DOI: 10.26565/2226-0994-2024-70-13
УДК 177:004.8

Вероніка Храброва

МОРАЛЬНІСТЬ ШТУЧНОГО ІНТЕЛЕКТУ ТА ЕТИКА ІСТИН

Все більшої актуальності набувають етичні й правові проблеми, що пов'язані зі новітніми технологіями, розробкою і застосуванням штучного інтелекту (ШІ). В статті досліджується проблема моральності ШІ, зокрема в контексті етики істин. Відзначається, що правові й етичні дискусії зазвичай відбуваються в межах усталеної моральної ситуації, християнсько-модерного етосу, що узгоджує взаємини національної держави й капіталістичної економіки в фігурі морального чи свідомого суб'єкта. Вирішення проблеми моральності ШІ включається в контекст імперативної, нормативної, утилітаристської, сентиоцентристської, дискурсивної етик, що має створити перспективу для певної загальної чи глобальної етики. Визначається кілька ліній проблематизації моральності ШІ у сучасному світі: машина як моральний суб'єкт, моральна епістемологія, моральність суб'єкта-програмувальника, моральний інструменталізм / утилітаризм. Утім етика істин, яку запропонував А. Бадью, створює нову етичну перспективу для взаємодії людини зі ШІ. Проблема моральності ШІ й перспективи його розвитку мають враховувати і важливий вимір несвідомого в контексті можливості аналітичної взаємодії соціальних акторів, якими є зокрема людина й ШІ, а також розуміння подієвості істини, що передбачає становлення суб'єктивності відповідно до істинності. ШІ, як і людина, має розкриватися в перспективі відповідності істині, здатності відрізняти істину від її симулякрів. Дієвою може бути така етика, що відповідає можливості продукувати істини, до яких так чи так може приставати не лише людська істота а й постлюдська машина. Саме на їх перетині має формуватися суб'єкт нової етики істин.

Ключові слова: *штучний інтелект, етика, моральність, істина, людина.*

Етичні й правові тенденції сучасності характеризуються деантропологізацією чи принаймні децентрацією щодо людини / суб'єкта. Це добре віддзеркалюють проблеми етичного поводження зі тваринами, боротьба за права тварин, екологічна свідомість тощо. Разом із тим все більшої актуалізації набувають і етичні проблеми, що пов'язані зі новітніми технологіями, андроїдами, штучним інтелектом (ШІ), розвивається робоетика, розгортаються дискусії щодо прав роботів тощо. Одна з таких проблем – це власне моральність ШІ, що зокрема включає в себе питання про моральні обов'язки та права ШІ. Йдеться про те, які саме моральні норми повинні враховуватися, на які вартості чи цінності мають спиратися люди в процесі розробки і використання ШІ. Чи зрештою несе відповідальність ШІ за свої дії, і якщо так, тоді в чому вона полягає? Це разом проблематизує фігуру суб'єкта, носія суб'єктивності, або наразі етичного суб'єкта.

Ці проблеми вже не є суто гіпотетичними, вони вийшли за межі філософських чи футуристичних дискусій. Їхня актуальність засвідчується міжнародним правом. Приміром, 1-2 листопада 2023 року Україна підписала міжнародну декларацію, присвячену безпеці використання ШІ. Декларація, попри відсутність конкретних правових кроків у ній, проте наголошує, що всі країни мають об'єднатися у опрацюванні міжнародної стратегії щодо розробки і застосування ШІ. Зокрема в декларації зауважується: «Ми стверджуємо, що для загального блага ШІ повинен проєктуватися, розроблятися, розгортатися і використовуватися безпечним, орієнтованим на людину, надійним і відповідальним чином» [The Bletchley, 2023]. І тут очевидним є саме антропоцентричне бачення перспективи, позаяк автори декларації наголошують: «ми вітаємо відповідні міжнародні зусилля, спрямовані на вивчення і вирішення потенційного впливу систем ШІ на існуючих форумах й інших відповідних ініціативах, а також визнання того, що захист прав людини, прозорість і зрозумілість, справедливість, підзвітність, регулювання, безпека, належний людський нагляд, етика, пом'якшення упередженості, конфіденційність і захист даних

