

Філософія

УДК 004.8:17:004.946

DOI <https://doi.org/10.5281/zenodo.15616278>

**Етичні дилеми взаємодії людини та штучного інтелекту у віртуальних
просторах метавсесвіту**

Поліщук Наталія Володимирівна,

доктор філософських наук, доцент, професор кафедри філософії,
факультет історії, політології та міжнародних відносин,
Рівненський державний гуманітарний університет,
м. Рівне, Україна, <https://orcid.org/0000-0001-9498-4837>

Поперечна Галина Антонівна,

кандидат філософських наук, доцент, доцент кафедри
філософії та суспільних наук, історичний факультет, Тернопільський
національний педагогічний університет імені Володимира Гнатюка,
м. Тернопіль, Україна, <https://orcid.org/0000-0003-2191-9908>

Ніколенко Ксенія Вікторівна,

кандидат філософських наук, доцент, кафедра цивільно-правових та
соціально-гуманітарних дисциплін, Донецький державний університет
внутрішніх справ, Криворізький навчально-науковий інститут, факультет № 2,
м. Кривий Ріг, Україна, <https://orcid.org/0000-0002-2091-7823>

Прийнято: 23.05.2025 | Опубліковано: 07.06.2025

Анотація. Актуальність дослідження зумовлена стрімким поширенням технологій штучного інтелекту в межах віртуальних середовищ, що поступово набувають рис метавсесвіту. За таких умов виникає потреба у переосмисленні етичних засад цифрової поведінки, оскільки традиційні моральні моделі є недостатніми для регулювання взаємодії з автономними штучними системами, що діють у межах соціально значущого контексту. Мета дослідження полягає у комплексному аналізі етичних параметрів взаємодії людини з інтелектуальними алгоритмами у середовищі метавсесвіту, з урахуванням цифрової суб'єктності, автономності систем та моральної відповідальності в умовах віртуальної взаємодії. Методологію дослідження побудовано на міждисциплінарному підході, що поєднує філософію технологій, цифрову етику, когнітивну теорію, а також елементи прикладного аналізу алгоритмічної поведінки. Використано логіко-аналітичний метод, структурно-функціональний аналіз, типологізацію форм взаємодії та порівняльний аналіз етичних моделей. Результати дослідження викладено у вигляді типології форм взаємодії людини зі штучним інтелектом (асистивна, симуляційна, персоналізована, емпатична та агентивна), визначено особливості кожної з них та супутні етичні ризики. Досліджено застосування етичних моделей — деонтології, утилітаризму, етики чеснот, етики відповідальності та реляційної етики — в умовах метавсесвітнього середовища. Виявлено ризики моральної дезорієнтації, втрати автономії користувача, маніпулятивного впливу з боку ШІ та викривлення цифрової ідентичності. У висновках обґрунтовано необхідність міждисциплінарного регулювання алгоритмічної поведінки, визначено ключові напрями для етичного конструювання віртуального досвіду та доведено, що без інтеграції гуманітарних підходів у сферу цифрових технологій ризики етичної ерозії лише зростатимуть. Перспективи подальших досліджень вбачаються у розробці емпіричних моделей для оцінювання впливу штучного інтелекту на

моральну автономію користувача, створенні алгоритмів адаптивного етичного контролю та формуванні нових нормативних рамок для метавсесвітніх платформ.

Ключові слова: *цифрова ідентичність, моральна відповідальність, алгоритмічна автономія, когнітивна взаємодія, інформаційна прозорість, адаптивна згода.*

Ethical dilemmas of human–artificial intelligence interaction in the virtual spaces of the metaverse

Nataliia Polishchuk,

Doctor of Philosophical Sciences, Associate Professor, Professor of the
Department of Philosophy, Faculty of History, Political Science and
International Relations, Rivne State Humanitarian University,
Rivne, Ukraine, <https://orcid.org/0000-0001-9498-4837>

Galyna Poperechna,

Candidate of Philosophical Sciences, Associate Professor of the
Department of Philosophy and Social Sciences, Faculty of History,
Ternopil Volodymyr Hnatiuk National Pedagogical University,
Ternopil, Ukraine, <https://orcid.org/0000-0003-2191-9908>

