

УДК 340.12:004.89:17.022.1

DOI <https://doi.org/10.24144/2788-6018.2025.05.3.67>

ВІД АЛГОРИТМУ ДО СПРАВЕДЛИВОСТІ: ФІЛОСОФСЬКО-ЕТИЧНІ МЕЖІ АВТОМАТИЗАЦІЇ ПРАВОСУДДЯ

Кучук А.М.,

*доктор юридичних наук, професор,
професор кафедри права та публічного управління
Сумського державного педагогічного університету імені А.С. Макаренка,
ORCID: 0000-0002-5918-2035*

Завгородня Ю.С.,

*кандидат юридичних наук, доцент,
доцент кафедри теорії та історії держави та права,
конституційного права та прав людини
Дніпровського державного університету внутрішніх справ
ORCID: 0000-0001-9257-7269*

Кучук А.М., Завгородня Ю.С. Від алгоритму до справедливості: філософсько-етичні межі автоматизації правосуддя.

У статті досліджується проблема співвідношення справедливості та алгоритмічної автоматизації правосуддя. Питання про природу справедливості отримує нові виміри в умовах використання штучного інтелекту у судочинстві. Класичні концепції справедливості зіставляються з можливостями алгоритмів, які, попри видиму неупередженість, відображають упередження своїх творців і навчальних даних. Особлива увага приділяється процедурній справедливості та проблемі «справедливої процедури», що вимагає поєднання результату з належним процесом ухвалення рішень. Підкреслено, що алгоритмічні системи функціонують у парадигмі індуктивного узагальнення, що суперечить принципам правової визначеності та стандартам доказування. Герменевтична традиція демонструє неспроможність алгоритмів адекватно враховувати контекст і ціннісні аспекти. У роботі також аналізується концепція ціннісного плюралізму Ісайї Берліна, яка вказує на неможливість редукції моральних цінностей до математичних критеріїв, що загрожує редукціонізмом у правосудді. Розглядається проблема «прогалини відповідальності», оскільки алгоритмічні системи позбавлені свободи та моральної відповідальності, що є необхідними умовами легітимності судового рішення. Окремо наголошується на епістемологічних труднощах: непрозорість алгоритмів суперечить вимогам обґрунтованості й зрозумілості судових рішень. Увага приділена темпоральному аспекту справедливості, оскільки алгоритми, навчені на історичних даних, ризикують відтворювати застарілі упередження. На основі філософсько-правового аналізу обґрунтовується принцип субсидіарності алгоритмічного правосуддя: автоматизація можлива у сфері процедурних, аналітичних та стандартизованих рішень, але є неприйнятною для питань вини, покарання чи складних ціннісних дилем. Запропоновано принципи етичної архітектури алгоритмічного правосуддя: моральної агентності, герменевтичної прозорості, ціннісної повноти та темпоральної справедливості. Зроблено висновок, що автоматизація правосуддя може виступати лише допоміжним інструментом, тоді як остаточна відповідальність за правове рішення має належати людині-судді.

Ключові слова: верховенство права, етика, права людини, правова процедура, правосуддя, філософія права, штучний інтелект, юридична деонтологія, юридичні інновації.

Kuchuk A.M., Zavorodnia Yu.S. From algorithm to justice: philosophical and ethical limits of justice automation.

The article explores the issue of the correlation between justice and algorithmic automation of justice. The question of the nature of justice gains new dimensions in the context of the artificial intelligence use in judicial proceedings. Classical concepts of justice are compared with the capabilities of algorithms, which, despite their apparent impartiality, reflect the biases of their creators and training data. Special attention is paid to procedural justice and the problem of "fair procedure" requiring combining the result with the proper decision-making process. It is emphasized that algorithmic systems function in the paradigm of inductive generalization, which contradicts the principles of legal certainty and standards of proof. The hermeneutic tradition demonstrates the inability of algorithms to adequately take into account the context and value aspects. The paper also analyzes the concept of value pluralism of Isaiah Berlin, which indicates the impossibility of reducing moral values to mathematical criteria,

which threatens reductionism in justice. The issue of the “accountability gap” is considered, since algorithmic systems are deprived of freedom and moral responsibility that are necessary conditions for the legitimacy of a judicial decision. Epistemological challenges are particularly emphasized, namely the opacity of algorithms contradicts the requirements of the validity and comprehensibility of judicial decisions. Attention is paid to the temporal aspect of justice, since algorithms trained on historical data risk reproducing outdated prejudices. Based on philosophical and legal analysis, the principle of algorithmic justice subsidiarity is substantiated: automation is possible in the field of procedural, analytical and standardized decisions, but is unacceptable for issues of guilt, punishment or complex value dilemmas. The principles of the ethical architecture of algorithmic justice are proposed, in particular moral agency, hermeneutic transparency, value completeness and temporal justice. It is concluded that the automation of justice can only act as an auxiliary tool, while the ultimate responsibility for the legal decision should belong to the human judge.

