

УДК 111.1:004.8 + 17:004.8

Тіло / тілесність між смертю та життям у часи штучного інтелекту (до підґрунтя появи питання етики ШІ)

BODY / CORPOREALITY BETWEEN DEATH AND LIFE
IN THE AGE OF ARTIFICIAL INTELLIGENCE
(ON THE ORIGINS OF THE AI ETHICS ISSUE)

КОСС Андрій Миколайович – аспірант кафедри філософії, Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2, м. Харків, 61002, Україна

ORCID <https://orcid.org/0009-0001-6158-3540>

ДОЛЬСЬКА Ольга Олексіївна – доктор філософських наук, професор, професор кафедри філософії, Національний технічний університет «Харківський політехнічний інститут», вул. Кирпичова, 2, м. Харків, 61002, Україна

ORCID <https://orcid.org/0000-0002-9577-8282>

KOSS Andrey Mykolayovych – graduate student of the Department of Philosophy, National Technical University «Kharkiv Polytechnic Institute», 2 Kyrpychov St., Kharkiv, 61002, Ukraine

DOLSKA Olga Oleksiyivna – Doctor of Philosophical Sciences, Professor, Professor of the Department of Philosophy, National Technical University «Kharkiv Polytechnic Institute», 2 Kyrpychov St., Kharkiv, 61002, Ukraine

DOI <https://doi.org/10.54891/2786-7013/2026-1-4>

Анотація. Стаття присвячена дослідженню онтологічного та етичного підґрунтя впровадження штучного інтелекту (далі – ШІ) крізь призму концептів тіла та тілесності. **Об’єктом роботи** є теоретичне та практичне коріння виникнення питання етики ШІ, а **предметом** – етичні наслідки інтеграції ШІ у життєвий простір людини з урахуванням специфіки суб’єктивності в біологічних (тіло) та штучних структурах. **Методологія** трансдисциплінарності дослідження сформована на перетині різнопланових напрямів: нейробіології, когнітивних наук, фізики, психології технологій, обчислювальної соціальної науки, економіки, бізнесу, інженерії, медицини та права. **Метою** статті є обґрунтування переходу від абстрактних дискусій про свідомість до розробки практичних механізмів відповідального проектування технологій. У роботі проаналізовано фундаментальну відмінність між людиною та машиною: на основі праць Г. Нортгоффа, А. Буччеллато, Ф. Зіліо та С. Гувейя доведено, що ШІ бракує унікальної «просторо-часової динаміки» та нейропластичності, які є основою суб’єктивної точки зору. Спираючись на концепцію «отілесненого розуму» О. Дольської, обґрунтовано, що будь-яка раціональність обмежена біологічним втіленням, через що ШІ демонструє лише «компетентність без усвідомлення» (Д. Деннетт, Р. Хаклі, П. М’якеля, В. Мюллер). Особлива увага приділена питанню, чи отримує людина у часи ШІ шанс на якісно нове тілесне життя або смерть (еволюцію), чи потрапляє у зону екзистенційних ризиків. Розглянуто соціально-економічні загрози дегуманізації праці та «знецінення навичок» (Е. Брін’ольфссон, Д. Лі, Л. Реймонд, Ж. Мансо, А. Косс, Д. Аутор), а також ризики алгоритмічного контролю (К. Келлог, М. Валентайн, А. Крістін). На перетині ШІ та нейротехнологій виявлено виклики для приватності та свободи волі, про що попереджають Р. Юсте, С. Герінг, Б. Агуера і Аркас, Г. Бі, Х. Кармена, А. Картер, Дж. Фінс, Ф. Фрізен, Дж. Галлант, Дж. Гаггінс, Дж. Іллес, Ф. Кельмаєр, Е. Кляйн, А. Марблстоун, К. Мітчелл, Е. Паренс, М. Фам, А. Рубел, Н. Садато, Л. Спеккер Салліван, М. Тейхер, Д. Вассерман, А. Векслер, М. Віттакер та Дж. Волпо. Як ключовий інструмент мінімізації цих загроз обґрунтовано підхід «етика через дизайн» (Б. Фрідман, Ф. Брей, Б. Дайноу), що передбачає інтеграцію суспільних цінностей безпосередньо в архітектуру систем.
© КОСС Андрій Миколайович, ДОЛЬСЬКА Ольга Олексіївна

Це підкріплюється даними «Експерименту моральної машини» (Е. Авад, С. Дсоуза, Р. Кім, Д. Шульц, Д. Генріх, А. Шаріф, Ж.-Ф. Боннефон, І. Рахван). Юридичний аспект регулювання без визнання морального статусу ШІ представлений поглядами Д. Брайсона, М. Діамантіса, Т. Гранта, А. Бертоліні та Д. Айелло.

Ключові слова: людина, штучний інтелект, тіло та тілесність, тілесне життя або смерть, етика штучного інтелекту, методологія трансдисциплінарності, «етика через дизайн».

