

УДК 37.013.73

DOI <https://doi.org/https://doi.org/10.31392/cult.alm.2026.2.20>

Бутурліна Оксана Василівна,
*кандидат філософських наук,
доцент, завідувач кафедри управління
інформаційно-освітніми проектами
Комунального закладу вищої освіти
«Дніпровська академія неперервної освіти»
Дніпропетровської обласної ради»
orcid.org/0000-0002-9603-4752
obutur@dano.dp.ua*

Мірошніченко Андрій Анатолійович,
*кандидат філософських наук, доцент,
доцент кафедри управління інформаційно-освітніми проектами
Комунального закладу вищої освіти
«Дніпровська академія неперервної освіти»
Дніпропетровської обласної ради»
orcid.org/0000-0002-9973-018X
gmaa2000@gmail.com*

Кольцова Світлана Михайлівна,
*старший викладач кафедри управління
інформаційно-освітніми проектами
Комунального закладу вищої освіти
«Дніпровська академія неперервної освіти»
Дніпропетровської обласної ради»
koltsova@dano.dp.ua*

Тухтарова Тетяна Карімжанівна,
*методист навчально-методичної лабораторії
дистанційних технологій в освіті
обласного навчально-методичного центру
підвищення кваліфікації педагогічних працівників
Комунального закладу вищої освіти
«Дніпровська академія неперервної освіти»
Дніпропетровської обласної ради»
orcid.org/0000-0001-7460-0912
ttk10@ukr.net*

МАЙБУТНЄ ЛЮДСТВА В ЕПОХУ ШТУЧНОГО ІНТЕЛЕКТУ: ФІЛОСОФСЬКО-АНТРОПОЛОГІЧНИЙ АНАЛІЗ ТА ТЕХНОЛОГІЧНІ ЗАПОБІЖНИКИ

Стаття присвячена комплексному філософському та соціокультурному аналізу трансформації антропогенної цивілізації під впливом стрімкого розвитку систем штучного інтелекту в умовах так званої

© Бутурліна О. В., Мірошніченко А. А., Кольцова С. М., Тухтарова Т. К., 2026



Стаття поширюється на умовах
ліцензії відкритого доступу CC BY 4.0

«технологічної підлітковості» людства. Досліджуються теоретико-методологічні засади розуміння природи інтелекту, зокрема через критичну деконструкцію фізикалістських та натуралістичних підходів, які призводять до неправомірної антропоморфізації великих мовних моделей та «машинізації» людської свідомості. На противагу цьому запропоновано парадигму синергійної антропології як методологічну основу для осмислення людини як відкритої енергійної системи. Автори аналізують еволюційні етапи розвитку систем штучного інтелекту за концепцією М. Сулеймана: від фаз «класифікатора» та «креатора» до фази «діяча», що володіє агентністю та здатністю до автономного планування складних операцій у реальному світі. Особливу увагу приділено новітнім відкриттям у нейрофізіології, про геометричні *eigenmodes* мозку, яка ставить під сумнів коннектомну модель і може вимагати повного перегляду архітектури майбутніх інтелектуальних систем. У дослідженні систематизовано ключові цивілізаційні ризики: посилення техніко-технологічного відчуження за Л. Мамфордом, загрозу атрофії критичного мислення у мас, а також гострі політико-етичні виклики, пов'язані з цифровим масовим стеженням та розробкою повністю автономних систем озброєння. Розглянуто сучасні інструменти контролю та безпеки, такі як рівні «діяча» у політиці відповідального масштабування та технічний фреймворк Sophon, що впроваджує механізми не-настроюваного навчання для захисту моделей від шкідливого донавчання. У статті проведено порівняльний аналіз футурологічних сценаріїв різних авторів, зокрема М. Сулеймана, щодо майбутнього демократії та глобального ладу. У висновках обґрунтовано безальтернативність гуманістичного підходу, в якому людина визнається вищою цінністю, що є єдиним запобіжником від антропологічної деградації в епоху алгоритмів.