Key words: rule of law, ethics, human rights, legal procedure, justice, philosophy of law, artificial intelligence, legal deontology, legal innovations.

Постановка проблеми. Стрімкий розвиток штучного інтелекту та машинного навчання уже активно впливає на різні сфери суспільного життя, зокрема і на сферу права. Сьогодні алгоритми вже використовуються для аналізу доказів, прогнозування ризиків рецидиву, автоматизації документообігу та навіть прийняття рішень у деяких категоріях справ. Така технологічна революція потребує глибокого філософського осмислення меж та можливостей автоматизації в означеній галузі. Правосуддя торкається фундаментальних засад організації суспільного життя – верховенства права, права на справедливий суд, презумпції невинуватості, рівності перед законом та ін. Впровадження алгоритмічних рішень у судову систему може як покращити доступність та ефективність правосуддя, так і створити нові форми дискримінації та несправедливості. Особливо гостро це питання постає в контексті алгоритмічних упереджень, які можуть посилювати існуючі соціальні нерівності.

Наголосимо, що автоматизація правосуддя піднімає фундаментальні етичні питання: чи може алгоритм бути морально відповідальним за свої рішення? Як забезпечити прозорість та підзвітність автоматизованих систем? Яке співвідношення між ефективністю та справедливістю є прийнятним? Чи можна делегувати «машинам» право вирішувати долі людей? Вказані питання потребують не лише технічних, але й глибоких філософсько-етичних відповідей.

Додамо, що правові системи світу лише починають адаптуватися до викликів цифрової епохи. Відсутність чітких законодавчих меж для використання штучного інтелекту в правосудді, питання відповідальності за алгоритмічні рішення, захисту персональних даних та забезпечення права на оскарження автоматизованих рішень роблять це дослідження важливим для формування майбутньої правової політики.

Метою цього дослідження є проаналізувати філософсько-етичні засади автоматизації правосуддя та визначити концептуальні межі застосування алгоритмічних рішень у судовій системі.

Стан опрацювання проблематики. Проблематика автоматизації правосуддя та її філософсько-етичних аспектів перебуває на стику кількох дисциплінарних напрямів, що зумовлює фрагментарність її дослідження та необхідність міждисциплінарного синтезу.

У сфері філософії права фундаментальні питання справедливості та її співвідношення з технологічними рішеннями досліджували провідні мислителі сучасності. Джон Ролз у своїй теорії справедливості заклав основи для розуміння процедурної справедливості, що є важливим для оцінки алгоритмічних рішень [1]. Рональд Дворкін розробив концепцію права як цілісності, яка ставить під сумнів можливість механістичного застосування правових норм [2].

Лон Фуллер у роботі «Мораль права» сформулював принципи правового порядку, серед яких особливе значення для автоматизації має вимога зрозумілості та передбачуваності правових норм [3]. Його концепція створює теоретичну базу для аналізу етичності алгоритмічних рішень.

Густав Радбрух та його послідовники, розробили формулу співвідношення правової визначеності та справедливості, що має пряме відношення до проблеми автоматизації судових рішень [4-5].

Етичні аспекти застосування ШІ в правосудді активно досліджуються в рамках нової галузі – етики штучного інтелекту. Лучано Флоріді розробив концепцію інформаційної етики, яка пропонує новий погляд на моральні відносини в цифрову епоху [6].

Кетлін О'Ніл продемонструвала, як алгоритми можуть створювати та поглиблювати соціальну несправедливість, що має безпосереднє відношення до правосуддя [7].

Додамо, що підґрунтям цього дослідження є і попередні дослідження одного зі співавторів [8-9].

