Дотримання прозорості в обробці персональних даних за допомогою штучного інтелекту

Базалицький Віталій Ігорович[1]

Опубліковано	Секція	УДК
10.06.2024	Право	341.1/.8

DOI: : https://doi.org/10.5281/zenodo.11544316

Анотація. Стаття присвячена дослідженню проблеми дотримання прозорості в обробці персональних даних за допомогою штучного інтелекту (ШІ) в контексті Загального регламенту із захисту персональних даних (GDPR). Основна увага приділяється праву на пояснення автоматизованих рішень, яке є критичним для забезпечення справедливості та захисту прав суб'єктів даних. У статті аналізуються юридичні аспекти права на пояснення, технічні бар'єри забезпечення прозорості алгоритмів ШІ та можливі шляхи їх подолання. Дослідження базується на аналізі останніх наукових публікацій і нормативних актів, таких як GDPR, і включає пропозиції щодо вдосконалення правових рамок та технічних рішень для забезпечення прозорості ШІ. Підкреслена важливість інтеграції принципу « privacy by design/privacy by default» у розробку алгоритмів ШІ та впровадження технік пояснюваного ШІ (XAI). Запропонований комплексний підхід до вирішення проблеми прозорості, що включає технічні та правові аспекти, з метою забезпечення етичного та справедливого використання ШІ в обробці персональних даних.

Ключові слова: прозорість, захист персональних даних, штучний інтелект, GDPR, право на пояснення, автоматизовані рішення, пояснюваний ШІ, XAI.

Maintaining transparency in the processing of personal data using artificial intelligence

Annotation. The article is devoted to a detailed analysis of the issue of transparency in the processing of personal data using artificial intelligence (AI) under the General Data Protection Regulation (GDPR). The main focus is on the right to explain automated decisions, which is key to ensuring fairness and protecting the rights of data subjects. The article analyses the legal aspects of the right to explanation, technical barriers to ensuring transparency of AI algorithms, and possible ways to overcome them. The study is based on the analysis of recent scientific publications and regulations, such as the GDPR, and includes suggestions for improving the legal framework and technical solutions to ensure AI transparency. The importance of integrating the principle of 'privacy by design/privacy by default' into the development of AI algorithms and the introduction of explainable AI (XAI) techniques is emphasized. The author proposes a comprehensive approach to solving the transparency

[1] аспірант кафедри міжнародного права НН Інституту міжнародних відносин Київського національного університету імені Тараса Шевченка, ORCID: https://orcid.org/0009-0002-4225-7077

problem, including technical and legal aspects, in order to ensure ethical and fair use of AI in personal data processing.

Keywords: transparency, personal data protection, artificial intelligence, GDPR, right to explanation, automated decisions, explainable AI, XAI.

Вступ

Тема дотримання прозорості в обробці персональних даних за допомогою штучного інтелекту є надзвичайно актуальною в сучасному інформаційному суспільстві. Штучний інтелект (ШІ) стає все більш інтегрованим у різні аспекти нашого життя, від медицини до фінансів, і його використання в обробці персональних даних має суттєві переваги та виклики. Одним із головних викликів є забезпечення прозорості цих процесів. Прозорість у контексті обробки даних означає, що суб'єкти даних повинні мати змогу зрозуміти, як і чому їхні дані використовуються. Це особливо важливо, враховуючи складність і непрозорість багатьох алгоритмів ШІ, які часто описують як «чорні ящики». Багато алгоритмів машинного навчання можуть бути настільки складними, що їхня внутрішня логіка може бути не зрозумілою навіть розробникам, що викликає питання щодо їхньої прозорості та справедливості. Забезпечення прозорості є не лише етичним, але й юридичним зобов'язанням. Відповідно до Загального регламенту із захисту персональних даних (GDPR), принцип прозорості вимагає, щоб суб'єкти даних мали право на отримання чіткої інформації щодо обробки їхніх даних, включаючи логіку, яка використовується при автоматизованому прийнятті рішень. Це положення є критично важливим для забезпечення довіри до систем ШІ та захисту прав суб'єктів даних. Вирішення проблеми прозорості в обробці персональних даних за допомогою ШІ потребує мультидисциплінарного підходу, який включає в себе як технічні, так і правові аспекти. З одного боку, технічні рішення повинні забезпечувати можливість інтерпретації результатів алгоритмів ШІ. З іншого боку, правові рамки, такі як GDPR, повинні забезпечувати, щоб ці технічні рішення були впроваджені належним чином і забезпечували захист прав суб'єктів даних.

