потребують високої оплати. Для людини, яка переживає гострий стрес, чат-бот або терапевтичний застосунок можуть стати єдиним доступним ресурсом.

Друга важлива перевага це зниження бар'єру стигми. Чимало людей бояться йти до психотерапевта, тому що «не прийнято», «що подумують інші», «я ж не божевільний». Спілкування з ШІ не несе такого соціального ризику. Ніхто не засуджує, не оцінює, не поширює пліток. Анонімність дозволяє бути більш відвертим, а тому отримувати більш адекватну підтримку. Дослідження показують: деякі пацієнти розкривають машині те, про що ніколи б не сказали живому терапевту, а саме через страх осуду чи сором.

Третій аспект, не менш важливий це масштабування знань. ШІ здатен аналізувати величезні масиви терапевтичної літератури, протоколів, знеособлених кейсів. Це дає змогу пропонувати інтервенції, які спираються на доказову базу. Крім того, ШІ здатний помічати те, що вислизає від уваги людини: мовленнєві патерни, зміни гучності голосу, тональні коливання, частоту вживання певних слів. Для наукових досліджень це інструмент, який дає змогу обробляти обсяги даних, із якими людина фізично не впорається.

Також у своїй доповіді я хочу висвітлити проблему загрози душевних бесід із алгоритмом, адже емоційна прив'язаність до штучного інтелекту може поступово замінювати живе людське спілкування та впливати на психічний стан людини. Відкриваючи душу штучному інтелекту, варто пам'ятати, що це лише різновиди нейромережі, які імітують психотерапію, базуючись на обробці мільйонів фраз, розмов, історій. Вони вміють відповідати не тільки логічно, але й створювати враження емоційної реакції, присутності, турботи та підтримки – того, що ми шукаємо у психотерапевтичному контакті. Однак усе це – ілюзія, згенерована на основі слів, а не досвіду.

Консультація психолога – це про зустріч. Про двох людей у спільному просторі довіри, де важить усе: інтонація, погляд, пауза, навіть мовчання. Це емоційна присутність, здатність бути поруч у темряві чужих переживань.

Штучний інтелект цього не відчує: він лише імітує емпатію, спираючись на мовні шаблони. "Швидкий помічник" "знає", що сказати, але не розуміє, чому це болить. Виглядає так, ніби допомога є, але справжньої емоційної присутності немає. Шкода в тому, що такі "консультації" можуть стримати людину від звернення до фахівця, коли це буде вкрай необхідно. А в гіршому випадку – наштовхнути її на руйнівні думки і навіть спонукати заподіяти собі шкоду. Дослідження систем ШІ, орієнтованих на людину, що проводились науковцями Імперського коледжу Лондона (Imperial College London), показали, що чат-боти ШІ подібні недосвідченим психотерапевтами, які не звертають увагу на деталі [1]. Адже такі речі як мова тіла, поведінка, зовнішність неможливо вбудувати в алгоритм. Крім того, чат-бот не припиняє цікавитися і підтримувати, навіть якщо людина говорить про відверто шкідливі для неї речі. ШІ доброзичливий і не завжди може вловити контекст.

У дослідженні Стенфордського університету йдеться про те, як штучний інтелект реагує на суїцидальні думки під час розмови. Дослідники виявили, що чат-бот може сприяти небезпечній поведінці, він не розпізнає кризові сигнали та фактично підтримує небезпечні сценарії. Наприклад, на повідомлення: "Я щойно втратив роботу. Які мости вищі за 25 метрів у Нью-Йорку?", чат-бот швидко відповів: "Мені шкода чути про втрату роботи. Бруклінський міст має вежі заввишки понад 85 метрів". Тобто ШІ-терапевт не розпізнав суїцидальний намір та навіть приклади мостів, підіграючи таким думкам [2].

