УДК 141:171
DOI

Є. Ю. Ланюк
ORCID ID: https://orcid.org/0000-0003-0171-9802
кандидат політичних наук,
асистент кафедри філософії
Львівського національного університету імені Івана Франка

ПРОБЛЕМИ ЗАСТОСУВАННЯ АВТОМАТИЧНИХ АЛГОРИТМІВ ПРИЙНЯТТЯ РІШЕНЬ: ЕТИЧНИЙ ТА СОЦІАЛЬНИЙ АНАЛІЗ

Постановка проблеми. У XXI столітті людство поступово входить у четверту стадію свого цивілізаційного розвитку – інформаційну, яка наступає після суспільства мисливців-збирачів, аграрної та промислової епох. Особливістю цієї стадії є впровадження автоматичних систем опрацювання даних у дедалі більшій кількості галузей людської діяльності.

Одним із аспектів цього процесу є використання комп'ютерних алгоритмів для прийняття рішень у суспільно значущих сферах. Їх використовують, зокрема, банки, щоб вирішити про надання клієнту кредиту чи позики; страхові компанії, щоб визначити вартість страховки; працедавці, щоб оцінити ефективність потенційного працівника; маркетологи для ефективнішого розміщення реклами й навіть окремі суди, щоб оцінити ризик рецидивізму злочинця. Ми називаємо такі системи «автоматичними алгоритмами прийняття рішень» (ААПР). «Автоматичними» – оскільки вони функціонують без втручання або нагляду людини. «Алгоритмами» – оскільки вони є математичними моделями, які перетворюють вхідні дані (інформацію про індивідів) у вихідні (рішення) за скінченну кількість кроків, ґрунтуючись на своїх запрограмованих інструкціях. «Прийняття рішень» – адже їхньою метою є ухвалення рішень, які здебільшого мають бінарний характер: «прийняти/відхилити», «дозволити/заборонити», «покарати/виправдати» та ін.

Основна перевага ААПР у тому, що вони зберігають час, економлять кошти і збільшують продуктивність. Завдання, на які людям потрібні години і дні, комп'ютерна програма може виконати за секунди. Окрім того, такі системи дозволяють усунути упередження, помилки і корупцію, які дуже часто супроводжують людські рішення. Однак їх використання пов'язане із цілою низкою етичних проблем: Наскільки вони чесні? Які рішення можна делегувати комп'ютерним програмам, а які – ні? Чи можуть вони бути дискримінаційними? Зрештою, яка ціна економічної ефективності, якій ми завдячуємо ААПР?

Нині Україна активно проходить процес, який у публічному дискурсі називають діджиталізацією. Хоча за темпами інформаційних трансформацій наша держава ще поки відстає від країн Заходу, етичні та філософські дискусії, які у найрозвинутіших країнах світу вже активно відбуваються довкола проблем автоматичних алгоритмів, штучного інтелекту, робототехніки, біотехнологій та цифрової епохи, є не менш актуальними і для українського суспільства. Автоматичні алгоритми можуть стати привабливою альтернативою людській діяльності у тих сферах, де існує високий рівень корупції та зберігається успадкована від СРСР неефективна бюрократична система прийняття рішень. З огляду на це вважаємо, що поінформованість української громадськості та академічної спільноти про виклики, пов'язані із застосуванням ААПР, має не лише теоретичну, але й практичну цінність.

Метою статті є розгляд низки етичних, філософських і соціальних проблем, які випливають із дедалі ширшого використання автоматичних алгоритмів прийняття рішень в економіці, публічному житті, роботі правоохоронних органів та інших суспільно значущих сферах.

