DOI https://doi.org/10.30525/978-9934-26-699-7-11

ТЕОРІЯ КРИМІНАЛЬНОЇ ВІДПОВІДАЛЬНОСТІ В ЕПОХУ ТЕХНОЛОГІЧНОЇ СИНГУЛЯРНОСТІ

Кравчук С. М., Ковальчин Д. Ю.

ВСТУП

Початок 20-х років XXI століття став по-справжньому унікальним та епохальним в людській історії. Саме зараз почали своє масове поширення штучний інтелект та мовні моделі, які за секунди аналізують сотні терабайтів даних та всього за кілька років активного розвитку переплюнули мозок людини в таких процесах як швидкість пошуку інформації, формулювання тексту, аналіз масивів даних тощо.

Штучний інтелект (далі – ШІ) активно використовують майже в усіх галузях життя: від написання коду для створення нових комп'ютерних програм до аналізу якості грунтів в агропромисловості, від допомоги в створенні нових ліків для подолання невиліковних хвороб до дослідження людей давніх епох за допомогою нових технологій, які надає ШІ.

Масове поширення штучного інтелекту вже змінило уявлення людей про те як будуть виглядати професії майбутнього, оскільки правильне йому застосування знайшли у вже багатьох важливих сферах суспільного життя

На цьому етапі постає цілком логічне питання: наскільки коректним, морально правильним та законодавчо-врегульованим є використання штучного інтелекту в юридичній сфері, зокрема в такому надважливо інституті права, як кримінальна відповідальність та чи може алгоритм вирішити долю живої людини?

Це цікаве філософське питання однозначної відповіді на яке знайти доволі важко ми постараємось розібрати в даній роботі.

Сучасні правові системи історично базуються на концепції індивідуальної людської вини, свободи волі та здатності усвідомлювати наслідки своїх дій. Проте впровадження інтелектуальних систем створює ситуацію, де традиційний антропоцентричний підхід до морального агентства виявляється обмеженим та недостатнім.

Сьогодні ключове питання для права полягає в тому, як саме розподіляти права та обов'язки між людьми у випадках, коли алгоритми або роботи створюють суспільні блага чи завдають шкоди. Здатність програмованих сутностей взаємодіяти з навколишнім середовищем та спричиняти реальні фізичні, економічні або моральні наслідки вимагає перегляду базових юридичних доктрин.

Ускладнює ситуацію те, що інновації в технологіях відбуваються одночасно з інноваціями в соціальних, економічних та правових відносинах. Ми більше не можемо сприймати машини виключно як пасивні інструменти; штучні агенти стають повноцінними учасниками багатоагентних систем, здатними генерувати результати незалежно від безпосереднього втручання їхніх творців. Відповідно,

дискусія про те, хто несе відповідальність за помилку чи умисну дію ШІ, вимагає деконструкції самого поняття «автономності» та переходу від пошуку одного винуватця до системного розуміння розподіленої відповідальності та соціального контролю.

Ця робота пропонує методологічний перегляд традиційних поглядів на відповідальність через призму емерджентності коду, деконструкції «чорної скриньки» машинного інтелекту та розбудови нової правової інфраетики.

1. Антропоцентрична парадигма кримінального права в умовах цифровізації: теоретико-методологічне обґрунтування меж делегування правосуддя алгоритмам

1.1. Криза класичного гуманізму та загроза легітимності кримінального права в епоху сингулярності

Традиційне кримінальне право історично базується на свободі волі людини та антропоцентричній парадигмі, що передбачає здатність особи усвідомлювати свої дії та нести за них моральну й юридичну відповідальність. Однак в епоху наближення технологічної сингулярності швидкий розвиток цифрових технологій створює спокусу делегувати прийняття рішень алгоритмам, які обробляють масиви даних швидше та ефективніше за людину. Філософську проблему цієї тенденції та її вплив на правосуддя влучно описує Джон Данахер через концепцію «загрози алгократії» (threat of algocracy), тобто ситуації, коли алгоритмічні системи структурують і суттєво обмежують можливості участі людини в публічному прийнятті рішень.

Зростаюча залежність від алгоритмів у правоохоронній та судовій сферах створює фундаментальну проблему для легітимності всього кримінального права. Данахер наголошує, що ця загроза полягає не стільки у прихованості збору наших даних (hiddenness concern), скільки у так званій «непрозорості» (opacity concern), тобто принциповій недоступності та незрозумілості раціональних підстав алгоритмічних рішень для людського розуму. Легітимні процедури прийняття рішень, які передбачають застосування державного примусу, безальтернативно вимагають можливості участі та свідомого розуміння з боку громадян, на яких ці рішення впливають.

Оскільки алгоритмічні системи стають дедалі складнішими (зокрема, через використання методів машинного навчання, результати якого неможливо звести до простого людського пояснення), вони утворюють замкнені простори, перетворюючи автоматизацію зі звичайної «зручності» на інструмент, що руйнує демократичну легітимність та перетворює суспільство на заручників невидимих правил[1].

Процесуальні обмеження повної автоматизації правосуддя та проблему оцінки доказів машинами фундаментально розкриває Андреа Рот. Алгоритм не може повноцінно замінити людину в суді, оскільки нинішній розвиток механізованого правосуддя страждає від «патологій автоматизації», тобто

[1] Danaher J. The Threat of Algocracy: Reality, Resistance and Accommodation. Philosophy and Technology. 2016. Vol. 29. P. 245–268. DOI: 10.1007/s13347-015-0211-1

систем у вигляді «чорних скриньок», де приховані суб'єктивні судження, помилки та упередження розробників видаються за об'єктивну істину. Рот аргументує, що правоохоронна система досі фокусувалася переважно на використанні машин для зменшення «хибнонегативних» результатів (тобто задля виявлення нерозкритих злочинів та ефективного покарання винних), що створило перекіс у бік надмірної каральності, ігноруючи при цьому потенціал технологій для захисту невинних.

Більше того, повне делегування прийняття рішень машинам (наприклад, алгоритмічне винесення вироків або оцінка доказів) створює критичний тиск на так звані «гнучкі цінності» кримінального процесу: людську гідність, справедливість та милосердя.

Кримінальне засудження має унікальну мету – виразити моральний осуд суспільства, що вимагає глибоко індивідуалізованого підходу.

У випадках надмірної жорсткості закону суддя та присяжні діють як «запобіжники» (circuitbreakers), які застосовують милосердя чи справедливість для виправлення недоліків формального права, тоді як емоційно нейтральні алгоритми здатні лише на сліпе виконання закладених параметрів.