Summary. *The article is devoted to the research of the ontological and ethical basis of the implementation of artificial intelligence (hereinafter – AI) through the prism of the concepts of body and corporeality. The object of the work is the theoretical and practical roots of the emergence of the issue of AI ethics, and the subject is the ethical consequences of the integration of AI into human living space, taking into account the specifics of subjectivity in biological and artificial structures. The interdisciplinary research methodology is formed at the intersection of diverse directions: neurobiology, cognitive sciences, physics, psychology of technology, computational social science, economics, business, engineering, medicine, and law. The aim of the article is to substantiate the transition from abstract discussions about consciousness to the development of applied mechanisms for responsible technology design. The paper analyzes the fundamental difference between humans and machines: drawing on the works of G. Northoff, A. Buccellato, F. Zilio, and S. Gouveia, it is demonstrated that AI lacks the unique «spatiotemporal dynamics» and neuroplasticity, which form the basis of a subjective point of view. Drawing on O. Dolska's concept of the «embodied mind», it is argued that all rationality is constrained by biological embodiment, which is why AI demonstrates only «competence without comprehension» (D. Dennett, R. Hakli, P. Mäkelä, V. C. Müller). Special attention is paid to the question of whether humans in the era of AI receive a chance for a qualitatively new bodily life or death (evolution), or fall into the zone of existential risks. The socio-economic threats of dehumanization of labor and «deskilling» (E. Brynjolfsson, D. Li, L. Raymond, J. Manseau, A. Koss, D. Autor) are considered, as well as the risks of algorithmic control (K. Kellogg, M. Valentine, A. Christin). At the intersection of AI and neurotechnologies, challenges to privacy and free will are identified, as warned by R. Yuste, S. Goering, B. Agüera y Arcas, G. Bi, J. Carmena, A. Carter, J. Fins, P. Friesen, J. Gallant, J. Huggins, J. Illes, P. Kellmeyer, E. Klein, A. Marblestone, C. Mitchell, E. Parens, M. Pham, A. Rubel, N. Sadato, L. Specker Sullivan, M. Teicher, D. Wasserman, A. Wexler, M. Whittaker, and J. Wolpaw. The approach of «ethics by design» (B. Friedman, P. Brey, B. Dainow) is substantiated as a key tool for minimizing these threats, which implies the integration of societal values directly into the system architecture. This is supported by data from the «Moral Machine Experiment» (E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, I. Rahwan). The legal aspect of regulation without recognizing the moral status of AI is presented by the views of J. Bryson, M. Diamantis, T. Grant, A. Bertolini, and D. Aiello.*

Key words: human, artificial intelligence, body and embodiment, bodily life or death, AI ethics, transdisciplinarity methodology, «ethics by design».

Вступ. Інтенсивний прогрес нейротехнологій та алгоритмів штучного інтелекту (ШІ), які майстерно відтворюють людську автономію, кидає виклик класичним антропоцентричним парадигмам. Соціум постає перед необхідністю глобального перегляду базових моральних орієнтирів. Нагальність цього питання посилюється тим, що розгортання інтелектуальних систем знаменує старт чергової промислової революції, яка невідворотно трансформує всі галузі нашого буття.

Актуальність роботи зумовлена необхідністю переходу від теоретичних дебатів про свідомість ШІ до практичного формування відповідальної політики та норм його інтеграції. Ключова проблема полягає в тому, щоб: 1) визначити напрями, що створюють трансдисциплінарну методологію для дослідження ШІ; 2) продемонструвати, як саме ця методологія, представлена цими напрямками, формує фундамент порушення питання про етику ШІ. Це зумовлено тим, що моральний статус вимагає здатності до суб'єктивних переживань, а сучасний ШІ позбавлений живої нейропластичності та суб'єктивного світовідчуття, що втілює парадокс «сліпої

майстерності». Він демонструє бездоганну виконавчу здатність, за якою, проте, немає справжнього осягнення суті. Наукові дослідження та юридична доцільність пропонують розглядати ШІ суто через правові рамки, а не через моральне визнання. Така концептуальна суперечність зумовлює прикладну дилему: як регулювати ШІ, чий статус досі залишається дискусійним? Крім того, надмірна залежність від ШІ провокує професійне та соціальне знецінення навичок, що відкриває шлях до дегуманізації професійного середовища.

Аналіз останніх досліджень. Специфіка джерельної бази цієї роботи полягає в тому, що вона спирається на праці учених, чия освіта часто далека від академічної філософії (нейробіологи, фізики, економісти та інженери). Однак, спираючись на емпіричні дані, ми все частіше бачимо, що їхні роботи безпосередньо формують порядок денний етики ШІ. Це створює цікавий прецедент, коли поєднання висновків різних наук набуває філософського звучання. Особливу увагу слід звернути на те, що у традиційній філософії ставлення людини до питання життя та смерті завжди було трансцендентним. Натомість коли у дискусію вступають техніки, нейробіологи, соціологи та психологи, фокус зміщується на те, що відбувається з людиною як носієм тілесності. Виникає закономірне питання: ці технології покращують тіло / тілесність людини, ведуть до еволюції чи занурюють у нові ризики? Зокрема, Г. Нортгофф (досліджує психіатрію та нейробіологію), А. Буччеллато (досліджує когнітивну науку, психологію та біоінформатику), Ф. Зілію (досліджує етику) та С. Гувейя (досліджує філософію, етику та нейронауку) вважають, що суб'єктивність формується завдяки особливій конфігурації взаємодії простору, часу та нейропластичності біологічного мозку, яка забезпечує його адаптацію до зовнішнього середовища та надає можливість формувати власну перспективу сприйняття. На відміну від біологічних систем, сучасним архітектурам ШІ бракує нейропластичності, оскільки їхня структура стає статичною одразу після завершення етапу навчання [16, р. 29], [17, р. 1097, 1111–1112]. Як зазначають Д. Деннетт (досліджує філософію та когнітивні науки), Р. Хаклі (досліджує філософію, логіку та ШІ), П. Мякеля (філософ соціальних наук) та В. Мюллер (досліджує філософію та етику ШІ), штучні алгоритми мають «компетентність без осмислення» [9], [11], [15, р. 13]. Це знаходить обґрунтування у концепції «отілесненого розуму» О. Дольської, яка в межах феноменології та філософії тілесності наголошує, що фундаментом для формування суб'єктивності є специфічна біологічна природа людини та її здатність пізнавати світ через єдність сприйняття, відчуття та дії [1, р. 7, 9]. Прикладна етика зміщує фокус на практичні наслідки інтеграції ШІ в суспільство і представлена центральною ідеєю «ціннісно-орієнтованого дизайну». Цю концепцію розвивають Б. Фрідман (досліджує психологію та інформатику), Ф. Брей (досліджує етику, соціальну та політичну філософію) та Б. Дайноу (досліджує етику та ШІ), стверджуючи, що технології не є нейтральними, а їхні архітектури повинні від самого початку проектуватися з урахуванням етичних цінностей [10, р. 17], [6, р. 1266]. Вагому доказову базу забезпечують Е. Авад (досліджує економіку, бізнес, етику та ШІ), С. Дсоуза (досліджує програмну інженерію та обчислювальну соціальну науку), Р. Кім (досліджує обчислювальну соціальну науку та ШІ), Д. Шульц (досліджує економіку), Д. Генріх (досліджує культурну та генетичну еволюцію), А. Шаріфф (досліджує психологію), Ж.-Ф. Боннефон (досліджує економіку, етику ШІ та психологію технологій) та І. Рахван (досліджує обчислювальну соціальну науку та ШІ) в аналізі соціальних очікувань та вподобань, які можуть бути інтегровані в алгоритми ШІ [4, р. 59–60]. Паралельно з цим досліджуються соціально-економічні та екзистенційні ризики. Е. Брінгольфссон (досліджує економіку та ШІ), Д. Лі (досліджує економіку), Л. Реймонд (досліджує прикладну мікроекономіку) та Ж. Мансо (досліджує інформаційні системи та ШІ) розкривають парадокс ШІ у професійному середовищі. З одного боку, впровадження інтелектуальних систем значно підвищує продуктивність,

