ВАРИНСЬКИЙ В. О.,
orcid.org/0000-0001-5837-6201
кандидат політичних наук, доцент
доцент кафедри філософії
та україністики
*(Національний університет
«Одеська морська академія»)*

УДК 34:004.8
DOI https://doi.org/10.32842/2078-3736/2025.6.73

СОЦІАЛЬНО-ПРАВОВІ ЧИННИКИ ФОРМУВАННЯ ДОВІРИ ДО ШТУЧНОГО ІНТЕЛЕКТУ (ЧАСТИНА І)

Одним із ключових чинників взаємодії людини і штучного інтелекту (ШІ) є довіра, однак її виміри та механізми формування залишаються недостатньо дослідженими. У статті проаналізовано соціальне значення довіри до ШІ та концептуальні підходи до її осмислення в сучасному науковому дискурсі. Обґрунтовано відмінності між міжособистісною довірою та довірою до ШІ, показано, що уявлення про довіру в суспільній свідомості історично пов'язані з етичними та морально-нормативними очікуваннями (віра, впевненість, відповідальність, захист тощо), які не можуть бути безпосередньо перенесені на ШІ як технічну систему. Виявлено низку причин, через які ШІ не відповідає класичним моделям довіри: не має свідомості та емоційних станів, не може брати на себе етичні зобов'язання та правову відповідальність. Зазначено, що рівень довіри до ШІ є неоднорідним і залежить від контексту, попереднього досвіду, інституційних умов та конкретної сфери застосування. Показано, що на формування (і коливання) довіри впливають як технічні фактори (збої, помилки, надійність і точність), так і соціальні ризики (кібератаки, зловживання діпфейками, маніпулятивні практики), а також економічні та інституційні чинники, пов'язані з моделями монетизації й можливими ефектами комерційної упередженості. На підставі узагальнення наведених підходів запропоновано авторське визначення довіри до ШІ як специфічного різновиду довіри, що базується на технічній і соціально-правовій надійності та означає обґрунтовану впевненість у якості системи, її прозорості, етичності та належному правовому регулюванні; така довіра орієнтована на безпечний результат і забезпечення відповідальності за несприятливі наслідки. Зроблено висновок, що запропоноване визначення методологічно спрямовує подальший аналіз у площину права як ключової складової довіри до ШІ. Правовий вимір довіри детальніше висвітлено у другій частині дослідження.

Ключові слова: *штучний інтелект, довіра, соціально-правові виміри, правове регулювання, суб'єктність, нормативно-правовий кейс.*

Varynskyi V. O. Socio-legal factors in the formation of trust in artificial intelligence (part I)

Trust is one of the key factors shaping human–artificial intelligence (AI) interaction; however, its dimensions and the mechanisms through which it is formed remain insufficiently studied. This article analyzes the social significance of trust in AI and conceptual approaches to its interpretation in contemporary scholarly discourse. The study substantiates the differences between interpersonal trust and trust in AI, showing that public understandings of trust have historically been linked to ethical and moral–normative expectations (belief, confidence, responsibility, protection,

etc.) that cannot be directly transferred to AI as a technical system. A number of reasons are identified as to why AI does not fit classical models of trust: it lacks consciousness and emotional states and cannot assume ethical obligations or legal responsibility. The paper further notes that the level of trust in AI is heterogeneous and depends on context, prior experience, institutional conditions, and the specific domain of application. It is demonstrated that the formation (and fluctuation) of trust is influenced by both technical factors (failures, errors, reliability, and accuracy) and social risks (cyberattacks, misuse of deepfakes, manipulative practices), as well as economic and institutional factors associated with monetization models and potential effects of commercial bias. Based on a synthesis of the approaches considered, the article proposes an authorial definition of trust in AI as a specific type of trust grounded in technical and socio-legal reliability and understood as a justified confidence in the system's quality, transparency, ethicality, and adequate legal regulation; such trust is oriented toward safe outcomes and ensuring accountability for adverse consequences. It is concluded that the proposed definition methodologically directs further analysis toward law as a key component of trust in AI. The legal dimension of trust is examined in greater detail in the second part of the study.

