

ТЕОРІЯ ТА ФІЛОСОФІЯ ПРАВА

ОРЛОВСЬКИЙ ІЛЛЯ АНДРІЙОВИЧ

аспірант кафедри філософії Національного юридичного університету імені Ярослава Мудрого
ORCID 0009-0001-1738-6785

УДК 340.12:[004.8:001.101]

DOI 10.37772/2518-1718-2026-1(53)-11

ПРИНЦИПИ ШТУЧНОГО ІНТЕЛЕКТУ В ПРАВІ

У статті обґрунтовується, що для того, щоб уникнути шкоди, особливо в міру того, як системи штучного інтелекту на різних рівнях все більш інтегруються в соціум і безпосередньо співіснують з користувачами, необхідність вбудовування етичних та правових норм в автономних агентів стає все більш вираженою. Наголошено, що оскільки автономні системи вже не розглядаються лише як інструменти, а поступово переходять до функцій, які раніше могли виконувати тільки люди, навіть, до ролі опікунів та інтерактивних агентів, доречність їхніх дій та вибору необхідно включає міркування нормативного характеру. Це передбачає нормативну чутливість та певний рівень прийняття нормативних рішень, наслідки яких є далекосяжними. Зазначено, що для забезпечення безпечного та надійного функціонування автономних систем штучного інтелекту важливо звертати пильну увагу на правовий, етичний та культурний контекст, у якому вони існують. Вказано широкий спектр етичних принципів та цінностей, які представлені різними авторитетними організаціями й на які слід спиратися при розробці та впровадженні автономних роботів, а також правових принципів, серед яких зазначені право на приватність, повагу до людської гідності, прозорість та право на належну правову процедуру, право на інформацію, право на самовизначення та недискримінацію, а також соціально-економічні права, права на безпеку та соціальне забезпечення. Наголошено, що для того, щоб уникнути шкоди, особливо в міру того, як технології штучного інтелекту стають все більш інтегрованими на більш тісних рівнях з їхніми користувачами, необхідно не тільки здійснити узагальнення базових принципів штучного інтелекту, які можна звести до робочих правил, яких буде дотримуватися автономний агент, а розробити методи вбудовування соціальних норм в автономних агентів. Зазначено, що ці норми часто виражаються як абстрактні високорівневі принципи, які нелегко звести до робочих правил, яких має дотримуватися автономна система. Припущено, що один із методів впровадження цих високорівневих нормативних принципів у робочі правила поведінки автономної системи, якому можна довіряти таким чином, щоб це задовольнило користувачів, полягає в поєднанні обов'язкових вимог і добровільних стандартів.

Ключові слова: право та технології; принципи штучного інтелекту; етика штучного інтелекту; правове регулювання штучного інтелекту, штучна соціальність.

Постановка проблеми. Технологічні досягнення, які ми спостерігаємо останнім часом, дозволили автономним системам штучного інтелекту ставати дедалі складнішими. Це надає великі переваги як для окремих користувачів, так і для суспільства. Однак, для цього важливо, щоб такі системи не порушували соціальних норм, які є «фундаментальними принципами, що регулюють питання того, як ми повинні жити та що ми повинні робити морально» [1, с. 32] або «звичайними правилами, що регулюють поведінку в групах та суспільствах» [2].

Враховуючи те, що такі системи перестають виконувати лише інструментальні функції, а їхні ролі, дії та вибір розширюються, вони безпосередньо співіснують з користувачами, часто в особистому про-

сторі та беруть участь у особистій взаємодії з користувачами, то дедалі стає важливішим створення та впровадження систем, чутливих до правових, етичних, традиційних, корпоративних норм. Це пояснюється тим, що системи такого типу не функціонують у нормативному вакуумі, а існують у певному соціальному середовищі. Оскільки автономні системи вже не розглядаються лише як інструменти, а поступово переходять до функцій, які раніше могли виконувати тільки люди, навіть, до ролі опікунів та інтерактивних агентів, доречність їхніх дій та вибору необхідно включає міркування нормативного характеру. Це передбачає нормативну чутливість та певний рівень прийняття нормативних рішень, наслідки яких є далекосяжними.

Формулювання цілей. Метою дослідження є розгляд базових принципів штучного інтелекту, які можна звести до робочих правил, яких буде дотримуватися автономний агент, і які будуть залишатися відповідними сучасним вимогам правового регулювання.