Незважаючи на зростаючу увагу до проблематики, залишаються значні лакуни у дослідженні комплексної філософсько-етичної теорії автоматизації правосуддя, яка б поєднувала класичні теорії справедливості з сучасними викликами, пов'язаними зі штучним інтелектом. Загалом пробле-

матика автоматизації правосуддя знаходиться на початковій стадії наукового осмислення. Існуючі дослідження здебільшого зосереджені на окремих аспектах проблеми та не пропонують цілісного бачення філософсько-етичних меж застосування штучного інтелекту в правосудді, що створює нагальну потребу в комплексному міждисциплінарному дослідженні, яке б синтезувало досягнення різних наукових напрямів та запропонувало концептуальну основу для етично обґрунтованого розвитку автоматизованого правосуддя.

Виклад основного матеріалу. Фундаментальне питання про природу справедливості набуває нових вимірів у контексті автоматизації правосуддя. Класичне аристотелівське розуміння справедливості як пропорційності та відповідності стикається з принциповою проблемою: чи може алгоритм, позбавлений морального суб'єктивізму, адекватно втілювати ідеал справедливості?

Концепція дистрибутивної справедливості Джона Ролза, заснована на принципі «завіси незнання» (the veil of ignorance), парадоксальним чином перегукується з алгоритмічною логікою. Алгоритм, теоретично позбавлений упереджень, міг би наближатися до ролзівського ідеалу неупередженого вибору. Однак це припущення руйнується при детальному аналізі: алгоритми є продуктом людської діяльності і неминуче відображають упередження своїх творців та навчальних даних. У цьому аспекті згадаємо, що як вказують окремі зарубіжні автори, результати досліджень свідчать, що великі мовні моделі виявляють поведінкові реакції, подібні до людських, коли стикаються з техніками переконання, навіть попри те, що вони не є людьми. Зокрема, застосування класичних принципів впливу – авторитету, послідовності та відчуття єдності – істотно підвищує ймовірність того, що штучний інтелект виконуватиме запити, від яких мав би відмовитися. Для GPT-4o-mini ефект виявився особливо значним: рівень виконання зріс майже удвічі (72,0% проти 33,3% у контрольній групі) [10].

Процедурна справедливість, як її розуміє Джон Ролз, вимагає не лише справедливого результату, але й справедливої процедури його досягнення. Алгоритмічні системи можуть забезпечити консистентність процедур, проте вони піднімають питання про сутність «справедливої процедури»: чи включає вона людський елемент розсуду як необхідну складову?

Правосуддя традиційно розглядається як процес встановлення істини через судове пізнання. Алгоритмічні системи пропонують альтернативний епістемологічний підхід, заснований на статистичній індукції та розпізнаванні паттернів, що створює фундаментальну напругу між істиною, що базується на доказах та аргументації, і алгоритмічною «істиною», що ґрунтується на кореляціях у великих масивах даних.

Карл Поппер у своїй філософії науки наголошував на принциповій відмінності між індуктивним узагальненням та дедуктивною фальсифікацією [11, с. 170–192]. Алгоритми машинного навчання функціонують переважно в парадигмі індуктивного узагальнення, що робить їх висновки принципово імовірнісними. У контексті правосуддя означене створює проблему: як узгодити імовірнісний характер алгоритмічних висновків з вимогами правової визначеності та стандартом доказування «поза розумним сумнівом»?

Герменевтична традиція, від Гадамера до Дворкіна, підкреслює інтерпретативний характер правового пізнання. Правова норма не є простою логічною конструкцією, але потребує інтерпретації в контексті конкретної життєвої ситуації. Алгоритмічні системи, при всій своїй складності, залишаються принципово механістичними та позбавленими герменевтичного розуміння контексту.

До цього слід додати, що правосуддя як соціальний інститут втілює множину цінностей: справедливість, рівність, ефективність, передбачуваність, індивідуальність підходу. Означені цінності часто перебувають у конфлікті одна з одною, що вимагає від судді здійснення ціннісного вибору. Ісайя Берлін у своїй концепції ціннісного плюралізму стверджував, що базові людські цінності є несумісними та їх неможливо звести до єдиної шкали вимірювання [12, с. 177–231].

Алгоритмічні системи, навпаки, вимагають операціоналізації цінностей через математичні функції. Означене створює проблему редукціонізму: складні моральні цінності мають бути зведені до числових параметрів. Процес такої операціоналізації неминуче спотворює сутність цінностей, перетворюючи їх на технічні критерії оптимізації.