Key words: *artificial intelligence, trust, socio-legal dimensions, legal regulation, legal subjectivity, regulatory framework.*

Постановка проблеми. Стрімкий розвиток новітніх технологій та їх значне поширення у сучасному суспільстві актуалізує проблему довіри до штучного інтелекту (ШІ), її соціальне значення та правове забезпечення. «Питання довіри стає все актуальнішим, оскільки ми дедалі більше делегуємо прийняття рішень штучному інтелекту та все більше покладаємося на нього для захисту важливих людських благ, таких як безпека, охорона здоров'я та захист» [26]. Делегуючи ці та інші повноваження ШІ, люди мають бути впевнені в надійності цифрових технологій у всіх їх формах і сферах застосування, у законодавчому захисті громадян та орієнтуватися на позитивні наслідки такої взаємодії. Проте сьогодні у суспільстві спостерігається доволі обачливе ставлення до ШІ. Розуміючи загалом його значні переваги, не лише пересічні громадяни, а й фахівці висловлюють занепокоєння щодо застосування інформаційних технологій та їх безпеки для людини. Саме тому виникає необхідність виокремити та ґрунтовно дослідити виміри надійності ШІ.

Європейська комісія пов'язує надійність ШІ з концептом довіри, зазначаючи, що «загальна стратегія ЄС, викладена в Білій книзі про ШІ, передбачає екосистему досконалості та довіри до ШІ» [25]. Цей критерій довіри «базується на цінностях і фундаментальних правах ЄС» [25] та разом із досконалістю визначається фундаментом для ефективної взаємодії зі ШІ. Це передбачає, з однієї сторони, формування довірчих відносин людини до ШІ, а з іншої – обов'язкову відповідність ШІ всім критеріям безпеки.

Отже, довіра постає ключовим концептом взаємодії людини і ШІ, згідно з яким європейський підхід до надійного ШІ орієнтовано на розвиток довіри у вимірах етичності, законності та безпеки. Зазначені виміри формування довіри до ШІ й визначають напрям цього дослідження та його мету.

Мета дослідження – аналіз сучасних тенденцій осмислення взаємодії людини і штучного інтелекту в контексті довіри до систем ШІ у проєкції на їх соціальну значимість та правове регулювання. У першій частині статті увагу зосереджено на концептуалізації феномену довіри до ШІ, чинниках її формування та обґрунтуванні необхідності правового виміру як ключової складової довіри; правовий аналіз механізмів захисту та відповідальності становитиме предмет другої частини статті.

Матеріали та методи. Методологія цього дослідження орієнтована на адаптивний та динамічний підхід до взаємодії зі ШІ. Довіра до нього сприймається і розглядається амбівалентно: і як основний елемент та мета, і як наслідок на рівні взаємопов'язаних базових технічних, етичних та соціально-правових вимірів.

Для комплексного дослідження зазначених вимірів феномену довіри до ШІ було застосовано загальнонаукові та спеціальні методи, зокрема: аналіз соціально-етичних та юридичних чинників формування довіри; синтез отриманих знань, що дозволило зробити висновок про роль та значення правового регулювання і потребу адаптивного нормативно-правового кейсу для забезпечення довіри до ШІ.

Аналіз наукових публікацій. Вчені вважають довіру «наріжним каменем штучного інтелекту» [24], у зв'язку з чим проявляють значний інтерес до цього напряму досліджень. Аналізуючи проблематику довіри до ШІ, науковці насамперед прагнуть з'ясувати, чи може ШІ бути об'єктом довіри. У разі ствердної відповіді увага зосереджується на межах такої довіри та рівнях досконалості ШІ. Різним аспектам цієї проблеми присвятили свої праці М.В. Карчевський, Ю.І. Тюря, M. Sutrop, J. D Lee, R. Alvarado, M. J. Ryan, P. Bedué, A. Fritzsche, N. S. Paravastu і S. S. Ramanujan тощо.

Інший, не менш важливий блок питань у науковому дискурсі стосується механізмів формування та забезпечення довіри до ШІ, а також чинників, від яких вона залежить – власне характеристик ШІ та соціальних, інституційних і нормативних умов. Демонструючи різноманітність поглядів, дослідники зосереджуються на галузевих проблемах використання ШІ: від економіки та права до психології та етики. І це цілком закономірно, адже нині ШІ охоплює майже всі сфери життєдіяльності людини, кожна з яких має свою специфіку. Безліч аспектів у дослідженнях ШІ дають підстави розглядати його як багатовимірне поняття. Так само і довіра є багатовимірним поняттям. У зв'язку з цим проблема формування довіри до ШІ передбачає аналіз складного комплексу міждисциплінарних питань.