Ключові слова: людина, свідомість, штучний інтелект, агентність, цивілізаційні ризики, Sophon, гуманістична перспектива, майбутнє людства.

Buturlina Oksana,

*Candidate of Philosophical Sciences,
Associate Professor, Head of The Informational Educational Projects
Management Department
Communal Institution of Higher Education
“Dnipro Academy of Continuing Education”
of Dnipropetrovsk Regional Council
orcid.org/0000-0002-9603-4752
obutur@dano.dp.ua*

Miroshnychenko Andrii,

*PhD in Philosophy,
Associate Professor of Department of Information
and Educational Projects Management
Communal Institution of Higher Education
“Dnipro Academy of Continuing Education”
of Dnipropetrovsk Regional Council
orcid.org/0000-0002-9973-018X
gmaa2000@gmail.com*

Koltsova Svitlana,

*Senior Lecturer at the Department of Information
and Educational Projects Management
Communal Institution of Higher Education
“Dnipro Academy of Continuing Education”
of Dnipropetrovsk Regional Council
koltsova@dano.dp.ua*

Tukhtarova Tetiana,
*Methodologist of the Educational and Methodological Laboratory
of Distance Technologies in Education of the Regional Educational
and Methodological Center for Advanced Training of Pedagogical Workers
Communal Institution of Higher Education
“Dnipro Academy of Continuing Education”
of Dnipropetrovsk Regional Council
orcid.org/0000-0001-7460-0912
ttk10@ukr.net*

THE FUTURE OF HUMANITY IN THE ERA OF ARTIFICIAL INTELLIGENCE: PHILOSOPHICAL-ANTHROPOLOGICAL ANALYSIS AND TECHNOLOGICAL SAFEGUARDS

The article is devoted to a comprehensive philosophical and socio-cultural analysis of the transformation of anthropogenic civilization under the influence of the rapid development of artificial intelligence systems in the context of humanity's so-called “technological adolescence”. It examines the theoretical and methodological foundations for understanding the nature of intelligence, specifically through the critical deconstruction of physicalist and naturalist approaches that lead to the improper anthropomorphization of large language models and the «mechanization» of human consciousness. In contrast, the paradigm of synergetic anthropology is proposed as a methodological basis for conceptualizing the human being as an open energetic system. The authors analyze the evolutionary stages of AI development according to Mustafa Suleyman's concept: from the «classifier» and «creator» phases to the «doer» phase (Artificial Capable Intelligence), characterized by agency and the capacity for autonomous planning of complex operations in the real world. Special attention is paid to the latest discoveries in neurophysiology concerning the brain's geometric eigenmodes, which challenge the connectome model and may require a complete revision of the architecture of future intelligent systems. The study systematizes key civilizational risks: the intensification of technical and technological alienation according to Lewis Mumford, the threat of critical thinking atrophy among the masses, and acute political and ethical challenges related to digital mass surveillance and the development of fully autonomous weapons systems. Modern control and safety instruments are considered, such as the safety levels within the Responsible Scaling Policy and the Sophon technical framework, which implements non-fine-tunable learning mechanisms to protect models from harmful fine-tuning. The article provides a comparative analysis of futurological scenarios of various authors, in particular M. Suleiman, regarding the future of democracy and global order. The conclusions substantiate the lack of alternatives to the humanistic approach, in which the human being is recognized as the highest value and the sole safeguard against anthropological degradation in the age of algorithms.

Key words: human, consciousness, artificial intelligence, agency, civilizational risks, Sophon, humanistic perspective, future of humanity.

