Філософія

УДК 177

DOI https://doi.org/10.5281/zenodo.19957733

Етика технологій: філософські засади відповідального використання штучного інтелекту

Патлайчук Оксана Віталіївна,
кандидат філософських наук, доцент,
доцент кафедри соціально-гуманітарних дисциплін та філософії,
Національний університет кораблебудування імені адмірала Макарова,
Миколаїв, Україна, https://orcid.org/0000-0002-1448-3360

Костроміна Ганна Михайлівна,
кандидат філософських наук, доцент,
доцент кафедри філософії,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна, https://orcid.org/0000-0002-4822-4914

Бабіна Світлана Іванівна,
кандидат філософських наук,
старший викладач кафедри філософії,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Київ, Україна, https://orcid.org/0000-0002-7825-3628

Прийнято: 11.04.2026 | Опубліковано: 30.04.2026

Анотація: ***Мета.*** *Актуальність мети дослідження зумовлена активною трансформацією соціальної взаємодії під впливом динаміки інформаційно-комунікативного простору. Нові виклики сучасності, серед яких і штучний інтелект, зумовлюють необхідність переосмислення значимості віртуального світу для ефективної колаборації. Метою статті є розширений аналіз філософських засад відповідального використання штучного інтелекту та його впливу на соціокультурну динаміку й відповідальне використання технологій.* ***Методи.*** *Методологічна основа статті формувалась із врахуванням пріоритетних принципів реалізації системних досліджень, на основі комплексного підходу. Із метою максимально повного розкриття проблематики було використано низку теоретичних методів дослідження, зокрема, історико-філософський та філософсько-культурологічний методи, цілісний та системний аналіз, синтез, методи порівняння, узагальнення та деякі інші.* ***Результати.*** *У дослідженні виокремлено дотичні ризики та потенційні переваги глобальної цифровізації суспільного простору. Акцентовано, що сучасні алгоритми та віртуальна реальність впливають на розуміння реальності, свободи, індивідуальності та суспільної взаємодії, що потенційно може призвести до деструктивної трансформації особистості, її надмірного поглинання цифровим середовищем, втрати ідентичності. Розглянуто ключові етичні дилеми використання штучного інтелекту, проаналізовано філософські засади відповідальності у сфері цифрових технологій.* ***Висновки.*** *Основний науковий внесок дослідження полягає у дослідженні питань зміни людської ідентичності та поняття свободи, переосмислення реальності, динаміки соціальній структурі та епістемологічних проблем. У статті обґрунтовано, що на сучасному етапі відбувається інтеграція штучного інтелекту в усі сфери суспільного життя, що зумовлює необхідність синергії соціокультурної сфери та ідентичності,*

 ISSN: 3083-5712

цифрової етики і критичного мислення, розвитку культури відповідального використання нейронних мереж.

Ключові слова: *штучний інтелект, цифровізація, філософські засади, етика технологій, соціокультурний розвиток, віртуальна взаємодія, відповідальне використання, етика штучного інтелекту, цифрова відповідальність, філософія технологій.*

Technology Ethics: Philosophical Principles for Responsible Use of Artificial Intelligence

Oksana Patlaichuk,
Candidate of Philosophical Sciences, Associate Professor,
Associate Professor of the Department of Psychology, Philosophy and Social Sciences and Humanities, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine, https://orcid.org/0000-0002-1448-3360

Hanna Kostromina,
Candidate of Philosophical Sciences, Associate Professor,
Associate Professor of the Department of Social and Humanities and Philosophy, National Technical University of Ukraine «Ihor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, https://orcid.org/0000-0002-4822-4914

Svitlana Babina,
Candidate of Philosophical Sciences,
Senior Lecturer at the Department of Philosophy,
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, https://orcid.org/0000-0002-7825-3628