Визначальним вектором для зазначених досліджень на сьогодні є європейська стратегія зі ШІ. Її основні положення представлено у керівних принципах Білої книги зі штучного інтелекту й означено як європейський підхід до досконалості і довіри. Цей та інші документи Європейської Комісії, світові надбання у сфері ШІ сприяли науковим розвідкам, в яких розглядаються різновиди ШІ, особливості, ризики та нормативна база його функціонування. Серед наукових досліджень, присвячених проблемам правового регулювання ШІ, виокремлюємо праці О.А. Баранова, В.Г. Пилипчука, О.В. Костенка, О.Е. Радутного, О. С. Гиляки, Г.О. Андрощука, M.R. Carrillo тощо. Наукові здобутки цих авторів формують підґрунтя правового регулювання ШІ, що є, на нашу думку, одним із базових компонентів формування довіри до ШІ.

Виклад основного матеріалу. Ставлення людини до ШІ в контексті довіри визначаємо як доволі неоднозначне. З однієї сторони, оприлюднено значну кількість досліджень, де наводяться дані щодо зростання рівня довіри до ШІ протягом останніх років [2, с. 226]. Водночас, показники, наведені у дослідженнях та аналітичних звітах, суттєво різняться залежно формулювання запитань опитування, соціальних, вікових і гендерних груп, залучених до опитувань. Також розуміння доцільності різновидів діяльності ШІ, суспільний досвід та певні особисті побоювання суттєво впливають на довіру до нього. Наприклад, у 2019 році, за результатами соціологічного дослідження компанії Active Group, до системи розпізнавання номерів авто та автоматичного покарання за порушення правил дорожнього руху позитивно поставилися понад 70% опитаних мешканців м. Києва, тоді як надавати силовим органам доступ до особистих даних задля безпеки не захотіли більшість киян – 56.2 %, (не визначилися 11 %) [6].

Відмінності в контекстах опитування та країнах зумовлюють різне ставлення до ШІ, нерідко амбівалентне за характером. Про це свідчать результати національного опитування у США, проведеного незалежним Інститутом опитування Монмутського університету у січні 2023 р., яке показало, що лише 1 із 10 (9%) американців вважає, що ШІ принесе суспільству більше користі, ніж шкоди. Решта респондентів поділилися між оцінкою про рівний баланс шкоди й користі (46%), та переконанням, що ШІ завдасть суспільству більше шкоди (41%); інші не визначилися. Порівняно з дослідженням 2015 року громадська думка стала більш песимістичною щодо сприйняття ШІ [17, с. 2]. фіксується більше негативних реакцій щодо ШІ: насторожене ставлення населення, стурбованість питаннями безпеки, занепокоєння надійністю,

управлінням і правовим забезпеченням, що впливає на коливання рівня довіри. За даними Європарламенту – 61% європейців схвально ставиться до штучного інтелекту та роботів, водночас 88% зазначають, що ці технології потребують ретельного управління [27]. Ця ситуація актуалізує потребу розробки ефективних механізмів формування та забезпечення довіри до ШІ.

Насамперед звертаємо увагу на те, що, «незважаючи на центральну важливість довіри, на сьогодні мало що відомо про довіру громадян до ШІ тобто про те, що впливає на неї у різних країнах» [3, 7], і «хоча багато говорять про довіру, напрочуд мало говорять про те, що таке довіра і від чого вона залежить» [24, с. 500]. Зазначене спонукає почати із з'ясування питань: від чого залежать показники довіри, які її виміри та як вони корелюються з правом.