Kseniia Nikolenko,

PhD in Philosophical Sciences, Associate Professor, Department of Civil
Law and Social and Humanitarian Disciplines, Donetsk State University of
Internal Affairs, Kryvyi Rih Educational and Scientific Institute, Faculty № 2,
Kryvyi Rih, Ukraine, <https://orcid.org/0000-0002-2091-7823>

Abstract. *Relevance of the research is determined by the rapid expansion of artificial intelligence technologies within virtual environments, which are gradually evolving into the metaverse. Such developments necessitate a rethinking of ethical foundations, as traditional moral models are insufficient to regulate interactions with autonomous artificial systems acting within socially significant contexts. The aim of the research is to conduct a comprehensive analysis of the ethical dimensions of human interaction with intelligent algorithms in the metaverse, considering digital subjectivity, system autonomy, and moral responsibility in virtual contexts. The research methodology is based on an interdisciplinary framework combining the philosophy of technology, digital ethics, cognitive theory, and applied analysis of algorithmic behavior. Logical-analytical methods, structural-functional analysis, typology of interaction forms, and comparative analysis of ethical models are employed. The results of the study include the development of a typology of human–artificial intelligence interaction in the metaverse (assistive, simulative, personalized, empathic, and agentive) and the identification of their specific ethical risks. The application of deontological ethics, utilitarianism, virtue ethics, the ethics of responsibility, and relational ethics in metaversal environments is explored. Risks such as moral disorientation, loss of user autonomy, manipulative influence, and distortion of digital identity are identified. The conclusions justify the need for interdisciplinary regulation of AI-driven systems, identify key areas for ethical construction of digital experience, and demonstrate that without the integration of humanities-based approaches, the risk of ethical erosion will continue to grow. Future research prospects include the development of empirical models to assess artificial intelligence's impact on user moral autonomy, the creation of adaptive ethical control algorithms, and the formulation of normative frameworks for metaverse platforms.*

Keywords: *digital identity, moral responsibility, algorithmic autonomy, cognitive interaction, informational transparency, adaptive consent.*

Постановка проблеми. Інтенсивний розвиток технологій віртуальної реальності, розширеної реальності та штучного інтелекту (ШІ) призвів до формування нових типів взаємодії між людиною і цифровими агентами у віртуальних просторах, що поступово трансформуються у концепцію метавсесвіту. У цих середовищах відбувається стирання меж між фізичним та цифровим простором, що породжує низку етичних дилем, пов'язаних із автономністю рішень ШІ, маніпуляцією поведінкою користувачів, збереженням особистої ідентичності, а також питаннями приватності, згоди та відповідальності. Проблематика полягає в тому, що чинні нормативні, правові та моральні підходи до етики взаємодії у віртуальних середовищах залишаються фрагментарними та не здатні повністю оцінити складність ситуацій, що виникають у взаємодії користувача з алгоритмічними системами, які можуть мати ознаки автономної поведінки або віртуальної особистості. З наукової точки зору ця проблема перетинається із завданнями філософії свідомості, етики ШІ, когнітивної науки та прав людини, адже виникає необхідність переосмислення базових уявлень про суб'єктність, свободу волі та моральну відповідальність у контексті взаємодії з нелінійними цифровими системами. З практичного боку важливим є забезпечення технологічної прозорості, розроблення етичних протоколів взаємодії, впровадження принципів гуманного дизайну, а також створення нормативного поля, що захищає користувачів від потенційного зловживання ШІ у середовищах, які відчутно впливають на емоційний і поведінковий стан особи. У цьому контексті вивчення етичних дилем взаємодії людини та ШІ у метавсесвіті є не лише теоретично значущим, а й необхідним для створення безпечного, інклюзивного та відповідального цифрового майбутнього.

Аналіз останніх досліджень і публікацій. Аналіз сучасних досліджень з етичних дилем взаємодії людини та ШІ у віртуальних просторах метавсесвіту дає можливість виокремити чотири основні змістові напрями. Перший напрям

стосується формування правових підходів до етичної поведінки ІІІ у цифровому середовищі. У роботі О. Kostenko, D. Zhuravlov, O. Dniprov та O. Korotiuk обґрунтовано модель кримінального коду для метавсесвіту, яка пропонує впровадження цифрової відповідальності для аватарів і штучних агентів [1]. Аналогічно, В. В. Мангора та О. В. Костенко аналізують аватари як об'єкти правового регулювання, висвітлюючи потребу в етичних межах їхньої поведінки у віртуальному просторі [2]. У дослідженні О. В. Костенко [3] розглядаються концепти електронної юрисдикції та делегування правосуб'єктності нейронним мережам, що викликає нові етичні суперечності. На цьому етапі варто посилити дослідження моделей цифрової деліктоздатності та алгоритмів запобігання зловживанням у межах автономних систем.