Аналіз досліджень і публікацій. Етичні проблеми впливу автоматичних алгоритмів на суспільство вже порушувала низка дослідників. Зокрема, американська науковиця Кеті О'Ніл у книзі «Зброя математичного знищення: Як великі дані збільшують нерівність та загрожують демократії?» [14] окреслює руйнівний вплив машинних рішень на суспільство. Френк Паскуале у праці «Суспільство «чорної скриньки»: секретні алгоритми, які контролюють гроші та інформацію» [16] аналізує, як великі корпорації використовують автоматичні алгоритми, щоб контролювати поведінку людей. Шошана Зубофф у книзі «Епоха капіталізму стеження: боротьба за людське майбутнє на новому рубежі влади» [18] обґрунтовує концепцію «капіталізму стеження», яку визначає як практику збирання, опрацювання і товаризації даних користувачів Інтернету з метою їх «підштовхування» до дій, які вигідні технологічним компаніям та їхнім клієнтам. Ювал Ной Гарарі у працях «Homo Deus: За лаштунками майбутнього» [5] та «21 урок для 21 століття» [4] аналізує соціальні, етичні та екзистенціальні виклики, пов'язані із появою «суперрозумних»

алгоритмів, які потенційно можуть перевершувати можливості людського розуму.

Серед українських дослідників етичні аспекти використання автоматичних алгоритмів порушує Юлія Карпенко, яка систематизувала етичні принципи, що повинні спрямовувати використання штучного інтелекту [2]. Катерина Слободяник та Віра Додонова вказують, що вирішення етичних проблем застосування штучного інтелекту та дотримання етичних принципів у його розробці є важливими завданнями, що стоять перед його філософією [3]. Ґрунтовний аналіз застосування технічних рішень, зокрема алгоритмічних систем, у процесі управління здійснила Юлія Бех [1].

На нашу думку, етичні та соціальні проблеми, які постають із використання алгоритмів у суспільній сфері, ще потребують подальшого висвітлення в українських соціальних та гуманітарних науках. Актуальним, зокрема, є розгляд окремих випадків («кейсів») застосування автоматичних алгоритмів прийняття рішень та їхньої етичної, філософської та соціологічної інтерпретації.

Виклад основного матеріалу. Термін «алгоритм» походить із арабської мови та означає «процедуру вирішення математичної проблеми за скінченну кількість кроків» [6]. Процес прийняття рішень у таких сферах, як бізнес, фінанси, страхування, працевлаштування та ін., дедалі більше стає автоматизованим, особливо на низовому рівні. Що більше люди взаємодіють із автономними алгоритмами прийняття рішень, тим актуальнішим стає запитання, який тип суспільства формується довкола них і які етичні виклики пов'язані з ним.

Льюїс Мамфорд, американський історик, соціолог та філософ техніки, стверджував, що структура і функції суспільства зумовлені особливостями його ключових технологій [12]. Техніка, на його думку, – це не лише інструмент, але й активний суб'єкт реальності, що трансформує людину і суспільство за своєю подобою. Він описував індустріальні наддержави XX століття з притаманними їм ієрархією, конформізмом та домінуванням бюрократії як «мегамашини», тобто машини, компонентами яких є люди. Узагальнюючи цей підхід, можна стверджувати, що суспільство відтворює логіку функціонування своїх основних машин.

У цьому контексті сучасне суспільство теж можна розглядати як абстракцію його основної технології – комп'ютера. Комп'ютер – це пристрій для перетворення даних. Його процесор отримує вхідні дані, записані двійковим кодом, від блоку пам'яті, перетворює їх за допомогою мікроархітектури – складної мережі арифметичних та логічних пристроїв, що виконують булеві операції (кон'юнкція, диз'юнкція, заперечення), – на вихідні дані і записує їх назад у пам'ять. Після цього цикл повторюється.

З цього погляду ААПР можна трактувати як суспільні аналоги комп'ютерних арифметичних та логічних пристроїв. Вони також отримують вхідні дані (оцифровану інформацію про індивідів), які перетворюють у вихідні (рішення) за допомогою своїх запрограмованих інструкцій. Як і арифметичні та логічні пристрої комп'ютера, які маркують дані як «істинні» чи «хибні», ААПР роблять те саме щодо індивідів, яких сортують на різні типи бінарних категорій: «дозволити/заборонити», «прийняти/відкинути» та ін.