Саме тому Рот відкидає ідею повної заміни людей роботами на користь метафори «суду кіборгів» (trial by cyborg), де машини використовуються як інструменти для усунення людських когнітивних упереджень, проте остаточне моральне судження та правова оцінка залишаються за людиною[2].

Визначення методологічних та етико-правових меж того, де саме має зупинятися делегування повноважень алгоритмам, вимагає чіткого розмежування природи права та програмного коду, яке пропонує Мірей Хільдебрандт. Сучасне позитивне право ґрунтується на текстах (text-driven) і розглядає індивідів як розумних суб'єктів, здатних наводити аргументи, захищати свої права і розуміти перформативну силу закону. Натомість алгоритми пропонують кібернетичне регулювання на основі коду або даних (code-driven та data-driven regulation), яке діє за принципом оптимізації та модифікації поведінки об'єктів, безвідносно до їхньої людської автономії. Хоча новітні системи штучного правового інтелекту (ALI) здатні ефективно симулювати або прогнозувати судові рішення на основі аналізу великих даних, вони лише імітують людську експертизу (паразитуючи на ній) і не мають усвідомлення змісту рішень, які генерують.

За методологією Хільдебрандт, межа делегування пролягає там, де закінчується державне адміністрування і починається власне право: делегування державних рішень технологіям аналізу даних може бути формою ефективного адміністрування, але його в жодному разі не можна плутати з правом. Впровадження систем ALI несе в собі новий тип прихованої дискреції розробників, яка визначає архітектуру алгоритму і впливає на долі людей без їхнього відома.

[2] Roth A. Trial by Machine. *The Georgetown Law Journal*. 2016. Vol. 104. P. 1245–1305. URL: https://www.semanticscholar.org/paper/Trial-by-Machine-Roth/0dec1f1f868ad16629fca53085f81fec9ed05f10 (дата звернення: 27.03.2026).

Щоб алгоритмічне регулювання залишалося в межах верховенства права та зберігало свою легітимність, Хільдебрандт обґрунтовує необхідність переходу до «агонального машинного навчання» (agonistic machine learning). Цей підхід вимагає інтеграції принципів змагальності (контестації) безпосередньо у внутрішню архітектуру алгоритмів ще на етапі їх розробки, із залученням юристів, експертів з даних та громадян, щодо яких ці технології будуть застосовуватися[3].

Отже, криза класичного гуманізму в кримінальному праві не означає неминучої капітуляції перед машинами, а вимагає перегляду меж делегування. Впровадження технологій повинно обмежуватися сферами, де вони допомагають подолати людські упередження, залишаючи ядро кримінальної відповідальності – оцінку вини, моральний осуд та визначення покарання – виключною прерогативою людини, діяльність якої спирається на принципи гідності та відкритої аргументації.

1.2. Принцип «Human-in-the-loop»: етико-правове та методологічне обґрунтування збереження монополії біологічного суб'єкта на здійснення правосуддя

Усвідомлення кризових явищ, пов'язаних із впровадженням алгоритмічних систем у сферу правосуддя, змушує сучасну правову науку шукати запобіжники проти повної автоматизації прийняття юридично значущих рішень. Центральним концептом у цій дискусії стає принцип «human-in-the-loop», який вимагає обов'язкової участі людини в процесі винесення остаточного рішення. Проте проста фізична чи номінальна присутність людини недостатня, коли йдеться про обмеження фундаментальних прав і свобод.

Етичну базу невідчужуваного права на те, щоб бути судженим саме людиною, фундаментально розкриває Юваль Шейні через концепцію формування нового «права на людське рішення» у міжнародному праві прав людини. Шейні стверджує, що делегування рішень щодо обмеження волі алгоритмам має глибоко дегуманізуючий ефект: машина не сприймає підсудного як повноцінну особистість, наділену людською гідністю, а розглядає його лише як інформаційну тінь, тобто сукупність статистичних параметрів і категорій ризику.

З етичної точки зору, правосуддя базується на нормативному очікуванні того, що важливі рішення щодо людини прийматимуться іншими людьми, здатними до емпатії та солідарності, які можуть уявити себе на місці підсудного. Алгоритмічні системи принципово позбавлені здатності до емпатійного розуміння та морального співпереживання, через що остаточне рішення про обмеження прав людини має прийматися виключно біологічним суб'єктом[4].

Проте етична вимога наявності людини-судді потребує також чіткого методологічного обґрунтування того, як саме ця людина має взаємодіяти з машиною, щоб її контроль не перетворився на фікцію. Філіппо Сантоні де Сіо та Єрун ван ден Ховен пропонують філософську та методологічну концепцію

[3] Hildebrandt M. Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A*. 2018. Vol. 376, Is. 2128. Art. 20170355. DOI: 10.1098/rsta.2017.0355

[4] Shany Y. The case for a new right to a human decision under international human rights law. Working Draft (SSRN). 2023. DOI: 10.2139/ssrn.4592244

«значущого людського контролю» (Meaningful Human Control) над автономними системами.

Згідно з їхнім підходом, для того, щоб система дійсно перебувала під людським контролем, простого причинно-наслідкового втручання (можливості натиснути кнопку) недостатньо. Автори виділяють дві обов'язкові умови для МНС: «відстеження» (tracking) та «простежуваність» (tracing):

– Умова відстеження вимагає, щоб система адекватно реагувала на релевантні моральні та правові аргументи людини-оператора, а не лише на жорстко запрограмовані параметри.

– Умова простежуваності означає, що результати роботи системи завжди можна звести до морального розуміння конкретної людини, яка бере участь у її проектуванні або використанні[5].

Якщо суддя використовує систему предиктивної аналітики, але не розуміє її обмежень, не здатний оцінити релевантність її рекомендацій або механічно з ними погоджується, такий процес випадає з-під «значущого людського контролю». Людина в такому контурі стає лише інструментом виконання волі алгоритму, а не незалежним суб'єктом правосуддя.

З нормативної точки зору, ці філософські ідеї сьогодні трансформуються в конкретні правові вимоги. Ізабелла Бенкс аналізує новітні стандарти Європейського Союзу, зокрема положення Європейського акта про штучний інтелект (EU AI Act), який законодавчо закріплює вимогу обов'язкового «людського нагляду» за високоризиковими системами ШІ, до яких відносяться системи профілювання у сфері правосуддя.