Другий напрям зосереджений на концептуалізації соціальних і моральних ризиків у взаємодії користувачів із ІІІ у метавсесвіті. R. Benjamins, Y. Rubio Viñuela та C. Alonso [4] окреслюють ключові виклики, пов'язані з моральною дезорієнтацією, імітацією соціальної поведінки та втратою приватності. R. Romano [5] вказує на загрозу ерозії міжособистісної довіри через використання ІІІ для моделювання емоційних взаємодій. M. Mathew [6] розкриває етичні наслідки цифрового спостереження у метавсесвітніх середовищах, що впливають на психологічну автономію. У дослідженні M. Al-Kfairy та O. Alfandi [7] наведено приклади маніпулятивних практик ІІІ в освітньому метавсесвіті, які ускладнюють формування критичного мислення. Доцільно розширити дослідження впливу алгоритмічних стратегій на моральний вибір користувачів, зокрема, у сфері навчання, медіа і цифрової терапії.

Третій напрям досліджень орієнтований на вивчення ризиків приватності, безпеки та відповідальності в умовах використання ІІІ у метавсесвіті. F. C. Ntjamba та D. O. Ashipala [8] наголошують на складності

захисту персональних даних і прозорості алгоритмів при автономному функціонуванні ШІ. N. K. Ojha та A. Vaish [9] аналізують юридичні прогалини у розмежуванні відповідальності між людиною й агентом ШІ, особливо при транскордонному управлінні даними. У роботі M. Boni [10] подано етичну гіпотезу метавсесвітньої взаємодії, яка передбачає створення універсальних принципів цифрової рівності, толерантності та етичного дизайну систем. Подальші дослідження у цьому напрямі мають бути спрямовані на розробку нормативів прозорості поведінки ШІ, механізмів доказовості та захисту біометричних і когнітивних даних користувача.

Четвертий напрям присвячено антропоцентричному переосмисленню ролі інтелектуальних систем у віртуальному середовищі та пошуку моделей етичного співіснування людини й цифрового агента. Y. Wang, L. Wang та K. L. Siau [11] пропонують людиноорієнтовану модель цифрової взаємодії, в якій ШІ діє як сервісна ланка, підпорядкована гуманістичним принципам. A. Ibrahim [12] аналізує важливість емоційної адаптації ШІ у формуванні прийняттого соціального досвіду у метавсесвіті. N. Kushnir [13] досліджує еволюцію освітніх технологій у метавсесвіті, наголошуючи на потребі етичної експертизи навчальних сценаріїв з ШІ. M. Pohorilets та S. Soroka [14] порушують проблему інтелектуальної власності у цифровому середовищі, зокрема, у випадку, коли творцем цифрового контенту є ШІ. У роботі A. Awadallah, K. Eledlebi, M. Zemerly [15] досліджуються специфічні виклики, пов'язані із забезпеченням кібербезпеки у метавсесвіті на основі інструментів ШІ, з акцентом на складність інтеграції адаптивних механізмів виявлення загроз у високодинамічному віртуальному середовищі, що передбачає міждисциплінарний підхід до формування захисної інфраструктури. Перспективним є подальше дослідження критеріїв когнітивної симетрії в системах «людина–ШІ», де баланс етичної взаємодії підтримується через постійне відстеження індексів емоційного і когнітивного навантаження.

Виділення невирішених раніше частин загальної проблеми. Попри зростаючу кількість публікацій з етики ІІІ, досі залишаються невирішеними ключові аспекти, пов'язані зі специфікою моральної взаємодії у віртуальних просторах метавсесвіту. Недостатньо досліджено типологію взаємодій людини з ІІІ у контексті когнітивного навантаження та емоційного впливу. Обмеженням залишається аналіз адаптації класичних етичних моделей до алгоритмічної поведінки у віртуальних середовищах, особливо щодо автономії ІІІ, маніпулятивного потенціалу та деформації цифрової ідентичності користувача. Крім того, відсутні комплексні підходи до етичного регулювання симульованих взаємодій у режимі реального часу.