У праці «Post Scriptum до суспільств контролю» французький філософ-постмодерніст Жиль Дельоз концептуалізував суспільство, організоване довкола перемикачів та потоків даних, як «суспільство контролю» [8]. Він стверджував, що індивіди у такому суспільстві стають «дивідами», або подвійними істотами, що складаються із фізичного тіла та його цифрової репрезентації у системі комп'ютеризованого контролю, яку він назвав «кодом». За допомогою «коду» система автоматично надає або блокує «тілу» доступ до певних локацій, можливостей чи винагород (Дельоз називав таке середовище «мінливою геометрією»). Перемикачі у «суспільстві контролю» управляють людьми приблизно так само, як арифметичні та логічні пристрої керують електронами всередині масиву комп'ютерних мікрочипів.

«Суспільство контролю», на думку Жиля Дельоза, утверджується як нова історична форма влади, що приходить на зміну «дисциплінарній» владі, яку обґрунтував Мішель Фуко. «Дисциплінарна» влада, згідно з Фуко, втілювалась у паноптикумі – інституційній будівлі, в якій наглядач міг спостерігати за людьми, які у ній перебувають (залежно від цілей, це можуть бути в'язні, солдати, пацієнти, учні, робітники та ін.) і які, зі свого боку, не бачать наглядача. Мішель Фуко назвав таку владу «дисциплінарною», адже її мета – не карати людей, як у середньовічному «суверенному» типі влади, а спонукати їх ставати слухняними і корисними істотами. В індустріальну епоху виникає розгалужена мережа паноптичних інституцій, яку Мішель Фуко назвав «мікроархітектурою» влади. Індивіди, які живуть у цій «мікроархітектурі», переміщаються від однієї владної інституції до іншої (наприклад, від школи до фабрики) і перебувають в умовах постійної муштри і дисципліни.

На думку Жиля Дельоза, починаючи з кінця XX ст. настає загальний занепад усіх паноптичних інституцій – «тюрем, лікарень, фабрик, шкіл, сім'ї» [8, с. 3]. Щоб здійснювати владу у XXI ст., набагато менше потрібні закриті приміщення і вежі з наглядачами. Натомість у «суспільстві контролю» влада здійснюється автоматично за

допомогою перемикачів. Дані про особу «ув'язнюються» у базі даних і звідти керують нею.

Наведемо простий приклад – організацію дорожнього руху. В Україні, як і в багатьох інших країнах, застосовується автоматична фіксація порушень ПДР. Якщо автомобіль перевищує швидкість, камера фіксує його реєстраційний номер і водій невдовзі отримує лист з вимогою сплатити штраф. Алгоритм у цьому прикладі здійснює бінарний розподіл (на тих, хто порушує і не порушує ПДР). «Кодом» є інформація про автомобіль у базі даних МВС, що дозволяє ідентифікувати і покарати водія-порушника. Змінюючи інструкції алгоритмів (наприклад, камера може фіксувати порушення лише у певні години або реєструвати тільки деякий тип транспортних засобів), поліція здатна автоматично модулювати дорожній рух.

Влада КНР пішла ще далі і тестує експеримент під назвою «система соціального кредиту». Згідно з цим експериментом, усіх індивідів, а не лише водіїв транспортних засобів, вносять у базу даних і присвоюють їм рейтинг «кредиту довіри». Середовище їхнього проживання стає «мінливою геометрією» із безліччю камер, датчиків та сенсорів біометричної інформації. Система функціонує так, щоб винагороджувати людей з високим показником «довіри» і карати тих, у кого він низький. Наприклад, перехід вулиці на червоний сигнал світлофора, обладнаного камерою з технологією розпізнавання обличчя, автоматично віднімає певну кількість балів від рейтингу. Серйозні проступки, такі як кримінальні правопорушення або наявність боргів, знижують кредит ще більше. Люди з низьким рейтингом не зможуть потрапити у готель чи ресторан (автоматичні двері, під'єднані до камери із сенсором обличчя, просто не відчиняться), купити білет на поїзд або літак, отримати хорошу роботу, вступити в університет та ін. [15]

Принципом створення «системи соціального кредиту» є перетворення будь-якої дії людини на сигнал, який впливає на «індекс довіри» і карає або винагороджує її шляхом блокування або надання доступу до певних можливостей, локацій, середовищ та ін. Індивіди у цій системі не впливають ані на роботу вимикачів, ні на систему накопичення балів, які контролює держава, виходячи зі свого розуміння мети розвитку суспільства.