Abstract: ***Aim***. *The relevance of the research goal is due to the active transformation of social interaction under the influence of the dynamics of the information and communication space. New challenges of modernity, including artificial intelligence, necessitate a rethinking of the significance of the virtual world for effective collaboration. The purpose of the article is an extended analysis of the philosophical foundations of the responsible use of artificial intelligence and its impact on socio-cultural dynamics and the responsible use of technologies.* ***Methods***. *The purpose of the article is an extended analysis of the philosophical foundations of the responsible use of artificial intelligence and its impact on socio-cultural dynamics and the responsible use of technology.* ***Results.*** *The study highlights the relative risks and potential benefits of the global digitalization of public space. It emphasizes that modern algorithms and virtual reality affect the understanding of reality, freedom, individuality, and social interaction, which can potentially lead to a destructive transformation of the individual, its excessive absorption by the digital environment, and loss of identity. It examines key ethical dilemmas in the use of artificial intelligence, and analyzes the philosophical principles of responsibility in the field of digital technologies.* ***Conclusions.*** *The main scientific contribution of the research is to investigate the issues of changing human identity and the concept of freedom, rethinking reality, the dynamics of social structure, and epistemological problems. The article substantiates that at the present stage, artificial intelligence is being integrated into all spheres of social life, which necessitates the synergy of the socio-cultural sphere and identity, digital ethics and critical thinking, and the development of a culture of responsible use of neural networks.*

Keywords: *artificial intelligence, philosophical basis, digitalization, technology ethics, socio-cultural development, virtual interaction, responsible use, AI ethics, digital responsibility, philosophy of technology.*

 ISSN: 3083-5712

Постановка проблеми. Розвиток штучного інтелекту (ШІ) стимулює філософську дискусію про відмінності у цілях та мотиваціях, ризики технологічної сингулярності, загрози втрати контролю через стрімку еволюцію нейронних мереж. Виникає необхідність закладення критеріальності в алгоритми ШІ, які були б максимально комплементарними людським етичним цінностям. Тому філософське осмислення відповідального використання ШІ є викликом, що змушує переосмислити фундаментальні питання етики технологій.

Темпи розвитку нейронних мереж та їх інтеграції в усі сфери життєдіяльності суспільства формують нові вимоги до сучасної людини. Адаптивність, критичне мислення, резильєнтність, медіаграмотність стали невід'ємними передумовами її якісної взаємодії з цифровою реальністю. Особливої ваги набувають питання гармонійного поєднання рішень на основі ШІ та традиційних основ суспільного розвитку, що підтверджує актуальність дослідження філософських засад відповідального використання штучного інтелекту. Зважаючи на зазначене, особливої актуальності набуває проблема етики технологій. Філософські засади відповідального використання штучного інтелекту потребують глибшого наукового дослідження.

Аналіз останніх досліджень і публікацій. Проблематиці присвячена низка публікацій сучасних авторів, зокрема, Б. де Брюйн, Л. Флоріді (B. de Bruin, L. Floridi) [3, с. 25–34], Т. Хагендорф (T. Hagendorff) [8, с. 100–116] та інших. Авторами проаналізовано етику використання ШІ та хмарних технологій, виділено ключові проблеми у сфері. Як переконують дослідники, у сучасних умовах трансформації інформаційного поля дезінформація набуває системного характеру та перетворюється на один із ключових інструментів маніпулятивного впливу. Нові можливості для маніпуляції громадською думкою створюються стрімким розвитком цифрових технологій, поширенням соціальних мереж, алгоритмізацією інформаційних потоків.

ISSN: 3083-5712

Філософське осмислення трансформації взаємодії суспільства та цифрового інтелекту здійснено в роботах О. Гіль (O. Gil) [7], Т. Симоніт (T. Simonite) [13], де науковці актуалізують гострі етичні питання конфіденційності, ризиків втрати контролю над алгоритмами, кіберзлочинності, відповідальності за дії в цифровому середовищі.

Аналізуючи потенціал системи цінностей, ідентичності та громадянської позиції, М. Тадео, Л. Флоріді (M. Taddeo, L. Floridi) [14, с. 751–752] детермінують соціально-філософську сутність інформаційної культури майбутнього. Як стверджують автори, сучасні нейронні мережі чинять значний вплив на формування особистості, пропонуючи нові форми самовираження та водночас – ризики для стабільності людського «Я». М. Тегмарк (M. Tegmark) [15] переосмислюючи дійсність під впливом ШІ, наголошує, що розмивання меж між фізичним та віртуальним світом актуалізує питання ідентичності людини в кіберпросторі.

Як бачимо, науковці висвітлюючи окремі аспекти проблеми, акцентують увагу на розвитку стійких етичних переконань у сучасної людини, утвердженні її ідентичності, відповідальному використанні ШІ.

Виділення невирішених раніше частин загальної проблеми. Не зважаючи на значний науковий інтерес до проблематики ШІ, питання його відповідального використання залишаються дослідженими фрагментарно. Філософська дилема етики технологій залишається малодослідженою нішою. Особливої уваги потребують етичні аспекти у контексті інтерактивності та впливу віртуальної взаємодії на людину, зокрема, механізми інтеграції етики в сучасні алгоритми.