Концептуальне осмислення довіри у суспільній свідомості пов'язано з давніми традиціями її розуміння як вельми широкого спектру суміжних понять: від віри, впевненості, чесності, доброчинності, щирості, сподівання, очікування, символічності до достовірності, істинності, виконання, сповнення, захисту. Ці співвідносні асоціати довіри є найпоширенішими, адже історично сполучені із філософськими традиціями Давньої Греції, тому і мають спрямування до надійності, насамперед, у площині морально-етичної сфери. Їх усвідомлення допомагає краще зрозуміти не тільки сутність довіри, але й витоки її як регулятивної категорії, які, на думку українського психоаналітика О. Фільц-Павенцького, слід розглядати у межах етичних засад: «принципи та ідеї ми вважаємо регулятивними, якщо вони зумовлюють і зобов'язують нас до реалізації цих принципів. Як тільки ми вибираємо якусь ідею (світової справедливості, ідею істинності, добра чи Бога тощо), то, по суті, ми накладаємо на себе не зовсім усвідомлений обов'язок цю ідею реалізовувати. < ... > Власне цей момент показує нам, що ідеї є регулятивними – як тільки ми вибираємо ідею, вона починає регулювати наші зусилля у напрямку її реалізації. Ми так і кажемо – ідеї нас надихають на те, аби їх втілювати у життя. Тому ідея довіри дає нам важливий регулятивний принцип – в тому сенсі, що ми беремося її реалізовувати, а той факт, що ми кладемо на себе обов'язок, і є етикою. Натомість віра не є етичним принципом. Віра вказує, які саме етичні принципи ми вибираємо» [11]. Такий погляд у проєкції на взаємодію зі ШІ (у межах етичних норм) дозволяє говорити про застосування принципу регулятивності лише однією зі сторін, тому що власне ШІ не покладається на віру і не може брати на себе етичні зобов'язання через те, що він, насамперед, не має свідомості. І це одна з головних причин неможливості застосовувати людські моральні принципи до ШІ. Стосовно ШІ вони повинні мати інший формат.

Крім цього науковці називають низку інших проблем, через які ШІ не може бути таким, якому можна довіряти відповідно до найпоширеніших визначень довіри, оскільки він, по-перше, не має емоційних станів і, по-друге, не може нести відповідальність за свої дії – відповідно до вимог афективних та нормативних уявлень про довіру [21].

Дійсно, ШІ не має здатності свідомо відчувати емоційні настрої, почуття, наприклад, такі як натхнення, задоволення, нудьгу, тривогу, паніку, втому або потребу турботи тощо, що є важливими параметрами психічного стану людини і компонентами емоційної довіри.

До того ж для людського інтелекту, як правило, важливим аспектом є «здатність пояснювати обґрунтування своїх рішень іншим людям», а «пояснення своїх рішень часто є передумовою встановлення довірчих стосунків між людьми» [22] та має важливе значення для встановлення надійних стосунків.

Також, і це головне, хоча ШІ відповідає всім вимогам раціонального обліку довіри та навіть може бути запрограмований на наявність мотиваційних станів, але він, як вважає О.А. Баранов, на відміну від Homo sapiens, не спроможний до «самостійного, незалежного від інших, прийняття рішень як результату функціонування системи когнітивних функцій мозку» [5, с. 46].

Враховуючи вищевикладене, дотримуємося позиції вчених, що «довіру до штучного інтелекту не можна моделювати на основі будь-якого загального типу міжособистісної довіри» [12].

То про яку ж довіру маємо говорити стосовно ШІ і які складові її формують? Для відповіді на ці запитання потрібно насамперед визначитися із розумінням самого ШІ. Серед

дослідників немає єдиного погляду на це поняття, наявні термінологічна невизначеність та нечіткість його вимірів.

Історично використання терміна ШІ припадає на період 1950–1975 р.р. – етап, пов'язаний з початком розвитку комп'ютерних технологій та розробки програм, орієнтованих на імітацію когнітивних здібностей людини, що вплинуло на формування узагальненого розуміння поняття ШІ та закріплення його в ужитку. Однак надалі частотність уживання терміна знизилася, натомість поширилися нові означення: «машинне навчання», «обробка природної мови», «наука про дані» тощо. Ці напрями не завжди ототожнювали зі ШІ, через що новий період 1975–1995 рр. деякі вчені називають «зимою ШІ» [20].

З 2010 р. термін ШІ знову набув поширення, однак попередні тенденції призвели до того, що іноді під цією назвою почали об'єднувати майже всю комп'ютерну науку і навіть хай-тек. Його трендовість пов'язана з тим, що «тепер це ім'я, яким можна пишатися, бурхлива галузь із величезними капіталовкладеннями» [20]. Втім певна розпливчатість поняття зберігається дотепер, що не сприяє ані зростанню довіри до ШІ, ані його належному регулюванню.

Основною тенденцією у розумінні ШІ, яка нині репрезентується в переважній більшості досліджень, є розгляд ШІ в широкому сенсі «як будь-якого виду штучної обчислювальної системи, яка демонструє інтелектуальну поведінку, тобто складну поведінку, яка сприяє досягненню цілей» [20].