Постановка проблеми. Системи штучного інтелекту (далі – СШІ) сьогодні виступають не просто як черговий етап науково-технічного прогресу, а як фундаментальний виклик антропогенній цивілізації. Бурхливий розвиток алгоритмів та їхня тотальна інтеграція в усі сфери людського буття, від побуту до прийняття стратегічних державних рішень, провокують глобальну дискусію щодо майбутнього виду Homo Sapiens. Витоки цієї дискусії сягають середини минулого століття, коли у класичній праці *Computing Machinery and Intelligence* було вперше концептуально поставлено радикальне питання про можливість мислення машин (Turing, 1950, с. 433). Сучасна ситуація

загострює це питання до меж. Якщо частина експертного співтовариства впевнена у немисливості створення цифрового бога у формі надрозуму, то інша частина вказує на нездоланну онтологічну прірву між логікою коду та живим феноменом свідомості. Ретельне вивчення філософських та етичних аспектів СШІ є критично необхідним для ідентифікації цивілізаційних ризиків.

Актуальність дослідження зумовлена експоненціальним темпом розвитку обчислювальних потужностей, які ставлять людство перед необхідністю негайного переосмислення меж власної суб’єктності. Ми опинилися у стані «технологічної підлітковості», згідно

з концепцією Д. Амодей, де людська цивілізація отримала майже безмежну інструментальну силу. Проте ще не досягла моральної та інтелектуальної зрілості для безпечного керування нею (Amodei, 2024). Поява потужних автономних систем наближає крапку сингулярності, де США стане здатним до самовдосконалення, що може призвести до радикального прискорення прогресу або, навпаки, до повної втрати людиною статусу суб'єкта історії.

Аналіз останніх досліджень і публікацій. Сучасний науковий дискурс формується у напруженому полі між техно-оптимізмом та алармізмом. Фундаментальні логічні основи імітації інтелекту, закладені А. Тюрінгом, сьогодні втілюються у складних архітектурах нейромереж, які демонструють здібності, що раніше вважалися суто людськими (Turing, 1950, с. 433). М. Сулейман у праці *The Coming Wave* обґрунтовує неминучість переходу від креаторів (генеративних моделей) до діячів (Artificial Capable Intelligence), що володіють агентністю – здатністю автономно діяти у фізичному та цифровому просторі (Suleyman, 2023, с. 216). У той же час інші автори попереджають про ерозію ліберальних цінностей під тиском алгоритмічного маніпулювання, а дослідження Дж. Панга доводять, що когнітивна функція людини жорстко обмежена геометрією мозку, що ставить під сумнів можливість повного копіювання розуму на кремнієві носії (Pang, 2023, с. 566).

Мета статті – здійснення цілісного філософсько-антропологічного аналізу сценаріїв цивілізаційної трансформації в епоху тотальної алгоритмізації та теоретичне обґрунтування гуманістичних і технічних заходів, здатних запобігти антропологічній деградації та втраті людської ідентичності.

Методологія. Дослідження базується на міждисциплінарному підході, що інтегрує методи системного аналізу технологічних трендів, деконструкції фізикалістських концепцій свідомості та прогностичного моделювання соціальних систем у межах парадигми синергійної антропології.

Виклад основного матеріалу. Онтологічний статус свідомості та критика фізикалістської редукції. Сучасний інтелектуальний ландшафт перебуває під потужним впливом натуралістичних та фізикалістських підходів.

У межах цієї парадигми пропонується спрощене розуміння людини як складного «біологічного комп'ютера», а свідомості – як побічного продукту нейронних обчислень. Така редукція призводить до неправомірного отождолення людини та США. З одного боку, відбувається «машинізація» живого суб'єкта шляхом позбавлення його духовної унікальності, а з іншого – необґрунтована антропоморфізація великих мовних моделей (LLM), яким приписують наявність волі, розуму та почуттів.