Аналіз останніх досліджень і публікацій показує значну увагу до проблеми прозорості в обробці персональних даних за допомогою штучного інтелекту (ШІ). Основним нормативним документом, який регулює цю сферу, є Загальний регламент із захисту персональних даних, що визначає права суб'єктів даних на інформацію, доступ та втручання людини в автоматизовані процеси прийняття рішень. Дослідники, такі як М. Бркан [6], І. Мендоза, Л. Байгрейв [7], Б. Гудмен, С. Флаксмен [9], М. Камінські [10], С. Вотчер, Б. Міттельштадт, К. Рассел [11], Л. Едвардс, М. Віл [12], Д. Кролл, Д. Г'юї, С. Барокас [13] та інші, підкреслюють необхідність забезпечення пояснюваності алгоритмів і прозорості процесів прийняття рішень. Незважаючи на значний прогрес, залишається ряд невирішених питань, таких як технічні бар'єри для пояснюваності складних алгоритмів, юридичні аспекти реалізації права на пояснення та баланс між прозорістю і захистом комерційних таємниць. Подальші дослідження повинні зосередитися на розробці ефективних технік пояснюваного ШІ, удосконаленні правових рамок і врахуванні інтересів усіх зацікавлених сторін для забезпечення етичного та справедливого використання ШІ.

Метою даної статті є дослідити аспекти прозорості в обробці персональних даних за допомогою штучного інтелекту у контексті Загального регламенту із захисту персональних даних, зокрема права на пояснення автоматизованих рішень.

Завдання статті включають аналіз існуючих досліджень та нормативних актів, виявлення основних технічних і юридичних бар'єрів для забезпечення прозорості алгоритмів ШІ, а також розробку рекомендацій щодо подолання цих перешкод для забезпечення прав суб'єктів даних на справедливу і прозору обробку їхніх даних.

Результати

Одним із головних принципів, передбачених Загальним регламентом із захисту персональних даних (надалі – GDPR), є принцип прозорості [1]. Його роль полягає у забезпеченні суб'єктів даних справедливим і прозорим обробленням їх персональних даних. Прозорість, однак, є елементом, який значною мірою відсутній у системах штучного інтелекту. Загалом, алгоритми, що використовуються в системах штучного інтелекту, не повністю зрозумілі навіть тим людям, які їх розробляли. Це особливо стосується машинного навчання, коли програма аналізує великі обсяги даних і «самонавчається» під час цього процесу. Це означає, що логіка прийняття рішень виходить за межі контролю програміста. Характеристика цих систем як «чорних ящиків» вказує на непрозорість, яка їх описує.

Однак GDPR, прагнучи забезпечити прозорість в автоматизованому прийнятті рішень, надає суб'єктам даних право отримувати інформацію щодо логіки, якій слідують при прийняті рішень, та право знати, що цей процес означає і які наслідки цього оброблення (статті 13 п. 2 (f), 14 п. 2 (g) і 15 п. 1 (h) GDPR) [2-4]. Крім того, у випадку автоматизованого прийняття індивідуальних рішень, він надає суб'єктам даних право на втручання людини, право висловити свою думку та оскаржити рішення, прийняте шляхом автоматизованого оброблення персональних даних (стаття 22 GDPR) [5].