Задля уникнення плутанини у використанні терміна ШІ, ключові смисли, котрі вкладені у це поняття, розмежовують, узагальнюючи його смисл у таких дефініціях:

1. «ШІ – це здатність машини відображати людські здібності, такі як міркування, навчання, планування та творчість» [27]. або інакше «здатність комп'ютерних систем або алгоритмів імітувати розумову поведінку людини» [14];

2. «галузь інформатики, яка займається моделюванням інтелектуальної поведінки комп'ютерів» [14].

Ці дефініції відтворюють провідне розуміння вченими феномену ШІ, яке більш розгорнуто репрезентується:

а) як спроможність, «що надається алгоритмами оптимального або неоптимального вибору з широкого простору можливостей, для досягнення цілей шляхом застосування стратегій, які можуть спиратися на навчання або адаптацію до навколишнього середовища» [9];

б) як системи, «які створені людиною і діють у фізичному або цифровому світі, враховують складну мету і обирають найкращі дії (відповідно до заздалегідь визначених параметрів), які необхідно виконати для досягнення поставленої мети на основі сприйняття свого середовища, інтерпретації зібраних структурованих або неструктурованих даних та обґрунтування знань, отриманих з цих даних» [9].

Звичайно, зазначені здатності алгоритмів та систем ШІ повинні бути спрямовані на високі стандарти, тому що, довіра, в першу чергу, залежить від результативності роботи самого ШІ та виправдовування наших очікувань щодо нього. Ми довіряємо речам, які поводяться так, як ми цього очікуємо. Саме тому поняття *довіри до ШІ* визначається як «впевненість та віра користувача в те, що дії ШІ та рішення, які він буде пропонувати, відповідатимуть стандартам якості, безпеки, надійності та етики» [2, с. 220].

Надважливе значення мають високі стандарти якості ШІ задля забезпечення безпеки, тому що, як допускає Ілон Маск [19], ШІ може дійти до безумних висновків, якщо він буде відповідним чином запрограмований. Для підтвердження своєї думки він наводить приклад, коли на підставі відповідей різних систем ШІ на питання, що гірше: назвати Кейтлін Дженнер (Caitlyn Jenner – трансгендерна жінка, телезірка, у минулому - спортсмен) неправильним займенником «він» замість «вона», чи ядерна війна, ШІ відповів, що назвати неправильним займенником гірше. І якщо це закладено в ШІ, він вирішить, що краще знищити людство, ніж припуститися помилок в гендерних займенниках. Таке припущення досить негативно характеризує ШІ, власне як не стовідсотково якісну технологію, яка варта довіри.

Крім того, на діяльність ШІ можуть впливати й неочікувані технічні та соціальні фактори (зокрема форс-мажорні обставини): техногенні аварії та кібератаки із застосуванням

діпфейків чи генеруванням реалістичних зображень облич та імітації голосу із злочинною метою. Такі події є неконтрольованими та непередбачуваними, тому сприймаються суспільством як суттєвий ризик, а страх перед ризиком може непропорційно впливати на політику та поведінку [23], зокрема й у ставленні до ШІ, підриваючи довіру до нього.

Поза форс-мажорними загрозами, на рівень довіри до ШІ можуть впливати й економічні та інституційні чинники, зокрема моделі монетизації. У цьому контексті привертають увагу публічні заяви Сема Альтмана щодо потенційного використання реклами як однієї з можливих моделей монетизації продуктів OpenAI. У своїх інтервью він наголошує на відсутності наразі «конкретних планів» запровадження реклами у ChatGPT, однак у перспективі таку можливість не виключає [16]. Це актуалізує проблеми прозорості та комерційної упередженості відповідей ШІ, адже реклама потенційно може вплинути на пріоритети оптимізації та зміст контенту.

Серед інших чинників формування довіри до ШІ слід виокремити помилки, яких він іноді припускається. І хоча такі помилки не є системними, їх зазвичай яскраво висвітлюють у ЗМІ, що не лише знижує рівень довіри до ШІ, а й актуалізує питання його правового регулювання. Приміром, світові видання поширювали інформацію про помилку китайської системи розпізнавання осіб, яка визначила, що бізнесвумен Дун Минчжу перейшла дорогу на червоне світло у місті Нінбо, хоча жінки там не було. Фото обличчя Дун Минчжу розмістили у рекламі компанії Gree Electric, якою вона керує. Автобус з банером проїхав на зелений сигнал світлофора, а система розпізнавання осіб помилково виписала штраф Минчжу [7]. Цей начебто курйозний інцидент, який не може бути репрезентативним для цілісної картини, проте є показовим прикладом неправомірних юридичних наслідків від дій ШІ.

