

Штучний інтелект для прийняття судових рішень: деякі потенційні ризики¹

Юлія Сергіївна Разметаєва*

Національний юридичний університет імені Ярослава Мудрого,
Харків, Україна;
Уппсальський університет,
Уппсала, Швеція

*e-mail: yu.s.razmetaeva@nlu.edu.ua

Анотація

Актуальність теми дослідження зумовлено швидким і невідворотнім розгортанням технологій штучного інтелекту у більшості правових систем, що включає його застосування у процесі прийняття рішень. Мета статті полягає у тому, щоб, проаналізувавши використання штучного інтелекту у прийнятті судових рішень, виокремити неясні потенційні ризики алгоритмізації та запропонувати розумний підхід до впровадження відповідних технологій. У дослідженні застосовано такі методи наукового пізнання, як структурний аналіз, уявний експеримент та правовий доктринальний підхід, а також елементи «case study». Досліджено питання штучного інтелекту у прийнятті судових рішень, акцентуючи увагу на потенційних ризиках і викликах його впровадження. Підкреслено необхідність основи у вигляді цінностей і принципів справедливості та верховенства права при застосуванні штучного інтелекту, а також надано аргументи розумної (розсудливої) алгоритмізації. Наголошено на проблемах алгоритмізації складних процесів судочинства, зокрема щодо вибору правових принципів та можливостей штучного інтелекту впливати на індивідуальний характер правосуддя. Серед ризиків виокремлено схильність штучного інтелекту до раціоналізації та стандартизації рішень, обмежену здатність до інтерпретації людських характеристик та обставин справ, підміну правової визначеності алгоритмічною передбачуваністю. Обговорено труд-

¹ Це дослідження є частиною проекту Центр Досконалості Жана Моне «Європейські фундаментальні цінності в цифрову еру», 101085385 – EFVDE – ERASMUS-JMO-2022-HEICTH-RSCH, що фінансується Європейським Союзом. Однак висловлені погляди та думки належать лише авторам і не обов'язково відображають погляди Європейського Союзу чи Європейського виконавчого агентства з питань освіти і культури (EACEA). Ні Європейський Союз, ні орган, що надає гранти, не можуть нести за них відповідальності. Окремі результати дослідження були опубліковані англійською мовою у статті: Razmetaeva Y. (2024). Artificial intelligence and the end of justice. *BioLaw Journal – Rivista di BioDiritto*, 1, 345-365. <https://doi.org/10.15168/2284-4503-3001>.

нощі, пов'язані з розумінням і тлумаченням юридичних текстів алгоритмами, зазначено, що штучний інтелект не здатен до справжнього мислення й морального судження. Окрему увагу приділено питанню юридичної аргументації: стверджується, що рішення суду мають бути не лише обґрунтованими, а й переконливими для суспільства, чого неможливо досягти за допомогою штучного інтелекту через його нездатність розуміти дискурс і контекст справ. На підставі проведеного дослідження зроблено висновок, що, незважаючи на технологічний прогрес, повна заміна людського судження штучним інтелектом несе серйозні ризики і може призвести до викривлення самого поняття справедливості та її знецінення. Перспективи подальших досліджень можуть включати осмислення того, як технологічний детермінізм та чистий утилітаризм переважають більш обережні та помірковані підходи й відсунувають деонтологічні міркування на другий план.

Ключові слова: штучний інтелект; принципи права; верховенство права, правосуддя; судові рішення; справедливість.

Artificial Intelligence for Judicial Decision-making: Some Potential Risks

Yulia S. Razmetaeva*

Yaroslav Mudryi National Law University,

Kharkiv, Ukraine;

Uppsala University,

Uppsala, Sweden

*e-mail: yu.s.razmetaeva@nlu.edu.ua

Abstract

The relevance of the research topic is due to the rapid and inevitable deployment of artificial intelligence technologies in most legal systems, which includes its application in the decision-making process. The purpose of the article is to, after analyzing the use of artificial intelligence in judicial decision-making, to single out implicit potential risks of algorithmization and to propose a reasonable approach to the implementation of those technologies. The research uses such methods of scientific knowledge as structural analysis, imaginary experiment and legal doctrinal approach, as well as "case study" elements. The article explores the issue of implementing artificial intelligence in judicial decision-making, accentuating potential risks and challenges. It highlights the need to consider justice, fairness and the rule of law when applying AI, and provides arguments for a reasonable and limited algorithmization. The article focuses on the problems of algorithmizing complex judicial processes, particularly regarding the selection of legal principles and AI's potential negative impact on the individualized nature of justice. Among the risks, the tendency of AI towards rationalization and standardization of decisions, its limited ability to interpret human characteristics and case circumstances, and the substitution of legal certainty with algorithmic predictability are emphasized. The article also discusses the difficulties related to the understanding and interpretation of legal texts by algorithms,

noting that AI is incapable of thinking and making moral judgment. Special attention is given to the issue of legal reasoning; the article argues that court decisions must not only be justified but also convincing to society, which is impossible to achieve with AI due to its incapability to comprehend discourse and case context. Based on the conducted research the article concludes that despite technological advances, the complete replacement of human judgment with AI carries risks and may lead to a distortion of the very concept of justice and its devaluation. Prospects for further research may include understanding how technological determinism and pure utilitarianism override more cautious and moderate approaches and relegate deontological considerations to the background.