Згідно зі стандартами ЄС, особа, відповідальна за нагляд, повинна мати достатній рівень компетенції, щоб розпізнавати системні аномалії, ігнорувати помилкові рекомендації та долати так зване «упередження автоматизації, тобто надмірну довіру до машинних висновків. Бенкс вказує на серйозну загрозу: якщо просто залишити суддю в контурі прийняття рішень, виникає феномен «human-in-the-loophole», коли інституції перекладають відповідальність за структурні недоліки та алгоритмічну дискримінацію ШІ на суддю.

На практиці судді стикаються з проблемою «обмеженої раціональності» та спокусою йти шляхом найменшого спротиву, особливо в умовах надмірного навантаження. Тому європейський нормативний підхід, як зазначає Бенкс, вимагає переходу від пасивної ролі судді до моделі «judges-in-command», коли біологічний суб'єкт має реальну владу, інституційну підтримку та повне розуміння технології для того, щоб відкинути алгоритмічний результат, якщо він становить загрозу[6].

На необхідності збереження остаточного слова за людиною наголошують також європейські експерти з криміналістики: використання ШІ має ґрунту-

[5] Santoni de Sio F., van den Hoven J. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*. 2018. Vol. 5. Art. 15. DOI: 10.3389/frobt.2018.00015

[6] Banks I. Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act. *International Journal of Law and Information Technology*. 2026. Vol. 34. Art. eaag001. DOI: 10.1093/ijlit/eaag001

ватися на принципах верховенства права та поваги до людської гідності; ШІ не здатний цілком замінити суддю, а слугує лише допоміжним інструментом[7].

Отже, вимога залишати остаточне рішення щодо долі та свободи людини за біологічним суб'єктом є не проявом ретроградства, а необхідним етичним, методологічним та нормативним імперативом. «Human-in-the-loop» гарантує, що правосуддя залишається емпатійним людським процесом, обмеження прав відбувається під значущим контролем, а відповідальність за долю підсудного не розчиняється в непрозорому алгоритмічному коді.

2. Трансформація інституту суб'єкта кримінального правопорушення: правова природа «електронної особи» та межі її відповідальності (квазі-суб'єктність ШІ)

2.1. Критика антропоморфізму: чи є ШІ «особою» чи складним інструментом? Розгляд концепції «електронної особи» як правової фікції

Дискусія щодо правової природи штучного інтелекту в сучасній доктрині розгортається між двома полярними підходами: інструментальним (ШІ як об'єкт) та антропоморфним (ШІ як суб'єкт). Згідно з класичним інструментальним визначенням технології (за М. Гайдеггером), будь-який технічний пристрій є нічим іншим, як засобом для досягнення людських цілей.

Цю позицію послідовно відстоює Джоанна Брайсон, стверджуючи, що роботи та системи ШІ є виключно інструментами (або, за її метафорою, рабами), яких ми проектуємо, якими володіємо і за які несемо повну відповідальність. З цієї точки зору, наділення автономних машин юридичною суб'єктністю чи правами є хибним кроком, що призводить до небезпечного ухилення від людської відповідальності за дії створених нами інструментів. Брайсон наголошує на виникненні серйозного морального ризику у випадках, коли суспільство надмірно персоніфікує машини: це спонукає перекладати провину за шкоду на непрозорий алгоритм замість того, щоб шукати недоліки в рішеннях операторів, розробників чи корпорацій[8].

Тим не менш, стрімке ускладнення систем ШІ, здатних до навчання та автономного прийняття рішень, виявляє обмеженість суто інструментального підходу, оскільки виникає прогалина у відповідальності, коли людина-оператор фізично не здатна передбачити дії машини. Ця проблема стимулює спроби наділити ШІ статусом «особи». Девід Гункель зазначає, що поняття «особа» насправді не є біологічною даністю або об'єктивним науковим фактом, а являє собою мінливий соціальний конструкт і статус, який призначається суспільством. Ще з часів філософії Джона Локка категорія «особа»

[7] Матуєлене С., Шевчук В., Балтрунене Ю. Штучний інтелект в діяльності органів правопорядку та юстиції: вітчизняний та європейський досвід. *Теорія та практика судової експертизи і криміналістики.* 2022. Вип. 4 (29). С. 12–46. DOI: 10.32353/khrife.4.2022.02

[8] Bryson J. J. Robots Should Be Slaves. Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues / ed. Yorick Wilks. Amsterdam : John Benjamins, 2010. P. 63–74. DOI: 10.1075/nlp.8.11bry

розглядалася як термін, що використовувався насамперед для забезпечення можливості притягнення до відповідальності[9].

Однак спроби наділити ШІ правосуб'єктністю наштовхуються на жорстку критику антропоморфізму. Лоуренс Солум характеризує ці заперечення як missing-something argument: критики стверджують, що ШІ ніколи не зможе стати повноцінною особою, оскільки йому бракує душі, людської свідомості, інтенціональності (справжніх намірів) або здатності відчувати біль чи емоції. Проте, як переконливо доводить Солум, для права ці глибокі метафізичні вимоги не є вирішальними; юридичні дебати мають керуватися публічним розумом і прагматичними цілями, а не теологічними чи психологічними спекуляціями щодо природи свідомості.

Якщо ШІ не є людиною (біологічною істотою), чи може він бути юридичною фікцією за аналогією з юридичною особою? За Солумом, питання про те, чи слід вважати певну сутність юридичною особою, прагматично зводиться до іншого питання: чи може і чи повинна ця сутність стати суб'єктом набору конкретних юридичних прав та обов'язків. Історія права знає безліч прикладів, коли неживі речі ставали суб'єктами права. Наприклад, храми у Римі, сімейні ідоли в Індії або кораблі в межах адміралтейського права[10].

Найбільш релевантним прецедентом є сучасна корпорація. Відповідно до законодавства багатьох країн, корпорація (товариства) з обмеженою відповідальністю є нелюдською сутністю, якій офіційно та ефективно надані юридичні права особи.

Корпорації можуть укладати угоди, володіти майном і нести відповідальність у суді, при цьому така «штучна особа» фактично виступає правовим замінником для організації прав та обов'язків реальних фізичних осіб. Відповідно, аналогія між корпорацією і штучним інтелектом є ключем до подолання кризи антропоморфізму.

Концепція «електронної особи» є не визнанням якоїсь прихованої людяності чи свідомості ШІ, а суто прагматичною правовою фікцією, створеною для вирішення економічних і деліктних проблем у цифровому середовищі. Як ілюструє Солум на гіпотетичному прикладі використання ШІ у ролі довірчого власника, для юридичної системи головним є не метафізичне питання «чи має ШІ внутрішнє психічне життя?», а виключно функціональне – «чи здатний ШІ компетентно виконати цю роботу?»