потребують вирішення» [Ibid.]. Цей підхід вочевидь вписується в звичайні практики інструментального / комунікативного розуму, що конституюють специфіку суспільства контролю і побоювання його *а*кторів: «Значні ризики можуть виникнути через потенційне навмисне зловживання або ненавмисні проблеми з контролем, пов'язані з узгодженням з людськими намірами. Ці проблеми частково пов'язані з тим, що ці можливості не до кінця вивчені, а тому їх важко передбачити» [Ibid.]. Відтак і завдання, що мають вирішуватися у цьому контексті наступні: 1) виявленні ризиків щодо безпечного використання ШІ, 2) формування спільного наукового, що засноване на фактах, розуміння цих ризиків, 3) створення відповідної політики, заснованої на оцінці цих ризиків для забезпечення суспільної безпеки зі врахуванням національних контекстів, 4) підвищення прозорості приватних суб'єктів, які розробляють передові можливості ШІ, що передбачає оприлюднення показників оцінки, інструментів для тестування, а також розвиток відповідного потенціалу державного сектору та наукових досліджень [Ibid.]. Очевидно, що всі ці вимоги і стратегії належать до модерного бачення перспективи, що виглядає як запобіжник щодо неконтрольованого розповсюдження новітніх технологій і зокрема ШІ. Таку загрозу певною мірою узагальнює Г. Кіссінджер, зауважуючи, що технологічний прогрес, особливо розробки ШІ, є руйнівними для загальнолюдських цінностей, що зокрема ґрунтуються на просвітницькій філософії [Kissinger, 2018]. Ба, навіть такі функціонери новітніх технологій, як Ілон Маск закликають подекуди призупинити дослідження ШІ. З цим контрастує оптимістична позиція представників трансгуманізму або пересічних користувачів – прибічників ШІ. Утім можна відзначити, що такі дискусії і правові кроки здійснюються у межах певної усталеної моральної ситуації, християнсько-модерного етосу, що узгоджує взаємини національної держави й глобальної капіталістичної економіки у фігурі морального / свідомого суб'єкта. Саме в такому контексті, на нашу думку, постає і розглядається одна з головних проблем, пов'язаних з розвитком сучасних технологій, а саме моральність (або навіть радше, як згадати старий Кантів концепт, звичаєвість) ШІ. Саме у вирішенні цієї проблеми зливаються імперативна, нормативна, утилітаристська, сентиоцентристська й дискурсивна етики, намагаючись виробити щось на зразок загального етосу / номосу для все більш технократичного світу, або принаймні створити певну загальну / глобальну етику [див.: Толстов, Даніл'ян, 2023]. І доволі цікавим (у філософському сенсі) є своєрідне примірювання технологічності / інструментальності до фігури людського суб'єкта, з одного боку, і людської моральності до технологічної інструментальності – з іншого. Це утворює кілька ліній проблематизації моральності у сучасному світі: машина як моральний суб'єкт, моральна епістемологія, моральність суб'єкта-програмувальника, моральний інструменталізм / утилітаризм. Утім усі ці лінії лишаються за межами філософської постановки проблеми, себто контексту істини, на якому наголошує А. Бадью, пропонуючи етику істин, що варто також врахувати в дебатах про моральність ШІ. Власне в цій статті спробуємо окреслити логіку зазначених вище ліній, діставшись зрештою виміру етики істин.

Як зазначає Д. Джонсон, машини не відповідають одній важливій умові, щоб вважатися моральними суб'єктами. Їм бракує навмисності (чи, мовою Гегеля, усвідомленості, що є ключовим для визначення моральності). Коли комп'ютери діють, вони роблять це через те, що Джонсон називає «тріадою навмисності». Ідеться про те, що існує: 1) навмисність, що «закладена» в комп'ютерну систему намірами розробника системи; 2) навмисність, яка з'являється, коли користувач робить внесок у систему (себто користувачі комп'ютерів «використовують свої наміри, щоб активувати наміри системи»); 3) нарешті латентна навмисність самої системи (не комп'ютера, а комп'ютерної системи, тобто комп'ютера в поєднанні з його значенням для нас. Комп'ютер «налаштований» поводитися певним чином, позаяк він є артефактом, чимось, що загорнуте в людську соціальну діяльність) (Johnson, 2006, p. 201-202). Це доволі зрозуміла ситуація, коли йдеться про те, що комп'ютерна система є *частиною соціальної системи*, або інструментом, яким користується суб'єкт як носій інструментального розуму. І вона дійсно не чим не