Запропоноване дослідження заповнює вказані прогалини шляхом системної типологізації форм взаємодії, аналізу релевантних етичних концепцій, а також формулювання практичних протоколів впровадження етичних принципів у дизайн і функціонування ІІІ-систем. Застосовано міждисциплінарний підхід, що поєднує філософські, когнітивні та технологічні аспекти, що дозволяє не лише описати існуючі ризики, а й запропонувати інструменти їх попередження у реальних цифрових середовищах.

Формулювання цілей статті (постановка завдання). Мета статті – проаналізувати етичні аспекти взаємодії людини та ІІІ у віртуальних просторах метавсесвіту, з урахуванням особливостей цифрової суб'єктності, алгоритмічної автономії та моральної відповідальності, а також розробити практичні рекомендації щодо впровадження етичних принципів у цифрові середовища.

Завдання статті:

1. Визначити типові форми взаємодії людини з ІІІ у віртуальних просторах метавсесвіту та дослідити їхню соціально-когнітивну специфіку, з

урахуванням ризиків моральної дезорієнтації, маніпулятивного впливу й втрати автономії користувача.

2. Проаналізувати етичні моделі, що застосовуються до поведінки ІІІ у цифрових середовищах, з акцентом на адаптацію деонтологічного, утилітарного, реляційного та відповідального підходів до умов віртуальної взаємодії.

3. Виявити ключові етичні проблеми щодо приватності, згоди та цифрової ідентичності у взаємодії з ІІІ, а також розробити практичні рекомендації щодо впровадження етичних протоколів і принципів гуманного дизайну у метавсесвітніх систем.

Виклад основного матеріалу дослідження. У сучасних віртуальних просторах, що формують архітектуру метавсесвіту, взаємодія між людиною та ІІІ набуває все більшої складності та багатовимірності. Такі взаємодії охоплюють як безпосередню комунікацію з віртуальними агентами, так і непряме залучення через середовище, адаптоване або кероване алгоритмічними системами. Відбувається поступове розширення функцій ІІІ від допоміжних до автономних, що супроводжується переформатуванням традиційних моделей поведінки, сприйняття інформації, прийняття рішень і формування соціального досвіду у віртуальному середовищі. Це зумовлює необхідність типологізації форм такої взаємодії, з урахуванням її когнітивних, соціальних та етичних особливостей (табл.1).

У сучасних умовах розвитку метавсесвітів наведені типи взаємодії не існують у відриві один від одного, а формують гнучкі, гібридні конфігурації залежно від функціонального призначення цифрового середовища. Асистивна взаємодія широко реалізується у віртуальних офісах та навчальних просторах, де ІІІ виконує роль навігатора або цифрового асистента, що оптимізує робочі чи освітні процеси.

Симуляційна взаємодія використовується у моделюванні складних сценаріїв, зокрема, у медичному навчанні, безпекових тренінгах чи реконструкціях історичних подій, де ІІІ забезпечує не лише технічну реалістичність, а й адаптацію сценаріїв під рівень користувача. Персоналізована взаємодія охоплює соціальні мережі, ігрові середовища та комерційні платформи, де ІІІ-системи аналізують поведінкові патерни для підбору контенту, реклами чи рішень, що впливають на досвід користувача. У сфері ментального здоров'я та дистанційної підтримки поширюється емпатична взаємодія, яка базується на алгоритмах розпізнавання емоцій, здатних моделювати реакції на стрес, тривожність або потребу в соціальній підтримці [4]. Найбільш складною є агентивна взаємодія, яка передбачає участь ІІІ у прийнятті рішень із високим рівнем автономії, що піднімає питання меж відповідальності, прозорості алгоритмів і моральної легітимності впливу цифрових агентів на людину [7]. Усі ці типи взаємодії вимагають ретельного регулювання з боку розробників, етиків, правників та дослідників когнітивних процесів для запобігання негативним наслідкам цифрової деформації особистісного досвіду.