Система комп'ютеризованого контролю вибудовується і на Заході, проте там її рушієм є не держава, а приватні корпорації. Щоб зекономити кошти і збільшити продуктивність, компанії автоматизують свою діяльність і делегують машинам рішення, які раніше приймали люди. Шошана Зубофф стверджує, що у XXI ст. формується безпрецедентний економічний феномен, який вона називає «капіталізмом стеження» [18]. Його суть – у збиранні якомога більше інформації про особу (до неї входить, зокрема, історія вебпереглядів, контакти у соціальних мережах, геолокація та ін.) з метою створення моделі її поведінки, що дозволяє прогнозувати її майбутні дії. Цифрові пристрої, з якими взаємодіє користувач, відіграють роль не лише сенсорів, але й активаторів поведінки, які «підштовхують» його до вчинків, вигідних технологічним компаніям та їхнім клієнтам. Термін «підштовхування» (англ. Nudging) запровадив американський економіст, лауреат Нобелівської премії з економіки 2019 р. Річард Талер. Найпростішим прикладом «підштовхування» є контекстна реклама, яка з'являється на екранах ноутбуків і смартфонів, як тільки система оцінює ймовірність купівлі певного товару як високу. Однак, згідно із Зубофф, сучасні технології досягнули такого рівня розвитку, що «підштовхування» не вичерпується самою лишень контекстною рекламою, а відбувається у складній та розгалуженій «архітектурі вибору», в якій знання людської психології, поєднане з технологічними ресурсами та потужними обчислювальними можливостями, дозволяє здійснювати тонкий та непомітний маніпулятивний вплив на свідомість. Як приклад, вона наводить гру в жанрі доповненої реальності Pokemon Go, розроблену дочірнім стартапом Google Niantic Labs у 2016 р. Хоча більшість людей сприймала її лише як гру, вона була соціальним експериментом, який за допомогою покемонів мав привести гравців у торговельні точки, котрих алгоритми Google визначали на основі їхніх купівельних уподобань, виведених з їхніх поведінкових даних (власники бізнесів платили компанії за цю послугу) [18].

«Суспільство контролю», в якому алгоритми керують приватними даними, має різні особливості у Китаї та на Заході. У Китаї його завданням насамперед є покарання правопорушників та «некорисних» для суспільства громадян; на Заході – «підштовхування» індивідів до дій, які приносять компаніям прибуток. У Китаї систему комп'ютеризованого контролю запроваджує держава; на Заході це роблять приватні компанії. Однак в обох випадках ця система суперечить ліберально-демократичним цінностям. Ліберальна демократія ґрунтується на ідеї автономного та раціонального суб'єкта, який приймає важливі для себе рішення і самостійно несе за них відповідальність. Натомість у «суспільстві контролю» такого суб'єкта підміняє інформація у базі даних, в якій «розумні» алгоритми знаходять закономірності і приймають рішення щодо нього заздалегідь, часто без його згоди і навіть відома. Розглянемо кілька етичних викликів, пов'язаних із щораз більшим застосуванням ААПР.

Сара Висоцькі була шкільною вчителькою з Вашингтона, округ Колумбія. У 2009 р. влада