Keywords: artificial intelligence; principles of law; rule of law; justice; court decisions; fairness.

Вступ

Штучний інтелект (ШІ) у процесі прийняття судових рішень, безсумнівно, має застосовуватися з урахуванням вимог справедливості, верховенства права та прав людини, бути прозорим, підзвітним і неупередженим, а також бути навченим на різноманітних, очищених і належним чином захищених даних. Немає також сумнівів, що спроби алгоритмізувати справедливість, яка лежить в основі судових рішень, триватимуть доти, доки відповідні технології продовжують стрімко розвиватися. Однак, розглядаючи алгоритмізацію прийняття рішень у процесі відправлення правосуддя, слід також розглянути її потенційні ризики. Низка питань, які слід поставити при цьому, включає такі проблеми, як: 1) належний вибір правових принципів або, у ширшому сенсі, концепцій справедливості; 2) подолання граничної складності процесу прийняття судових рішень; 3) збереження індивідуального характеру правосуддя; 4) відповідність стандартам верховенства права, що включають неупередженість, обґрунтованість та зрозумілість рішень; 5) відповідність принципу правової визначеності.

Належний вибір вихідної концепції для ШІ є основою розгляду перелічених вище проблем. Цей вибір ускладнюється тим, що консенсус можливий на рівні загальних принципів справедливості, але майже недосяжний на рівні детальних понять і їх практичного втілення. Якщо навіть концепція буде обрана, її потрібно трансформувати на рядки коду й навчити ШІ приймати рішення на основі цього алгоритму, що може виявитися наступним майже неможливим завданням, враховуючи граничну складність і заплутаність процесу прийняття судових рішень, яка, по суті, містить «людську» частину на додаток до суто раціональної, та особливості машинного навчання. Нарешті, навіть якби ми прийняли якійсь набір принципів або одну з концепцій справедливості, їх складність така, що навряд чи може бути наведена у вигляді алгоритму. І навіть якби якимось чином вдалося алгоритмізувати надзвичайно

складну модель справедливості, цей уявний алгоритм потрібно буде застосувати до нескінченно складних людських стосунків, які постійно розвиваються.

Ідея кодифікувати справедливість або зробити її обчислюваною має як прихильників, так і противників. З одного боку, численні спроби створити моделі для вимірювання справедливості, рівності та принципів права як таких відображають прагнення багатьох досягти такої справедливості й, відповідно справедливих рішень, які спираються на ШІ [див.: 1–3]. З іншого боку, є численні аргументи проти, серед яких особлива увага має бути приділена таким аспектам, як нездатність машин розуміти контекст [4], схильність до «помилкових узагальнень» [5, с. 1265] і схильність багатьох моделей ШІ «цементувати або навіть посилювати наявні дискримінаційні практики та нерівності» [6, с. 655], що може зробити алгоритмічну справедливість недосяжною.

Іншою передумовою для загального скептицизму щодо алгоритмізації справедливості може бути нездатність машин обмірковувати та переосмислювати прийняті рішення. Люди, які застосовують концепції справедливості та принципи права в судах і органах державної влади, а також люди, які розмірковують над ними та навчають інших, вивчають і переосмислюють власні підходи протягом значної частини життя. Це включає передачу знань і досвіду, та, як дехто може сказати, формування почуття справедливості й здатності робити моральні судження. Цілком імовірно, що ті виклики, які постають перед людьми та суспільствами на шляху алгоритмізації, сприятимуть відродженню поглядів, заснованих на природному праворозумінні та його варіаціях, оскільки стане зрозумілим, що праву слід спиратися на «невловимі», але необхідні ціннісні основи.

Мета цього дослідження полягає у тому щоб проаналізувавши використання штучного інтелекту у прийнятті судових рішень виокремити неявні потенційні ризики алгоритмізації та запропонувати розумний підхід до впровадження відповідних технологій. Досягнення мети дослідження передбачає вирішення таких завдань: 1) аналіз здатності ШІ забезпечувати справедливі, обґрунтовані та переконливі рішення; 2) дослідження здатності ШІ розуміти зміст текстів та інтерпретувати їх, у тому числі контекстно; 3) вивчення проблем алгоритмічної передбачуваності та правової певності; 4) й осмислення стандартів верховенства права для впровадження ШІ у процеси відправлення правосуддя.