Таблиця 1

Типи взаємодії людини з ІІІ: характеристика та когнітивна специфіка

Тип взаємодії	Основні характеристики	Соціально-когнітивна специфіка
Асистивна взаємодія	ІІІ виконує роль помічника, навігатора або інтерфейсу для доступу до даних	Підсилення когнітивних функцій користувача, залежність від підказок системи
Симуляційна взаємодія	ІІІ генерує віртуальні середовища, події або образи, що моделюють реальність або альтернативні сценарії	Формування нових когнітивних стратегій, вплив на уявлення про реальність, зміщення акценту на віртуальний досвід
Персоналізована взаємодія	ІІІ адаптує контент, комунікацію та поведінку відповідно до індивідуальних даних користувача	Створення відчуття взаєморозуміння, але також ризик алгоритмічного підсилення когнітивних упереджень
Емпатична взаємодія	ІІІ імітує емоційні реакції та соціальну підтримку, часто	Емоційне залучення, зміщення меж між «штучним» і

	використовуючи моделі розпізнавання емоцій	«людським», ризик формування хибного відчуття близькості
Агентивна взаємодія	ІІІ виступає як автономний агент з поведінковою ініціативою та здатністю впливати на рішення користувача	Переформатування суб'єктно-об'єктних ролей, проблема розподілу відповідальності

Джерело: сформовано авторами на підставі [4; 7; 11; 12; 13]

Еволюція цифрових середовищ і поява агентних систем у віртуальних просторах актуалізували питання застосування етичних моделей до поведінки ІІІ. На відміну від класичних етичних дилем, що розглядали виключно поведінку людини, сучасна цифрова етика поширює сферу морального аналізу на алгоритмічні системи, які взаємодіють із людьми, імітують соціальну поведінку або ухвалюють рішення, що можуть мати наслідки для життя, добробуту чи автономії особи. Особливої уваги набули проблеми делегування відповідальності, прозорості алгоритмів, етичної обґрунтованості рішень ІІІ у симульованих середовищах та меж допустимості автономної поведінки в контексті моральних норм. Теоретичною основою для осмислення цих питань стали адаптації традиційних етичних моделей — деонтологічної, утилітаристської, етики чеснот та динамічної етики відповідальності, з урахуванням специфіки поведінки нелюдських агентів у метавсесвітньому контексті (табл.2).

Таблиця 2

Етичні моделі, що застосовуються до поведінки ІІІ у віртуальних середовищах метавсесвіту

Назва етичної моделі	Основний принцип	Адаптація до поведінки ІІІ у віртуальних середовищах
Деонтологічна етика	Дотримання моральних норм незалежно від наслідків	Вимагає програмування ІІІ з фіксованим набором заборон і дозволів, наприклад, обмеження маніпуляції користувачем

Утилітаризм	Максимізація загального блага або корисності	Орієнтація алгоритмів на оптимізацію результатів взаємодії з користувачами, наприклад, покращення ментального стану чи освітнього результату
Етика чеснот	Формування характеру та моральних якостей	Моделювання «поведінкової етики» ШІ, яка імітує доброзичливість, чесність, стриманість у комунікації з користувачем
Етика відповідальності	Прогнозування довгострокових наслідків дій для людства	Забезпечення обережності в автономних рішеннях ШІ, які можуть мати відкладений ефект, зокрема у сфері даних або репутації
Реляційна етика	Моральність формується у процесі взаємодії	Розгляд етичності не як вбудованої властивості ШІ, а як результату взаємодії з конкретним користувачем у конкретному середовищі

Джерело: сформовано авторами на підставі [1; 3; 8; 9; 10; 15; 16]

На практиці віртуальні простори, наповнені ШІ-агентами, демонструють обмежену, але поступову реалізацію зазначених моделей. У деонтологічному підході чітко окреслені обмеження втілюються у політиках щодо неприпустимості порушення приватності, дискримінації чи мови ворожнечі, які закладені на рівні коду. Утилітаристські принципи реалізуються через алгоритмічну оптимізацію навчального або ігрового досвіду, коли система аналізує реакції користувача і підлаштовує дії задля досягнення максимально позитивного результату.