Формулювання цілей статті (постановка завдання). Мета статті – на основі обґрунтування філософських засад відповідального використання технологій ШІ виділити конкретні етичні принципи їх інтеграції в сучасні алгоритми.

Виклад основного матеріалу. Суспільні трансформації призводять до змін у світогляді, цінностях та способі взаємодії людей, адже інформаційні технології впливають на розуміння індивідуальності, реальності, свободи та суспільної взаємодії, формуючи цифрову ідентичність, розмиваючи межі між віртуальним та реальним світами. Ці ж технології актуалізують питання щодо свободи та приватності, цифрового нерівноправ'я, маніпуляції та етики. Характерними для сучасного суспільства стають затьмарення емпатичного відношення до оточуючих, культивування свободи слова. При цьому «homo virtualis» орієнтований лише на віртуальність та володіє фреймовим характером світобачення [10, с. 1–3].

Філософські концептуальні засади відповідального використання ШІ. Існує чимало точок зору, іноді – полярних, щодо сумісності понять етики та штучного інтелекту. Так, наприклад, Ж. Брісон (J. Bryson) вважає, що «створення роботів такими, щоб вони заслужили бути моральними індивідуумами, саме по собі може бути витлумачено як аморальна дія, особливо якщо врахувати, що цього, очевидно, можна уникнути, оскільки створення таких технологій – це завжди власне вибір людини» [1]. Дж. Генріхс (Jh. Heinrichs) переконаний, що «оскільки відповідальність може взяти на себе лише той, кого можна покарати, сама «машина» не підходить до ухвалення відповідальності [9]. Оскільки вона не відчуває страждань та її поведінку не може бути виправлено покаранням, не кажучи вже про похвалу або засудження, покладати відповідальність на машину просто неможливо або недоречно».

Концептуальним ядром філософії та культурного розвитку сучасного соціуму є підхід множинної інтерпретації сенсів, новітній онтологічний смисл суспільного поступу, в якому людина ідентифікується, одночасно, основним творінням культури та її творцем. Процес трансформації зосереджений на досягненні максимальної відкритості, свободи вибору, невичерпності. Через

трансфер соціально й культурно значущих рис людини у віртуальний простір формується віртуальна особистість, яка є символічною, множинною та безтілесною та наділена певною автономністю дій у віртуальному середовищі. Зазначене репрезентує нову типологію розколу: межа з'являється не між індивідом і зовнішнім світом, як раніше, а всередині самої особистості. Несинергійне співіснування духовної та фізичної іпостасей у людській особі зазнає загрозливих ризиків, в тому числі щодо утримання цілісного образу власного «Я».

Віртуальна комунікація надає більше свободи. Ця свобода відноситься до етико-ціннісних характеристик, що передбачає відсутність бар'єрів у культурних пріоритетах; когнітивних характеристик, що дозволяє ширше використовувати інформаційні ресурси; емоційних характеристик, що передбачає зниження стресової напруги за допомогою гейміфікації, використання «потоку свідомості» в письмі, трансферу страхів та комплексів у віртуальну площину [2]. Найважливішим аспектом спілкування стає представлення себе Іншому, презентація свого образу, переконань, інтересів, почуттів тощо. Самопрезентація стає онтологічно важливою, оскільки в інформаційному світі життя функціонує за принципом: якщо тебе немає онлайн, тебе не існує.

Етичні ризики ШІ. Дослідники детермінують штучний інтелект радикальною технологією, що трансформує не тільки суспільство, але й саму людську природу. Це зумовлює необхідність розроблення та впровадження нових етичних та регуляторних норм, адже ШІ потенційно спроможний досягати зміни його статусу як технології.

Науковці визначили основні етичні ризики ШІ та підходи в даній сфері: принципи, процеси та етичну свідомість. Загалом, стрімкий розвиток штучних нейронних мереж стимулює дискурс про ризики технологічної сингулярності. Під цим феноменом розуміється гіпотетична неконтрольованість розвитку ШІ

 ISSN: 3083-5712

з серйозними для суспільства наслідками: самостійного вдосконалення та виходу за межі людського розуміння, відмінність у мотиваціях та цілях з людськими очікуваннями, ризику використання для ведення війни, втрата контролю над економікою та технологіями, політична маніпуляція. Зниження таких ризиків потребує закладення визначених чітких критеріїв – етичних і гуманістичних – у алгоритми ШІ, серед яких: гуманність, ненасильство та безпека, прозорість і чесність, взаємодія [11, с. 945–953].