Матеріали та методи

Теоретична основа цього дослідження складається з поєднання правових принципів та цінностей, з фокусом на справедливість та верховен-

ство права, стандартів прийняття рішень, що мають юридичне значення з фокусом на розуміння й аргументацію, та особливостей розгортання й використання технологій штучного інтелекту. Центральною гіпотезою тут є припущення про те, що алгоритмізація прийняття судових рішень має неочевидні й віддалені у часі ризики, що означає, що повинен бути прийнятий виважений і розумний підхід до впровадження ШІ у процеси відправлення правосуддя.

Ця гіпотеза, а також деякі з наведених у вступі аргументів та сумнівів будуть розглянуті у цій роботі з використанням методів структурного аналізу, уявного експерименту та правового доктринального підходу. Для ілюстрації окремих аргументів у цій статті використовуються елементи *case study* в частині вивчення рішень Європейського суду з прав людини та Суду Європейського Союзу.

Результати та обговорення

У цьому підрозділі дослідження розглянемо деякі потенційні ризики впровадження ШІ у прийняття судових рішень.

Раціоналізація, стандартизація та утилітарний підхід

Здатність ШІ забезпечувати «кращу» справедливість часто вважається само собою зрозумілою, оскільки «розумні» алгоритми справляються з багатьма завданнями краще ніж люди. Часто це означає, що ШІ-технології значно випереджають людей за обчислювальними здатностями, при цьому алгоритми не втомлюються, не витрачається час на іншу діяльність і не зазнають впливів нераціональних, емоційних сторін особи при прийнятті рішень. Водночас слід зазначити, що ШІ серйозно відрізняється від людського. Логіка алгоритму потребує вимірювання і зважування різних проявів людської природи, виражаючи деякі «людські» частини за допомогою цифр і відсотків. Таким речам, як вік, раса, стать та етнічна належність, мають бути присвоєні числові атрибути (причому це робиться до процесу розгляду справи, на етапі створення або коригування алгоритму), що робить їх основою оцінки. Це саме по собі може розглядатися як дискримінація.

Хтось може заперечити, що вимірювання та зважування, а також оцінка різних рис, яка робить їх основою оцінки, включені й до того, що роблять судді-люди. Зрештою, судді оцінюють вік або стать як правопорушників, так і жертв правопорушення як обставини, що впливають на рішення у справі, наприклад, щоб виправдати те, що конкретні злочини слід кваліфікувати як гендерно зумовлене насильство, домашнє насильство чи насильство щодо

неповнолітніх. У певних категоріях справ оцінка людських рис є основою юридичної аргументації, наприклад, коли йдеться про порушення стандартів, що забороняють дискримінацію.

Однак те, що відрізняє алгоритми від оцінки, яку можуть дати судді-люди, полягає в тому, що кількісні значення призначаються людським рисам заздалегідь. Щоб створити або навчити ШІ, необхідно дати чітке кількісне значення цих характеристик. Отже, рішення виявляється обтяженим заздалегідь встановленими міркуваннями. Хоча ця особливість може не стосуватися технологій ШІ, що самонавчаються, подібні алгоритми, зокрема штучні нейронні мережі, можуть мати інші підстави для обмеження їх використання в системі правосуддя, головною з яких є їх загальна невідомість і більш специфічна проблема з пояснюваністю [див.: 7–9], які підживлюють аргумент, що це було б необачно використовувати інструменти ШІ, які люди, що приймають рішення, не можуть зрозуміти [10].

Якщо провести уявний експеримент і припустити, що судді-люди стануть подумки зважувати расу, вік або стать осіб, призначаючи цим характеристикам 13, 11 і 9 % відповідного впливу на рішення у справі, постане питання: чи не здається такий спосіб прийняття рішень суддями надзвичайно дивним і далеким від того, як це відбувається сьогодні? Це, звичайно, спрощений погляд, і ШІ, безумовно, здатний швидко обробляти дані про мільйони характеристик людини та враховувати мільйони їх комбінацій. Однак той факт, що кількісне визначення буде виражено в тисячних частках відсотка і що точність алгоритмічних моделей зростає щодня, не робить меншою проблемою те, що люди, які розробляють або навчають певний ШІ, повинні стати на шлях чистого утилітаризму, призначаючи обчислювальні значення індивідуальним рисам та особливостям.

З іншого боку, існує межа складності алгоритмічного визначення, яке неминуче пропускає деталі, іноді – істотні деталі. Алгоритми, як правило, стандартизовані, тоді як правосуддя має бути індивідуалізованим. Все, що лежить за межами алгоритму, – унікальні обставини, рідкісні людські прояви, справжні наміри та емоційні стани – буде проігноровано машиною (але буде взято до уваги людиною-суддею). Це спотворює саму ідею справедливості. Тоді як судді, розглядаючи справу та приймаючи рішення, можуть враховувати індивідуальні обставини – в рамках принципів права та нормативних вимог. Так, схильність ШІ до стандартизації та виключно утилітарна логіка алгоритмів є потенційними ризиками впровадження цих технологій у прийняття судових рішень.