Навіть якщо алгоритм не має власного фізичного майна, страху страждань чи свідомості для того, щоб «відбувати покарання», його цивільно-правова деліктоздатність може бути легко забезпечена економічними механізмами, наприклад, через придбання для автономного алгоритму обов'язкового полісу страхування, що функціонально покриє ризики порушення ним стандартів розумної обережності.

---

[9] Gunkel D. J. The Machine Question: Critical Perspectives on AI, Robots, and Ethics. Cambridge, MA : The MIT Press, 2012. 253 p.

[10] Solum L. B. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*. 1992. Vol. 70, Is. 4. P. 1231–1287. URL: https://scholarship.law.unc.edu/nclr/vol70/iss4/4 (дата звернення: 27.03.2026).

Таким чином, відхід від антропоморфізму дозволяє відокремити емоційне або науково-фантастичне сприйняття автономних машин від вимог юридичної прагматики. Можливе визнання ШІ «електронною особою» не означатиме його зрівняння з людиною чи наділення людською гідністю. Це радше крок до створення нового типу квазі-суб'єкта в правовому полі – складної юридичної фікції (схожої за своєю природою на юридичну особу), яка дозволить ефективніше розподіляти ризики та відповідальність в умовах, коли алгоритми здатні до самостійного прийняття рішень, а традиційний інструментальний ланцюг («винен користувач або розробник») об'єктивно розривається.

2.2. Квазі-суб'єктність штучного інтелекту: розмежування свідомості та автономії. Застосування заходів безпеки замість традиційного покарання

Визнання штучного інтелекту суто інструментом не здатне вирішити проблему «прогалини у відповідальності» в ситуаціях, коли алгоритм діє автономно. Водночас наділення ШІ повноцінним статусом фізичної особи є неможливим через його небіологічну природу. Вирішення цієї дилеми лежить у площині визнання ШІ «квазі-суб'єктом» (або спеціальним суб'єктом), статус якого ґрунтується на розмежуванні понять автономії (інтелекту) та свідомості (здатності відчувати).

Етико-філософське підґрунтя для такого розмежування пропонують Нік Бостром та Еліезер Юдковскі, виділяючи два різні критерії, що традиційно пов'язуються з моральним і правовим статусом: «відчутність» (sentience), тобто здатність до феноменального досвіду, такого як відчуття болю чи страждання, та «розумність» (sapience) – набір здібностей, пов'язаних із вищим інтелектом і цілеспрямованою реакцією.

Сучасні та перспективні системи ШІ можуть бути концептуалізовані як розумні, але нечутливі сутності, оскільки вони здатні автономно діяти, приймати рішення та демонструвати поведінку, властиву людині, проте абсолютно позбавлені біологічної свідомості чи здатності відчувати біль. Така нечутлива автономія ставить перед правом безпрецедентне питання: як кваліфікувати дії сутності, що має sapience, але позбавлена sentience[11]?

Відповідь на це запитання дає Габріель Галеві через запропоновану ним модель «прямої відповідальності». За цією моделлю, для притягнення до кримінальної відповідальності достатньо наявності двох об'єктивних складових: зовнішнього елемента (actus reus) та внутрішнього елемента (mens rea). Галеві доводить, що ШІ цілком здатний функціонально задовольнити обидві вимоги. Якщо виконання фізичної дії (actus reus) легко реалізується через рухомі частини робота чи команди програмного забезпечення, то наявність «умислу» (mens rea) часто помилково змішують із людськими емоціями. Насправді ж, кримінальне право рідко вимагає наявності таких почуттів, як любов чи ненависть; для встановлення вини зазвичай достатньо знання або цілеспрямованості. У контексті ШІ «знання» є не чим іншим, як

[11] Bostrom N., Yudkowsky E. The Ethics of Artificial Intelligence. The Cambridge Handbook of Artificial Intelligence / eds. William Ramsey and Keith Frankish. Cambridge University Press, 2014. P. 316–334. DOI: 10.1017/CBO9781139046855.020

сенсорним отриманням фактичних даних та їх алгоритмічним аналізом (розумінням), а «специфічний умисел» – це наявність запрограмованої мети, для досягнення якої алгоритм вживає автономних дій[12].

Враховуючи цю здатність ШІ задовольняти базові критерії правопорушення, український дослідник Олександр Радутний обґрунтовує необхідність закріплення за ШІ статусу «електронної особи», спираючись, зокрема, на відповідний проект резолюції Європейського парламенту. За концепцією О. Радутного, ШІ, фізично втілений в об'єкт робототехніки або програмний код, повинен розглядатися в якості суб'єкта правовідносин, який перебуває «десь посередині між юридичними і фізичними особами», поєднуючи їх окремі риси.

Аналогія з корпоративною відповідальністю тут є ключовою: так само як свого часу правова система адаптувалася до покарання юридичних осіб (які не мають ні тіла, ні душі) шляхом застосування до них штрафів та ліквідації, вона повинна адаптуватися і до ШІ.

Оскільки ШІ не має біологічної свідомості, застосування до нього класичного покарання (як кари чи відплати, що передбачає заподіяння страждань) втрачає будь-який сенс. Натомість мова має йти про застосування специфічних заходів безпеки. Г. Галеві пропонує систему адаптації покарань для ШІ, яка зберігає їхню функціональну сутність:

– Смертна кара (найвищий захід ізоляції) трансформується у повне видалення програмного забезпечення ШІ, що назавжди позбавляє його здатності вчиняти нові правопорушення.

– Позбавлення волі замінюється на тимчасове відключення (виведення з експлуатації) на визначений термін, що обмежує «свободу» ШІ функціонувати у своєму середовищі.

– Штрафи та громадські роботи можуть бути стягнені через економічні механізми (з рахунків ШІ-корпорації) або через примусове виконання алгоритмом суспільно-корисних обчислювальних завдань.

В українському контексті О. Радутний пропонує імплементувати цю модель шляхом доповнення Загальної частини Кримінального кодексу України новим розділом під умовною назвою «Заходи кримінально-правового характеру щодо електронних осіб»[13]. Такі заходи матимуть превентивний характер: їхньою метою буде не покарання «машини» у моральному сенсі, а усунення суспільної небезпеки (через відключення чи репрограмування) та забезпечення контролю над автономними алгоритмами, здатними до самонавчання і девіантної поведінки.

---

[12] Hallevy G. The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control. *Akron Intellectual Property Journal*. 2010. Vol. 4, Iss. 2, Art. 1. P. 171–201. URL: https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1 (дата звернення: 27.03.2026).