Концепція етичних понять Бернарда Вільямса демонструє, що багато моральних категорій поєднують дескриптивний та нормативний аспекти в неподільній єдності. Поняття «жорстокості», «справедливості», «гідності» не можуть бути адекватно формалізовані без втрати своєї моральної сутності [13].

Загалом фундаментальне питання полягає в онтологічному статусі алгоритмічного рішення: чи може воно розглядатися як автентичний акт правосуддя? Традиційне розуміння судового рішення передбачає наявність морально відповідального агента, здатного до рефлексії та обґрунтування своїх дій.

Алгоритмічні системи, незважаючи на свою складність, позбавлені інтенціональності у феноменологічному розумінні. Вони не мають свідомості, волі чи моральної відповідальності, що піднімає

питання: чи може рішення, прийняте системою без моральної агентності, мати моральну легітимність?

Ганс Йонас у «Принципі відповідальності» наголошував на нерозривному зв'язку між свободою, знанням та відповідальністю [14]. Алгоритмічні системи можуть мати «знання» (у вигляді даних) та певну «свободу» (у межах своїх параметрів), але вони принципово позбавлені відповідальності. Наведене створює «прогалину відповідальності» в алгоритмічному правосудді.

Сутність правової держави вимагає, щоб правові рішення були обґрунтованими та зрозумілими. Традиційне судове рішення включає мотивувальну частину, яка пояснює логіку суду. Складні алгоритми машинного навчання, особливо глибокі нейронні мережі, функціонують як «чорні ящики», процес прийняття рішень у яких є непрозорим навіть для їх творців. Означене створює фундаментальну епістемологічну проблему: як можна довіряти рішенням, логіку якого неможливо зрозуміти? Вітгенштайнівське розуміння правил як соціальних практик підкреслює, що дотримання правила передбачає можливість його пояснення та обґрунтування [15]. Алгоритмічне рішення, позбавлене можливості раціонального пояснення, ставить під сумнів свою правову природу.

Спроби створення «пояснюваних» алгоритмів стикаються з фундаментальною дилемою між точністю та інтерпретованістю. Більш прості, зрозумілі алгоритми часто менш точні, тоді як найбільш ефективні системи залишаються непрозорими, що створює необхідність вибору між ефективністю та легітимністю.

Окрім цього, правосуддя має фундаментально темпоральний характер: воно розгортається в часі та враховує історичний контекст. Алгоритмічні системи, навчені на історичних даних, ризикують увічнити минулі упередження та несправедливості, що створює темпоральний парадокс: система, покликана покращити правосуддя, може законсервувати його недоліки.

Концепція «алгоритмічної справедливості» як статичної властивості системи протиставляється динамічному розумінню справедливості як постійного процесу. Деррідівська деконструкція протиставляє справедливість як нескінченне завдання праву як системі правил [16]. Алгоритмічні системи, будучи втіленням правил, можуть віддаляти нас від справедливості як відкритого горизонту.

Відтак, на основі філософського аналізу можна сформулювати принцип субсидіарності для алгоритмічного правосуддя: алгоритми можуть застосовуватися лише там, де людський розсуд не є необхідним для досягнення справедливості. Наведене передбачає, що автоматизація допустима для:

1. Процедурних рішень – розподіл справ, календарне планування, перевірка формальних вимог.
2. Аналітичних завдань – пошук прецедентів, аналіз великих обсягів документів, виявлення паттернів.
3. Стандартизованих рішень – справи з чітко визначеними критеріями та обмеженим ціннісним контекстом.

Водночас алгоритмічна автоматизація є неприйнятною для:

1. Рішень про вину та покарання – які вимагають морального судження.
2. Складних ціннісних дилем – де необхідне балансування конкуруючих принципів.
3. Унікальних життєвих ситуацій – що потребують індивідуального підходу.

Філософсько-правовий аналіз дозволяє окреслити принципи етичної архітектури систем автоматизованого правосуддя:

Принцип моральної агентності: остаточна відповідальність за правове рішення завжди має належати людині-судді, алгоритм може лише підтримувати процес прийняття рішення.