Аналіз останніх досліджень і публікацій. Наукові праці про необхідність вбудовування етичних та правових норм в автономні системи є міждисциплінарними, які розробляють сучасні філософи, інженери, правознавці. Принципи штучного інтелекту, які спрямовані на запобігання зловживанням та сприяння сталому розвитку протягом усього його життєвого циклу, зосереджені на розробці безпечних, справедливих, прозорих та підзвітних систем, що приносять користь людству. Самі принципи можуть стати аргументом на підтримку тези, що системі можна довіряти в її роботі відповідно до очікуваної поведінки.

Останні розробки в галузі машинного навчання та інженерії управління стрімко наближають час, коли роботи зможуть забезпечувати догляд та підтримку людей у їхньому повсякденному житті [3; 4], що не лише потребує відповідного нормативного регулювання, а й особливої уваги до подолання розриву між нормативними принципами та практикою. Питанням моралі і технології присвячені дослідження Б. Латура, К. Венна [5] та ін., створенню етичного інтелектуального агента – Л. Флориді, Дж. Коулз, М. Белтраметті [6; 7] та ін., ціннісно-орієнтованому дизайну автономних систем штучного інтелекту для забезпечення принципів соціального блага – С. Умбрелло та І. Ван де Поел [8] та ін.

Широкий спектр етичних принципів та цінностей, на які слід спиратися при розробці та впровадженні автономних роботів представлено різними авторитетними організаціями. Зокрема, ці принципи були сформульовані Європейською Комісією в 2019 р. в Етичних рекомендаціях для надійного штучного інтелекту [9], в Монреальській декларації про відповідальний штучний інтелект, яка розроблена під егідою Монреальського університету за результатами Форуму з питань соціально відповідального розвитку штучного інтелекту в листопаді 2017 р., в принципах штучного інтелекту ASILOMAR, які розроблені під егідою Інституту майбутнього життя у співпраці з учасниками конференції високого рівня ASILOMAR у січні 2017 р.

В Україні регулювання штучного інтелекту знаходиться на початковому етапі. В 2020 р. в нашій країні була схвалена Концепція розвитку штучного інтелекту в Україні [10], що має на меті запропонувати підхід до регулювання технологій штучного інтелекту, а в грудні 2024 р. був ухвалений Добровільний кодекс поведінки з етичного та відповідального використання штучного інтелекту, який скеро-

вує компанії на впровадження етичних принципів у внутрішні процеси, передбачає оцінку ризиків і адаптивне впровадження заходів та створення саморегуляторного органу для підтримки та обміну досвідом. Першу версію Стратегії розвитку штучного інтелекту України до 2030 року було представлено в листопаді 2025 р. на WINWIN Summit. Це покроковий план, як за допомогою технології покращити державне управління, освіту, оборону, медицину та інші сфери життя в Україні.

Українські науковці в основному присвячують свої дослідження аналізу актуальних юридичних питань розробки та впровадження штучного інтелекту (С. Вознюк, В. Гришко, Н. Руденко, Ж. Удовенко та ін.) [11; 12], визначенню меж використання технології в різних галузях (Д. Белов М., М. Белова, Я. Берназюк та ін.) [13; 14], захисту персональних даних, зібраних системами штучного інтелекту, визначенням правил щодо їх збереження та використання тощо (І. Бочкова, К. Врублевська-Місюна та ін.) [15]. При цьому аналізі принципів штучного інтелекту в праві, згідно з якими високоавтономні системи штучного інтелекту повинні бути розроблені таким чином, щоб їхні цілі та поведінка могли гарантовано відповідати людським цінностям через їхню роботу та бути сумісними з ідеалами людської гідності, прав, свобод та культурного різноманіття, також з можливістю цьому довіряти, у вітчизняних наукових джерелах не приділяється достатньої уваги.

Виклад основного матеріалу. Впровадження штучного інтелекту несе водночас як переваги, так і загрози для світу. Ця технологія швидко стала невід’ємною частиною повсякденного життя людини на роботі, включно з автоматизацією повторюваних завдань, пришвидшенням процесів та зменшенням ймовірності людських помилок, а також у побуті, позбавляючи від домашньої рутини, наприклад, завдяки використанню автономних роботів для прибирання оселі. За статистикою понад мільярд людей на планеті використовують інструменти штучного інтелекту [16]. Доля підприємств, які впровадили технології штучного інтелекту за останній рік зросла на 10% – з 78% до 88% [17].