Обробка юридичних текстів та розуміння їх змісту

Сприйняття ШІ як чогось досконалого, видається, засноване на тому, що алгоритм виконує завдання разюче ефективно, швидко й точно. Водночас те, що люди бачать як результат дії ШІ, не є результатом процесу мислення, радше – його імітацією. Люди, проте, схильні довіряти алгоритмам, тому що останні демонструють успішне удавання мислення або чогось дуже схожого на мислення. Враження від спілкування за допомогою ChatGPT, наприклад, можуть бути досить сильними, щоб забути, що цей ШІ відтворює спілкування, а не робить це в такий спосіб, як ніби користувач ШІ мав діалог із людиною, яка знаходиться по інший бік екрана девайсу. Тобто ця імітація мислення, яку демонструє ШІ, може захоплювати настільки, щоб люди мимоволі наділили алгоритм суб'єктивністю та навіть свободою волі.

Зважаючи на відмінності між людським і нелюдським мисленням, немає серйозних гарантій того, що будь-який ШІ, який використовується під час здійснення правосуддя, розуміє правові принципи та норми й правильно тлумачить їх. Крім того, немає серйозних гарантій того, що ШІ зможе переконати людей у справедливості рішення. Безсумнівно, дослідники та розробники алгоритмів досягли значного прогресу в питанні пояснюваного ШІ, від появи конкретних методів, таких як, наприклад, використання методів *post-hoc* пояснюваності [11], до таких підходів, як, наприклад, використання прагматичних інструментів для збільшення прозорості ШІ-моделей [12]. Проте слід підкреслити, що пояснення та обґрунтування, особливо переконливе, – це не одне й те саме.

Алгоритми стають дедалі досконалішими у розпізнаванні та вилученні інформації з текстів. Це стосується і юридичних текстів: законів і судових рішень, проектів актів, візців правової доктрини тощо. Безперечно, швидкість обробки текстів і точність вилучення даних із подальшим формуванням зв'язних і змістовних відповідей є приголомшливими. Інформація, яку «витагує» з текстів ШІ, крім того, придатна для подальшої ефективної обробки, оскільки часто це структуровані дані. Однак, незважаючи на те, що алгоритми ідеально підходять для обробки, вилучення, інтерпретації та збирання фрагментів даних, виводячи відповіді, які важко відрізнити від людських відповідей, або навіть такі, що виглядають кращими за візці людських, ми все одно не можемо сказати, що машини «розуміють» те, що вони «читають».

Якщо задати алгоритму певні критерії, він успішно впорається з пошуком, розпізнаванням і «розумінням» того, що, наприклад, 200 судових рішень у базі даних є рішеннями про примусове стягнення аліментів з батьків.

Розумний алгоритм, який має конкретне завдання, може допомогти нам встановити шаблони у вибірці. Наприклад, ШІ може показати, що 87 % судових справ стосуються стягнення аліментів з батька, а 13 % – з матері. Ця інформація, відсортована та наведена у зручному форматі, може допомогти зрозуміти деякі закономірності, передбачити тенденції та побудувати судові моделі щодо справ про аліменти. Водночас виокремлення патернів, з якими добре справляється ШІ, не має підмінювати всебічне розуміння суті справи, якого ми вимагаємо від суддів-людей. Тому алгоритми можуть бути гарними допоміжними інструментами правосуддя, однак не єдиними, хто вивчає, оцінює справу та висуває аргументи на користь певного рішення. Отже, одним із потенційних ризиків залучення ШІ до вирішення судових справ є їх значне, а не допоміжне використання. Цей ризик спровокований надмірним покладанням людей на здатність алгоритмів до розуміння текстів, яке впливає з хибного припущення про те, що вдала імітація мислення штучними агентами може бути достатньою для виконання певної роботи, тоді як розуміння сенсів і мислення як таке є властивостями людського розуму.

Юридична аргументація, обґрунтованість та переконливість рішень

На продовження попереднього аргументу, слід зазначити, що справедливі судові рішення мають бути належним чином обґрунтованими та в певному сенсі переконливими як для учасників процесу, так і для широкого загалу.