Принцип герменевтичної прозорості: алгоритмічні системи мають бути здатними не лише до статистичного прогнозування, але й до смислового пояснення своїх рекомендацій у термінах, зрозумілих для правової спільноти.

Принцип ціннісної повноти: алгоритм має враховувати весь спектр релевантних правових цінностей, а не лише ті, що піддаються простій квантифікації.

Принцип темпоральної справедливості: система має включати механізми постійного оновлення та корекції для запобігання увічненню історичних несправедливостей.

Наведені принципи визначають концептуальні межі автоматизації правосуддя, утверджуючи пріоритет людської моральної агентності над алгоритмічною ефективністю у сфері, що торкається прав і свобод людини.

Висновки. Таким чином, алгоритмічні системи, попри їхню ефективність та здатність до обробки великих обсягів даних, не можуть повноцінно замінити людський розсуд у сфері правосуддя. Їхня обмеженість полягає у відсутності моральної агентності, інтенціональності та здатності до герменевтичної інтерпретації. Справедливість у правовому контексті завжди має багатовимірний характер, поєднуючи рівність, індивідуальний підхід і ціннісний вибір, що принципово не піддається редукції до математичних моделей. Алгоритмічні рішення здатні відігравати допоміжну роль – у стандартизованих, процедурних та аналітичних завданнях – але остаточне судове рішення має

залишатися у площині людської відповідальності. Саме тому автоматизація правосуддя можлива лише на умовах субсидіарності, коли алгоритм доповнює, а не замінює людину. Такий підхід дозволяє зберегти баланс між технологічною інновацією та ціннісними засадами правовладдя.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Ролз Д. Теорія справедливості / Пер. з англ. О.Мокровольський. Київ: Основи, 2001. 822 с.
2. Дворкін Р. Серйозний погляд на права / Пер. з англ. Андрій Фролкін. Київ: Основи, 2000. 519 с.
3. Фуллер Лон Л. Мораль права / пер. з англ.. Н. Комарова. Київ: Сфера, 1999. 232 с.
4. Радбрух Г. Законне неправо та надзаконне право: пер. з нім. В.С. Бігуна. *Проблеми філософії права*. 2004. Том II. С. 83–94.
5. Середюк В.В., Дудар С.К. Формула Густава Радбруха: концептуальний аналіз та практичне значення. *Науковий вісник Ужгородського національного університету. Серія: Право*. 2024. Вип. 84. Ч. 4. С. 389–396.
6. Floridi L. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford University Press, 2023. 272 p.
7. О'Ніл К. BIG DATA. Зброя математичного знищення. Як великі дані збільшують нерівність і загрожують демократії. BookChef, 2019. 336 с.
8. Kuchuk A.M., Andrusyshyn B.I., Bilozorov Y.V., Minchenko O.V., Filianina L.A. Lawyer's emotional intelligence and professional ethics development: aspects of the widespread introduction of legal clinic. *Linguistics and Culture Review*. 2021. №5(S2). P. 608-618.
9. Perepolkin S., Kuchuk A., Lehka, O., Labenska L., Stryzhak I. Artificial Intelligence and international law: from recommendations to conventional regulation. *Brazilian Journal of International Law*. 2024. Vol. 21. №3. P. 319–337.
10. Meincke L., Shapiro D., Duckworth A., Mollick E., Mollick L., Cialdini R. Call Me A Jerk: Persuading AI to Comply with Objectionable Requests (July 18, 2025). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5357179.
11. Поппер К. Логіка наукового відкриття. Огляд деяких фундаментальних проблем / Пер. з англ. під кер. О.Є. Оліфер. Актуальні проблеми духовності. 2021. № 22. С. 170–192.
12. Берлін І. Чотири есеї про свободу / перекл. з англ. О. Коваленка. Київ: Основи, 1994. 272 с.
13. Williams B. *Philosophy as a Humanistic Discipline*. Princeton University Press, 2008. 264 p.
14. Йонас Г. Принцип відповідальності. У пошуках етики для технологічної цивілізації. Пер. з нім. Київ: Лібра, 2001. 400 с.
15. Вітгенштайн Л. *Tractatus Logico-Philosophicus*. URL https://chtyvo.org.ua/authors/Wittgenstein_Ludwig/Lohiko-filosofskyi_traktat.
16. Деріда Ж. Письмо та відмінність / перекл. В. Шовкун. Київ: Основи, 2004. 602 с.