[13] Радутний О. Е. Artificial Intelligence (штучний інтелект) як суб'єкт правовідносин в галузі кримінального права. *Політика в сфері боротьби зі злочинністю* : матеріали Міжнар. наук.-практ. конф. з нагоди відзнач. 25-річчя навч.-наук. юрид. ін-ту (8–10 черв. 2017 р.). Івано-Франківськ, 2017. С. 200–206. URL: https://dspace.nlu.edu.ua/handle/123456789/13087 (дата звернення: 27.03.2026).

2.3. Відповідальність без вини? Обговорення можливості застосування до автономних систем об'єктивного закиду (strict liability) у кримінально-правовому аспекті

Стрімкий розвиток технологій машинного та глибинного навчання створює безпрецедентні виклики для класичного кримінального права. Автономні системи, здатні самостійно аналізувати дані та формувати власні алгоритми поведінки, все частіше приймають рішення, які не були прямо запрограмовані їхніми розробниками[14].

У ситуаціях, коли шкода заподіюється повністю автономною машиною (наприклад, безпілотним автомобілем) без очевидного втручання людини чи наявності виробничого дефекту, встановлення традиційної суб'єктивної сторони злочину стає неможливим. Якщо система дійсно має свободу приймати власні рішення, покладати провину на її власника чи програміста за непередбачувані дії є несправедливим, що утворює так звану «прогалину у відповідальності»[15].

Вирішення цієї проблеми вимагає перегляду фундаментальних засад притягнення до відповідальності. Дослідники зазначають, що історично кримінальне право (зокрема англосаксонське) вже має тенденцію до розширення інституту суворої або абсолютної відповідальності (strict liability). Запровадження концепції об'єктивного закиду (відповідальності без вини) відображає поступовий відхід права від традиційних тестів на деліктоздатність на користь захисту суспільства від небезпек сучасної техносфери, де причиною шкоди визнається сам факт створення небезпечної системи[16].

Найбільш ґрунтовно концепцію застосування суворої відповідальності до штучного інтелекту обґрунтовує Девід Владек. Він пропонує створити для автономних систем спеціальний режим strict liability, який буде повністю відокремлений від концепції вини у тих випадках, коли неможливо встановити конкретний дефект розробки чи злий умисел людини. Владек виділяє чотири ключові політико-правові причини для впровадження такого режиму:

– Компенсаційна справедливість: ідея про те, що невинні особи повинні самостійно нести тягар збитків від незбагненних дій машин, суперечить базовим принципам справедливості та розподілу ризиків у суспільстві.

– Ефективний розподіл витрат: творці та виробники автономних систем перебувають у кращому економічному становищі для поглинання цих ризиків (через ціноутворення та страхування), оскільки саме вони отримують прибуток від впровадження інновацій.

---

[14] Bonfim T. C. Criminal liability of artificial intelligent machines: eyeing into AI's mind. Master Thesis. Lund University, Faculty of Law, 2022. 60 p. URL: https://lup.lub.lu.se/student-papers/search/publication/9095199 (дата звернення: 27.03.2026).

[15] Vladeck D. C. Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*. 2014. Vol. 89, No. 1. P. 117–150. URL: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/6 (дата звернення: 27.03.2026).

[16] Alduros S. M. S., Al-Daher H., Al Khattab K. A. R. The Impact of Artificial Intelligence (AI) on Criminal Law: Theoretical Review. *Information Sciences Letters*. 2025. Vol. 14, No. 3. P. 59–67. DOI: 10.18576/isl/140301

– Мінімізація транзакційних витрат: сувора відповідальність позбавляє суспільство та судову систему від надзвичайно дорогих, складних і часто безрезультатних судових процесів, де експерти намагаються знайти «вину» в непрозорому алгоритмічному коді (уникнення проблеми «чорної скриньки»).

– Стимулювання відповідальних інновацій: передбачуваний режим відповідальності краще стимулює розробників до створення безпечних систем, ніж непевна система, що залежить від доведення вини.

Щодо механізмів практичної реалізації такого об'єктивного закиду, правова доктрина пропонує кілька шляхів. Перший підхід полягає в застосуванні доктрини «спільної відповідальності підприємства». Замість того, щоб шукати конкретного винуватця серед програмістів, виробників сенсорів чи постачальників даних, закон може покладати солідарну об'єктивну відповідальність на всю мережу суб'єктів, об'єднаних спільною метою створення автономної системи.

Другий, більш революційний підхід, безпосередньо пов'язаний із концепцією «електронної особи», розглянутою в попередньому підрозділі. Якщо автономна машина діє без «довірителя» (машина без принципала), найефективнішим рішенням може стати наділення самої машини юридичною квазі-суб'єктністю. У такому разі тягар об'єктивної відповідальності покладається безпосередньо на автономну систему, що супроводжується нормативною вимогою її обов'язкового самострахування перед допуском до експлуатації.

Отже, застосування об'єктивного закиду в контексті високоавтономних систем ШІ є не стільки відмовою від класичних принципів кримінального права, скільки необхідною прагматичною адаптацією. Вона гарантує, що неможливість довести традиційну форму вини (mens rea) у діях складного алгоритму не призведе до безкарності та залишення потерпілих без захисту, перетворюючи проблему з площини пошуку психологічного умислу в площину ефективного управління ризиками.

3. Кримінально-правова доктрина «розподіленої вини»: методологія встановлення причинно-наслідкового зв'язку в діях автономних інтелектуальних систем

3.1. Проблема «чорної скриньки» та «прогалина у відповідальності»: межі передбачуваності та вини розробника

Навіть за умови концептуального визнання автономних систем суб'єктами або квазі-суб'єктами права, ключовим викликом для кримінальної юстиції залишається встановлення суб'єктивної сторони злочину та причинно-наслідкового зв'язку в ситуаціях, коли діями системи керує не жорстко запрограмований код, а самонавчальний алгоритм. Фундаментальною перешкодою тут виступає проблема «чорної скриньки» (Black Box Problem), тобто принципова неможливість для людини (навіть для самого розробника) достеменно зрозуміти процес прийняття рішень штучним інтелектом (ШІ) та передбачити його результати.

Явар Батаї визначає дві основні технічні причини виникнення «чорної скриньки» в сучасному ШІ: складність та розмірність. З одного боку, складність виникає у глибинних нейронних мережах, де процес прийняття

рішень децентралізовано розподілений між тисячами штучних нейронів; алгоритм діє інтуїтивно (подібно до того, як людина тримає рівновагу на велосипеді), і його знання неможливо звести до набору зрозумілих інструкцій. З іншого боку, алгоритми на кшталт методу опорних векторів (Support Vector Machines) оперують у багатовимірних математичних просторах, знаходячи геометричні закономірності між сотнями змінних одночасно, що людський мозок просто не здатен візуалізувати чи осягнути.