Ключові принципи та механізми етичного регулювання використання штучних нейронних мереж. В останні роки світова спільнота зробила досить серйозний крок уперед у питаннях встановлення та контролю етики застосування технологій ШІ. Під етичним застосуванням розуміється систематичне нормативне осмислення етичних аспектів ШІ на основі комплексної, всеосяжної та багатокультурної системи взаємопов'язаних ціннісних установок, принципів та процедур, здатне орієнтувати суспільство у питаннях відповідального обліку відомих та невідомих наслідків застосування ШІ-технологій. До етичних принципів використання технологій ШІ загалом можуть бути віднесені наступні характеристики:

- повага, захист прав людини й основних свобод та людської гідності – ШІ не має провокувати залежність, будь-яку шкоду чи обмеження свобод людини;

- запобігання загроз знеособлення, дискримінації, утиску;

- забезпечення різноманітності та можливості вибору – технології ШІ мають пропонувати людині різні варіанти використання, зміну параметрів конкретних технологій, адаптацію їх під власні потреби;

- безпека і захищеність – за наявності будь-якої, нехай навіть потенційної, загрози ШІ повинен оцінити можливі ризики та вжити максимально можливих дій для їх запобігання;

ISSN: 3083-5712

- захист персональних даних – на будь-якому етапі роботи технологій ШІ будь-яка обробка даних має здійснюватись у повній відповідності до чинних нормативних документів;

- стійкість розвитку суспільства, прозорість та підзвітність.

Як стверджує Ясон Габріель (I. Gabriel), «головним завданням для теоретиків є не визначення «справжніх» моральних принципів для ШІ, а визначення справедливих принципів узгодження, які отримують рефлексивне схвалення, попри широку різноманітність моральних переконань людей» [5]. Водночас Д. Гіффорд (D. Gifford) [6] переконує у необхідності запровадження стандарту суворої колективної відповідальності в рамках загального права, який не вимагає доведення чи індивідуального причинно-наслідкового зв'язку як умови відповідальності.

Наскільки алгоритми ШІ можуть бути подібними до мислення людини? М. Кокельберг (M. Coeckelbergh) описує впливові наративи ШІ, починаючи від монстра Франкенштейна і закінчуючи трансгуманізмом та технологічною сингулярністю. Він розглядає відповідні філософські дискусії: питання про фундаментальні відмінності між людьми та машинами та дебати щодо морального статусу ШІ [4]. Однією з ознак людини, окрім мислення, є воля, почуття, прагнення, бажання. Мозок може помилятися, у безпосередній ситуації дії якісь рішення можуть здаватися рівнозначними, але людина може приймати рішення під впливом емоцій. Дж. Серль (J. Searle) намагався вирішити цю проблему, ввівши поняття розриву. Розрив, на його думку, нічим не заповнений і пов'язаний із поняттям свободи волі [12, с. 28]. Рішення тому приймаються в умовах неповної інформації та є імовірнісними. Але наші неправильні чи не до кінця правильні рішення можуть створювати нові ситуації. Таким чином, фалібілізм (принципова можливість помилки) також лежить в основі свободи волі.

ISSN: 3083-5712

Окрім того, свобода волі пов'язана із саморефлексією. З наявністю у людини волі пов'язане поняття моральної та правової відповідальності. Наскільки це можна віднести до штучного інтелекту? Для цього, навіть якщо йдеться про якісь елементарні форми людиноподібного штучного інтелекту, повинні бути виконані певні умови: у ШІ мають бути закладені критерії вибору оптимальних рішень разом із вбудованими етичними обмеженнями, можливість, спираючись на різні теоретичні концепції, оцінити ризики, можливість помилки і пов'язана з цим рефлексія, аналіз індивідуально неповторних ситуацій, аналог моральних переживань, пов'язаний із оцінками інших. Для цього треба буде створити щось на зразок емоційного центру, пов'язаного з задоволенням, збудженням та гальмуванням нервових реакцій, занепокоєнням та заспокоєнням, а в розвиненому варіанті – також здатність до емпатії. У такому випадку найімовірнішим перспективним сценарієм суспільного розвитку є паралельний поступ двох цивілізацій: побудованої на природному біологічному носії і цивілізації технічної, причому остання, не зважаючи на потенційно більшу досконалість, буде нездатна повністю замінити людську цивілізацію.