З огляду на швидкі темпи впровадження штучного інтелекту в різні сфери нашого життя, уряди намагаються терміново врегулювати ці технології, оскільки злочинці все частіше використовують штучний інтелект для незаконної діяльності, зокрема найчастіше для шахрайства та відмивання грошей. Щоб захищати основні права людей від зловживань, пов’язаних з цими технологіями, таких як дискримінація, порушення конфіденційності та непрозорі автоматизовані рішення, потрібне ефективне нормативне регулювання, яке базується на принципах штучного інтелекту та врахує те, що штучний інте-

лект може бути життєво важливим інструментом для тих, хто бореться з незаконною діяльністю, а також може мати численні легальні бізнес-застосування. Наприклад, технології штучного інтелекту можуть допомогти в боротьбі з відмиванням грошей різними способами, від автоматизації аналізу даних для виявлення підозрілої поведінки транзакцій до допомоги співробітникам з питань дотримання законодавства за допомогою помічників штучного інтелекту. Зокрема, системи боротьби з відмиванням грошей на основі цих технологій, які можуть навчатися на моделях фінансових злочинів для виявлення нових ризиків, дозволяють установам реагувати на загрози, перш ніж вони завдадуть серйозної шкоди [18].

Очікується, що діяльність та результати автономних роботів відповідатимуть принципам штучного інтелекту. Якщо узагальнити, то згідно з основними етичними рекомендаціями для надійного штучного інтелекту, системи штучного інтелекту повинні втілювати три базові принципи: бути законними, етичними та технічно надійними. З цими принципами узгоджуються ключові вимоги до того, що вважається надійним штучним інтелектом і які мають оцінюватися та враховуватися протягом усього життєвого циклу системи. До них належать: людська діяльність та нагляд; технічна надійність та безпека; конфіденційність та управління даними; прозорість; різноманітність, недискримінація та справедливість; екологічне та суспільне благополуччя; а також підзвітність. Корпорація Microsoft визначає шість принципів, які, на їхню думку, мають регулювати розвиток і використання штучного інтелекту, серед яких справедливість, надійність, конфіденційність і безпека, інклюзивність, прозорість, підзвітність. Ці принципи передбачають, що система штучного інтелекту має розподіляти можливості, ресурси та інформацію справедливо стосовно людей, які її використовують; функціонувати належним чином для людей у різних умовах і контекстах використання, включно з тими, для яких вона спочатку не була призначена; мати функції підтримки конфіденційності та безпеки; бути інклюзивною для всіх, незалежно від можливостей користувачів; зрозумілою для всіх людей щодо можливостей системи; мати систему нагляду з функціями підзвітності та керування.

Дослідники зазначають серед основних етичних принципів благодійність, оскільки автономний робот повинен прагнути принести користь користувачеві; незавдання шкоди, оскільки він повинен уникати заподіяння шкоди користувачеві як психологічно, так і фізично; автономію, оскільки особи мають ефективну здатність самостійно приймати рішення, які мають практичне значення для їхнього життя, і що система повинна поважати бажання користувача та отримувати згоду (або погодження) залежно від

обставин; та справедливість [19]. Деякі вчені вважають, що у автономних систем основні принципи мають включати ознаку пояснимості (Флориді та ін.) [6; 7]. Коли питання стосується автономної системи, яка призначена для догляду та підтримки людини, рамки основних принципів розширюються принципами сфери чеснот, таких як терпіння, турбота, доброта та толерантність, а також емпатичних норм, що заохочують просоціальну поведінку, таку як допомога, співчуття та співпраця [20]. Серед правових принципів називають право на приватність, повагу до людської гідності, прозорість та право на належну правову процедуру, право на інформацію, право на самовизначення та недискримінацію, а також соціально-економічні права, права на безпеку та соціальне забезпечення. Ці принципи можна розглядати як принципи високого рівня для виведення робочих правил для системи. При цьому одним із шляхів подолання розриву між нормативними принципами та практикою, для виведення робочих правил, є сенс розглянути принципи, що визначені документами з етики штучного інтелекту та відповідними нормами права, як оціночні стандарти. Це дозволить системі з набору можливих дій вибрати та виконати найбільш нормативно сприятливу дію в передбачуваному контексті, спираючись на низку основних нормативних принципів.