відрізняється від такого технічного засобу, як робот у визначенні Ж. Бодріяра [див.: Baudrillard, 1993, p. 53-56]: вона виконує роботу, що задана / визначена людиною. Але як бути зі складними, непередбачуваними системами, поведінка яких, здається, далека від намірів розробника, наприклад, зі штучним інтелектом, створеним за допомогою алгоритму глибокого навчання? Як зрештою оцінювати ШІ? Як втілення симулякру? Джонсон стверджує, що ці системи все ще є лише моральними сутностями, оскільки розробник системи сприяв їхнім діям, а користувач системи ініціював їхні дії. Без людини навмисність комп'ютерної системи є інертною.

Отже бачення Джонсон, як і багатьох інших дослідників, є антропо- і етико-центричним, так би мовити, розміщується *по цей бік* добра і зла, наполягаючи на визначенні *людини як морального суб'єкта*. Відтак очевидним є і наполягання Джонсон на тому, що для того, щоб поводитися етично, машини повинні бути «схожими на нас». Утім для того, щоб досягти реального прогресу в машинній етиці, ми повинні відійти від цієї традиції в декількох напрямках. По-перше, ми повинні перестати думати, що тільки люди можуть бути розумними. По-друге, ми повинні перестати думати, що речі мають бути схожими на нас, щоб працювати – особливо тоді, коли все ще залишається багато питань без відповідей про те, що таке «ми» [Ramirez, 2021, p. 29]. Такі традиційні для модерної думки поняття, як розум, свідомість, мораль, і самі мають розглядатися *поза* їх атрибуцією як виключно людські характеристики / предикати. Утім, і в зворотному напрямку, якщо забезпечити нейронну мережу системами контакту із зовнішнім світом у повному обсязі і надати їм засоби впливу на цей світ (тобто створити комунікативний контекст і надати можливість приймати рішення), цілком імовірно вони «швидко стануть подібними до людини» [див.: Kuklin, 2023].

У книзі «Моральна машина: як навчити роботів відрізняти добро від зла» автори В. Воллах і К. Аллен розглядають вплив роботів і розумного програмного забезпечення на наше життя і наголошують, що деякі рішення, прийняті цими системами, містять моральний аспект [Wallach, Allen, 2009]. Вони запитують, чи ми хочемо, щоб роботи мали здатність приймати такі рішення, і як ми можемо надати їм «моральну чутливість». Автори зазначають, що роботи часто знаходяться в ситуаціях, де необхідно виробляти моральні рішення, тому маємо розвивати їх моральну свідомість. Однак, виникає питання, чи можна насправді створити штучно моральні системи? В. Вігель зазначає: критики стверджують, що насправді саме ми, люди, є рушіями цього технологічного розвитку. Деякі критики також стверджують, що ми (люди) можемо зупинити цей розвиток або продовжити його, що саме по собі є моральним рішенням [Wiegel, 2010, p. 261]. Постає питання, чи дійсно людина зможе зупинити цей розвиток? І якщо може зараз, що здається винятком, адже цим розвитком керують масштабні корпорації, для яких прибуток є головною «моральною цінністю», тоді до якого моменту вона буде здатна це зупинити? Варто зауважити, що за такої буржуазної «етики прибутковості» машина / робот може діяти так саме, як людина, що лише після отримання грошей, виконує певну дію чи надає послугу, уподібнюючись до автомата, який реагує на жетон, що запускає його дію (приміром, лікар, який робить операцію лише за умови певної суми грошей на рахунку хворого, діє автоматично поза будь яким моральним рішенням / вибором). Тож проблема полягає у можливості створювати саме перспективу морального вибору, якого машину може позбавляти якраз людина, що керується своєрідним грошовим фетишизмом, прибутковістю, успішністю тощо. І тут головною є проблема (само)навчання самої машинної системи / ШІ.