На противагу цьому, синергійна антропологія пропонує розглядати свідомість як відкриту динамічну систему, що функціонує в крапках біфуркації, моментах вільного вибору, які неможливо прорахувати алгоритмічно. Свідомість не є просто «мовним додатком» до мозку. Це цілісна форма психічної рецепції світу, що має культурно-історичну та колективну основу. Вона нерозривно пов'язана з інтенційністю, а саме із спрямованістю на об'єкт, що має для суб'єкта особистісний сенс. Фізикалізм, намагаючись перекласти все багатство людського духу на мову фізики та бінарного коду, ігнорує той факт, що людина – це передусім істота, яка створює сенси, а не просто обробляє дані.

Прірва між формальною логікою алгоритмів та живою інтуїцією. Дослідження людської ментальності, її витоків та значення для соціуму залишається засадничим питанням філософії. Кожна історична епоха пропонувала свою інтерпретацію цієї загадки, проте сьогодні дискурс набув особливої гостроти через прогрес у галузі нейротехнологій. Психофізіологічна дилема, окреслена ще у XVII сторіччі, сьогодні фокусується на пошуку «містка» між ментальними станами та фізичними процесами в нейронних мережах.

Людське мислення принципово відрізняється від «мислення» США наявністю творчої інтуїції та здатності до емоційного резонансу. Сучасні LLM демонструють феноменальні результати в логічному перетворенні знаків, проте вони залишаються замкненими у просторі синтаксису та намагаються отримати доступу до семантики. Інформація, якою оперує машина, не має для неї жодного значення чи життєвої ваги. Натомість людський розум об'єднує в собі логіку, креатив, уяву та фантазію, створюючи ідеальні моделі, що випереджають свій час. Як зауважував Л. Мамфорд, причина багатьох

соціальних бід криється у розриві між технологічною могутністю та моральним розвитком особистості (Mumford, 1934, с. 419). Якщо людство продовжить делегувати функцію цілепокладання алгоритмам, воно ризикує опинитися у стані технічного відчуження, де людина стає лише додатком до глобальної обчислювальної мережі.

Еволюційні фази США: від класифікатора до автономного діяча. Генезис інтелектуальних систем пройшов шлях від механічних калькуляторів Паскаля та Лейбніца до сучасних нейромереж із трильйонами параметрів. Н. Вінер у фундаментальній праці *Cybernetics* (Wiener, 1948) вказував на небезпеку неконтрольованого самовідтворення машин та наголошував на необхідності збереження ведучої ролі людини у симбіозі людина-машина. Сьогодні ми спостерігаємо реалізацію сценаріїв, описаних М. Сулейманом, який виділяє три ключові фази розвитку штучного інтелекту. Це класифікатор, коли система, що розпізнає патерни (обличчя, тексти, діагнози). Креатор – сучасна фаза генеративного ШІ, що створює контент. І це діяч або ACI (Artificial Capable Intelligence, штучний спроможний інтелект), майбутня фаза, де США зможе самостійно планувати та виконувати складні ланцюжки дій для досягнення поставленої мети (Suleyman, 2023).

Експерименти професора Хунцзін Лу довели, що сучасні моделі вже демонструють «емергентне аналогічне міркування», що дозволяє їм розв'язувати задачі шляхом пошуку неявних паралелей (Webb, 2023, с. 1534). Проте, результати Дж. Панга щодо «геометричних обмежень функцій мозку» (Geometric constraints) свідчать, що людська когніція зумовлена фізичною формою та хвильовою динамікою мозку, а не лише коннектомом (Pang, 2023, с. 571). Це створює природний бар'єр для повної цифрової копії людської особистості, проте не виключає появи іншого, позалюдського типу інтелекту.

Технологічні огорожі та фреймворки цивілізаційної безпеки. Для того, щоб людство могло безпечно співіснувати з системами, які наближаються до рівня надрозуму, необхідне впровадження жорстких технічних та правових обмежень. Наразі провідні дослідницькі центри пропонують кілька рівнів захисту.