Висновки. Швидкий прогрес у розвитку автономних агентів може призвести до появи застосувань, які можуть значно покращити добробут, але водночас мають потенціал для заподіяння шкоди. Для забезпечення їх безпечного та надійного функціонування важливо звертати пильну увагу на правовий, етичний та культурний контекст, у якому вони існують. Оскільки правове регулювання технології знаходиться на початковому етапі, то для забезпечення подальшого прогресу норми мають орієнтуватися на етичні принципи, а не заборони і залишатися відповідними сучасним вимогам правового регулювання.

Щоб уникнути шкоди, особливо в міру того, як ці технології стають все більш інтегрованими на більш тісних рівнях з їхніми користувачами, необхідно не тільки здійснити узагальнення базових принципів штучного інтелекту, які можна звести до робочих правил, яких буде дотримуватися автономний агент, а й розробити методи вбудовування соціальних норм в автономних агентів. Ці норми часто виражаються як абстрактні високорівневі принципи, які нелегко звести до робочих правил, яких має дотримуватися автономна система. Один із методів впровадження цих високорівневих нормативних принципів у робочі правила поведінки автономної системи, якому можна довіряти таким чином, щоб це задовольнило користувачів полягає в поєднанні обов'язкових вимог і добровільних стандартів.

ЛІТЕРАТУРА

1. Driver J. Normative ethics. In Jackson, Frank & Smith, Michael. The Oxford handbook of contemporary philosophy. Oxford University Press UK. 2007. P. 31-62.
2. Bicchieri C., Ryan M., Sontuoso A. Social Norms. The Stanford Encyclopedia of Philosophy. Zalta E. N., Nodelman U. (eds.). 2023. URL: <https://plato.stanford.edu/archives/win2023/entries/social-norms/>.
3. Zhang F., Cully A., Demiris Y. Probabilistic real-time user posture tracking for personalized robot-assisted dressing. *IEEE Transactions on Robotics*. 2019. № 35(4). P. 873–888.
4. Coşar S., Fernandez-Carmona M., Agrigoroaie R., et al. Enrichment: Perception and interaction of an Assistive Robot for the Elderly at Home. *International Journal of Social Robotics*. 2020. № 12(3). P. 779–805.
5. Latour B., Venn C. Morality and technology. *Theory, culture & society*. 2002. № 19.5-6. P. 247-260.
6. Floridi L., Cowls J., Beltrametti M., et al. AI4People — An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*. 2018. № 28(4). P. 689-707. <https://doi.org/10.1007/s11023-018-9482-5>.
7. Floridi L., Cowls J., Beltrametti M. et al. An ethical framework for a good AI Society: Opportunities, risks, principles, and recommendations. In: *Ethics, Governance, and Policies in Artificial Intelligence* / [ed] Luciano Floridi. Springer Nature, 2021. P. 19-39. https://doi.org/10.1007/978-3-030-81907-1_3.
8. Umbrello S., Van de Poel I. Mapping value sensitive design onto AI for social good principles. *AI and Ethics*. 2021. № 1(3). P. 283-296.
9. European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. Ethics guidelines for trustworthy AI, Publications Office. 2019. URL: <https://data.europa.eu/doi/10.2759/346720>.
10. Про схвалення Концепції розвитку штучного інтелекту в Україні. Розпорядження Кабінету Міністрів України від 2 грудня 2020 р. № 1556-р. URL: <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text>.
11. Гришко В.І., Вознюк С.С. Проблемні аспекти впровадження штучного інтелекту у сфері юриспруденції. *Електронне наукове видання «Аналітично-порівняльне правознавство»*. № 2. 2024. С. 29–34. URL: <http://journal-app.uzhnu.edu.ua/article/view/302954/294963>.
12. Удовенко Ж.В., Руденко Н.В. Переваги та недоліки впровадження системи штучного інтелекту у правосуддя України. *Актуальні питання у сучасній науці*. 2023. № 4(10). С. 252–262. [https://doi.org/10.52058/2786-6300-2023-4\(10\)-252-262](https://doi.org/10.52058/2786-6300-2023-4(10)-252-262).
13. Белов Д.М., Белова М.В. Штучний інтелект в судочинстві та судових рішеннях, потенціал та ризики. *Науковий вісник Ужгородського національного університету. Серія: Право*. 2023. Вип. 78. Ч. 2. С. 315–320. <https://doi.org/10.24144/2307-3322.2023.78.2.50>.
14. Берназюк Я. Штучний інтелект у правосудді: конституційно-правовий аспект. *Електронне наукове видання «Аналітично-порівняльне правознавство»*. № 2. 2025. С. 130–136. <https://doi.org/10.24144/2788-6018.2025.02.16>.
15. Бочкова І.І., Врублевська-Місюна К.М. Доступ до даних та конфіденційність в аспекті штучного інтелекту. *Науковий вісник Ужгородського національного університету. Серія: Право*. 2025. Вип. 90. Ч. 3. С. 139–146. <https://doi.org/10.24144/2307-3322.2025.90.3.19>.
16. Kemp S. Digital 2026: more than 1 billion people use AI. *DataReportal*. Oct. 15, 2025. URL: <https://datareportal.com/reports/digital-2026-one-billion-people-using-ai>.
17. McKinsey. The state of AI in 2025: Agents, innovation, and transformation. *QuantumBlack*. Nov. 5, 2025. URL: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.
18. How AI is Revolutionizing Anti-Money Laundering and Compliance (2026). *The Sumsuiber*. Nov. 10, 2025. URL: <https://sumsub.com/blog/ai-in-anti-money-laundering-and-compliance/>.
19. Ross W., Stratton-Lake P. The right and the good. *Canadian Journal of Philosophy*. 2002. № 25. P. 571-594.
20. Olderbak S., Sassenrath C., Keller J., Wilhelm O. An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in psychology*. 2014. № 5. P. 1-14. <https://doi.org/10.3389/fpsyg.2014.00653>.