GDPR (як у преамбулі, так і в основному тексті) не згадує алгоритмічне оброблення персональних даних як таке. Європейський законодавець головним чином враховував випадки автоматизованого оброблення, які включають «профілювання», як у випадку автоматизованої оцінки кредитоспроможності позичальників або оцінки робочих показників працівників. Проте що відбувається у випадку, коли автоматизоване оброблення з чи без мети «профілювання» здійснюється за допомогою складніших технологічних засобів, таких як системи штучного інтелекту, які, як вже згадувалося, є найбільш непрозорими? З огляду на те, що порушення положень GDPR, пов'язаних із правами суб'єктів даних, підлягають вищому адміністративному штрафу, ніж порушення інших положень GDPR, відповідь на зазначені питання має вирішальне практичне значення. Особливо у випадку оброблення персональних даних за допомогою систем штучного інтелекту, відповідні питання повинні бути ефективно вирішені до або під час етапу розробки систем штучного інтелекту, щоб мати систему штучного інтелекту, що відповідає вимогам GDPR.

Надання інформації суб'єктам даних може набувати різних форм у контексті GDPR. Коли рішення ґрунтуються виключно на автоматизованому обробленні персональних даних і це оброблення має значні юридичні наслідки для суб'єктів даних, останні також мають право бути поінформованими. У статті 22 GDPR суб'єктам даних за певних обставин надається право на неавтоматизоване прийняття рішень. Відповідно до цього права, суб'єкти даних мають право на втручання людини, висловлення своєї думки та оскарження рішення. Однак реалізація цих прав з боку суб'єктів даних передбачає, що їм було надано відповідну інформацію [5].

Стаття 22 GDPR насамперед піднімає два важливі питання: коли оброблення рішень є «виключно» автоматизованим і які юридичні наслідки та подібно значущі наслідки роблять це оброблення забороненим або потребують заходів безпеки, коли застосовуються винятки статті 22 п. 1 (a-c) [5]. Контролер не може уникнути заборони статті 22, лише посилаючись на певну людську участь в обробленні рішень. Навпаки, ця людська участь повинна бути значущою (не просто символічною), здійснюваною особою, яка має як повноваження, так і компетенцію змінити рішення [5]. «Значущість» означає людське втручання, здатне змінити результат рішення. Тому простий перегляд рішення людиною, який має більш-менш формальний характер, не є значущим

людським втручанням. Відповідно, таке оброблення рішень дорівнює виключно автоматизованому обробленню даних.

Деякі автори, які підтримують існування права на пояснення, пропонують чотири типи інформації, які можуть бути надані як «значуща інформація». Ця інформація повинна містити: а) вхідні дані, що використовувалися для автоматизованого рішення, б) фактори, що впливають на рішення, в) важливість цих факторів г) текстову інформацію, що розумно пояснює підстави для прийняття певного рішення [6].

Інші автори зосереджуються більше на видах пояснень, що надаються з точки зору їх змісту та часу у відношенні до процесу прийняття рішень. Відповідно до цього підходу, існують пояснення, що стосуються функціональності системи (загальна функціональність автоматизованої системи прийняття рішень, наприклад, дерева рішень, передвизначені моделі тощо) та пояснення, що стосуються конкретних рішень (логіка конкретних рішень, такі як визначені машиною правила прийняття рішень у конкретних випадках тощо). З точки зору часу надання пояснень у відношенні до процесу прийняття рішень, можна виділити два типи пояснень: а) попередні пояснення, тобто пояснення, надані до процесу прийняття рішень, б) післяпроцесові пояснення, тобто пояснення, надані після процесу прийняття рішень [7].

Коли йдеться про вираз «юридичні наслідки» або «подібно значущі» наслідки, необхідне додаткове пояснення. GDPR не надає жодних роз'яснень щодо цих термінів. У преамбулі GDPR, в статті 71 згадуються лише показові ситуації, які можуть мати юридичні наслідки, такі як відмова в онлайн-заявці на кредит або практика електронного рекрутингу [8]. Однак юридичним наслідком також може бути щось, що впливає на юридичний статус особи або її права за контрактом, як, наприклад, коли автоматизоване оброблення рішень призводить до права або скасування контракту (банківський кредит, договір оренди) або до відмови в праві (громадянське право, соціальні пільги). Це також може призвести до значного обмеження або навіть відмови у фундаментальному праві, такому як право на доступ до правосуддя.