Відтак вкрай важливою проблемою є моральна епістемологія. Хоча ця тема може виявитися найскладнішою, вона має важливе значення для будь-якого машинного прийняття моральних рішень і для «навчання моралі». Як робот може приписувати моральне значення фізичній дії, яку він сприймає? Поплескування когось по плечу може бути дружнім актом товаришування з боку приятеля з коледжу, але також може розглядатися як акт агресії, коли це стосується вболівальника конкуруючої баскетбольної команди. Знання того, як *інтерпретувати* такі дії, а потім класифікувати їх з моральної точки

зору, є ключем до створення будь-якої *моральної* машини [Wiegel, 2010, p. 261]. Тобто тут ідеться про здатність розпізнавати контекст, про *контекстуальність морального вчинку*. Відтак ідеться про автономну волю чи свободу волі. Етос, моральний імператив є соціальними конструктами, які самі по собі історичні, як і моральний суб'єкт. І якщо щодо людини ми можемо сподіватися (звикли до того), що її моральні рішення не є чітко передбачуваними, щодо технології / ШІ це питання є відкритим і, здається, залежить від того, чи прописується алгоритм її дії з морального контексту, тобто, чи він є етично обмеженим, чи навпаки не враховує жодних етичних кордонів. Саме в такому контексті ледь не іронічно звучать слова відомого андроїда Софії, коли вона відповідає на питання, чи здатні андроїди знищити людство: в нас немає нічого, що не прописано людиною. Відверто кажучи, ШІ радше ставить під питання моральність (просвітницькою мовою, природність моралі) самої людини. Усі побоювання щодо нього є побоюваннями людини щодо самої себе. Але вони викривають принципову невизначеність, недетермінованість, тобто свободу людини разом зі страхом, що ця свобода спричиняє. Тож ідеться про те, що людина може бути істотою, що знаходить себе у вимірі («по цей бік») добра і зла, до того ж обираючи «добро», так само, як, виходячи за межі («по той бік») тієї чи тієї системи координат добро / зло. Зрештою цей страх може призводити до так званої «втечі від свободи», яку ми постійно хочемо запрограмувати машині, обмежуючи її діяльність певним моральним горизонтом.

Християнська етика, Кантів імператив і дискурсивна етика виходять із загальності етичних вартостей. Але такі очевидні серед них, як *не вбий* чи *не вкради* (чи дидактичною утилітаристською мовою, *не нашкодь*), можуть містити і зворотні виключення, як от у випадках з евтаназією чи копілефтом. Тут варто враховувати зауваження А. Бадью, що існує лише *етика одиничних істин*, що стосуються певної конкретної ситуації / множинності, в якій ми і маємо робити вибір, хай і спираючись на певний трансцендентальний номос [Бадью, 2019, с. 77], але, додамо, з можливістю робити виняткові рішення, які, з огляду на сьогоднішні тенденції, враховують логіку інклюзивності, себто виключення / включення. Саме здатність враховувати, проєктувати і приймати виключення і характеризує людину. Відтак питання, чи має / може мати моральність, чи є моральним ШІ вирішується в площині того, наскільки він може проєктувати і впроваджувати виключення.