Висновки. Суспільство перебуває наразі на етапі радикальних трансформацій, що може змінити траєкторію розвитку в контексті більшої інтерактивності та впливу цифрових технологій. Ключовими аспектами цих процесів стають: зміна людської ідентичності та поняття свободи, переосмислення реальності, зміни у соціальній структурі, епістемологічні проблеми. У філософському сенсі, потенціал ШІ в контексті впливу на розвиток суспільства наділений дуалістичним характером. Він формує одночасно як простір можливостей, так і ~~як~~ простір загрози.

Вочевидь, що сучасні алгоритми та віртуальна реальність потенційно можуть призвести до деструктивної трансформації особистості, надмірного поглинання цифровим середовищем, втрати ідентичності. На основі

ISSN: 3083-5712

обґрунтування філософських засад відповідального використання технологій ШІ можна виділити конкретні етичні принципи їх інтеграції в сучасні алгоритми: повага, захист прав людини й основних свобод та людської гідності; запобігання загроз знеособлення, дискримінації, утиску; забезпечення різноманітності та можливості вибору; безпека і захищеність; захист персональних даних; стійкість розвитку суспільства, прозорість та підзвітність.

Перспективні наукові розробки мають бути зосереджені на розробленні етико-правової концепції використання штучного інтелекту у добу тотальної цифрової інтеграції.

Список використаних джерел

1. Bryson J. J. Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*. 2018. №20. Pp. 15–26. DOI: https://doi.org/10.1007/s10676-018-9448-6

2. Campolo A., Sanfilippo M., Whittaker M., Crawford K. AI now 2017 report. *AINow*, 2018. URL: https://ainowinstitute.org/publications/ai-now-2017-report-2 (дата звернення 07.03.2026)

3. de Bruin B., Floridi L. The Ethics of Cloud Computing. *Science and Engineering Ethics*. 2017. №23 (1). Pp. 21–39. DOI: https://doi.org/10.1007/s11948-016-9759-0

4. Coeckelbergh Mark AI Ethics MIT Press, Cambridge MA, 2020. 248p.

5. Gabriel I. Artificial Intelligence, Values, and Alignment. *Minds & Machines*. 2020. №30. Pp. 411–437. DOI: https://doi.org/10.1007/s11023-020-09539-2

6. Gifford D. G. Technological triggers to tort revolutions: steam locomotives, autonomous vehicles, and accident compensation. *Journal of Tort Law*. 2018. №11 (1). Pp. 71–143. URL: https://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=2594&context=fac_pubs (дата звернення 07.03.2026)

7. Gil O. AI Philosophy: Sources of Legitimacy to Analyze Artificial Intelligence. Paper presented to the Sixteenth International Conference on Advances in Computer-Human Interactions, ACHI 2023 & The Seventeenth International Conference on Digital Society, ICDS 2023 (April 24-28, Venice, Italy). DOI: https://doi.org/10.31219/osf.io/39njx

8. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines.* 2020. №30. Pp. 99–120. DOI: https://doi.org/10.1007/s11023–020–09517–8

9. Heinrichs J. H. Responsibility assignment won't solve the moral issues of artificial intelligence. *AI Ethics*. 2022. №2. Pp. 727–736. DOI: https://doi.org/10.1007/s43681-022-00133-z

10. Inglehart R. F. Cultural evolution: People's motivations are changing, and reshaping the world. *Social Forces*. 2020. №98(4). Pp. 1–3. DOI: https://doi.org/10.1093/sf/soz119

11. Pappas I. O., Mikalef P., Dwivedi Y. K., Jaccheri L., Krogstie J. Responsible digital transformation for a sustainable society. *Information Systems Frontiers*. 2023. №25(3). Pp. 945–953. DOI: https://doi.org/10.1007/s10796-023-10406-5

12. Searle J. Rationality in Action. M.: Progress-Tradition, 2004.

13. Simonite T. AI experts want to end «black box» algorithms in government. Wired Business, 2017 (October, 18). URL: https://www. wired.com/story/ai-experts-want-toend-black-box-algorithms-in-government (дата звернення 07.04.2026)

14. Taddeo M., Floridi L. How AI can be a force for good. *Science*. 2018. №361 (6404). Pp. 751–752. DOI : *10.1126* /science.aat5991

15. Tegmark M. Life 3.0: Being human in the age of artificial intelligence. New York: Alfred A. Knopf, 2017.

ï