Вираз «подібно значущі» наслідки є відкритим для ширших інтерпретацій. Що саме потрапляє в цю категорію можна розглядати лише на основі кожного конкретного випадку, оскільки немає конкретного права, яке порушується в цих випадках. Загалом, можна визнати, що суб'єкт даних зазнає значущого впливу від автоматизованого прийняття рішень, коли останнє впливає на його/її вибір або поведінку, має тривалий або постійний вплив на суб'єкта даних або навіть призводить до виключення або дискримінації особи. Поріг «значущості» досягається, коли рішення може вплинути на фінансові обставини особи (відповідність вимогам щодо надання кредиту), доступ до медичних послуг, доступ до освіти (прийом до університету або коледжу) або доступ до працевлаштування.

Виняток із заборони статті 22 може застосовуватися у трьох випадках: а) коли автоматизоване рішення необхідне для підготовки або виконання договору між контролером даних і суб'єктом даних, б) коли держава-член дозволяє таке прийняття рішень, в) коли суб'єкт даних надає явну згоду [5]. У першому випадку контролер повинен надати достатнє обґрунтування для вибору цього типу оброблення даних, яке втручається в приватність, замість менш втручаючого, як, наприклад, коли відбувається автоматизоване оброблення заявок на роботу через великий обсяг заявників. З іншого боку, держава-член повинна мати можливість дозволити таке оброблення особливо для моніторингу та запобігання шахрайства і ухилення від сплати податків. Третій виняток, заснований на явній згоді суб'єкта даних, може на практиці виявитися найризикованішим для приватності суб'єктів даних і відповідно найскладнішим винятком для контролерів даних для застосування, особливо коли в обробленні використовуються додатки ШІ. Останнє, але не менш важливе, автоматизоване

оброблення спеціальних категорій персональних даних заборонено, якщо не виконуються дві умови одночасно: має місце виняток держави-члена і суб'єкт даних надає явну згоду на це оброблення або оброблення є необхідним з причин значного суспільного інтересу. Всі вищезазначені винятки, однак, йдуть рука об руку з відповідними заходами безпеки, які необхідно застосовувати, щоб контролери забезпечили справедливе та прозоре оброблення персональних даних.

Серед правознавців тривають дебати щодо існування та можливої правової основи права на пояснення при алгоритмічній обробці персональних даних. Існують думки, які заперечують існування права на пояснення в рамках GDPR, та погляди, що визнають юридичне існування такого права, хоча й на різних правових засадах.

С. Вотчер, Б. Міттельштадт і К. Рассел заперечують всі можливі правові підстави для існування права на пояснення, зокрема статті 22 [5], 13(2)f [2], 14(2)g [3] та 15(2)h [4] GDPR. За їхнім підходом, стаття 22 GDPR не згадує явно про право на пояснення, як це робить стаття 71, отже, це не має обов'язкової їх сили [9]. Статті 13 [2] та 14 [3] GDPR, на їхню думку, не можуть підтримати статтю 22 GDPR [5] і вимогу щодо обов'язку надання пояснення з боку контролера, оскільки вони вимагають лише попереднього пояснення функціональності системи, що передує прийняттю рішень. Нарешті, що стосується статті 15(2)h GDPR [4], автори доходять до того ж висновку, але з інших підстав. На їхню думку, стаття 15(2)h GDPR [4] не має проблеми з «часовою шкалою», однак термін «передбачуваний» знову ж таки відноситься до попереднього пояснення функціональності системи, а не до права на післяпроцесове пояснення.

Л. Едвардс і М. Віл не вважають статтю 22 GDPR [5] корисною для отримання прозорого пояснення роботи системи машинного навчання, вони вважають статтю 15 [4] більш відповідною, хоча вона не стосується безпосередньо обробки даних автоматизованим прийняттям рішень. Вони стверджують, що права доступу, зазначені в статті 15 [4], застосовуються після обробки і, таким чином, можуть надати суб'єктам даних знання про «логіку або раціональність, причини та індивідуальні обставини конкретного автоматизованого рішення». І. Мендоза та Л. Байгрейв не заперечують право на пояснення, яке може мати свою правову основу в статтях 22 і 15 GDPR [4-5]. На їхню думку, формулювання статті 15 не виключає можливості права на післяпроцесове пояснення автоматизованих рішень, тоді як термін «оспорення» в статті 22(3) означає більше, ніж «заперечення» чи «протидія», швидше це схоже на право на апеляцію [5].