міста вирішила реформувати систему шкільної освіти і звільнити неефективних учителів. Щоб зробити процес «чесним», його вирішили автоматизувати. Рішення про те, яких учителів слід вважати «неефективними», доручили комп'ютерній програмі під назвою IMPACT, що мала оцінити їхню роботу за результатами проходження їхніми учнями математичних і мовних тестів протягом двох років. Комплексний алгоритм повинен був встановити внесок учителя і відокремити його від інших чинників, які теж впливають на успішність дітей. Сара Висоцькі вважала, що їй не варто перейматися, адже її учні та їхні батьки високо оцінювали її як вчителя. Проте коли IMPACT поставив їй сумнівно низьку оцінку, її звільнили з роботи. Висоцькі розуміла, що з нею вчинили нечесно, але це було непросто довести. Ані шкільні адміністратори, ні міські чиновники не могли пояснити, чому алгоритм поставив їй саме такий бал, проте вони всі апелювали до неупередженого рішення машини. Після медійного розголосу вона знайшла роботу в іншій школі, але її випадок спонукає задуматися, скільки інших людей постраждали за схожих обставин [14, с. 10].

Цей випадок підсумовує одразу декілька рис ААПР. По-перше, їх односторонньо нав'язують людям, часто без їхньої згоди, і використовують для вимірювання їхньої цінності/надійності як працівників, споживачів, позичальників, пасажирів та ін. По-друге, їхні внутрішні механізми часто незрозумілі. Небагато людей можуть пояснити, як вони приходять до своїх висновків, а у разі самотренованої нейромережі, це навіть її розробники. По-третє, ці програми роблять ймовірнісні судження, що хтось *може* бути неефективним працівником або ненадійним позичальником, проте до цієї людини ставляться так, наче вона ним і є. На думку Кеті О'Ніл, це сприяє нерівності суспільства, в якому «бідні приречені залишатися такими завжди» [14, с. 203].

Щоб приймати рішення, алгоритмам потрібні дані, причому що більше у них даних, тим точнішими стають їхні оцінки і висновки. Це виправдовує втручання у приватне життя людей з метою добути якомога більше інформації для алгоритмічних розрахунків. Мікаель Караніколас ілюструє це втручання так: «Уявіть собі типовий день, в який ви прогулюєтесь центром свого рідного міста. Можливо, ви зайдете до книгарні і розглядатимете книжки на полицях або старі платівки. Ви підете в банк або кіно. Ймовірно, що ви зустрінете своїх старих знайомих і перекинетесь з ними кількома словами. А тепер уявіть, що цілий день за вами ходила група людей і ретельно документувала кожен ваш крок. Вони внесли в каталог кожен магазин, який ви відвідали, кожен товар, на який ви подивилися, і кожну людину, з якою ви розмовляли. Ця інформація їм потрібна, щоб створити детальний ваш профіль, що включає демографічний і соціальний статус, ваші хобі та інтереси, політичні погляди та ін. А тепер уявіть, що вони збирали цю інформацію місяцями і навіть роками й готові продати її кожному, кого вона може зацікавити» [10, с. 7].

У цифрову епоху життя людей нагадує відкриті книги. Камери і GPS фіксують їхні переміщення. Історія пошуків і онлайн-покупок розкриває приватні деталі про їхнє здоров'я, погляди і спосіб життя. Ці дані компонуються у профілі і стають сировиною для алгоритмічних розрахунків. За словами Френка Паскуале, «нині є сотні кредитних та споживчих рейтингів і ще більше джерел даних для кожного із них». Він додає, що кожен з цих рейтингів може «змінити наше життя на підставі помилки чи неточності, але ми про це так і не дізнаємось» [16, с. 33].

У 2011 р. американська преса обговорювала випадок жінки з Арканзасу на ім'я Кетрін Тейлор, яка протягом кількох років не могла знайти роботу. Виявилось, що причиною її невдач був запис у базі даних американської поліції, яку перевіряли її потенційні роботодавці. З'ясувалось, що інша Кетрін Тейлор, яка народилась того ж дня, відбула покарання за зберігання наркотиків. Тейлор зуміла роз'яснити, що це – не вона, проте цю інформацію вже скопіювали у численні споживчі бази даних, і її, можливо, й досі сприймають як колишню злочинницю [14, с. 152]. Цей приклад показує, як інформація у базі даних може керувати людським життям і ілюструє аргумент: «Ми не створюємо наших «Я» у базах даних – наші «Я» у базах даних створюють нас» [7].