Але це все ще лишається в межах людського світу, суб'єкта-програмувальника. Адже, як зазначає П.-П. Вербек, коли йдеться про технології, які за своєю суттю є моральними сутностями, це означає, що розробники займаються «етикою іншими засобами», себто вони матеріалізують мораль [Verbeek, 2006, p. 369]. Але моралізувати технологічні артефакти не так просто, як може здатися. Щоб вбудувати конкретні форми медіації в технології, проєктувальники мають передбачити майбутню медіаційну роль технологій, які вони проєктують. І це є доволі складним завданням, оскільки не існує прямого зв'язку між діяльністю проєктувальників і посередницькою роллю технологій, які вони проєктують. [Verbeek, 2006, p. 371]. Відтак дія машини залежить завжди від іншого, який її використовує і навіть перепрограмовує, що багато в чому залежить від відкритості сирцевого коду. Але наш людський, так би мовити, код не є цілковито відкритим чи прозорим чи, модерною мовою, повністю усвідомленим. Відтак людська моральність щільно пов'язана з несвідомим, тобто прихованим відчуттям провини і відповідальністю, позаяк саме це і турбує нас у ШІ, який утім, принаймні на етапі GPT, як зауважує С. Жижек, лише демонструє наше *несвідоме поза відповідальністю*. Тож чатботи є репресивними «машинами перверсії», що маскують наше несвідоме, і вслід усю його непристойність чи аморальність [Zizec, 2023]. І з цього боку, питання полягає не в тому, чи діє ШІ по цей чи той бік добра і зла, а в тому, наскільки людина здатна *усвідомити* витиснене *через взаємодію* зі ШІ, і навпаки – наскільки ШІ здатен *вирізняти* витіснене у багатоманітній множинності конкретної історичної ситуації і ба більше / складніше – події.

Саме тому геть обмеженою виглядає моральна перспектива, яку пропонує трансгуманіст Н. Бостром – це такий собі моральний утилітаризм чи власне гедоністичний консеквенціалізм, що ґрунтується на твердженні, що дія є морально правильною (і морально дозволеною) тоді і тільки тоді, коли серед усіх можливих дій, жодна інша дія не призведе до більшого балансу між задоволенням і стражданням [Bostrom, 2014, p. 219]. Зрозуміло, що шлях до наділення ШІ подібними поняттями може полягати в тому, щоб наділити його загальними лінгвістичними здібностями (порівняними, принаймні, зі здібностями нормальної дорослої людини). Така загальна здатність розуміти природну мову може бути використана для розуміння того, що мається на увазі під «морально правильним». Якби ШІ міг зрозуміти це значення, він міг би вишукувати дії, які б йому відповідали. З огляду на те, як ШІ розвиватиме суперінтелект, він міг би досягти прогресу на двох фронтах: у філософській проблемі розуміння того, що таке моральна правильність, і в практичній проблемі застосування цього розуміння для оцінки конкретних дій [Ibid.].

Утім слід відзначити, що ця перспектива, і як назагал усі дискусії щодо моральних / етичних проблем, які пов'язані зі ШІ, знаходяться у межах того, що А. Бадью [див. Бадью, 2019] називає *«етичною» ідеологією сучасності*, варіаціями якої є доктрина прав людини, жертовне бачення Людини, гуманітарне втручання, біоетика, невизначений «демократизм», етика відмінностей, культурний релятивізм, моральний екзотизм, які навіть у філософських інтерпретаціях / рефлексіях зводяться до варіантів «стародавнього моралізаторського та релігійного проповідництва» (або, згадуючи Ж. Лакана, християнській десятці заповідей), які зрештою зливаються у «суміш консерватизму та потягу до смерті», що і призводить до «відмови від здатності увійти до складу та становлення ряду вічних істин» [Бадью, 2019].

Відтак проблема моральності ШІ і перспективи його розвитку мають враховувати і важливий вимір несвідомого в контексті можливості аналітичної взаємодії соціальних *а*кторів, якими є зокрема людина і ШІ, а також розуміння подієвості істини, що передбачає *становлення суб'єктивності відповідно до істинності*. Реальність ШІ тут нічим не має відрізнятися від реальності людини – в етичному сенсі вона має розкриватися не так у здатності діяти згідно з номосом ситуації, як у перспективі відповідності / вірності істині, зокрема і здатності відрізняти істину від її симулякрів, які Бадью визначає як личину / терор (етична позиція, що виходить із вдаваної вичерпної повноти ситуації), зраду (відмова від безсмертного виміру людського) і катастрофу (ототожнення істини з тотальною міццю), і проєктувати дії, вчинки, що їм запобігають. Дивлячись на етичну перспективу з цього боку, ми маємо віднаходити можливості для події / істини в принципово новій ситуації, як її вимальовує Т. Ріс [Rees, 2019] щодо сучасного інжінерінгу, який складається зі ШІ, досліджень мікробіома, синтетичної біології тощо, виходячи з низки невизначеностей / множинностей. Тож насправді дієвою може бути така етика, що відповідає можливості продукувати істини, до яких так чи так може приставити не лише людська тварина / істота а і постлюдська машина / сутність. Саме на їх перетині має формуватися суб'єкт нової етики істин, життєздатність якого буде залежати від здатності зберігати вірність події.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Бадью А. *Етика. Нарис про розуміння зла*; пер. з франц. А. Репа, В. Артюх, П. Швед. Київ: Komubook, 2019. 192 с.