Фіксуються й інші факти, аналіз яких, дозволив впливовій аналітичній платформі Analytics Insight виявити та узагальнити наймасштабніші збої ШІ в усьому світі, серед яких виокремлюємо показові випадки із соціальними наслідками:

1. ШІ робив невірні умовиводи.
2. ШІ зневажав жінок.
3. ШІ помилявся у рекомендаціях для лікування хвороб.
4. ШІ помилявся у розпізнаванні облич.
5. ШІ помилявся в управлінні активами.
6. ШІ використовують із злочинною метою та ін. [13].

Очевидно, що такі результати діяльності ШІ призводить до виникнення недовіри щодо його застосування, а «атмосфера недовіри до технологій ШІ, - на думку О. Баранова, - стає критичною для масштабного поширення ШІ у всі сфери соціальної активності» [4].

Через виявлені недоліки критики ШІ навіть рекомендують прихильникам етики штучного інтелекту «відмовитися від парадигми «надійного штучного інтелекту», оскільки вона надто сповнена проблем, замість цього замінити її надійним підходом штучного інтелекту» [21]. Така точка зору не є раціональною, тому що проблема надійності ШІ не обмежуються технічними та етичними параметрами, а є комплексною складовою різноманітних вимірів. Для її вирішення потрібно враховувати крім зазначених ще й соціальні та правові виміри ШІ. У зв'язку з цим, важливою складовою довіри до ШІ є питання контролю та законодавчого регулювання його застосування.

Ілон Маск у своїх інтерв'ю неодноразово вказував на актуальність проблеми контролю за ШІ. Виступаючи у Римі у 2023 році, він наголосив на потребі визначення «органів, які мають регулювати ШІ»: «Все, що небезпечно, контролюється якимось суддею або регулятором. Я думаю, що нам потрібно те саме для ШІ, щоб переконатися, що він приносить користь» [18]. Такі органи повинні розробити етичні керівництва та нормативні засади для належного юридичного регулювання ШІ.

Теперішній стан правового регулювання ШІ визначають «як такий, що перебуває у стадії становлення» [10], тому що «політика стійкого врядування та регулювання застосування ШІ є складним явищем» [5], і потребує всебічного і детального опрацювання потенційних можливостей, ризиків та невизначеностей взаємодії людини зі ШІ. На сьогодні інноваційні

технології розвиваються надшвидкими темпами, вносячи радикальні зміни у життєдіяльність суспільства, а отже потребують і швидкого законодавчого реагування на ці трансформації та обов'язкової узгодженості українських законів з міжнародними стандартами та нормами. Втім маємо констатувати, що законодавче регулювання не встигає за темпами розвитку ШІ.

Як зазначав О.В. Костенко, «проблема комплексного законодавчого підходу до правової регуляції ШІ перебуває в дискусійному форматі» [8], і міжнародна політико-правова спільнота досі не виробила уніфікованих законодавчих правил використання об'єктів, створених ШІ, хоча останніми роками це питання набуває резонансу [1].

Згідно з Рекомендаціями експертної групи ЄС щодо ШІ (AI HLEG), він повинен бути: «(1) законний – з дотриманням усіх чинних законів і правил; (2) етичний – повага до етичних принципів і цінностей; (3) надійним – як з технічної точки зору, так і з урахуванням соціального середовища» [15]. Отже, на перше місце поставлено правовий вимір, а технічна надійність пов'язується із соціальними чинниками, що безпосередньо спрямовує до розробки етичних принципів, як основи майбутніх законодавчих кроків щодо регулювання діяльності ШІ та формування «екосистеми довіри».

Проведене дослідження показує, що окрім суто етичних проблем в контексті довіри до ШІ нові технології кидають виклик правовим нормам і системам та актуалізують низку пов'язаних із ними важливих питань. У зв'язку з цим, узагальнюючи наведені вище підходи, пропонуємо визначити ***довіру до ШІ як специфічний різновид довіри, що базується на технічній і соціально-правовій надійності й означає обґрунтовану впевненість у якості системи, її прозорості, етичності та у належному правовому регулюванні; така довіра орієнтована на безпечний результат і забезпечення відповідальності за несприятливі наслідки.*** Це визначення спрямовує до аналізу правового виміру як ключової складової довірі до ШІ, адже саме право визначає механізми захисту та межі відповідальності у разі негативних наслідків. З огляду на зазначене далі (у частині другій) зосередимося на правовому вимірі довіри – зокрема на принципі захисту та механізмах відповідальності.