М. Бркан пропонує цікаву концепцію. Вона передбачає комбіноване тлумачення положень GDPR, а саме поєднання статей 22 [5] (у світлі статті 71), 13(2)f, 14(2)g та 15(2)h [2-5], з метою надання суб'єктам даних права на післяпроцесове пояснення автоматизованого рішення. Б. Гудмен та С. Флаксмен не надають конкретної правової основи для права на пояснення в рамках GDPR, однак визнають необхідність такого права, особливо коли обробка стосується конфіденційних даних, разом із необхідністю подолання технічних бар'єрів, які можуть бути пов'язані з таким правом [9].

Підхід М. Камінскі до права на пояснення суттєво відрізняється. М. Камінскі не зосереджується на конкретному положенні GDPR як правовій основі права на пояснення, однак виступає за «систематичну прозорість», яка йде далі, ніж режим індивідуальної прозорості. На думку автора, цей режим «систематичної прозорості» повинен включати оцінки впливу на захист даних для автоматизованої обробки, доступ до інформації про алгоритми з боку регуляторів та прийняття компаніями внутрішніх режимів відповідальності та розкриття інформації [10].

Стаття 22 GDPR дійсно не згадує прямо право на пояснення. Однак це право передбачене в статті 22(3) GDPR [5]. Як суб'єкт даних може оскаржити рішення без будь-якої інформації про раціональність конкретного рішення? Що б могло стати «відповідними заходами» для захисту прав і свобод суб'єкта даних, якщо інформація про

те, як було досягнуто конкретного рішення, не надається? Отже, очевидно, що суворе граматичне тлумачення статті 22(3) GDPR необґрунтовано звужує захист суб'єкта даних. Навпаки, телеологічне тлумачення цього положення здається більш відповідним [5]. Цей підхід також підтримується самим формулюванням статті 22(3) [5], оскільки контролер даних повинен «запровадити відповідні заходи для захисту прав і свобод та законних інтересів суб'єкта даних, принаймні (виділено) право на отримання втручання людини з боку контролера, висловлення своєї точки зору та оскарження рішення». Показовий і мінімальний поріг захисту, наданого суб'єкту даних, підтримує існування і правову обґрунтованість права на пояснення в автоматизованому прийнятті рішень щодо обробки персональних даних.

Статті 13(2)f та 14(2)g дійсно накладають інформаційні обов'язки на контролерів до прийняття конкретного рішення [2-3]. В автоматизованій обробці рішень контролер зобов'язаний надати суб'єкту даних значущу інформацію про логіку, що використовується, та інформацію про значущість і передбачувані наслідки такої обробки для суб'єкта даних. Однак у випадку зазначених статей ця інформація повинна бути доступною суб'єкту даних у момент отримання персональних даних. Отже, статті 13(2)f та 14(2)g [2-3] не можуть підтримувати аргумент на користь обов'язку надання післяпроцесового пояснення для контролера, оскільки це встановлено в статті 22(3) [5]. Однак, як ми можемо прийняти той факт, що GDPR встановлює право на попереднє пояснення, але не на післяпроцесове пояснення, коли обробка вже відбулася і конкретне рішення має юридичні наслідки для суб'єкта даних? Було б нераціонально доходити до такого висновку при тлумаченні GDPR, особливо з урахуванням його статтей. Тому зазначені положення діють доповнюючи одне одного і підвищують захист прав і свобод суб'єктів даних, надаючи суб'єктам даних як попереднє, так і післяпроцесове право на пояснення відповідно.