REFERENCES

1. Driver, J. (2007). Normative Ethics. In Frank Jackson & Michael Smith, The Oxford Handbook of Contemporary Philosophy. New York: Oxford University Press UK [in English].

2. Bicchieri, C., Ryan, M., Sontuoso, A. (2023). Social Norms. *The Stanford Encyclopedia of Philosophy*. Zalta, E. N., Nodelman, U. (eds.). URL: <https://plato.stanford.edu/archives/win2023/entries/social-norms/> [in English].
3. Zhang, F., Cully, A., & Demiris, Y. (2019). Probabilistic real-time user posture tracking for personalized robot-assisted dressing. *IEEE Transactions on Robotics*, 35(4), 873–888 [in English].
4. Coşar, S., Fernandez-Carmona, M., Agrigoroaie, R., et al. (2020). Enrichment: Perception and interaction of an Assistive Robot for the Elderly at Home. *International Journal of Social Robotics*, 12(3), 779–805 [in English].
5. LATOUR, Bruno; VENN, Couze. Morality and technology. *Theory, culture & society*, 2002, 19.5-6: 247-260 [in English].
6. Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5> [in English].
7. Floridi, Luciano & Cows, Josh & Beltrametti, Monica & Chatila, Raja & Chazerand, Patrice & Dignum, Virginia & Lütge, Christoph & Madelin, Robert & Pagallo, Ugo & Rossi, Francesca & Schafer, Burkhard & Valcke, Peggy & Vayena, Effy. (2021). *An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. https://doi.org/10.1007/978-3-030-81907-1_3 [in English].
8. Umbrello, S., & Van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, 1(3), 283-296 [in English].
9. European Commission: Directorate-General for Communications Networks, Content and Technology and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji. *Ethics guidelines for trustworthy AI, Publications Office*. 2019. URL: <https://data.europa.eu/doi/10.2759/346720> [in English].
10. Pro shkvalennia Kontseptsii rozvytku shtuchnoho intelektu v Ukraini. Rozporiadzhennia Kabinetu Ministriv Ukrainy vid 2 hrudnia 2020 r. № 1556-r. URL: <https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text>.
11. Hryshko, V. I., Vozniuk, S. S. (2024). Problemni aspekty vprovadzhennia shtuchnoho intelektu u sferi yurysprudentsii. *Elektronne naukove vydannia «Analytychno-porivnialne pravoznavstvo» – Electronic scientific publication “Analytical and Comparative Jurisprudence”*, 2, 29–34. URL: <http://journal-app.uzhnu.edu.ua/article/view/302954/294963> [in Ukrainian].
12. Udovenko, Zh. V., Rudenko, N. V. (2023). Perevahy ta nedoliky vprovadzhennia systemy shtuchnoho intelektu u pravosuddia Ukrainy. *Aktualni pytannia u suchasni nauki – Current issues in modern science*, 4(10), 252–262. [https://doi.org/10.52058/2786-6300-2023-4\(10\)-252-262](https://doi.org/10.52058/2786-6300-2023-4(10)-252-262) [in Ukrainian].
13. Bielov, D. M., Bielova, M. V. (2023). Shtuchnyi intelekt v sudochynstvi ta sudovykh rishenniakh, potentsial ta ryzyky. *Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu. Serii: Pravo – Scientific Bulletin of Uzhhorod National University. Series: Law*, 78(2), 315–320. <https://doi.org/10.24144/2307-3322.2023.78.2.50> [in Ukrainian].