проте з іншого – вони несуть загрозу «завданневого та соціального знецінення навичок», що зрештою призводить до дегуманізації праці [7], [13, р. 1, 3]. Зазначені тенденції підкріплюються напрацюваннями Д. Аутора (досліджує економіку), котрі фокусуються на автоматизації експертних оцінок та управлінських стратегій [3, р. 23]. К. Келлог (досліджує соціологію організацій), М. Валентайн (досліджує ШІ та організаційний дизайн) та А. Крістін (досліджує соціологію, медіа та технології) доводять, що алгоритмізація менеджменту фактично нівелює роль людського посередництва. Як наслідок, посилюється жорсткий раціональний контроль, що суттєво обмежує можливості для апеляції чи перегляду прийнятих системою рішень [12, р. 368, 387]. Р. Юсте (досліджує нейробиологію), С. Герінг (досліджує філософію), Б. Агуера і Аркас (досліджує обчислювальну нейронауку, соціальний інтелект та ШІ), Г. Бі (досліджує біофізику та нейробиологію), Х. Кармена (досліджує нейротехнології, нейробиологію та ШІ), А. Картер (досліджує нейробиологію), Дж. Фінс (досліджує етику), Ф. Фрізен (досліджує філософію), Дж. Галлант (досліджує нейробиологію), Дж. Гаггінс (досліджує біоінженерію та інтерфейс мозок-комп'ютер), Дж. Іллес (досліджує нейроетику та неврологію), Ф. Келльмаер (досліджує взаємодію людей і технологій, нейробиологію, неврологію та ШІ у медицині), Е. Кляйн (досліджує нейроетику, неврологію, спеціалізується на деменції), А. Марблстоун (досліджує біофізику, займається розробкою інтерфейсів мозок-комп'ютер), К. Мітчелл (досліджує філософію, етику, богослов'я та сестринську справу), Е. Паренс (досліджує біоетику та соціально-поведінкову геноміку), М. Фам (досліджує нейроетику, епістемологію, філософію науки), А. Рубел (досліджує філософію, інформаційну етику, політику та право), Н. Садато (досліджує нейрофізіологію та магнітно-резонансну томографію), Л. Спеккер Салліван (досліджує філософію та біоетику), М. Тейхер (досліджує алгебраїчну геометрію кривих та поверхонь, комп'ютерний зір, математичні моделі мозкової діяльності та нейронні обчислення), Д. Вассерман (досліджує біоетику, нейронауки, інвалідність, репродукцію, генетику та охорону здоров'я), А. Векслер (досліджує біоетику, нейроетику та соціологію медицини), М. Віттaker (досліджує ШІ) та Дж. Волпо (досліджує пластичність центральної нервової системи, нейрореабілітацію, інтерфейси мозок-комп'ютер) акцентують на вразливості когнітивного суверенітету особистості в умовах конвергенції ШІ та нейротехнологій [18, р. 159–160]. Д. Брайсон (досліджує поведінкову екологію, ШІ та етику), М. Діамантіс (досліджує конфіденційність, ШІ та експериментальну філософію), Т. Грант (досліджує міжнародне публічне право), А. Бертоліні (досліджує право, етику ШІ, страхування та управління ризиками) та Д. Айелло (досліджує політологію, соціальні науки та міжнародні відносини) пропонують інтегрувати ШІ у нормативно-правову площину без визнання моральної суб'єктності [8], [5], [15, р. 13]. Попри фрагментованість сучасних досліджень та значну кількість публікацій, ця стаття пропонує цілісну методологічну рамку для етичного та соціально-емпіричного проектування ШІ. **Об'єктом дослідження** є підґрунтя появи питання етики ШІ. **Предметом дослідження** є етичні наслідки інтеграції ШІ в життя людини (тіло / тілесне) з урахуванням природи суб'єктивності в біологічних (тіло) та штучних системах.