Стаття 15(2)h [5], з іншого боку, має те саме формулювання, що й статті 13(2)f [2] та 14(2)g [3], отже, вона передбачена в межах встановленого права на доступ, яке може бути здійснене суб'єктом даних у будь-який час обробки. Тому пояснення, яке вимагається від контролера через природу права на доступ і той факт, що воно не визначено у формулюванні статті 15 [4], може бути як попереднім, так і післяпроцесовим. Стаття 15(2)h [4] може не бути правовою основою для післяпроцесового права на пояснення в автоматизованій обробці рішень, оскільки вона служить різним правовим цілям, однак підтримує існування такого права в значенні, що післяпроцесове пояснення завжди повинно бути надане суб'єкту даних.

Загалом, на думку автора, GDPR у статті 22(3) [5] встановлює sui generis право на післяпроцесове пояснення, яке надається суб'єкту даних при обробці рішень алгоритмічним шляхом. Правовою основою цього права на пояснення є стаття 22(3) GDPR [5], а не необов'язковий. Стаття 71 [8], отже, останній підтримує телеологічне тлумачення статті 22(3) [5].

Правове встановлення права на пояснення тісно пов'язане з проблемою пояснюваності алгоритмів. Як вже зазначалося, системи штучного інтелекту можуть бути надзвичайно непрозорими не тільки для сторонніх, але й для самих їх творців. Наскільки прозорою може бути «чорна скринька» і, відповідно, наскільки вона може відповідати вимогам GDPR щодо прозорості? Пояснюваність і, таким чином, прозорість алгоритмічного прийняття рішень виходить за рамки захисту даних та положень GDPR. Вона впливає на права і свободи суб'єктів даних. Алгоритмічне прийняття рішень може дискримінувати суб'єктів даних або навіть позбавляти їх певних фундаментальних прав, таких як право на доступ до правосуддя, право на доступ до медичної допомоги тощо. Ключове питання, таким чином, полягає в тому, чому пояснюваність у штучному інтелекті стала проблемою і як ми можемо її подолати.

Алгоритмічна прозорість може стикатися з трьома типами перешкод: технічними перешкодами, перешкодами, пов'язаними з інтелектуальною власністю, та перешкодами, пов'язаними з секретною та конфіденційною інформацією державних органів. Інші автори додають до цієї класифікації четверту категорію, а саме юридичні бар'єри, що виникають через аргумент, що право на пояснення не існує в рамках GDPR. Останнє не слід вважати перешкодою, оскільки це передбачає, що право на пояснення не існує юридично і що це широко прийнято, тоді як, навпаки, це питання є надзвичайно спірним. Що стосується перешкод конфіденційності, пов'язаних з персональними даними, свободами та правами третіх осіб, то існування таких перешкод на практиці також є спірним. Стаття 22(3) GDPR [5], тлумачена у світлі статті 71 [8] та статей 13, 14 і 15 GDPR [2-4], вимагає розкриття контролером інформації про «логіку, що використовується» у процесі автоматизованого прийняття рішень, а не розкриття конкретних даних. Таким чином, небезпека розкриття персональних даних осіб, які використовувалися як навчальні дані, здається малоймовірною. Однак розкриття комерційних таємниць є дуже ймовірним. Це може фактично стримувати контролерів від розкриття «надто багатьох» відомостей про логіку, що використовується в їхніх алгоритмах, оскільки реверс-інжиніринг завжди є можливим сценарієм.

С. Вотчер, Б. Міттельштадт і К. Рассел пропонують контрфактичні пояснення як підхід для надання уявлення про внутрішню логіку алгоритмів без відкриття «чорної скриньки» [11]. Пропозиція моделі контрфактичних пояснень виглядає цікавою, особливо те, що пояснення на основі контрфактів можуть бути як зрозумілими, так і корисними для суб'єктів даних. Однак контрфактичні пояснення можуть спричинити більше проблем, ніж ті, які вони намагаються вирішити.