14. Bernaziuk, Ya. (2025). Shtuchnyi intelekt u pravosuddi: konstytutsiino-pravovyi aspekt. *Elektronne naukove vydannia «Analytychno-porivnialne pravoznavstvo» – Electronic scientific publication “Analytical and Comparative Jurisprudence”*, 2, 130–136. <https://doi.org/10.24144/2788-6018.2025.02.16> [in Ukrainian].
15. Bochkova I.I., Vrublevska-Misiuna K.M. . 2025. Dostup do danykh ta konfidentsiinst v aspekti shtuchnoho intelektu. *Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu. Serii: Pravo – Scientific Bulletin of Uzhhorod National University. Series: Law*, 90(3), 139–146. <https://doi.org/10.24144/2307-3322.2025.90.3.19> [in Ukrainian].
16. Kemp, S. (2025). Digital 2026: more than 1 billion people use AI. *DataReportal*. Oct. 15, 2025. URL: <https://datareportal.com/reports/digital-2026-one-billion-people-using-ai> [in English].
17. McKinsey (2025). The state of AI in 2025: Agents, innovation, and transformation. *QuantumBlack*. Nov. 5, 2025. URL: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai> [in English].
18. How AI is Revolutionizing Anti-Money Laundering and Compliance (2026). *The Sumsuber*. Nov. 10, 2025. URL: <https://sumsub.com/blog/ai-in-anti-money-laundering-and-compliance/> [in English].
19. Ross, W., Stratton-Lake, P. (2002). The right and the good. *Canadian Journal of Philosophy*, 25, 571-594 [in English].
20. Olderbak, S., Sassenrath, C., Keller, J., Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. *Frontiers in psychology*, 5, 1-14. <https://doi.org/10.3389/fpsyg.2014.00653> [in English].

ORLOVSKY ILYA

Graduate Student of the Department of Philosophy of the Yaroslav Mudryi National Law University

PRINCIPLES OF ARTIFICIAL INTELLIGENCE IN LAW

Problem setting. The technological advances we have seen in recent years have allowed autonomous AI systems to become increasingly complex. This offers great benefits for both individual users and society. However, it is important that such systems do not violate social norms.

As autonomous systems are no longer viewed as mere tools, but are gradually taking on functions that previously could only be performed by humans, even to the role of guardians and interactive agents, the appropriateness of their actions and choices necessarily includes normative considerations. This implies regulatory sensitivity and a certain level of regulatory decision-making, the consequences of which are far-reaching.

Analysis of recent researches and publications. A wide range of ethical principles and values to be relied upon in the development and deployment of autonomous robots has been presented by various authoritative organizations. In particular, these principles were formulated by the European Commission in 2019 in the Ethics Guidelines for Trustworthy AI, in the Montreal Declaration for a Responsible Development of AI, developed under the auspices of the University of Montreal following the Forum on the Socially Responsible Development of AI in November 2017, and in the ASILOMAR AI Principles, developed under the auspices of the Future of Life Institute in cooperation with participants of the high-level ASILOMAR Conference in January 2017.