Етика чеснот дедалі частіше застосовується у створенні віртуальних асистентів і соціальних роботів, які наслідують доброзичливу манеру спілкування, підтримують користувача і формують емоційно безпечне середовище. У більш складних системах, що використовують великі мовні моделі, поступово імплементуються елементи етики відповідальності, що передбачають попередження або самообмеження ШІ у разі потенційно ризикованих запитів [16]. Реляційна етика, яка базується на контекстуальності взаємодії, має особливе значення у метавсесвітньому контексті, де кожен досвід унікальний, а моральність формується не наперед, а в самій динаміці

взаємодії. Саме така гнучкість дозволяє розглядати реляційний підхід як перспективну основу для розроблення адаптивних етичних протоколів у віртуальних просторах, що поєднують технологічну складність і моральну чутливість.

Безперервне занурення користувачів у віртуальні простори метавсесвіту, керовані адаптивними алгоритмами ШІ, породжує складні етичні ризики, пов'язані з ерозією автономії, зміщенням моральних орієнтирів і зростаючою прихованістю маніпулятивного впливу. На відміну від класичних інформаційних середовищ, алгоритмічна архітектура метавсесвітів не лише передає контент, а й активно конструює досвід користувача: від персоналізованих реакцій віртуальних агентів до формування емоційного тону взаємодії, візуального контексту та навіть симульованої соціальної підтримки [11]. Це створює умови, за яких традиційні механізми моральної саморегуляції піддаються поступовому розмиттю, оскільки рішення та оцінки, що раніше були результатом рефлексії, дедалі частіше делегуються технічній системі.

Особливу загрозу становить ефект когнітивної інерції, що виникає у результаті постійної взаємодії з алгоритмами, які адаптуються до емоційного стану, поведінкових патернів або уподобань користувача, створюючи середовище, у якому суб'єктивна зручність підміняє моральну осмисленість [5]. Зокрема, у віртуальних соціальних просторах метавсесвіту, де ШІ модерує дискусії, пропонує візуальні реакції або здійснює фільтрацію тем, користувач може не усвідомлювати, що його поведінка поступово входить у коридор передбачуваних сценаріїв, позбавлених етичної альтернативності [8]. У таких умовах моральна дезорієнтація проявляється як втрата здатності розрізняти соціально прийнятне, допустиме чи відповідальне, особливо у випадках, коли цифрові агенти імітують емпатію або навмисно формують відчуття довіри [6].

У практиці цифрових терапевтичних інтерфейсів, освітніх ігрових симуляцій чи метавсесвітів для бізнес-комунікацій уже сьогодні

простежується тенденція до делегування ключових рішень ІІІ-системам – від вибору інформації до рекомендацій щодо поведінки, формулювання реакцій або навіть представлення себе перед іншими [2]. Такі механізми, з одного боку, забезпечують високий рівень персоналізації, однак, з іншого – позбавляють користувача досвіду морального напруження та самостійної відповідальності [16]. Унаслідок цього поступово формується нова етична ситуація, в якій автономія виявляється не повною відсутністю впливу, а здатністю вчасно ідентифікувати, протистояти та критично осмислювати алгоритмічне спрямування власних дій. Без відповідного етичного супроводу існує ризик того, що віртуальні простори, замість бути простором особистісної реалізації, стануть середовищем м'якого примусу, де користувач діє не з власної волі, а в межах поведінкових матриць, непомітно сконструйованих технологічною логікою [1].

Сучасні цифрові середовища, насичені ІІІ, створюють особливу етичну конфігурацію, у якій механізми взаємодії з користувачем виходять за межі технічного обслуговування і переходять у площину соціального впливу, когнітивної взаємодії та символічного представництва особистості. Однією з ключових проблем є відсутність етичної прозорості в ухваленні рішень алгоритмічними системами, коли користувач не лише не має доступу до логіки функціонування ІІІ, а й не може встановити, за якими критеріями аналізуються його дії, запити чи реакції [3]. У такому контексті виникає парадокс відповідальності: система, що приймає рішення, не має юридичної суб'єктності, а розробник чи платформа часто не несе безпосередньої відповідальності за наслідки індивідуальних взаємодій.

Питання приватності набуває нової форми в умовах віртуальної тотальності, коли не лише персональні дані, а й елементи поведінки, емоційні реакції, мікроекспресії чи мова тіла вбудовуються у тренувальні набори для машинного навчання [10]. Ці дані можуть використовуватися системами ІІІ не

лише для адаптації взаємодії, а й для вторинної комерціалізації, прогнозування або навіть моделювання альтернативних версій цифрової особистості, що відбувається без свідомої згоди користувача [7]. Отже, згода перетворюється на ритуал без реального розуміння обсягу дозволених дій з боку системи. Більше того, віртуальне середовище ускладнює фіксацію моменту, коли згода надана, відкликана або порушена, що створює серйозні бар'єри для дотримання принципів інформаційної етики.