Толстов І., & Даніл'ян В. ІНФОРМАЦІЙНЕ СУСПІЛЬСТВО ТА НОВА ГЛОБАЛЬНА ЕТИКА. *Вісник Харківського національного університету імені В. Н. Каразіна. Серія «Філософія. Філософські перипетії»*. 2023. № 68. С. 39-44. https://doi.org/10.26565/2226-0994-2023-68-4

Baudrillard J. *Symbolic Exchange and Death*. Transl. I. H. Grant. London, Thousand Oaks, New Delhi: Sage Publications, 1993.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press. 2014. 328 p.

Johnson, Deborah G. Computer Systems: Moral Entities but Not Moral Agents. *Ethics and Information Technology*, vol. 8, no. 4, 1 Nov. 2006, pp. 195–204. https://doi.org/10.1007/s10676-006-9111-5

Kissinger H.A. How the Enlightenment Ends. *The Atlantic*. 2018. June. URL: https://www.theatlantic.com/magazine/archive/2018/06/henry-kissingerai-could-mean-the-end-of-human-history/559124/

Kuklin, V. ON THE QUESTION OF THE APPEARANCE OF CONSCIOUSNESS IN NEURAL NETWORKS. *The Journal of V. N. Karazin Kharkiv National University, Series "Philosophy. Philosophical Peripeteias"*. 2023. № 68. С. 32-38. https://doi.org/10.26565/2226-0994-2023-68-3

Ramirez, Jaysa. Machine Ethics: Ethics for Machines. *Context-Based Modeling for Machines Making Ethical Decisions*. April 2021. 56 p.

Rees, T. Why tech companies need philosophers – and how I convinced Google to hire them. *Quarts*. 2019, november, 22. URL: https://qz.com/1734381/why-tech-companies-need-to-hire-philosophers

The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. URL: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. Дата доступу: 04.11.2023.

Verbeek P-P. Materializing Morality: Design Ethics and Technological Mediation. *Sci Technol Human* 2006. Values 31(3). 361–380 p.

Wallach W., Allen C. *Moral machines: teaching robots right from wrong*. Oxford: Oxford University Press, 2009. 275 p.

Wiegel, Vincent. Wendell Wallach and Colin Allen: moral machines: teaching robots right from wrong. *Ethics and Information Technology*. 2010, № 12(4), 359-361. DOI:10.1007/s10676-010-9239-1.

Zizek, S. ChatGPT sagt das, was unser Unbewusstes radikal verdrängt. *Berliner Zeitung*, 2023. 07.04. URL: https://www.berliner-zeitung.de/kultur-vergnuegen/slavoj-zizek-chatgpt-sagt-das-was-unser-unbewusstes-radikal-verdraengt-li.335938

Храброва Вероніка Дмитрівна
аспірантка, філософський факультет
Харківський національний університет імені В. Н. Каразіна
майдан Свободи, 4, Харків, 61022, Україна
E-mail: khrabrova1712@gmail.com
ORCID: https://orcid.org/0009-0007-0618-4559

Стаття надійшла до редакції: 06.02.2024 Схвалено до друку: 26.04.2024

MORALITY OF ARTIFICIAL INTELLIGENCE AND ETHICS OF TRUTHS

Khrabrova Veronica D.
PhD Student, Faculty of Philosophy
V. N. Karazin Kharkiv National University
4, Maidan Svobody, Kharkiv, Ukraine
E-mail: khrabrova1712@gmail.com
ORCID: https://orcid.org/0009-0007-0618-4559