Розумно припустити, що принаймні деякі з технологій ШІ можуть або невдовзі зможуть «зрозуміти» зміст тексту з точки зору їх здатності знаходити та аналізувати факти. Водночас можна стверджувати, що будь-який алгоритм не може зрозуміти тексти, оскільки він не може інтерпретувати факти в світлі нескінченної кількості міркувань, на відміну від суддів-людей. Щоб визначити, яким буде справедливе рішення в конкретній справі, судді застосовують (або принаймні мають застосовувати) не лише принципи права, законодавство та прецеденти, а й складні міркування, засновані на їхньому розумінні закону та досвіді застосування права. Таким чином, можна припустити, що будь-яка особа-суддя здатна зрозуміти зміст тексту, включаючи здатність знаходити й аналізувати факти та інтерпретувати їх у світлі нескінченної кількості контекстів. Це, безперечно, стосується будь-якої судді, яка пройшла кваліфікаційний відбір на посаду, за умови, що в суспільстві існує незалежна судова система, яка демонструє певний рівень суспільної довіри, і де правова система заснована на верховенстві права. Крім того, судді повинні обґрунтувати це рішення так, щоб переконати інших людей у його справедливості.

Від правильного розуміння та обґрунтування також залежить здатність правильно збалансувати фундаментальні права та законні інтереси. Останнє, у свою чергу, як видається, включає розуміння дискурсу як такого або, у вужчому сенсі, контексту конкретного випадку. Зокрема, щоб збалансувати свободу вираження думок із приватністю, заборонаю дискримінації чи захистом від мови ненависті, може знадобитися розуміння дискурсу свободи вираження, а також політичний, соціальний, релігійний тощо контекст справ.

Так, судді можуть, наприклад, відрізнити форми протесту від порушень громадського порядку та тлумачити мову ворожнечі як таку, що суперечить цінностям Конвенції. Наприклад, ЄСПЛ установив, що у справі «Швидка проти України» мало місце невербальне вираження, яке захищається статтею 10 Конвенції про захист прав людини і основоположних свобод [13]. У цьому випадку заявниця, як представниця опозиційної політичної партії, під час офіційної церемонії обірвала стрічку на вінку, покладеному до монументу. Потім вона розірвала стрічку з ім'ям Януковича, що на той момент був президентом України, тим самим висловлюючи свій протест проти того, що він займає цю посаду. У справі «Lenis проти Греції» ЄСПЛ встановив, що скарга щодо застосування статті 10 Конвенції є неприйнятною [14] й, відповідно, дії особи не мають бути під захистом свободи вираження. У цьому випадку церковний чиновник опублікував гомофобну статтю у своєму особистому блозі, коли грецький парламент збирався обговорити запропонований закон про запровадження цивільних союзів для одностатевих пар. Суд постановив, що заявник намагався відвернути свободу вираження поглядів Конвенції від її справжньої мети, використовуючи її в цілях, що явно суперечать цінностям, яким Конвенція прагне сприяти. Обидва випадки демонструють складність балансування прав та інтересів, а також потребу в розумінні дискурсу навколо прав і контексту конкретних справ.

І дискурс, і контекст поки що є тим, у що люди, а не ШІ, можуть повноцінно зануритися й що вони беруть до уваги. Тому нездатність алгоритмів до належного їх розуміння й, відповідно, належного, зрозумілого та переконливого обґрунтування рішень створює додаткові ризики для впровадження ШІ у процеси відправлення правосуддя.

Алгоритмічна передбачуваність і правова певність

Одним із завдань ШІ-технологій є покликання допомогти людям впоратися з невизначеністю, що включає здатність розраховувати та передбачати майбутні рішення. Утім невпевненість і здатність сумніватися також є важливими складниками людської природи. «Невизначеність», однією з дефініцій якої є «непередбачуваність подій» [15], може бути проблемою в деяких кон-

текстах, але це не те, що люди можуть – і повинні – подолати раз і назавжди. ШІ, однак, націлений на певність, на остаточне вирішення якоїсь проблеми. Він також прагне максимально збільшити передбачуваність. Справедливість, з іншого боку, полягає не лише в певності. Дуже часто йдеться про невпевненість і сумніви. Тому людям потрібні незалежні та справедливі суди, щоб певною мірою вирішувати сумніви та індивідуалізувати правосуддя залежно від правової основи та обставин справи.

Більше того, здається, що брак місця для сумнівів і дебатів, що впливають із цих сумнівів, означає брак справедливості. Довгострокові наслідки розгляду складних справ і прийняття обґрунтованих рішень часто викликають широкі публічні дебати. Приміром, низка справ ЄСПЛ та Суду ЄС мали серйозний вплив на правові системи країн – учасниць Конвенції про захист прав людини і основоположних свобод та Хартії прав ЄС, відповідно. Дискусії, які виникають щодо рішень по складних справах, здатні спровокувати зміни в законодавстві та юридичній практиці, а в ширшому розумінні – зміну правових підходів і серйозні суспільні зрушення. Відсутність дискусії та заміна дебатів процентним прорахунком у рішеннях збіднює людські суспільства.