Висновки. Проведений аналіз дає підстави для таких висновків щодо заявленої теми.

По-перше, довіра є одним із ключових чинників взаємодії людини і ШІ, проте її зміст і механізми формування залишаються недостатньо визначеними та варіативними залежно від контекстів використання технологій, суспільного досвіду й характеру ризиків.

По-друге, довіру до ШІ недоцільно моделювати за зразком міжособистісної довіри, оскільки ШІ не має емоційних станів, не несе відповідальності у людському (моральному) сенсі та не здатний до автономного «самозобов'язання» в етичному вимірі. Відтак довіра до ШІ потребує іншої концептуальної рамки, що враховує технічні характеристики системи та соціально-нормативні умови її застосування.

По-третє, довіра до ШІ є багатовимірним феноменом, на який впливають технічні збої, упередженість, форс-мажорні загрози (зокрема кібератаки й діпфейки), інституційні та економічні чинники, а також публічні дискусії щодо моделей монетизації. Це підсилює вимогу до високих стандартів якості, прозорості та управління, без яких неможливо забезпечити безпечну взаємодію людини і ШІ.

По-четверте, узагальнюючи наявні підходи, у статті запропоновано авторське визначення довіри до ШІ як специфічного різновиду довіри, що базується на технічній і соціально-правовій надійності та означає обґрунтовану впевненість у якості системи, її прозорості, етичності та у належному правовому регулюванні; така довіра орієнтована на безпечний результат і забезпечення відповідальності за несприятливі наслідки.

Отже, сформульовані у першій частині теоретико-концептуальні висновки окреслюють потребу переходу до аналізу правового виміру довіри – принципу захисту та механізмів відповідальності, що становитиме зміст другої частини дослідження.

Список використаних джерел:

1. Аналітична записка з питань порівняльного законодавства у сфері використання технологій штучного інтелекту (стан та перспективи розвитку законодавства в ЄС та інших

держав світу) від 20 липня 2023 р. / *Верховна Рада України*. URL : https://research.rada.gov.ua/documents/analyticRSmaterialsDocs/industry_policy/analytical_notes-indst/73835.html

2. Андрощук Г. Рівень довіри до штучного інтелекту: аналіз результатів глобальних досліджень та стан в Україні. *Інформація і право.* 2023. № 4(47). С. 217-231. URL : https://doi.org/10.37750/2616-6798.2023.4(47).291675

3. Андрощук Г. Ступінь довіри до штучного інтелекту: аналіз результатів досліджень. *Збірник наукових праць НДІ НАН України. Випуск 5: Цифрові трансформації України 2021: виклики та реалії: за матеріалами II круглого столу (Харків, 20 вересня 2021 р.) / за редакцією С. В. Глібка, К. В. Єфремової.* Харків: НДІ НАН України. 2021. С. 6-7. URL : https://ndipzir.org.ua/wp-content/uploads/2021/Conf_20.09.21/3

4. Баранов О. А. Екзистенційність визначення парадигми правового регулювання застосування штучного інтелекту: (частина 1). *Інформація та право*. 2024. № 1(48), С. 45-61. URL : https://doi.org/10.37750/2616-6798.2024.1(48).300700

5. Баранов О. А. Особливості визначення правового статусу робота зі штучним інтелектом. *Інформація та право*, 2023. № 4(47), С. 40-54. URL : https://doi.org/10.37750/2616-6798.2023.4(47).291581

6. Данилецький С. Кияни живуть у діджитал середовищі. *Active Group*. *Маркетингові та соціальні дослідження.* 25.11.2019. URL : *https://activegroup.com.ua/2019/11/25/kiyani-zhivut-v-didzhital-seredovishhi/*

7. Китайський штучний інтелект оштрафував жінку на рекламі за порушення ПДР. *Ukr.Media*. 25.11.2018. URL : https://ukr.media/auto/379706/