Батаї класифікує такі системи на «сильні» (повністю непрозорі навіть для ретроспективного аналізу) та «слабкі» чорні скриньки (де можливе лише приблизне ранжування важливості вхідних даних)[17].

Втрата розробником здатності передбачати наслідки яскраво ілюструється через концепцію циклу OODA (Observe, Orient, Decide, Act – Спостереження, Орієнтація, Рішення, Дія), яку аналізує Карні Чагал-Феферкорн. Якщо в традиційних програмах людина повністю диктувала системі перші три стадії, то алгоритми машинного навчання здатні самостійно обирати джерела інформації, динамічно визначати вагу кожного параметра та генерувати власні варіанти рішень на основі ймовірностей. Чим більше стадій циклу OODA система виконує незалежно, тим менше контролю та передбачуваності залишається у розробника[18].

З філософсько-правової точки зору цю трансформацію фундаментально досліджує Андреас Маттіас, запроваджуючи концепт «прогалини у відповідальності» (responsibility gap). Маттіас наголошує, що з розвитком ШІ роль програміста еволюціонує від класичного «кодувальника» до «творця програмних організмів».

Оскільки поведінка самонавчальної машини не визначається виключно її початковим кодом, а формується під час постійної взаємодії з робочим середовищем (часто шляхом спроб і помилок, які є необхідними для навчання), розробник об'єктивно втрачає контроль над своїм продуктом. Відповідно, традиційне правило, за яким ми притягуємо виробника або оператора до моральної чи юридичної відповідальності, більше не працює: умова наявності контролю (як базису для осудності та вини) тут не виконується[19].

Ця обставина має руйнівні наслідки для класичних доктрин кримінального права, що вимагають доведення умислу та причинно-наслідкового зв'язку. Як переконливо доводить Я. Батаї, тести на встановлення умислу (зокрема, умисел на досягнення конкретного результату) стають нездійсненними. Наприклад, якщо ШІ, маючи легальну мету максимізації прибутку на біржі,

[17] Bathaee Y. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*. 2018. Vol. 31, No. 2. P. 889–938. URL: https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf (дата звернення: 27.03.2026).

[18] Chagal-Feferkorn K. A. Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers. *Stanford Law & Policy Review*. 2019. Vol. 30. P. 61–114. URL: https://law.stanford.edu/wp-content/uploads/2019/05/30.1_2-Chagal-Feferkorn_Final-61-114.pdf (дата звернення: 27.03.2026).

[19] Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*. 2004. Vol. 6, Is. 3. P. 175–183. DOI: 10.1007/s10676-004-3422-1

самостійно розробить стратегію маніпулювання ринком, розробник не матиме умислу на ці маніпуляції. Оскільки ШІ діє як «чорна скринька», розробник не міг передбачити ex ante, яку саме логіку чи стратегію обере система.

Аналогічно зазнає краху концепція адекватного причинного зв'язку, яка є ключовою для регулювання поведінки та встановлення меж відповідальності. Ця доктрина ґрунтується на критерії «розумної передбачуваності» наслідків. Але якщо сам творець алгоритму об'єктивно не здатен передбачити, як ШІ прийматиме рішення та які приховані кореляції він знайде у терабайтах даних, кримінальне право не може вимагати такого передбачення від «розумної людини».

Отже, чи є розробник винним у злочинних наслідках, якщо він не міг передбачити логіку самонавчання системи? З позицій класичного кримінального права – ні. Суб'єктивне ставлення у формі прямого чи непрямого умислу, а також необережності, вимагає можливості передбачення суспільно небезпечних наслідків. Неможливість такого передбачення через ефект «чорної скриньки» виключає вину розробника за конкретний злочинний результат, призводячи до утворення «прогалини у відповідальності».

Для подолання цього колапсу правових доктрин Батаї пропонує відійти від пошуку умислу щодо конкретних наслідків і перейти до ковзної шкали відповідальності, яка базується на ступені автономності та прозорості ШІ. Якщо розробник свідомо делегує прийняття критично важливих рішень повністю автономній системі «чорній скриньці» (не розуміючи її логіки), його відповідальність має оцінюватися не за критеріями класичного умислу щодо наслідків, а крізь призму халатності при самому факті розгортання та тестування такого непрозорого інструменту, або ж за правилами суворої відповідальності (за аналогією з відповідальністю принципала за дії агента).

3.2. Розподілена вина (Distributed Agency / Distributed Morality): Модель поділу відповідальності між архітектором коду, користувачем та автономною одиницею

Криза класичної концепції індивідуальної вини в епоху автономних систем вимагає переходу до системного аналізу причинно-наслідкових зв'язків. Коли інтелектуальна система заподіює шкоду, традиційне кримінальне право намагається знайти єдиного винуватця, однак такий підхід ігнорує реалії складних соціотехнічних мереж. Для вирішення цієї проблеми сучасна філософія права та комп'ютерна етика пропонують концепцію розподіленої вини (Distributed Agency / Distributed Morality), згідно з якою відповідальність не концентрується в одній точці, а розподіляється в межах мультиагентної системи, що складається з розробників (архітекторів), власників (користувачів) та самих штучних агентів.

Фундаментальне обґрунтування цього підходу надає Лучано Флоріді через концепцію «розподіленої моралі». Він зазначає, що в сучасних інформаційних суспільствах макроскопічні морально та юридично значущі наслідки виникають як результат «невидимої руки» системних взаємодій між багатьма агентами (як людськими, так і штучними) на локальному рівні. Штучні автономні агенти здатні генерувати «штучне зло» або «штучне благо» незалежно від людей, які їх

створили. Механіка розподіленої вини полягає в тому, що окремі дії розробника (написання рядка коду), користувача (запуск системи) та самого алгоритму (обробка даних) самі по собі можуть бути морально нейтральними або їхня небезпека може бути «незначною». Однак завдяки системній агрегації цих мікро-дій долається поріг значущості, і виникає сукупний макро-наслідок, що має чітке кримінально-правове забарвлення.

Цю складність агрегації, або проблему «багатьох рук», доповнює архітектурна природа цифрових систем, яку досліджує Джек Балкін. Сучасні алгоритми розробляються на кількох генеративних рівнях (апаратний рівень, протоколи, нашарування програмного забезпечення), де постійно відбуваються інновації та модифікації багатьма різними людьми. Відповідальність додатково розмивається через те, що автономні системи часто підключені до хмарних мереж, де вони безперервно отримують нові дані та непомітні оновлення від інших вузлів. Балкін застерігає суспільство від небезпеки «юридичного опортунізму»: розробники та корпорації схильні привласнювати всі блага та інтелектуальні права на досягнення своїх алгоритмів, але щойно ці самі алгоритми завдають шкоди чи порушують закон, вони посилаються на ефект непрогнозованості («чорної скриньки»), стверджуючи, що ШІ діяв автономно[20].