Responsible Scaling Policy (RSP). Це політика відповідального масштабування, впроваджена

компанією Anthropic, передбачає чотири рівні безпеки (ASL). При переході на кожен наступний рівень потужності розробники зобов'язані впроваджувати все суворіші заходи контролю, щоб уникнути витоку критичних знань або втрати контролю над системою (Anthropic, 2024).

Фреймворк Sophon (Non-Fine-Tunable Learning). Це інноваційний підхід до стримування шкідливої трансформації моделей. Sophon вирішує проблему перенавчання, коли зловмисники намагаються адаптувати законслухняну модель для створення зброї чи кібератак. Система використовує спеціальний модуль придушення, який знижує продуктивність моделі в обмежених та небезпечних областях знань під час процесів тонкого налаштування, залишаючи її ефективною у дозволених сферах (SOPHON, 2024).

Ці інструменти є критично важливими, оскільки, за прогнозами Е. Тоффлера, техногенна цивілізація вже створила реальні загрози своєму існуванню через екологічні та генно-інженерні виклики (Toffler, 2000, с. 429). США може стати або каталізатором цих загроз, або головним інструментом їх вирішення. Залежно від того, які червоні лінії ми закладемо в його архітектуру сьогодні.

Висновки. Проведене дослідження дає змогу сформулювати наступні узагальнення, що мають як теоретичне, так і практичне значення:

Доведено, що ключовою антропологічною загрозою епохи США є не фізичне знищення людства машинами, а тиха деградація людської суб'єктності через редукцію свідомості до алгоритмічних моделей та втрату навичок критичного мислення.

Обґрунтовано онтологічну відмінність між США та людським розумом. Якщо перший базується на логічному перетворенні знаків, то другий є цілісним біосоціальним феноменом, що володіє інтуїцією та інтенційністю. Геометричні обмеження мозку, виявлені Дж. Пангом, підтверджують унікальність біологічного інтелекту.

Визначено, що перехід США до фази діяча створює безпрецедентні ризики для соціальної стабільності, що вимагає негайного впровадження технічних запобіжників, таких як фреймворк Sophon та політики відповідального масштабування.

Наукова новизна роботи полягає у синтезі нейрофізіологічних даних про хвильову

природу мозку з футурологічними сценаріями трансформації цивілізації, що дозволяє виділити гуманістичне ядро, яке не підлягає алгоритмізації.

Практичне значення результатів полягає у можливості використання запропонованих підходів для розробки етичних кодексів та технічних стандартів безпеки США,

де людина залишається єдиним суб'єктом цілепокладання.

Збереження цивілізаційної суб'єктності можливе лише за умови подолання розриву між технологічним поступом та морально-етичним розвитком людства. США, на нашу думку, має залишатися інструментом розширення людських можливостей, а не заміником людської сутності.

Список використаних джерел:

- Amodei, D. (2024). The adolescence of technology. Essay. URL: <https://www.darioamodei.com/essay/the-adolescence-of-technology>.
- Anthropic. (2023–2026). Responsible scaling policy: Version 3.0. URL: <https://www.anthropic.com/news/responsible-scaling-policy-v3>.
- Deng, J., Pang, S., et al. (2024). SOPHON: Non-fine-tunable learning to restrain task transferability for pre-trained models. IEEE Symposium on Security and Privacy (S&P). URL: <https://ieeexplore.ieee.org/document/10646821>.
- Mumford, L. (1934). *Technics and civilization*. Harcourt, Brace and Co. URL: <https://search.worldcat.org/title/Technics-and-civilization/oclc/560667>.
- Pang, J. C., Aquino, K. M., et al. (2023). Geometric constraints on human brain function. *Nature*, 618, 566–574. URL: <https://www.nature.com/articles/s41586-023-06098-1>.
- Suleyman, M. (2023). The coming wave: Technology, power, and the twenty-first century's greatest dilemma. Crown. URL: https://www.researchgate.net/publication/387730680_The_Coming_Wave_Technology_Power_and_the_21st_Century's_Greatest_Dilemma_By_Mustafa_Suleyman_With_Michael_Bhaskar_New_York_Crown_2023.
- Трифлер, Е. (2000). Третя хвиля (А. Євса, пер.з англ.). Видавничий дім «Всесвіт». 480 с. URL: <https://lib.in.ua/100507-tretia-hvylia-vyd-2000/>.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. URL: https://www.researchgate.net/publication/263384944_Alan_Turing's_Computing_Machinery_and_Intelligence.
- Webb, T., Holyoak, K., & Lu, H. (2023). Emergent analogical reasoning in large language models. URL: <https://www.nature.com/articles/s41562-023-01659-w>.
- Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. The Technology Press and John Wiley & Sons, Inc., New York, 229 p. URL: <https://search.worldcat.org/title/320032-cybernetics-or-control-and-communication-in-the-animal-and-the-machine/oclc/255863>.
- Wiener, N. (1950). *The human use of human beings: Cybernetics and society*. Houghton Mifflin. Houghton Mifflin Company, Boston, 1950. 224 p. URL: <https://search.worldcat.org/title/human-use-of-human-beings-cybernetics-and-society/oclc/318388>.

References:

- Amodei, D. (2024). The adolescence of technology. Essay. URL: <https://www.darioamodei.com/essay/the-adolescence-of-technology> [in English].
- Anthropic. (2023–2026). Responsible scaling policy: Version 3.0. URL: <https://www.anthropic.com/news/responsible-scaling-policy-v3> [in English].
- Deng, J., Pang, S., et al. (2024). SOPHON: Non-fine-tunable learning to restrain task transferability for pre-trained models. IEEE Symposium on Security and Privacy (S&P). URL: <https://ieeexplore.ieee.org/document/10646821> [in English].
- Mumford, L. (1934). *Technics and civilization*. Harcourt, Brace and Co. URL: <https://search.worldcat.org/title/Technics-and-civilization/oclc/560667> [in English].
- Pang, J. C., Aquino, K. M., et al. (2023). Geometric constraints on human brain function. *Nature*, 618, 566–574. URL: <https://www.nature.com/articles/s41586-023-06098-1> [in English].
- Suleyman, M. (2023). The coming wave: Technology, power, and the twenty-first century's greatest dilemma. Crown. URL: https://www.researchgate.net/publication/387730680_The_Coming_Wave_Technology_Power_and_the_21st_Century's_Greatest_Dilemma_By_Mustafa_Suleyman_With_Michael_Bhaskar_New_York_Crown_2023 [in English].
- Toffler, E. (2000). *Tretia Khvylia [The third wave]* (A. Yevsa, Trans.). Vyd. dim “Vsesvit”. 480 p. URL: <https://lib.in.ua/100507-tretia-hvylia-vyd-2000/> [in Ukrainian].

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. URL: https://www.researchgate.net/publication/263384944_Alan_Turing's_Computing_Machinery_and_Intelligence [in English].

Webb, T., Holyoak, K., & Lu, H. (2023). Emergent analogical reasoning in large language models. URL: <https://www.nature.com/articles/s41562-023-01659-w> [in English].

Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. The Technology Press and John Wiley & Sons, Inc., New York, 229 p. URL: <https://search.worldcat.org/title/320032-cybernetics-or-control-and-communication-in-the-animal-and-the-machine/oclc/255863> [in English].

Wiener, N. (1950). *The human use of human beings: Cybernetics and society*. Houghton Mifflin. Houghton Mifflin Company, Boston, 1950. 224 p. URL: <https://search.worldcat.org/title/human-use-of-human-beings-cybernetics-and-society/oclc/318388>. [in English].

Дата першого надходження статті до видання: 25.03.2026

Дата прийняття статті до друку після рецензування: 20.04.2026

Дата публікації (оприлюднення) статті: 29.05.2026