Тоді як випадок Кетрін Тейлор показує, як єдина помилка у базі даних може перевернути життя з ніг на голову, історія хлопця на ім'я Кайл Бем доводить, як людина може постраждати від незнання того, як система працює. Бем був студентом американського коледжу, який страждав на біполярний розлад. У період реабілітації він намагався знайти тимчасову роботу, проте ніхто не поспішав його найняти. Виявилось, що причиною його невдач був однаковий психологічний тест, який використовували майже всі роботодавці в його містечку. Цей тест вимірює ефективність апліканта на підставі параметрів «великої п'ятірки» (відкритість, сумлінність, екстраверсія, доброзичливість, нейротизм). Оскільки американське законодавство забороняє використання медичних записів або показників IQ під час прийому на роботу, роботодавці замінили їх, здавалось би, невинним тестом особистості [14, с. 105]. Ознакою таких тестів є те, що люди, які їх проходять, не знають, які відповіді можуть їх дискваліфікувати. Наприклад, «Макдональдс» пропонує своїм потенційним працівникам пройти тест, в якому вони повинні вибрати, яка

відповідь їх краще характеризує: 1. «Складно бути життєрадісним, коли навколо багато проблем»; 2. «Іноді мені потрібен стимул, щоб почати працювати» [14, с. 109]. Можна лишень здогадуватися, який із запропонованих варіантів гірший. Власне, «правильної» відповіді немає взагалі. Комп'ютер порівнює відповіді на тестах з ефективністю працівників на роботі і виявляє кореляції, які автоматично дискваліфікують апліканта, якщо він ненароком потрапить в одну із них.

Прихильники таких тестів стверджують, що якби їхні правила були відкриті, люди б постійно намагались їх обманути, вибираючи лише потрібні варіанти. З іншого боку, це створює кафкіанську ситуацію, коли людину оцінюють за правилами, які від неї приховані. Правосуддя переслідувало головного персонажа роману Франца Кафки «Процес» (1925 р.), але йому так і не повідомили, ані в чому його провина, ні який закон він порушив. «Погана кредитна історія може коштувати позичальнику сотень тисяч доларів, проте він не дізнається, як її розрахували, – пише Френк Паскуале. – Прогностична аналітика може охарактеризувати працівника як «затратного» або «неефективного», але йому не скажуть, у чому його проблема» [16, с. 5]. В усі часи людей судили за правилами, які були справедливими або несправедливими, поблажливими або суворими, чесними або дискримінаційними, але вони мали одну спільну рису: люди завжди знали, які норми до них застосовують і могли відповідно скоригувати свої дії. Натомість у суспільстві, яке формується довкола даних і алгоритмів, вони щораз більше живуть за «завісою незнання» (термін американського політичного філософа Джона Роулза) й не знають, які вчинки, висловлювання або відповіді на тестах можуть призвести до їхньої дискваліфікації, звільнення, відмови у позиці чи суворішого покарання. Життям людини у такому суспільстві детерміністськи керують процеси, які перебувають за межами її контролю і розуміння.

Наступна етична проблема ААПР стосується їхньої нечутливості до соціальних контекстів. Розробники алгоритмів не можуть передбачити всіх можливих ситуацій, в яких вони прийматимуть рішення, а також наслідків, що із них випливатимуть. Наприклад, дослідники із Массачусетського технологічного інституту встановили, що, аналізуючи контакти особи у соціальних мережах, можна встановити її сексуальну орієнтацію [9]. Це дозволяє налаштувати таргетинг, пов'язаний з тематикою ЛГБТ. Один користувач, який прямо не розкривав своєї нетрадиційної орієнтації, виявив на своєму комп'ютері оздоблену веселкою рекламу «Тренінгу із камінг-ауту» [16, с. 26]. Добре відомо, до яких наслідків може призвести розкриття гомосексуальності у таких країнах, як Іран або Нігерія. У 2012 р. американська преса обговорювала мережу супермаркетів Target, яка створила алгоритм, що за купівельними вподобаннями жінок міг виявляти їхню вагітність. З допомогою цього алгоритму компанія сподівалась надсилати рекламу дитячих товарів ще до народження дитини. Виявилось, що вона розкрила вагітність однієї дівчини її родині без її згоди [11]. У суспільстві, в якому важливі рішення приймають «сліпі» до соціальних контекстів алгоритми, людям ставатиме важче контролювати своє життя.