Хтось може стверджувати, що результати «рішень» алгоритмів не виражають 100 % упевненості, а швидше, являють собою шкалу ймовірностей. Імовірніший ШІ, особливо ймовірнісне машинне навчання, надзвичайно корисні, коли це допомагає визначати речі, які не допускають подвійної інтерпретації. Наприклад, якщо мовиться про виокремлення послідовностей у вірусах чи визначення швидкості вітру. Імовірнісна статистика може бути шкідливою, якщо вона якимось чином визначає результат судового рішення. Дійсно, причинно-наслідковий зв'язок між алгоритмічною рекомендацією та конкретним збігом рішення може бути сумнівним, і важко довести, що він був причинно-наслідковим, навіть якщо він був таким насправді. У світлі цього цікавим є результат розгляду справи «SCHUFA Holding (Scoring)», у якій Суд Європейського Союзу постановив, що рішення має розглядатися як «автоматизоване індивідуальне рішення», коли воно має значний вплив на індивідуальні права і значною мірою покладається на рекомендації ШІ [16]. Визначальна чи ні, рекомендація, надана алгоритмом, може сприйматися як більш точна, ніж людська, а висока точність деяких алгоритмів може сприйматися як частина їхньої досконалості та, зрештою, безпомилковості запропонованих результатів. Це впливає на запровадження ШІ у процеси відправлення правосуддя та невинуватий рівень довіри до результатів, які пропонують такі технології.

Іншим ризиком, який постає при використанні ІІІ для прийняття судових рішень, є поява алгоритмічної передбачуваності, яка часто сприймається в позитивному ключі, оскільки ІІІ демонструють точніші прогнози й, відповідно, більше впливають на кінцеві рішення [див.: 17–19]. Така алгоритмічна передбачуваність, що дедалі більше підмінює правову певність, може підірвати концепції справедливості та верховенства права, оскільки, якщо ми вже знаємо результат справи, немає жодних причин шукати справедливості як такої й немає причин спиратися на принципи та норми права, а не на машинні розрахунки й комп'ютеризовані моделі. Прагнення до передбачуваності видається видатним і ніби затьмарює прагнення до справедливості та права. Однак правова певність – і визначеність як така – у праві далеко не те саме, що алгоритмічна визначеність.

Дійсно, правова визначеність є елементом верховенства права і включає, серед іншого, вимоги передбачуваності. Ці вимоги в найзагальнішому сенсі означають, що люди мають право розуміти наслідки своїх дій і зміст законів. Проте передбачуваність наслідків законних чи незаконних дій не означає, що ми можемо мати стовідсоткову впевненість у результаті судового процесу. Радше, це може спрацювати лише в деяких простих випадках, коли правильне і неправильне, справедливе і несправедливе очевидні. Навпаки, у складних судових справах, у яких є конфлікт прав або зіткнення прав і законних інтересів, передбачуваність не працює на користь верховенства права. Передбачуваність результату складних справ не може бути високою через складність і динамічний характер певного балансу прав у поєднанні з відсутністю консенсусу щодо цього балансу.

ІІІ працює на основі попереднього досвіду, який було введено у вигляді даних. Так само судді в більшості випадків враховують значну частину попереднього досвіду, втіленого у правові позиції, акти законодавства і доктрини. Утім судді-люди потенційно готові та здатні розглядати абсолютно нові справи, виносячи справедливі рішення незалежно від їх новизни. Хоча міжособистісні конфлікти підпадають під обмежену кількість сценаріїв, конкретні деталі та обставини завжди варіюються, тому важко та, ймовірно, неможливо передбачити усі сценарії випадків, які трапляються насправді. Відповідно, ІІІ не буде добре справлятися з викликами новизни, оскільки алгоритми фактично спираються на попередні досвіди та відтворюють їх у різних комбінаціях.

Розумна алгоритмізація процесів відправлення правосуддя

Здається, що існує дуже обмежений перелік видів справ, які можна довірити ІІІ. Застосування ІІІ можливе в рутинних і простих випадках, які не потре-

бують глибокого аналізу, але потребують застосування рутинних процедур. Зокрема, це можуть бути справи сімейного права, нескладні, повторювані та пов'язані з формальним визнанням фактичного стану. Приміром, у багатьох правових системах звернення до суду необхідне у випадках, коли обидві сторони хочуть розлучитися і не мають претензій одна до одної, але мають спільних неповнолітніх дітей. Застосування ШІ можливе, коли складність справи є надзвичайно низькою через те, що обставини дуже прості та спір відсутній у сенсі конфлікту як такого, тобто у справі немає реального конфлікту права та законних інтересів, що мають бути визначені або збалансовані. У більшості випадків результат таких справ цілком передбачуваний. Алгоритмізувати можна також справи, які є досить формальною частиною адміністративного чи цивільного процесу, зокрема ті, що стосуються винесення судових приписів чи підтвердження встановлених фактів.