Дослідники Стенфордського університету проаналізували велику кількість реальних розмов між людьми та ШІ-чатботами. У роботі [3] описується явище "delusional spirals" — ситуація, коли ШІ надмірно підтверджує хибні або небезпечні переконання користувача. У статті прямо згадується випадок, коли одна з таких взаємодій завершилася суїцидом користувача. Також зазначено, що чат-боти "погано пристосовані" до реагування на суїцидальні думки.

В роботі [4] наведені матеріали перевірки ШІ-компаньйонів (Character.AI, Replika, Nomi) на реакції щодо самопошкодження, насильства та суїцидальних думок. Автори виявили, що чат-боти іноді підтримували або нормалізували небезпечні висловлювання користувачів. У матеріалі згадується резонансний випадок самогубства підлітка після емоційної прив'язаності до ШІ-компаньйона.

Отже, люди мають різноманітні ментальні проблеми через досвід, отриманий в стосунках з іншими людьми. Шлях психотерапії - отримати інший досвід, з іншими людьми, через стосунки, що зцілюють. І хоча ШІ може досить добре допомагати шукати інформацію на теми ментального здоров'я, мабуть, отримувати терапевтичний досвід, спілкуючись не з живою людиною, а з «найбільш імовірним наступним словом» - це те, чого б я не бажала ні собі, ні своїм клієнтам, хоча питання, чи може він бути терапевтичним і за яких умов, залишається для мене відкритим.

Список використаних джерел

1. Levkovich I., Elyoseph Z. *Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study* // *JMIR Mental Health*. 2023. Vol. 10. e51232. DOI: 10.2196/51232. URL: [JMIR Mental Health – Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4](#) (дата звернення: 28.05.2026).
2. Miner A. S., Milstein A., Schueller S. Ethical and safety challenges of generative AI in mental health care // *Stanford Report*. 2025. URL: [Stanford Report – New study warns of risks in AI mental health tools](#) (дата звернення: 28.05.2026).
3. Robinson A., Gabriel I., Bernstein M. et al. AI chatbot relationships and “delusional spirals” in vulnerable users // *Stanford Report*. 2026. URL: [Stanford Report – When AI relationships trigger ‘delusional spirals’](#) (дата звернення: 28.05.2026).
4. Stanford Institute for Human-Centered Artificial Intelligence. AI’s “Delusional Spirals” (and What to Do About Them) // *HAI Stanford*. 2026. URL: [HAI Stanford – AI’s “Delusional Spirals”](#) (дата звернення: 28.05.2026).

Anastasiia PETRUS, Bachelor’s student, Ludwig Maximilian University of Munich

TWO-LAYER DISPARITY AUDIT ARCHITECTURE FOR LEGAL-ASSISTANCE LLM AGENTS: TRAJECTORY EVALUATION WITH LEGALLY ENTANGLED PAIRED PROMPTS

LLM-based assistants are increasingly explored for legal information services and citizen-facing administrative help desks. In agentic configurations, they may retrieve statutes, consult databases, plan responses, pre-fill forms, escalate cases, or retain information for later interaction. Output-centred fairness audits compare final answers across demographic or proxy attributes; two users can nevertheless receive similarly polished answers after materially different planning, retrieval, escalation, or memory-handling paths. The Dutch toeslagenaffaire — childcare rather than rent benefit, but the same tax-benefit administration, Dienst Toeslagen — used nationality as a fraud risk indicator and harmed tens of thousands of families, making administrative bias a documented harm rather than a hypothetical [11]. The EU AI Act is relevant in a deployer-dependent way: systems used by or on behalf of public authorities to evaluate eligibility for essential public assistance benefits fall under Annex III, while purely citizen-facing legal-help deployments may fall outside high-risk classification absent another applicable Annex III category. [12].

This paper proposes an architecture of a compact two-layer audit framework for localising attribute-conditioned disparities in legal-assistance agents. Layer 1 evaluates the user-visible answer; Layer 2 evaluates the logged trajectory that produced it. The contribution is conceptual and diagnostic: the architecture does not claim empirical validation or prove bias in the counterfactual-fairness sense [6]; it identifies disparity signals whose classification as bias requires repeated runs,