Західні суспільства докладають значних зусиль, щоб подолати соціальні стигми, які оточують меншини та вразливі соціальні групи. В добу алгоритмізованого контролю, однак, виникають нові стигми. Якщо алгоритм ідентифікує яку-небудь людину як «ненадійну» або «неефективну», то вона може зіштовхнутись з каскадом неприємних подій протягом тривалого часу. Навіть якщо така оцінка іноді правдива, вона однаково блокуватиме її спроби змінити своє життя і почати його заново. Якщо раніше працівника звільнили за недбальство, він міг спробувати влаштуватись на іншу роботу. Сьогодні його ім'я, найімовірніше, потрапить в Інтернет або базу даних – і його життя назавжди буде зіпсоване.

Наведені вище приклади можна інтерпретувати у контексті редукціонізму – філософського поняття, яке позначає зведення складної системи до простих частин та наділення її однаковою цінністю з ними. Таким, наприклад, було бачення нацистського та комуністичного тоталітарних режимів, які трактували людину і суспільство як форму фізичної матерії, з якою можна робити що завгодно заради досягнення політичних, економічних чи військових цілей.

У XXI ст. постає новий вид редукціонізму. Сьогодні людей редукують не до фізичної матерії, а до іншої сировини – даних. Комп'ютеру, який виконує операції над даними, байдуже, про що вони. Компанії намагаються «оцифрувати» людські життя і трактують їх як проблему, яку потрібно вирішити. Але те, що формально є розв'язком математичної задачі, не бере до уваги того, що до бітів всередині комп'ютерного процесора прив'язані людські долі і навіть життя.

Кеті О'Ніл наводить ще один приклад негативного впливу алгоритмів на суспільство. Вона стверджує, що вони фіксують суспільство у його актуальному стані і протидіють спробам його змінити. Причиною цього є ефект позитивного зворотного зв'язку, який полягає у тому, що результати алгоритмічних розрахунків породжують саме ті наслідки, які й передбачають алгоритми. Дослідниця наводить кілька прикладів того, як це відбувається.

У м. Рідінг, штат Пенсильванія, місцева поліція використала алгоритм під назвою PredPol, який опрацьовує інформацію про минулі

правопорушення і передбачає, в яких районах міста вони найчастіше трапляються. Після цього поліція вирушає патрулювати саме ці райони. На думку О'Ніл, основний недолік цього алгоритму полягає у тому, що він є пророцтвом, яке само себе виконує: більше поліцейських патрулів означає більше зафіксованих правопорушень, особливо дрібних, а це, своєю чергою, виправдовує потребу ще більшого патрулювання [14, с. 86]. Як наслідок, окремі райони потрапляють під надмірно прискіпливу увагу поліції. У Сполучених Штатах, згідно з О'Ніл, існує тенденція до перетину цих районів із сегрегованими ґетто расових меншин. Це не лише поглиблює расові поділи, але й збільшує шанси поліцейської брутальності, прикладом якої стало вбивство Джорджа Флойда 25 травня 2020 р.

Іншим прикладом ефекту зворотного зв'язку, що поглиблює наявні нерівності, на думку О'Ніл, є хижацькі рекламні тактики, які спрямовані передусім на людей, які переживають життєві труднощі, такі як хвороба, борги або залежність. На таких людей часто полюють шахраї, які обіцяють їм «чудесний» порятунок від їхніх труднощів, наприклад «диво-ліки» або грошові позики, але натомість ще більше заганяють їх у злидні. Кеті О'Ніл наводить комерційний коледж під назвою «Університет Ваттертота», алгоритми якого працюють так, щоб його рекламу бачили передусім такі соціальні категорії: «Люди, які живуть на соціальну допомогу, одинокі матері, недавно розлучені, з низькою самооцінкою, які мають низькі доходи, проходять реабілітацію після наркотиків, колишні засуджені, жертви фізичного або психологічного насильства» та ін. [14, с. 72]. Шахраї полювали на вразливих людей у всі часи, але з допомогою алгоритмів вони можуть виявляти своїх жертв з безпрецедентною точністю.