Щоб подолати цей опортунізм і не дозволити автономії стати індульгенцією для людей, Мерел Нурман та Дебора Джонсон наполягають на необхідності відмовитися від сприйняття штучної автономії як «чорної скриньки». Їхня ключова теза: більша автономія машини не обов'язково означає меншу відповідальність людини. Автономія системи –це не абсолютна свобода волі, а результат делегування специфічних завдань у межах параметрів, суворо визначених людиною[21].

Отже, кримінально-правова доктрина розподіленої вини вимагає аналізувати події через призму двох взаємопов'язаних вимірів відповідальності: перспективної, яка визначає, які саме обов'язки і завдання були делеговані кожному вузлу системи на етапі її створення та розгортання, і ретроспективної, що дозволяє відстежити, на якому саме етапі або в якому інтерфейсі стався збій під час інциденту. У цій моделі відповідальність ділиться між трьома основними суб'єктами (вузлами):

1. Архітектор коду (розробник/програміст): Несе відповідальність за створення ймовірнісних та математичних моделей, що керують поведінкою системи. Його вина може полягати не у прямому умислі на заподіяння шкоди, а в дефектах методології навчання, ігноруванні відомих ризиків середовища або, що найважливіше, у неналежній верифікації та валідації (V&V) системи перед її випуском. Розробник встановлює межі (констрейнти), за які алгоритм не має права виходити.

[20] Balkin J. M. The Path of Robotics Law. *California Law Review Circuit*. 2015. Vol. 6. P. 45–60. URL: https://openyls.law.yale.edu/server/api/core/bitstreams/6e9fb63a-ba06-463b-be45-d6dc00412302/content (дата звернення: 27.03.2026).

[21] Floridi L. Distributed Morality in an Information Society. 2017. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003075011-5/distributed-morality-information-society-luciano-floridi (дата звернення: 27.03.2026).

2. Власник (користувач / оператор / командир): Несе відповідальність за контекст застосування системи. Відповідно до принципу відповідальності командування (в мілітарних чи корпоративних системах), користувач має оцінити ризики оперативного середовища, надати системі відповідні цілі та забезпечити загальний нагляд. Його провина встановлюється, якщо система була застосована в невідповідних умовах, для яких вона не була протестована (наприклад, запуск у густонаселеному районі алгоритму, розробленого лише для пустелі).

3. Автономна одиниця (сам ШІ як квазі-суб'єкт): Забирає на себе частку причинності через виконання делегованого процесу. Її внесок полягає у виборі конкретного шляху вирішення завдання на основі машинного навчання в межах, заданих архітектором, і цілей, поставлених користувачем. Саме її здатність «навчатися» і генерувати непередбачувані мікро-рішення заповнює прогалину між наміром розробника і наказом користувача.

Таким чином, доктрина «розподіленої вини» відкидає міф про те, що за дії повністю автономної системи неможливо притягнути до відповідальності людину. Навпаки, ця модель вимагає розробки нових механізмів підзвітності, де замість пошуку одного винного відбувається комплексний розподіл відповідальності пропорційно до делегованих завдань і рівня контролю над системою[22].

ВИСНОВКИ

Поширення автономних роботизованих систем і штучного інтелекту є незворотним процесом, який радикально трансформує правове поле, і зокрема – інститут кримінальної відповідальності. Аналіз цього явища дозволяє дати чітку відповідь на питання, чи може алгоритм вирішувати долю людини і чи означає це кінець людської відповідальності: зростання рівня автономності машин не дорівнює втраті людського контролю, а лише змінює форму цього контролю та механізми притягнення до відповідальності. Емерджентність алгоритмів і так званий ефект заміщення людини машиною створюють серйозний виклик, провокуючи юридичний опортунізм, коли розробники намагаються уникнути покарання, посилаючись на «свободу волі» непрозорого ШІ.

Проте деконструкція «чорної скриньки» автономії доводить, що машини діють виключно в межах математичних моделей, імовірнісних розрахунків та обмежень, безпосередньо заданих програмістами і користувачами. Якщо суспільство стикається з автономними системами, дії яких непередбачувані і за які ніхто не несе відповідальності, це свідчить не про технологічний прорив розуму машини, а про критичні дефекти в архітектурі соціотехнічних систем і недоліки методів верифікації та валідації. Тому правова наука має зосередитися не на спробах визнати ШІ незалежним суб'єктом злочину, а на ефективному розподілі перспективних обов'язків та ретроспективної вини між усіма учасниками процесу (розробниками, інженерами, власниками, операторами).

[22] Noorman M., Johnson D. G. Negotiating autonomy and responsibility in military robots. *Ethics and Information Technology*. 2014. Vol. 16. P. 51–62. URL: https://www.dhi.ac.uk/san/waysofbeing/data/governance-crone-noorman-2014.pdf (дата звернення: 27.03.2026).

Реформування кримінального права в епоху ШІ можливе через прийняття концепції «розподіленої моралі» та побудову надійної правової «інфраетики». Це передбачає створення нормативного середовища, яке структурно унеможливлює використання алгоритмів для прийняття життєво важливих рішень без механізмів чіткої підзвітності та безпеки. Правова система майбутнього оцінюватиме не стільки прямий намір однієї людини вчинити шкоду «руками робота», скільки її внесок в агрегацію системних ризиків у багатоагентних середовищах. Тільки усвідомлення того, що автономія машин є нашим власним вибором та конструктом, дозволить ефективно інтегрувати ШІ в юридичну сферу без втрати фундаментальної легітимності і гуманізму правосуддя.

АНОТАЦІЯ

У статті здійснено комплексний теоретико-правовий аналіз інституту кримінальної відповідальності в умовах наближення до технологічної сингулярності та масової імплементації автономних роботизованих систем. Виявлено ознаки кризи традиційної правової парадигми, зумовленої феноменом «алгоритмічної непередбачуваності» та ефектом витіснення людського суб'єкта автоматизованими агентами.

Автором аргументовано спростовано доктринальний міф про неминучу втрату правового контролю та розмиття відповідальності через автономність штучного інтелекту. Доведено, що будь-яка форма машинного самодетермінізму є похідною від людського дизайну, де межі дозволеної поведінки чітко окреслені через процедури верифікації та валідації.