Л. Едвардс і М. Віл пропонують два типи пояснень. Модельно-центровані пояснення (MCE) включають інформацію про налаштування, метадані навчання, показники продуктивності, оцінену глобальну логіку та інформацію про процес, тоді як суб'єктно-центровані пояснення (SCE) включають пояснення, орієнтовані на чутливі, випадкові, демографічні та продуктивні характеристики суб'єкта [12]. Факт полягає в тому, що будь-яка модель, обрана для забезпечення пояснюваності алгоритму, повинна бути ретельно спроектована, щоб надати настільки багато прозорості, скільки необхідно суб'єкту даних для реалізації прав, передбачених GDPR, незалежно від того, чи це право на інформацію, право на доступ або право на пояснення при автоматизованому прийнятті рішень. Вже зараз IT-фахівці працюють над різними техніками пояснюваного ШІ, такими як Local Interpretable Model-agnostic Explanations (LIME) та Shapley Additive exPlanations (SHAP), щоб покращити інтерпретованість ШІ, що є особливо корисним для аудиторів і дослідників.

У більшості випадків автоматизоване прийняття рішень включає технології ШІ в процес прийняття рішень. Пояснюваність систем ШІ, тобто здатність людей зрозуміти внутрішню логіку систем ШІ, безумовно, є ключовим елементом для сприяння довірі до ШІ. Особливо це важливо, коли обробка пов'язана з фундаментальними правами, наприклад, коли ШІ використовується в сфері правосуддя або медицини. У останньому випадку пояснюваність системи була б надзвичайно важливою і корисною не тільки для пацієнтів, але й для медичних працівників. Розуміння медичних прогнозів системи нейронних мереж дійсно сприяло б більш ефективному, а не просто точному прийняттю рішень.

Вищенаведений приклад відкриває іншу перспективу на проблему пояснюваності алгоритмів. Чи може система машинного навчання бути пояснюваною для всіх зацікавлених сторін, а саме для суб'єктів даних, науковців з даних, компаній і регуляторів? Чи може ця система забезпечити пояснюваність, бути точною і технічно здійсненною з точки зору розробки одночасно? Пояснюваність у ШІ може дійсно бути

складним завданням для розробників ШІ, але не неможливим. Вже зараз комп'ютерні науковці в галузі ШІ розробляють пояснювані моделі машинного навчання, які надають зрозумілі та точні набори пояснень.

Нарешті, вчені Д. Кролл, Д. Г'юї, С. Барокас надають іншу перспективу на проблему доказів у приватних системах, особливо коли ці системи використовуються в регульованій галузі, такій як автомобільна промисловість. У мережевих автомобільних системах (наприклад, автомобілі Tesla) майже будь-яка функціональність пов'язана з ПК або браузером для регулярного, іноді щоденного, оновлення системи. Надання поточної верифікації в таких випадках може стикатися з труднощами, проте це не є неможливим для досягнення [13].

Загалом, прозорість вихідного коду алгоритмічного прийняття рішень разом з відповідними вхідними та вихідними даними системи не обов'язково задовольняє критерій прозорості з багатьох причин, таких як випадковість, що бере участь у процесі, регулярні зміни в процесі прийняття рішень і несумісність системи з оцінкою та відповідальністю. Нарешті, але не менш важливо, повна прозорість процесу прийняття рішень може не бути оптимальною для всіх випадків, оскільки збереження секретності певних аспектів політики прийняття рішень може запобігти маніпулюванню процесом прийняття рішень.

Висновки

Визнання права на пояснення в рамках GDPR не є лише теоретично важливим питанням. Навпаки, право на пояснення може виявитися на практиці більш важливим, ніж будь-яке інше право, на яке має право суб'єкт даних згідно з GDPR. Це зумовлено тим, що право на пояснення пов'язане з юридичними наслідками, які суб'єкт даних може оскаржити та скасувати. Суб'єкту даних надається можливість оскаржити автоматизоване рішення, прийняте проти нього, а не просто отримати інформацію або доступ до певного типу інформації.

Забезпечення балансу між алгоритмічною прозорістю та правами третіх сторін, такими як захищені права інтелектуальної власності або комерційні таємниці, з урахуванням можливих технічних перешкод, є складним завданням. Це може бути однією з причин, чому Європейська комісія ухвалила остаточну редакцію статті 22(3) GDPR без явного посилання на право на пояснення, «описане» в статті 71 GDPR.