In Ukraine, the regulation of AI is at an early stage. In 2020, the Concept for the Development of AI in Ukraine was approved, aiming to propose an approach to regulating artificial intelligence technologies. In December 2024, the Voluntary Code of Conduct on the Ethical and Responsible Use of AI was adopted, guiding companies toward the implementation of ethical principles in their internal processes, providing for risk assessment and adaptive implementation of measures, as well as the establishment of a self-regulatory body to support cooperation and exchange of experience. The first version of the AI Development Strategy of Ukraine until 2030 was presented in November 2025 at the WINWIN Summit. It outlines a step-by-step plan on how technology can improve public administration, education, defense, healthcare, and other areas of life in Ukraine.

Ukrainian scholars primarily focus their research on the analysis of current legal issues related to the development and implementation of AI (S. Vozniuk, V. Hryshko, N. Rudenko, Zh. Udovenko, etc.), the determination of the limits of technology use in various fields (D. Bielov, M. Bielova, Ya. Bernaziuk, etc.), and the protection of personal data collected by AI systems, including the establishment of rules for their storage and use (I. Bochkova, K. Vrublevska-Misiuna, etc.). At the same time, insufficient attention is paid in domestic academic sources to the analysis of AI principles in law, according to which highly autonomous AI systems should be designed in such a way that their goals and behavior are reliably aligned with human values through their operation and remain compatible with the ideals of human dignity, rights, freedoms, and cultural diversity, while also being trustworthy.

The purpose of this research is to consider the basic principles of artificial intelligence, which can be reduced to working rules that an autonomous agent will follow, and which will remain compliant with modern legal requirements.

Article's main body. The article substantiates that, in order to prevent harm – especially as artificial intelligence systems at various levels become increasingly integrated into society and directly coexist with users – the need to embed ethical and legal norms into autonomous agents is becoming ever more pronounced. It is emphasized that since autonomous systems are no longer regarded merely as tools but are gradually assuming functions that previously could be performed only by humans – even taking on the roles of caregivers and interactive agents – the appropriateness of their actions and choices necessarily involves normative considerations. This presupposes normative sensitivity and a certain level of normative decision-making, the consequences of which are far-reaching.

It is noted that ensuring the safe and reliable functioning of autonomous artificial intelligence systems requires careful attention to the legal, ethical, and cultural context in which they operate. A broad range of ethical principles and values presented by various authoritative organizations, which should serve as a foundation for the development and deployment of autonomous robots, is identified. Legal principles are also outlined, including the right to privacy, respect for human dignity, transparency and the right to due process, the right to information, the right to self-determination and non-discrimination, as well as socio-economic rights and the rights to security and social protection.

Conclusions and prospects for development. The rapid progress in the development of autonomous agents may lead to the emergence of applications that can significantly enhance well-being, while at the same time having the potential to cause harm. To ensure their safe and reliable functioning, it is essential to pay close attention to the legal, ethical, and cultural context in which they operate. Since the legal regulation of this technology is still at an early stage, in order to ensure further progress, regulatory norms should be guided by ethical principles rather than prohibitions and remain aligned with contemporary legal requirements.

To prevent harm, especially as these technologies become increasingly integrated at closer levels with their users, it is necessary not only to generalize the fundamental principles of artificial intelligence, which can be reduced to operational rules to be followed by an autonomous agent, but also to develop methods for embedding social norms

into autonomous agents. These norms are often expressed as abstract high-level principles that are not easily translated into operational rules that an autonomous system must follow. One of the methods for implementing these high-level normative principles into operational rules of behavior of an autonomous system, in a way that can be trusted and satisfy users, lies in combining mandatory requirements and voluntary standards.

Key words: law and technology; principles of artificial intelligence; ethics of artificial intelligence; legal regulation of artificial intelligence, artificial sociality.

За ДСТУ 8302:2015 цю статтю слід цитувати:

Орловський І. А. Принципи штучного інтелекту в праві. *Право та інновації*. 2026. № 1 (53). С. 88–94. [https://doi.org/10.37772/2518-1718-2026-1\(53\)-11](https://doi.org/10.37772/2518-1718-2026-1(53)-11).

Дата подання автором: 26.02.2026

Дата прийняття після рецензування: 12.03.2026

Дата опублікування: 29.04.2026