Мета статті спрямована на зміщення фокусу з теоретичних дискусій про суб'єктність ШІ у площину розробки прикладних інструментів його етично-відповідального конструювання, що передбачає розв'язання таких **завдань**: 1) проаналізувати, якими напрямками представлена сучасна методологія трансдисциплінарності у дослідженні проблем ШІ; 2) з'ясувати, що стає підґрунтям для появи питання етики ШІ, де окремим вектором є спроба відповісти на питання: чи людина отримує шанс на якісно нове життя і подальшу еволюцію, чи виникає в еволюції тіла та тілесного неймовірна кількість ризиків? 3) на основі аналізу архітектури системи ШІ та його інтерактивних можливостей вивчити виклики імплементації у суспільному, юридичному

та етичному вимірах, зокрема ризику дегуманізації праці; 4) визначити екзистенційний вимір взаємодії людини з ШІ, відповівши на питання: чи є ці технології еволюційним стрибком, чи стали вони новим джерелом ризиків для людства; 5) обґрунтувати підхід етичного проектування технологій як практичний метод закріплення ціннісних орієнтирів суспільства; 6) окремим вектором мети є спроба дати відповідь на питання: чи людина як носій тіла та тілесності у часи ШІ отримує шанс на якісно нове життя і подальшу еволюцію, і чи має вона неймовірну кількість ризиків?; 7) окреслити перспективи подальших досліджень у сфері розробки методологій та стандартів ШІ для впровадження етичних принципів в інженерні практики.

Наукова новизна роботи полягає в: 1) розробці інтегрованого трансдисциплінарного підходу, який синтезує різнохарактерні та різноспрямовані позиції – від технічних до психологічних; 2) розкритті підґрунтя появи питання «Чому виникає етично-орієнтоване проектування?», «Де аналізуються технічні, економічні, соціально-емпіричні дискурси щодо ШІ?»; 3) уточненні онтологічних меж прирівнювання машини та людини, що покладено в основу питання етики ШІ, через призму тілесності та нейробіології. Обрана методологічна стратегія сприяє переходу від теоретико-філософського рівня до реалізації практичних інженерно-соціальних завдань.

Методологія. Сучасна трансдисциплінарна методологія синтезує три підходи. **По-перше**, методологія техніко-технологічного дизайну машин спирається на концепції: «просторо-часової динаміки з'єднання мозку та свідомості» (Г. Нортгофф, А. Буччеллато та Ф. Зіліо [16, р. 29]); «компетентності ШІ без усвідомлення» (Д. Деннетт [9], Р. Хаклі, П. Мякеля [11] та В. Мюллер [15, р. 13]); «нейропластичності як невід'ємної екзистенційно-феноменологічної унікальності суб'єктивності» (Г. Нортгофф, А. Буччеллато, Ф. Зіліо [16, р. 29] та С. Гувейя [17, р. 1097, 1111–1112]); «отілесненого розуму» (О. Дольська [1, р. 7, 9]). **По-друге**, методологія ціннісно-орієнтованого дизайну спирається на концепції: «етичних машинних агентів» (В. Мюллер [15, р. 13] та Д. Мур [14]); «етики через дизайн архітектури системи» (Б. Фрідман [10, р. 17], Ф. Брей та Б. Дайноу [6, р. 1266]); «ієрархії суспільних очікувань респондентів під час машинного навчання» (Е. Авад, С. Дсоуза, Р. Кім, Д. Шульц, Д. Генріх, А. Шаріф, Ж.-Ф. Боннефон та І. Рахван [4, р. 59–60]). **По-третє**, методологія економіко-правового використання ШІ спирається на концепції: «знецінення навичок через дегуманізацію робочих процесів» (Е. Брінгольфссон, Д. Лі, Л. Реймонд [7], Ж. Мансо [13, р. 1, 3] та А. Косс [2, с. 60]); «автоматизації експертного судження та ухвалення рішень ШІ» (Д. Аутор [3, р. 23]); «алгоритмічного управління як раціонального контролю, що усуває людське посередництво та посилює дегуманізацію» (К. Келлог, М. Валентайн та А. Крістін [12, р. 368, 387]); «екзистенційних ризиків на перетині ШІ та нейротехнологій» (Р. Юсте, С. Герінг, Б. Агуера і Аркас, Г. Бі, Х. Кармена, А. Картер, Дж. Фінс, Ф. Фрізен, Дж. Галлант, Дж. Гагінс, Дж. Іллес, Ф. Кельмаєр, Е. Кляйн, А. Марблстоун, К. Мітчелл, Е. Паренс, М. Фам, А. Рубел, Н. Садато, Л. Спеккер Салліван, М. Тейхер, Д. Вассерман, А. Векслер, М. Віттакер та Дж. Волпо [18, р. 159–160]); «інтеграції ШІ в правове поле» (Д. Брайсон, М. Діамантіс, Т. Грант [8], А. Бертоліні та Д. Айелло [5]) [15, р. 13]. Виходячи з наведеного розмаїття джерельної бази, методологічна рамка дослідження вибудовується на принципах глибокої трансдисциплінарності. Ми маємо справу з конвергенцією знань, де перетинаються питання медицини, біології, інженерії, економіки, управління ризиками, теології та права. Такий широкий спектр дисциплін є необхідним, оскільки етика ШІ перестала бути суто гуманітарною рефлексією і перетворилася на поле, де технічні параметри системи безпосередньо впливають на екзистенційний стан людини.