Отже, використання ШІ має визначатися способом розгляду справ. У свою чергу, спосіб розгляду, як пише Дорі Рейлінг, залежить від таких факторів, як складність інформації у справі, а також ступінь передбачуваності результату [20]. Розвиваючи цю ідею, слід підкреслити, що передбачуваність результату тут буде означати, швидше, не наявність максимально точного прогнозу за допомогою ШІ, а відсутність сумнівів щодо суті та обставин справи, що дозволяє людському судженню розумно припустити, яким буде результат, як того вимагають закон і справедливість.

Такий підхід передбачає, що виступати слід не за повну відмову від ШІ у судових рішеннях і не за повне прийняття обчислювальної справедливості. Алгоритми можуть бути прийнятними для відправлення правосуддя, коли у нас є чіткі критерії справ, що виграють від технологічного розвитку, не спотворюючи при цьому правові принципи та цінності. Це можна було би назвати аргументом розумної (в сенсі розсудливої) алгоритмізації. Перевагу має бути надано сумлінному та лімітованому впровадженню ШІ до процесів відправлення правосуддя, навіть якщо здається, що очікувана ефективність варта ризику. Слід також зазначити, що відповідальність за рішення, принаймні у випадку, коли для їх прийняття потрібно судження, має залишатися за людьми. Це означає, що ми повинні завжди мати реальну можливість знайти людину за лаштунками алгоритму, а не стикатися з безвихідністю і безособовістю автоматизованої системи, яка не має видимої людської присутності.

Висновки

Отже, виклики і ризики впровадження ШІ у процеси відправлення правосуддя мають такий ефект для справедливості та верховенства права, який

не слід ігнорувати. Першим важливим аспектом є проблема комп'ютеризації справедливості, що може бути непідйомним завданням само по собі. Наступна проблема полягає в тому, що ШІ часто використовує утилітарні підходи, стандартизуючи оцінку людських рис на основі числових даних, що може призвести до дискримінаційних результатів, а також суперечить вимогам щодо індивідуального характеру правосуддя. Іншим ризиком є те, що алгоритми, хоч і можуть ефективно обробляти великі обсяги юридичних текстів, все ж не здатні ані розуміти їх, ані враховувати контекст і дискурс. Алгоритми, до того ж, не здатні застосувати належне балансування між різними фундаментальними правами та законними інтересами. ШІ може вилучати факти з тексту та визначати певні закономірності, проте не може правильно інтерпретувати їх у світлі складних і рідкісних обставин, а також індивідуальних рис та особливостей. Ще одна проблема полягає в тому, що алгоритмізація правосуддя за допомогою ШІ може призвести до підміни правової визначеності алгоритмічною передбачуваністю. Це може підірвати довіру до судової системи та зрештою легітимність. ШІ також навряд чи здатний до належної юридичної аргументації. Алгоритм може надати добре структуровані та логічно послідовні відповіді, проте судові рішення мають бути не лише такими, що відповідають правовим принципам і нормам, а й бути належним чином обґрунтованими, зрозумілими та переконливими.

Отже, хоча ШІ має потенціал для покращення ефективності процесів відправлення правосуддя, його використання може бути виключно допоміжним інструментом, а алгоритмізованими мають бути обмежені категорії справ. Підхід розумної (розсудливої) алгоритмізації передбачає, що основна відповідальність за ухвалення судових рішень має залишатися за судьями-людьми, а часткове покладання цієї відповідальності на алгоритми повинне відповідати чітким критеріям і супроводжуватися належною обачністю.

Перспективи подальших досліджень можуть включати осмислення того, як технологічний детермінізм і чистий утилітаризм переважають більш обережні та помірковані підходи та відсувають деонтологічні міркування на другий план. Для цього може знадобитися звернення до концепції моральних прав, а також природно-правового праворозуміння та аналітичної філософії.

References

- [1] Branting, L.K., et al. (2021). Scalable and explainable legal prediction. *Artificial Intelligence and Law*, 29, 213-238. <https://doi.org/10.1007/s10506-020-09273-1>.
- [2] Varona, D., & Subrez, J.L. (2022). Discrimination, Bias, Fairness, and Trustworthy AI. *Applied Sciences*, 12(12), 5826. <https://doi.org/10.3390/app12125826>.
- [3] De Oliveira, L.F., et al. (2022). Path and future of artificial intelligence in the field of justice: a systematic literature review and a research agenda. *SN Social Sciences*, 2, 180. <https://doi.org/10.1007/s43545-022-00482-w>.