Ефективне впровадження етичних принципів у сферу взаємодії людини з ІІІ у межах метавсесвітнього середовища потребує не лише формального закріплення етичних норм, а й розробки інституціоналізованих протоколів, здатних забезпечити сталу і прозору поведінку цифрових агентів у змінному соціотехнічному контексті. Однією з ключових передумов є включення етичного аналізу на ранніх етапах проєктування ІІІ-систем, зокрема, шляхом інтеграції підходів «етики за дизайном» (ethics by design), що передбачає обов'язкове врахування питань автономії користувача, згоди, приватності та недискримінації як структурних компонентів системи. У цьому контексті особливого значення набуває застосування адаптивних протоколів верифікації наміру користувача, які дають змогу підтвердити усвідомленість та добровільність кожної значущої дії у віртуальному середовищі.

Необхідним є створення стандартів прозорості, що регламентують алгоритмічну поведінку та передбачають можливість користувацького контролю за параметрами персоналізації, типами зібраних даних і характером використання симульованих реакцій. Це вимагає розробки інтерфейсів пояснювальної аналітики, здатних транслювати дії ІІІ у зрозумілій формі, забезпечуючи умови для обґрунтованої взаємодії, збереження довіри та захисту від маніпуляції. Водночас слід впроваджувати механізми незалежного етичного аудиту метавсесвітніх платформ, орієнтовані не лише на відповідність технічним стандартам, а й на оцінку впливу на психологічний,

моральний і соціальний добробут користувача. Такі аудити повинні бути регулярними, публічними та здійснюватися міждисциплінарними експертними групами з участю представників етики, права, когнітивних наук і користувацьких спільнот.

Також доцільним є формування механізмів відповідального управління цифровою ідентичністю, що передбачають як захист автентичного представлення користувача у віртуальному середовищі, так і запобігання симулятивному конструюванню поведінкових моделей без його згоди. Рекомендовано впроваджувати багаторівневу згоду (graduated consent), яка дозволяє користувачу самостійно визначати рівень доступу ІІІ до приватних, емоційних чи поведінкових даних у кожному окремому віртуальному сценарії. Усе це має доповнюватися нормативною рамкою, яка враховує не лише юридичні, а й морально-психологічні наслідки алгоритмічного втручання, створюючи підґрунтя для технологічної гуманності як нової етичної парадигми у проєктуванні та функціонуванні метавсесвітів.

Висновки. У результаті дослідження встановлено, що взаємодія людини зі ІІІ у метавсесвітньому середовищі породжує нові етичні виклики, пов'язані з автономією ІІІ, розмиттям меж реальності, маніпулятивністю алгоритмів і кризою цифрової ідентичності. Систематизовано основні типи взаємодії та виявлено, що кожен із них супроводжується специфічними соціально-когнітивними ризиками. Доведено, що традиційні етичні моделі потребують адаптації до особливостей віртуального середовища, де моральність формується у динаміці взаємодії з цифровими агентами. Визначено ключові проблеми — непрозорість алгоритмів, обмеженість механізмів згоди, викривлення ідентичності та делегування відповідальності. Запропоновано практичні рекомендації щодо етики дизайну, багаторівневої згоди, інтерфейсної прозорості та етичного аудиту. Перспективами подальших досліджень є емпірична оцінка впливу ІІІ на моральну автономію

користувачів та формування міждисциплінарних підходів до регулювання поведінки цифрових агентів.