Виклад основного матеріалу. Здатність до переживань можна пов'язати з біологічними процесами в людині. Г. Нортгофф, А. Буччеллато та Ф. Зіліо у статті «З'єднання мозку та

свідомості через просторо-часову динаміку: до теорії спільної валюти» (2025) наголошують, що поява свідомості зумовлена «просторо-часовою динамікою, в якій мозок синхронізується зі світом і формує унікальну точку зору, без якої неможлива суб'єктивність, і саме цієї фундаментальної основи бракує ШІ» [16, р. 29]. Цей висновок гарно доповнюють погляди Д. Деннетта [9], Р. Хаклі та П. Мякелі [11], що наведені у статті В. Мюллера «Етика ШІ та робототехніки» (2020). Автори зазначають, що сучасні системи ШІ мають потенціал до реалізації когнітивних функцій, притаманних людському мисленню, однак запрограмовані агенти іноді не вважаються «повноцінними», оскільки вони «компетентні без розуміння», так само як нейрони в мозку [15, р. 13]. Г. Нортгофф та С. Гувейя у статті «Чи виявляє ШІ основну фундаментальну суб'єктивність? Нейрофілософський аргумент» (2024) наголошують, що ключовою відмінністю між людиною та інтелектуальною машиною є когнітивна гнучкість. Єдиний метод оновлення інформації для ШІ – це перенавчання всієї моделі, що призводить до нової фіксованої структури. Натомість нейропластичність людського мозку дозволяє адаптуватися до нової інформації, коригуючи внутрішні часові шкали для синхронізації із зовнішнім світом, а цей механізм повністю відсутній у сучасних моделях ШІ [17, р. 1111–1112]. Саме нейропластична основа є фундаментом свідомості та опосередкована взаємодією часових шкал мозку із зовнішнім світом [17, р. 1097]. Наведені положення доповнює О. Дольська в статті «Ідея отілесненого («embodied») розуму як обґрунтування тілесної раціональності» (2020), з одного боку, формування суб'єктивності потребує наявності відповідної біологічної основи, яка через контакт із довкіллям забезпечує індивідуальний сенсомоторний досвід: «Все, про що ми можемо думати чи розуміти, стає можливим завдяки тілу і обмежується тілом, мозком і людським втіленням у навколишній світ» [1, р. 9]; з іншого боку, наміри приписати свідомість безтілесним системам (ШІ) призводять до регресу в бік застарілого картезіанського дуалізму фізичного та ментального: «Тим самим ідея отілесненого / втіленого розуму підтверджує відмову від картезіанської дихотомії розум–тіло як антинаукової, а будь-яке розуміння раціональності повинне враховувати тілесну раціональність» [1, р. 7].

Для концептуалізації ролі ШІ в аксіологічному просторі варто скористатися типологією Д. Мура [14], яку наводить В. Мюллер у статті «Етика ШІ та робототехніки» (2020). Д. Мур пропонує розмежовувати чотири категорії машинної агентності: 1) агенти етичного впливу (до яких належать, наприклад, роботи-жokeї); 2) неявні етичні агенти, чия безпека (як у випадку автопілотів) закладена на рівні коду; 3) явні етичні агенти, що оперують формалізованими алгоритмами для обчислення корисності дій; 4) повноцінні етичні агенти, здатні до самостійного формулювання та обґрунтування моральних суджень [15, р. 13]. З огляду на це, сучасні інтелектуальні системи, унаслідок відсутності феноменологічної свідомості, не набувають статусу повноцінних суб'єктів моралі. Водночас вони виступають беззаперечними чинниками етичного впливу. Таке твердження трансформує дискурс з абстрактних роздумів у прикладний вимір: оскільки функціонування ШІ неминуче трансформує моральне середовище, суб'єктивна відповідальність за цей процес покладається на його творців та операторів. Б. Фрідман у статті «Ціннісно-чутливе проектування» (1996) доводить, що цінності іманентно притаманні як самій природі створених технологічних інструментів, так і способам їхнього подальшого застосування [10, р. 17]. Ф. Брей та Б. Дайноу у статті «Етика через дизайн для ШІ» (2024) розвивають цю ідею і вважають, що цей підхід дозволяє певною мірою вбудовувати ціннісні орієнтири безпосередньо в процес проектування [6, р. 1266]. Зазначена концепція ілюструє, як загальні моральні норми трансформуються у прикладні технічні вимоги, перетворюючи моральні аспекти на фундаментальний елемент системної архітектури. Отже, пріоритетним викликом сучасності постає не стільки декларування аксіологічних орієнтирів ШІ, скільки

створення та імплементація верифікованих алгоритмів, що забезпечують присутність цих цінностей на всіх етапах розробки та експлуатації технологічного продукту.

Незважаючи на те, що ШІ позбавлений свідомості, Е. Авад, С. Дсоуза, Р. Кім, Д. Шульц, Д. Генріх, А. Шаріф, Ж.-Ф. Боннефон та І. Рахван у статті «Експеримент моральної машини» (2018) дослідним шляхом довели, що процес навчання систем можна значно інтенсифікувати завдяки платформі масштабного моделювання моральних вподобань. Цей масив даних дозволив виробити глобальні, соціально прийнятні принципи для регулювання поведінки автономних систем завдяки аналізу 40 мільйонів рішень десятками мовами від мільйонів людей у 233 країнах та територіях [4, р. 59]. Аналіз даних виявив чітку інтуїтивну ієрархію вподобань, які можна вважати основними будівельними блоками для машинної етики: 1) пріоритет життя людини над твариною; 2) збереження більшої кількості життів; 3) надання переваги молодшим людям [4, р. 60]. На нашу думку, будь-які спроби регулювати ШІ повинні враховувати суспільні очікування.