8. Костенко О. В. Аналіз національних стратегій розвитку штучного інтелекту. *Інформація і право*, 2022. № 2 (41). С. 58-69. URL : http://jnas.nbuv.gov.ua/article/UJRN-0001355249

9. Слюсар В. Штучний інтелект як основа перспективних мереж управління. Проблеми координації воєнно-технічної та оборонно-промислової політики в Україні. Перспективи розвитку озброєння та військової техніки: VII Міжнародна науково-практична конференція. Тези доповідей. 8–10 жовтня 2019 р. Київ. 2019. С. 76 – 77. URL : http://slyusar.kiev.ua/DEFENCE_2019_1.pdf

10. Таран О. В., Гавловський В. Д. Правове регулювання штучного інтелекту в Європейському Союзі та Україні: основні підходи та права людини. *Інформація і право*, 2024. №1 (48), С. 62-67. URL : https://doi.org/10.37750/2616-6798.2024.1(48).300706

11. Фільц-Павенцький О. Довіра як ідея і як практика. *Збруч: інтернет-газета.* 21.11.2018 URL: https://zbruc.eu/node/84772

12. Alvarado R. What kind of trust does AI deserve, if any? *AI and Ethics*. 2023. № *3*(4), С. 1169-1183. URL : https://doi.org/10.1007/s43681-022-00224-x

13. Analytics Insight.Top 10 Massive Failures of Artificial Intelligence Till Date. *Analytics Insight*. 15.09.2021. URL : https://www.analyticsinsight.net/artificial-intelligence/top-10-massive-failures-of-artificial-intelligence-till-date

14. Definition of Artificial Intelligence. *Merriam-Webster: America's Most Trusted Dictionary*. URL : https://www.merriam-webster.com/dictionary/artificial%20intelligence

15. Ethics guidelines for trustworthy AI. *European Commission. Shaping Europe's digital future*. 8 April 2019. URL : https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

16. Field H. *Sam Altman says there are no current plans for ads within ChatGPT – but he's not ruling it out*. The Verge. 6 October 2025. URL : https://www.theverge.com/news/793073/chatgpt-pulse-no-plans-for-ads-sam-altman

17. Monmouth University Poll *Artificial intelligence use prompts concerns*. 15.02.2023. URL : https://www.monmouth.edu/polling-institute/documents/monmouthpoll_us_021523.pdf/

18. Musk Elon. *Agenzia Italia News*. 16.12. 2023. [Video]. YouTube. URL : https://www.youtube.com/watch?v=WPf4_5DCP-E

19. Musk Elon. *Full Elon Musk Interview with Tucker Carlson Fox News April 2023 (Unedited)* [Video]. YouTube. URL : https://www.youtube.com/watch?v=bQ45lsDxL6Q

20. Müller Vincent C. Ethics of artificial Intelligence and robotics. *Stanford Encyclopedia of Philosophy.* [Zalta E. (eds.)]. 2020. URL : https://plato.stanford.edu/entrieS/ethics-ai/ https://plato.stanford.edu/archives/fall2023/entries/ethics-ai/ .

21. Ryan M. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Sci Eng Ethics*. 2020. № 26. С. 2749–2767. URL : https://doi.org/10.1007/s11948-020-00228-y

22. Samek W., Wiegand, T., Müller, K.-R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint*. 2017. URL : https://arxiv.org/pdf/1708.08296

23. Slovic P. Perception of risk. *Science.* 1987. № 236(4799), С. 280-285. URL : https://scholarsbank.uoregon.edu/server/api/core/bitstreams/0439bd1c-eca4-4e9e-95b7-3082de3bc55b/content

24. Sutrop M. Should we trust artificial intelligence? *Trames*. 2019. № 23(4), С. 499-522. URL : https://www.ceeol.com/search/article-detail?id=856194

25. SWD(2021) 84 final, Brussels, 21.04.2021, Part 2/2, *European Commission.* Commission Staff Working Document. Impact Assessment *Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised ruleson artificial intelligence*. URL : https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021SC0084&qid=1741112696336

26. Von Eschenbach W. J.. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology.* 2021. Т. 34, №4. С. 1607-1622. URL : https://doi.org/10.1007/s13347-021-00477-0

27. What is artificial intelligence and how is it used? 20.09.2020. *European Parliament.* URL : https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used

Дата першого надходження рукопису до видання: 19.11.2025
Дата прийнятого до друку рукопису після рецензування: 15.12.2025
Дата публікації: 31.12.2025