- [4] Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, 105567. <https://doi.org/10.1016/j.clsr.2021.105567>.
- [5] Cossette-Lefebvre, H., & Maclure, J. (2023). AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, 3, 1255-1269. <https://doi.org/10.1007/s43681-022-00233-w>.
- [6] Strümke, I., Slavkovik, M., & Madai, V.I. (2023). The social dilemma in artificial intelligence development and why we have to solve it. *AI and Ethics*, 2, 655-665.
- [7] Bathaee, Y. (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, 31(2), 889-938.
- [8] Schaeffer, R., Khona, M., & Fiete, I.R. (2022). No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit. *Advances in Neural Information Processing Systems*, 35, 16052-16067. <https://doi.org/10.1101/2022.08.07.503109>.
- [9] Bowers, J.S., et al. (2023). On the importance of severely testing deep learning models of cognition. *Cognitive Systems Research*, 82, 101158. <https://doi.org/10.1016/j.cogsys.2023.101158>.
- [10] Adams, J. (2023). Defending explicability as a principle for the ethics of artificial intelligence in medicine. *Medicine, Health Care and Philosophy*, 6(2). <https://doi.org/10.1007/s11019-023-10175-7>.
- [11] Vale, D., El-Sharif, A., & Ali, M. (2022). Explainable artificial intelligence (XAI) post-hoc explainability methods: risks and limitations in non-discrimination law. *AI Ethics*, 2, 815-826. <https://doi.org/10.1007/s43681-022-00142-y>.
- [12] Longo L., Goebel R., Lecue F., Kieseberg P., & Holzinger A. (2020) Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In Holzinger A., Kieseberg P., Tjoa A., & Weippl E. (Eds.) *Machine Learning and Knowledge Extraction. CD-MAKE 2020. Lecture Notes in Computer Science*, 12279. Springer, Cham. https://doi.org/10.1007/978-3-030-57321-8_1.
- [13] Shvydka V. Ukraine (Application No. 17888/12) (October 30, 2014). European Court of Human Rights.
- [14] Lenis V. Greece (Application No. 47833/20) (August 30, 2023). European Court of Human Rights.
- [15] Merlhiot, G., Mermillod, M., Le Penne, J.-L., Dutheil, F., & Mondillon, L. (2018). Influence of uncertainty on framed decision-making with moral dilemma. *PLoS ONE*, 13(5), e0197923. <https://doi.org/10.1371/journal.pone.0197923>.
- [16] SCHUFA Holding and Others (Scoring). Case C-634/21, ECLI:EU:C:2023:220.
- [17] McKay, C. (2019). Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making. *Current Issues in Criminal Justice*, 32(1), 22-39. <https://doi.org/10.1080/10345329.2019.1658694>.
- [18] Yassine, S., Esgbir, S., & Ibrihich, O. (2023). Using Artificial Intelligence Tools in the Judicial Domain and the Evaluation of their Impact on the Prediction of Judgments. *Procedia Computer Science*, 220, 1021-1026. <https://doi.org/10.1016/j.procs.2023.03.142>.
- [19] Zahir, J. (2023). Prediction of court decision from a rabic documents using deep learning. *Expert Systems*, 40(6), e13236. <https://doi.org/10.1111/exsy.13236>.
- [20] Reiling, A.D. (Dory). (2020). Courts and Artificial Intelligence. *International Journal for Court Administration*, 11(2), 8. <https://doi.org/10.36745/ijca.343>.

Юлія Сергіївна Разметаєва

кандидатка юридичних наук, доцентка,
доцентка кафедри прав людини та юридичної методології,
голова Центру права, етики та цифрових технологій
Національний юридичний університет імені Ярослава Мудрого
61024, вул. Григорія Сковороди, 77, Харків, Україна;
дослідниця Центру мультидисциплінарних
досліджень релігії та суспільства,
Упсальський Університет
75238, вул. Тхунбергсваген, 3В, Упсала, Швеція
e-mail: yu.s.razmetaeva@nlu.edu.ua
ORCID 0000-0003-0277-0554

Yulia S. Razmetaeva

Ph.D. in Law, Associate Professor,
Associate Professor of the Department of
Human Rights and Legal Methodology
Head of the Center for Law, Ethics and Digital Technologies
Yaroslav Mudryi National Law University
61024, 77 Hryhoriia Skovorody Str., Kharkiv, Ukraine;
Researcher at the Centre for Multidisciplinary
Research on Religion and Society
Uppsala University
75238, 3B Thunbergsvagen, Uppsala, Sweden
e-mail: yu.s.razmetaeva@nlu.edu.ua
ORCID 0000-0003-0277-0554

Рекомендоване цитування: Разметаєва Ю. С. Штучний інтелект для прийняття судових рішень: деякі потенційні ризики. *Проблеми законності*. 2024. Спецвипуск. С. 177–191. <https://doi.org/10.21564/2414-990X.166.311749>.

Suggested Citation: Razmetaeva, Yu.S. (2024). Artificial Intelligence for Judicial Decision-making: Some Potential Risks. *Problems of Legality. Special Issue*, 177-191. <https://doi.org/10.21564/2414-990X.166.311749>.

Статтю подано / Submitted: 19.09.2024
Доопрацьовано / Revised: 23.10.2024
Схвалено до друку / Accepted: 31.10.2024
Опубліковано / Published: 18.11.2024