У будь-якому випадку, прозорість на практиці дійсно може стикатися з технічними перешкодами. Підхід «конфіденційність за дизайном» у вирішенні проблеми пояснюваності в системах ШІ видається найбільш доцільним, оскільки майбутнє впровадження технік пояснюваного ШІ у систему ШІ не виглядає життєздатним або принаймні ефективним. Комп'ютерні науковці проклали шлях до пояснюваності в системах ШІ, розробляючи техніки XAI. Пояснювані нейронні мережі можуть бути відповіддю на проблему пояснюваності ШІ, хоча вони можуть не бути доцільними для всіх видів систем ШІ. Наприклад, системи машинного навчання можуть виявитися більш «стійкими» до моделей пояснюваності або інтерпретованості. Проте, оскільки пояснюваність стає частиною технічних вимог до систем ШІ, технічні перешкоди для права на пояснення можуть не бути основним фокусом. Навпаки, більше уваги слід приділяти змісту цього права на пояснення з юридичної точки зору. Можливо, системи ШІ не є настільки «чорними скриньками», як хтось міг би вважати. Врешті-решт, непрозорість систем ШІ та будь-яких інших систем залежить від їхніх творців. Визнання права на пояснення для суб'єктів даних відповідно до GDPR стане позитивним кроком до більш прозорого режиму в системах ШІ для всіх зацікавлених сторін.

Список використаних джерел

1. General Data Protection Regulation (GDPR) – Legal Text. *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/ (date of access: 03.06.2024).
2. Art. 13 GDPR – Information to be provided where personal data are collected from the data subject - General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/art-13-gdpr/ (date of access: 02.06.2024).
3. Art. 14 GDPR – Information to be provided where personal data have not been obtained from the data subject - General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/art-14-gdpr/ (date of access: 01.06.2024).
4. Art. 15 GDPR – Right of access by the data subject - General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/art-15-gdpr/ (date of access: 02.06.2024).
5. Art. 22 GDPR – Automated individual decision-making, including profiling - General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/art-22-gdpr/ (date of access: 01.06.2024).
6. Brkan M. Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond. International Journal of Law and Information Technology. 2019. Vol.34. №1. http://dx.doi.org/10.1093/ijlit/eay017.
7. Mendoza I., Bygrave L. A. The Right Not to Be Subject to Automated Decisions Based on Profiling. Tatiani Synodinou, Philippe Jougleux, Christiana Markou, Thalia Prastitou (eds.), EU Internet Law: Regulation and Enforcement (Springer, 2017, Forthcoming), University of Oslo Faculty of Law Research Paper No. 2017. №20. PP.16-17. Available at SSRN: https://ssrn.com/abstract=2964855
8. Art. 71 GDPR – Reports - General Data Protection Regulation (GDPR). *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/art-71-gdpr/ (date of access: 03.06.2024).
9. Goodman B., Flaxman S. European Union Regulations on Algorithmic Decision-Making and a «Right to Explanation». AI Magazine. 2017. Vol 38(3). PP. 50-57.
10. Kaminski M. E. The Right to Explanation, Explained. Berkeley Technology Law Journal. 2019. Vol 34. №1. http://dx.doi.org/10.2139/ssrn.3196985.
11. Wachter S., Mittelstadt B., Russell C. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law Technology. 2018. Vol 31 (2). P.842. http://dx.doi.org/10.2139/ssrn.3063289
12. Edwards L., Veale M. Slave to the Algorithm? Why a «Right to an Explanation» Is Probably Not the Remedy You Are Looking For. Duke Law Technology Review. 2017. Vol 16. PP.51-60. http://dx.doi.org/10.2139/ssrn.2972855
13. Kroll J. A., Huey J., Barocas S. Accountable Algorithms. University of Pennsylvania Law Review. 2017. Vol. 165, 2765268. https://ssrn.com/abstract=2765268