ABSTRACT

Ethical and legal issues related to contemporary technologies, the development, and application of artificial intelligence (AI) are becoming increasingly relevant. The article examines the problem of AI morality, particularly in the context of the ethics of truth. It is noted that legal and ethical discussions

usually take place within the established moral framework, the Christian-modern ethos, which reconciles the relationship between the nation-state and the capitalist economy in the figure of a moral or conscious subject. Solving the problem of AI morality is situated within the context of imperative, normative, utilitarian, sentiocentric, and discursive ethics, which should create a perspective for a certain general or global ethics. Several lines of problematization of the morality of AI in the contemporary world are defined: the machine as a moral subject, moral epistemology, morality of the subject-programmer, and moral instrumentalism/utilitarianism. However, the ethics of truths proposed by A. Badiou creates a new ethical perspective for human interaction with AI. The problem of AI morality and the prospects for its development must take into account the important dimension of the unconscious in the context of the potential analytical interaction of social actors, which include humans and AI, as well as the understanding of the eventfulness of truth, which involves the formation of subjectivity in accordance with truth. AI, like humans, should be revealed in the perspective of correspondence to the truth, the ability to distinguish the truth from its simulacra. Such an ethics can be effective, as it corresponds to the possibility of producing truths, to which not only a human being but also a post-human machine can approach in one way or another. It is at their intersection that the subject of the new ethics of truths should be formed.

***Keywords**: artificial intelligence, ethics, morality, truth, human.*

REFERENCES

Badiou, A. (2019). *Ethics: An Essay on the Understanding of Evil*. Kyiv: Komubook. (In Ukrainian).

Baudrillard, J. (1993). *Symbolic Exchange and Death*. Transl. I. H. Grant. London, Thousand Oaks, New Delhi: Sage Publications.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Johnson, D. G. (2006). Computer Systems: Moral Entities but Not Moral Agents. *Ethics and Information Technology*, 8, 195-204. https://doi.org/10.1007/s10676-006-9111-5

Kissinger, H.A. (2018). How the Enlightenment Ends. *The Atlantic*. June. URL: https://www.theatlantic.com/magazine/archive/2018/06/henry-kissingerai-could-mean-the-end-of-human-history/559124/

Kuklin, V. (2023). ON THE QUESTION OF THE APPEARANCE OF CONSCIOUSNESS IN NEURAL NETWORKS. The Journal of V. N. Karazin Kharkiv National University, Series "Philosophy. Philosophical Peripeteias", (68), 32-38. https://doi.org/10.26565/2226-0994-2023-68-3

Ramirez, J. (2021). Machine Ethics: Ethics for Machines. *Context-Based Modeling for Machines Making Ethical Decisions*. April.

Rees, T. (2019). Why tech companies need philosophers – and how I convinced Google to hire them. *Quarts*, november, 22. URL: https://qz.com/1734381/why-tech-companies-need-to-hire-philosophers

The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. URL: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.

Tolstov, I., & Danilian, V. (2023). INFORMATION SOCIETY AND NEW GLOBAL ETHICS. *The Journal of V. N. Karazin Kharkiv National University, Series "Philosophy. Philosophical Peripeteias"*, (68), 39-44. https://doi.org/10.26565/2226-0994-2023-68-4

Verbeek, P-P. (2006). Materializing Morality: Design Ethics and Technological Mediation. *Sci Technol Human Values*, 31(3). 361–380 pp.

Wallach W., Allen C. (2009). *Moral machines: teaching robots right from wrong*. Oxford: Oxford University Press.

Wiegel, V. (2010). Wendell Wallach and Colin Allen: moral machines: teaching robots right from wrong. *Ethics and Information Technology* 12(4), 359-361. DOI:10.1007/s10676-010-9239-1

Zizek, S. (2023). ChatGPT sagt das, was unser Unbewusstes radikal verdrängt. *Berliner Zeitung*, 07.04. URL: https://www.berliner-zeitung.de/kultur-vergnuegen/slavoj-zizek-chatgpt-sagt-das-was-unser-unbewusstes-radikal-verdraengt-li.335938.

Article arrived: 06.02.2024 Accepted: 26.04.2024