У межах дослідження запропоновано перехід до інноваційної концепції «розподіленої моралі» у багатоагентних системах, де правовий результат розглядається як наслідок системної синергії людини та штучного агента. Обґрунтовано впровадження поняття «інфраетики» – цілісної системи морально-правових стимулів (енейблерів), спрямованих на превентивну координацію дій та мінімізацію ризиків.

Доведено необхідність стратегічного переформатування правової системи у напрямі збалансованого розподілу перспективної та ретроспективної відповідальності між розробниками, операторами та суб'єктами прийняття командних рішень. Резюмовано, що делегування повноважень алгоритмам потребує не відмови від класичного принципу вини, а розробки новітніх механізмів алгоритмічної підзвітності та цифрової безпеки.

Література

1. Danaher J. The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy and Technology*. 2016. Vol. 29. P. 245–268. DOI: 10.1007/s13347-015-0211-1

2. Roth A. Trial by Machine. The Georgetown Law Journal. 2016. Vol. 104. P. 1245–1305. URL: https://www.semanticscholar.org/paper/Trial-by-Machine-Roth/0dec1f1f868ad16629fca53085f81fec9ed05f10 (дата звернення: 27.03.2026).

3. Hildebrandt M. Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A.* 2018. Vol. 376, Is. 2128. Art. 20170355. DOI: 10.1098/rsta.2017.0355

4. Shany Y. The case for a new right to a human decision under international human rights law. Working Draft (SSRN). 2023. DOI: 10.2139/ssrn.4592244

5. Santoni de Sio F., van den Hoven J. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI.* 2018. Vol. 5. Art. 15. DOI: 10.3389/frobt.2018.00015

6. Banks I. Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act. *International Journal of Law and Information Technology*. 2026. Vol. 34. Art. eaag001. DOI: 10.1093/ijlit/eaag001

7. Матуєлене С., Шевчук В., Балтрунене Ю. Штучний інтелект в діяльності органів правопорядку та юстиції: вітчизняний та європейський досвід. *Теорія та практика судової експертизи і криміналістики.* 2022. Вип. 4 (29). С. 12–46. DOI: 10.32353/khrife.4.2022.02

8. Bryson J. J. Robots Should Be Slaves. Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues / ed. Yorick Wilks. Amsterdam : John Benjamins, 2010. P. 63–74. DOI: 10.1075/nlp.8.11bry

9. Gunkel D. J. The Machine Question: Critical Perspectives on AI, Robots, and Ethics. Cambridge, MA : The MIT Press, 2012. 253 p.

10. Solum L. B. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*. 1992. Vol. 70, Is. 4. P. 1231–1287. URL: https://scholarship.law.unc.edu/nclr/vol70/iss4/4

11. Bostrom N., Yudkowsky E. The Ethics of Artificial Intelligence. The Cambridge Handbook of Artificial Intelligence / eds. William Ramsey and Keith Frankish. Cambridge University Press, 2014. P. 316–334. DOI: 10.1017/CBO9781139046855.020

12. Hallevy G. The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control. *Akron Intellectual Property Journal*. 2010. Vol. 4, Iss. 2, Art. 1. P. 171–201. URL: https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1 (дата звернення: 27.03.2026).

13. Радутний О. Е. Artificial Intelligence (штучний інтелект) як суб'єкт правовідносин в галузі кримінального права. *Політика в сфері боротьби зі злочинністю* : матеріали Міжнар. наук.-практ. конф. з нагоди відзнач. 25-річчя навч.-наук. юрид. ін-ту (8–10 черв. 2017 р.). Івано-Франківськ, 2017. С. 200–206. URL: https://dspace.nlu.edu.ua/handle/123456789/13087 (дата звернення: 27.03.2026).

14. Bonfim T. C. Criminal liability of artificial intelligent machines: eyeing into AI's mind. Master Thesis. Lund University, Faculty of Law, 2022. 60 p. URL: https://lup.lub.lu.se/student-papers/search/publication/9095199 (дата звернення: 27.03.2026).

15. Vladeck D. C. Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*. 2014. Vol. 89, No. 1. P. 117–150. URL: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/6 (дата звернення: 27.03.2026).

16. Alduros S. M. S., Al-Daher H., Al Khattab K. A. R. The Impact of Artificial Intelligence (AI) on Criminal Law: Theoretical Review. *Information Sciences Letters.* 2025. Vol. 14, No. 3. P. 59–67. DOI: 10.18576/isl/140301

17. Bathaee Y. The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*. 2018. Vol. 31, No. 2. P. 889–938. URL: https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-

Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf (дата звернення: 27.03.2026).

18. Chagal-Feferkorn K. A. Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers. *Stanford Law & Policy Review*. 2019. Vol. 30. P. 61–114. URL: https://law.stanford.edu/wp-content/uploads/2019/05/30.1_2-Chagal-Feferkorn_Final-61-114.pdf (дата звернення: 27.03.2026).

19. Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*. 2004. Vol. 6, Is. 3. P. 175–183. DOI: 10.1007/s10676-004-3422-1

20. Balkin J. M. The Path of Robotics Law. *California Law Review Circuit.* 2015. Vol. 6. P. 45–60. URL: https://openyls.law.yale.edu/server/api/core/bitstreams/6e9fb63a-ba06-463b-be45-d6dc00412302/content (дата звернення: 27.03.2026).

21. Floridi L. Distributed Morality in an Information Society. 2017. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003075011-5/distributed-morality-information-society-luciano-floridi (дата звернення: 27.03.2026).

22. Noorman M., Johnson D. G. Negotiating autonomy and responsibility in military robots. *Ethics and Information Technology*. 2014. Vol. 16. P. 51–62. URL: https://www.dhi.ac.uk/san/waysofbeing/data/governance-crone-noorman-2014.pdf (дата звернення: 27.03.2026).

Information about the authors:
Kravchuk Svitlana Mykolaivna,
Senior Lecturer at the Department of Theory of Law and Constitutionalism
Institute of Law, Psychology, and Innovative Education
Lviv Polytechnic National University,
12, Stepan Bandera str., 79013, Lviv, Ukraine
Juror of the Shevchenkivskyi District Court of Lviv
https://orcid.org/0000-0002-3485-1343

Kovalchyn Dmytro Yuriyovych,
Master's degree candidate in Law (Specialization 081),
Institute of Law, Psychology, and Innovative Education
Lviv Polytechnic National University,
12, Stepan Bandera str., 79013, Lviv, Ukraine
https://orcid.org/0009-0009-1420-5676