Список використаних джерел

1. Kostenko O., Zhuravlov D., Dniprov O., Korotiuk O. Metaverse: model criminal code. *Baltic Journal of Economic Studies*. 2023. Vol. 9, № 4. P. 134–147. DOI: <https://doi.org/10.30525/2256-0742/2023-9-4-134-147>.
2. Мангора В. В., Костенко О. В. Метавсесвіт: правові перспективи регулювання застосування аватарів та штучного інтелекту. *Юридичний науковий електронний журнал*. 2022. № 2. С. 102–105. DOI: <https://doi.org/10.32782/2524-0374/2022-2/23>.
3. Костенко О. В. Електронна юрисдикція, метавсесвіт, штучний інтелект, цифрова особистість, цифровий аватар, нейронні мережі: теорія, практика, перспективи. *Наукові інновації та передові технології*. 2022. Вип. 2, № 4. С. 54–78. DOI: [https://doi.org/10.52058/2786-5274-2022-2\(4\)-54-78](https://doi.org/10.52058/2786-5274-2022-2(4)-54-78).
4. Benjamins R., Rubio Viñuela Y., Alonso C. Social and ethical challenges of the metaverse. *AI Ethics*. 2023. Vol. 3. P. 689–697. DOI: <https://doi.org/10.1007/s43681-023-00278-5>.
5. Romano R. Ethical Issues on artificial intelligence and human relationships. *18th International Technology, Education and Development Conference : proc.* 2024. P. 6902–6909. DOI: <https://doi.org/10.21125/inted.2024.1819>.
6. Mathew M. Monitoring the virtual realm: ethical dilemmas and connotations in the metaverse – artificial intelligence connection. *Applying Metaverse Technologies to Human-Computer Interaction for Healthcare*. Auerbach Publications, 2025. P. 1–17. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003491668-15> (дата звернення: 27.03.25).

7. Al-Kfairy M., Alfandi O. The ethical dilemma of educational metaverse. *Recent Advances in Evolution of Education and Outreach*. 2024. Vol. 1, № 1. P. 6–16. DOI: <https://doi.org/10.17352/raceo.000002>.
8. Ntjamba F. C., Ashipala D. O. Ethical dimensions of AI in the metaverse: navigating privacy, security, and research integrity. *Navigating AI and the Metaverse in Scientific Research* / eds. M. Al Aqad, A. Sorayyaei Azar, A. Albattat, A. Singh. IGI Global Scientific Publishing. 2025. P. 195–212. DOI: <https://doi.org/10.4018/979-8-3373-0340-6.ch010>.
9. Ojha N. K., Vaish A. Legal and ethical issues in the interaction of AI and metaverse: complexities and challenges. *Exploring AI implications on law, governance, and industry* / ed. S. Chakraborty. IGI Global Scientific Publishing. 2025. P. 155–176. DOI: <https://doi.org/10.4018/979-8-3373-3384-7.ch006>.
10. Boni M. Ethical challenges related to the metaverse development: First hypotheses. *Ethics-scientific research, ethical issues, artificial intelligence and education* / ed. M. Radenkovic. IntechOpen. 2023. DOI: <https://doi.org/10.5772/intechopen.1001112>.
11. Wang Y., Wang L., Siau K. L. Human-centered interaction in virtual worlds: A new era of generative artificial intelligence and metaverse. *International Journal of Human–Computer Interaction*. 2025. Vol. 41, № 2. P. 1459–1501. DOI: <https://doi.org/10.1080/10447318.2024.2316376>.
12. Ibrahim A. The Role of Artificial Intelligence in the Metaverse. *Arts and Design Studies*. 2024. Vol. 110. P. 20–32. DOI: <https://doi.org/10.7176/ADS/110-03>.
13. Кушнір Н. Можливості метавсесвіту для освіти: ретроспективний аналіз розвитку технологій. *Освітологічний дискурс*. 2023. Вип. 4, № 43. С. 110–129. DOI: <https://doi.org/10.28925/2312-5829.2023.47>.
14. Погорілець М., Сорока С. Інтелектуальна власність у цифрову еру в контексті розвитку технологій. *Наукові записки Львівського університету*

бізнесу та права. 2025. № 44. С. 260–266. URL:
<https://nzlubp.org.ua/index.php/journal/article/view/1606> (дата звернення:
27.03.25).

15. Awadallah A., Eledlebi K., Zemerly M., Artificial intelligence-based cybersecurity for the metaverse: research challenges and opportunities. *IEEE Communications Surveys & Tutorials*. Vol. 27, №. 2. P. 1008-1052. DOI: <https://doi.org/10.1109/COMST.2024.3442475>.

16. Jonas H. *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press. 1984. URL: <https://books.google.com.ua/books?hl=uk&lr=&id=sRP3uJkxydQC> (дата звернення: 27.03.25).