Попри відсутність свідомості, ШІ має значний вплив на людину. Наприклад, Е. Брінюлфссон, Д. Лі та Л. Реймонд у статті «Генеративний ШІ на роботі» (2023), дослідивши поетапне впровадження розмовного помічника серед 5179 агентів підтримки клієнтів, зазначають, що використання цього інструменту підвищує продуктивність (виміряну кількістю вирішених питань на годину) в середньому на 14 %. При цьому автори наголошують на нерівномірності цього впливу: ефективність початківців та низькокваліфікованих працівників зростає на 34 %, тоді як на результати досвідчених і висококваліфікованих фахівців технологія має мінімальний вплив [7]. Ж. Мансо у статті «Розуміння впливу завдань ШІ та соціальної декваліфікації: агентно-орієнтована модель» (2025) акцентує на тому, що надмірна залежність від інтелектуальних систем ризикує призвести до декваліфікації працівників та послаблення соціальних взаємодій, потенційно створюючи дегуманізувальні робочі середовища, які надають пріоритет ефективності над добробутом працівників [13, р. 3]. Вказана тенденція виступає релевантним продовженням генезису технологічних інновацій, що докорінно перетворює сутність праці та сприйняття тіла / тілесності людини. У межах цього процесу відбувається перехід від розуміння індивіда як цілісного суб'єкта до трактування його як системи дискретних функцій, що потребують вдосконалення або замінити. А. Косс у статті «Трансформації гуманізму: між наукою і філософією» (2023) зазначає: «Одним із основних показників, який відрізняє гуманістичну людину від постгуманістичної, є присутність емпатії» [2, с. 60]. Отже, прагнучи максимальної продуктивності, професійна діяльність поступово підпорядковується машинній логіці. У цьому процесі суб'єкт праці постає перед загрозою перетворення на суто інструментальний елемент, що нівелює його індивідуальну значущість та екзистенційний вимір.

Цей феномен виходить за межі рутинних операцій. Як зазначає Д. Аутор у статті «Вплив технологічних змін на ринок праці: від неприборканого ентузіазму до кваліфікованого оптимізму та величезної невизначеності» (2022), ШІ дедалі частіше залучатиметься до виконання завдань з ухвалення рішень середнього та високого рівнів, які історично були прерогативою менеджерів і професіоналів. Це явище вже спостерігається у сферах фінансів та інвестування, управління запасами, надання кредитів, виявлення шахрайства і навіть у деяких напрямках дизайну, адже він забезпечує вищу оперативність та економічну рентабельність [3, р. 23]. Відтак, під приціл автоматизації за допомогою ШІ потрапляють операції, що базуються на фаховій експертизі та стратегічному виборі, що ставить під загрозу ексклюзивність компетенцій спеціалістів у цих галузях. Процеси, що зумовлюють таку переорієнтацію управлінських методів, стають об'єктом ґрунтовного дослідження К. Келлог, М. Валентайн та А. Крістін у статті «Алгоритми

в роботі: нова суперечлива територія контролю» (2020). Дослідники підкреслюють, що алгоритми стають новою формою раціонального контролю, дозволяючи роботодавцям скеровувати дії персоналу через обмеження та рекомендації, оцінювати його шляхом фіксації даних і рейтингування, а також дисциплінувати за допомогою механізмів заміни або винагороди [12, р. 368]. Такі системи уможливають усунення посередництва менеджерів у процесах керування, оцінювання та накладання стягнень. На відміну від бюрократичних систем, які допускали певну свободу дій завдяки можливості апелювати до людини, алгоритми здатні повністю виключити процес людського ухвалення рішень зі структур контролю. В умовах повної автоматизації управлінських рішень у працівників залишається менше можливостей розраховувати на емпатію керівників та на отримання винятків із встановлених правил [12, р. 387]. Таким чином, впровадження алгоритмічного управління веде до фундаментального зсуву в природі праці: від заміни фізичних зусиль до заміни когнітивних та соціальних компетенцій. Відсутність прямого контакту з керівником, здатним зважити на специфіку ситуації чи певні пом'якшувальні чинники, нівелює соціальну складову праці. Це провокує стан відчуженості та позбавляє трудові відносини необхідної адаптивності й емпатії. Таким чином, контроль через алгоритми виходить за межі простого знецінення кваліфікації – він фактично блокує прояви суб'єктності працівника та його право на розгляд справи з урахуванням людського чинника.

Е. Бріньолфссон, Д. Лі, Л. Реймонд [7] та Ж. Мансо зазначають, що інтелектуальні системи можуть спонукати організації замінювати кваліфікованих працівників початківцями, підвищуючи їхню продуктивність, що може призводити до декваліфікації працівників [13, р. 1]. Р. Юсте, С. Герінг, Б. Агуера і Аркас, Г. Бі, Х. Кармена, А. Картер, Дж. Фінс, Ф. Фрізен, Дж. Галлант, Дж. Гагінс, Дж. Іллес, Ф. Келльмаєр, Е. Кляйн, А. Марблстоун, К. Мітчелл, Е. Паренс, М. Фам, А. Рубел, Н. Садато, Л. Спеккер Салліван, М. Тейхер, Д. Вассерман, А. Векслер, М. Віттакер та Дж. Волпо у статті «Чотири етичні пріоритети для нейротехнологій та ШІ» (2017) наголошують, що розвиток цих технологій створює безпрецедентні виклики для людства. ШІ та інтерфейси «мозок-комп'ютер» повинні поважати й зберігати приватність, ідентичність, свободу волі та рівність людей. Проте сучасні технологічні розробки свідчать про те, що людство на шляху до світу, в якому буде можливо декодувати психічні процеси та безпосередньо маніпулювати мозковими механізмами, що лежать в основі людських намірів, емоцій та рішень; світу, де індивіди зможуть спілкуватися з іншими силою думки [18, р. 159–160]. Цей підхід зміщує фокус на практичні загрози: питання вже не в тому, чи є ШІ тільки інструментом, а чи не перетвориться він на механізм управління людською свідомістю і чи здатний він повністю замінити людину? В. Мюллер у статті «Етика ШІ та робототехніки» (2020) зазначає, що Д. Брайсон, М. Діамантіс, Т. Грант [8], А. Бертоліні та Д. Айелло [5] розглядають доцільність надання роботам та іншим системам ШІ статусу правосуб'єктності. Це передбачає – за аналогією з державами, підприємствами чи організаціями – наявність у них фінансових прав і зобов'язань, але водночас виключає кримінальну відповідальність [15, р. 13]. Такий підхід уможливує інтеграцію ШІ в правове поле без визнання за ним морального статусу особистості. Ці міркування підводять до висновку, що головним етичним завданням сьогодні стають не стільки дебати про наявність свідомості у машин, скільки розробка системи цінностей, що закладаються в процесі їх створення.

Висновки. Узагальнюючи результати проведеного дослідження згідно з поставленою метою, можемо констатувати наступне: 1) проаналізовано сучасну трансдисциплінарну методологію дослідження ШІ та встановлено, що вона формується на основі напрочуд різнорідного дискурсу. Хоча на перший погляд ці напрями здаються фрагментованими, при глибшому аналізі вони демонструють конвергенцію, складаючись у комплексний методологічний синтез,

де технічні, біологічні та етичні аспекти взаємодоповнюють одне одного, утворюючи складний контекст для розуміння феномену ШІ. 2) З'ясовано, що підґрунтям для появи питання етики ШІ є фундаментальна невизначеність щодо впливу технологій на онтологічний статус людини. Етична проблематика виникає на перетині спроб зрозуміти: чи отримує тіло та тілесність імпульс до еволюційного розквіту, чи, навпаки, технологічне втручання несе ризики знецінення та атрофії людських якостей. Саме ця межа між потенціалом розвитку та загрозою втрати людського є простором формування етичного дискурсу. 3) На основі аналізу архітектури ШІ доведено, що сучасні системи характеризуються відсутністю базової суб'єктності, оскільки не мають біологічної основи (насамперед нейронної пластичності та властивої живим організмам просторово-часової організації), а отже, демонструють лише функціональну результативність без фактичного осмислення. Це призводить до специфічних соціальних наслідків: ШІ не є моральним суб'єктом, проте виступає потужним агентом впливу. Інтеграція таких систем несе ризики дегуманізації праці, нівелювання цінності професійних компетенцій та посилення алгоритмічного контролю, зміщуючи фокус відповідальності виключно на розробників та користувачів. 4) Визначено екзистенційний вимір взаємодії людини з ШІ, який характеризується діалектичною суперечністю: з одного боку, технології демонструють ознаки еволюційного стрибка, підвищуючи ефективність діяльності, а з іншого – стають джерелом нових загроз. Виявлено закономірність: масштабування інтеграції високих технологій у життєвий простір пропорційно збільшує потенційні ризики, зокрема стосовно приватності, свободи волі та збереження ідентичності. 5) Обґрунтовано, що ключовим практичним механізмом вирішення окреслених проблем є етичний дизайн технологій. Ця концепція передбачає відмову від сприйняття технологій як нейтральних інструментів та вимагає інтеграції суспільних цінностей безпосередньо в архітектуру систем ШІ ще на етапі їх проектування, що дозволяє мінімізувати ризики до моменту їх виникнення. 6) Розкриваючи вектор еволюції людини як носія тілесності, встановлено, що людство опинилося в ситуації, яку можна порівняти з міфом про Скіллу і Харібду: від дегуманізації сутності людини до загрози фізичного знищення тіла / тілесності (зокрема через мілітаризацію ШІ). Це підтверджує, що розвиток ШІ несе не лише нові можливості, а й екзистенційні загрози. 7) Окреслено перспективи подальших досліджень, які, на нашу думку, полягають у розробці конкретних методологій та стандартів для впровадження «етики через дизайн» в інженерні практики, а також у подальшому моніторингу соціально-економічних та екзистенційних наслідків інтеграції ШІ в суспільство задля збереження людської суб'єктивності.

Список використаних джерел

1. Дольська О. О. Ідея отілесненого («embodied») розуму в якості обґрунтування тілесної раціональності. *Вісник Харківського національного університету імені В. Н. Каразіна. Серія «Теорія культури і філософія науки»*. 2020. Вип. 62. С. 7–14. DOI: <https://doi.org/10.26565/2306-6687-2020-62-01>
2. Косс А. Трансформації гуманізму: між наукою і філософією. *Актуальні проблеми філософії та соціології*. 2023. Вип. 42. С. 57–63. DOI: <https://doi.org/10.32782/apfs.v042.2023.10>
3. Autor D. The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty. Cambridge: Nber Working Paper Series. 2022. 30074. DOI: <https://doi.org/10.3386/w30074>
4. The Moral Machine Experiment / E. Awad et al. *Nature*. 2018. Vol. 563. P. 59–64. DOI: <https://doi.org/10.1038/s41586-018-0637-6>
5. Bertolini A., Aiello G. Robot Companions: A Legal and Ethical Analysis. *The Information Society*. 2018. Vol. 34, № 3. P. 130–140. DOI: <https://doi.org/10.1080/01972243.2018.1444249>
6. Brey P., Dainow B. Ethics by Design for Artificial Intelligence. *AI and Ethics*. 2024. Vol. 4. P. 1265–1277. DOI: <https://doi.org/10.1007/s43681-023-00330-4>

7. Brynjolfsson E., Li D., Raymond L. R. Generative AI at Work. *NBER National Bureau of Economic Research*, 2023. 31161. DOI: <https://doi.org/10.3386/w31161>
8. Bryson J. J., Diamantis M. E., Grant T. D. Of, for, and by the People: The Legal Lacuna of Synthetic Persons. *Artificial Intelligence and Law*. 2017. Vol. 25, № 3. P. 273–291. DOI: <https://doi.org/10.1007/s10506-017-9214-9>
9. Dennett D. C. *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W.W. Norton, 2017.
10. Friedman B. Value-Sensitive Design. *Interactions*. 1996. Vol. 3. № 6. P. 16–23.
11. Hakli R., Mäkelä P. Moral Responsibility of Robots and Hybrid Agents. *The Monist*. 2019. Vol. 102. № 2. P. 259–275. DOI: <https://doi.org/10.1093/monist/onz009>
12. Kellogg K. C., Valentine M. A., Christin A. Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*. 2020. Vol. 14. № 1. P. 366–410. DOI: <https://doi.org/10.5465/annals.2018.0174>
13. Manseau J. Understanding The Effects of AI Tasks and Social Deskilling: An Agent-Based Model. *AMCIS 2025 Proceedings: Thirty-first Americas Conference on Information Systems, Montréal*, 2025. 1098. URL: <https://aisel.aisnet.org/amcis2025/sigadit/sigadit/4> (дата звернення: 28.09.2025).
14. Moor J. H. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*. 2006. Vol. 21. № 4. P. 18–21.
15. Müller V. C. Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. 2020. URL: <https://plato.stanford.edu/entries/ethics-ai/> (дата звернення: 28.09.2025).
16. Northoff G., Buccellato A., Zilio F. Connecting Brain and Mind through Temporo-Spatial Dynamics: Towards a Theory of Common Currency. *Physics of Life Reviews*. 2025. Vol. 52. P. 29–43. DOI: <https://doi.org/10.1016/j.plrev.2024.11.012>
17. Northoff G., Gouveia S. S. Does Artificial Intelligence Exhibit Basic Fundamental Subjectivity? A Neurophilosophical Argument. *Phenomenology and the Cognitive Sciences*. 2024. Vol. 23. P. 1097–1118. DOI: <https://doi.org/10.1007/s11097-024-09971-0>
18. Four Ethical Priorities for Neurotechnologies and AI. / Yuste R. et al. *Nature*. 2017. Vol. 551. P. 159–163. DOI: <https://doi.org/10.1038/551159a>

References

1. Dolska O. O. (2020). Ideia o tilesnenoho («embodied») rozumu v iakosti obgruntuvannia tilesnoi ratsionalnosti [The Idea of Embodied Mind as Substantiation of Corporal Rationality]. *Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina. Seriiia «Teoriia kultury i filozofia nauky» – Bulletin of V. N. Karazin Kharkiv National University. Series «Theory of Culture and Philosophy of Science»*, 62, 7–14. DOI: <https://doi.org/10.26565/2306-6687-2020-62-01> [in Ukrainian].
2. Koss A. (2023). Transformatsii humanizmu: mizh naukoiu i filozofieiu [Transformations of Humanism: Between Science and Philosophy]. *Aktualni problemy filozofii ta sotsiologii – Current problems of philosophy and sociology*, 42, 57–63. DOI: <https://doi.org/10.32782/apfs.v042.2023.10> [in Ukrainian].
3. Autor D. (2022). The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty. *National Bureau of Economic Research*. 30074. DOI: <https://doi.org/10.3386/w30074>
4. Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.-F., & Rahwan I. (2018). The Moral Machine Experiment. *Nature*, 563, 59–64. DOI: <https://doi.org/10.1038/s41586-018-0637-6>
5. Bertolini A., & Aiello G. (2018). Robot Companions: A Legal and Ethical Analysis. *The Information Society*, 34 (3), 130–140. DOI: <https://doi.org/10.1080/01972243.2018.1444249>
6. Brey P., & Dainow B. (2024). Ethics by Design for Artificial Intelligence. *AI and Ethics*, 4, 1265–1277. DOI: <https://doi.org/10.1007/s43681-023-00330-4>
7. Brynjolfsson E., Li D., & Raymond L. R. (2023). Generative AI at Work. *NBER National Bureau of Economic Research*, 31161. DOI: <https://doi.org/10.3386/w31161>
8. Bryson J. J., Diamantis M. E., & Grant T. D. (2017). Of, for, and by the People: The Legal Lacuna of Synthetic Persons. *Artificial Intelligence and Law*, 25 (3), 273–291. DOI: <https://doi.org/10.1007/s10506-017-9214-9>

9. Dennett D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W.W. Norton.
10. Friedman B. (1996). Value-Sensitive Design. *Interactions*, 3(6), 16–23.
11. Hakli R., & Mäkelä P. (2019). Moral Responsibility of Robots and Hybrid Agents. *The Monist*, 102 (2), 259–275. DOI: <https://doi.org/10.1093/monist/onz009>
12. Kellogg K. C., Valentine M. A., & Christin A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, 14 (1), 366–410. DOI: <https://doi.org/10.5465/annals.2018.0174>
13. Manseau J. (2025). Understanding the Effects of AI Tasks and Social Deskilling: An Agent-Based Model. *AMCIS 2025 Proceedings*. 4, 1098. URL: <https://aisel.aisnet.org/amcis2025/sigadit/sigadit/4> (accessed: 28.09.2025).
14. Moor J. H. (2006). The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, 21 (4), 18–21.
15. Müller V. C. (2020). Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. URL: <https://plato.stanford.edu/entries/ethics-ai/> (accessed: 28.09.2025).
16. Northoff G., Buccellato A., & Zilio F. (2025). Connecting Brain and Mind through Temporo-Spatial Dynamics: Towards a Theory of Common Currency. *Physics of Life Reviews*, 52, 29–43. DOI: <https://doi.org/10.1016/j.plrev.2024.11.012>
17. Northoff G., & Gouveia S. S. (2024). Does Artificial Intelligence Exhibit Basic Fundamental Subjectivity? A Neurophilosophical Argument. *Phenomenology and the Cognitive Sciences*, 23, 1097–1118. DOI: <https://doi.org/10.1007/s11097-024-09971-0>
18. Four Ethical Priorities for Neurotechnologies and AI. / Yuste R. et al. *Nature*. 2017, 551, 159–163. DOI: <https://doi.org/10.1038/551159a>

Отримано 20.02.2026.

Отримано в доопрацьованому вигляді 05.03.2026.

Прийнято до друку 25.03.2026.

Опубліковано 08.04.2026.

Received 20.02.2026.

Received in revised form 05.03.2026.

Accepted for publication 25.03.2026.

Published 08.04.2026.