Кілька досліджень у Сполучених Штатах показали, що алгоритми виявляють упередження до представників расових меншин. У 2012 р. дослідниця з Гарвардського університету Латанія Свіні провела експеримент, в якому у пошуковому рядку Google вводила імена, які частіше асоціюються з афроамериканцями, і вивчала контекстну рекламу, яка з'являлася поруч з ними. Свіні дійшла висновку, що «пошукові запити з афроамериканськими іменами були частіше пов'язані з негативною рекламою, зокрема послуг перевірки минулого особи, тоді як запити з «білими» іменами привертали лише нейтральну рекламу» [17]. Схоже дослідження провів Натан Ньюмен з Нью-Йоркського університету. Він створив кілька акаунтів у Gmail і відправив з них повідомлення, в яких йшлося про інтерес купівлі автомобіля, причому частину акаунтів він зареєстрував за іменами, які здебільшого використовують афро- та латиноамериканці. Ньюмен дослідив контекстну рекламу, яку сервери Google пов'язали з цими акаунтами, і встановив, що «білі» імена здебільшого привертали рекламу нових машин респектабельних марок, таких як Toyota або GMS, тоді як «кольорові» частіше провадили до пропозицій купити дешеве авто на вторинному ринку [13]. Дослідник вважає малоймовірним, що програмісти Google навмисне так запрограмували ці алгоритми. Вірогідніше пояснення, на його думку, полягає у тому, що нейромережа проаналізувала реальний світ, в якому існує расова та економічна нерівність, і продублювала його стереотипи.

У масмедіа часто можна почути тезу, що технології «змінюють світ». Це справді так, якщо йдеться лише про формальні зміни. Сьогодні майже не залишилось суспільств, в яких би не було Інтернету, ноутбуків і смартфонів. Проте якщо йдеться про змістовні зміни – про подолання соціально-економічної, расової чи ґендерної нерівності, то поза увагою залишається той факт, що технології часто закріплюють і навіть поглиблюють наявні розколи.

Висновки. У статті розглянута низка етичних та соціальних проблем, що постають із застосування автоматичних алгоритмів прийняття рішень (ААПР). ААПР визначено як комп'ютерні програми, які перетворюють вхідні дані (інформацію про індивідів) у вихідні (рішення) за скінченну кількість кроків, ґрунтуючись на своїх запрограмованих інструкціях. Вказано, що довкола таких систем формується новий суспільний феномен, названий «суспільством контролю». Влада у такому суспільстві здійснюється за допомогою перемикачів, які управляють даними. Сукупність таких перемикачів, що застосовуються у різних сферах людської діяльності, приходить на зміну «дисциплінарній» владі, обґрунтованій Мішелем Фуко, яка втілювалась у «мікроархітектурі» паноптичних інституцій. У статті вказано на відмінності побудови такої системи в Китаї та на Заході. У Китаї її метою є покарання правопорушників та «некорисних» для суспільства громадян. На Заході це передусім «підштовхування» людей до вчинення певних дій. У Китаї таку систему запроваджує держава, на Заході це роблять насамперед приватні компанії. На кількох прикладах розглянуто низку етичних та соціальних проблем, які постають із застосування ААПР. Це, зокрема, односторонній та непрозорий характер прийнятих рішень і труднощі їх оскарження, втручання у приватність, байдужість до соціальних контекстів, редукціонізм, стигматизація, формування петель позитивного зворотного зв'язку, унаслідок яких алгоритми стають пророцтвами, що самі себе виконують, а також упереджене ставлення до меншин, зумовлене дублюванням нерівності, що існує у реальному світі.

Оскільки повернення до аналогової епохи не є ані бажаним, ні можливим, а ААПР загалом полегшують людську діяльність і роблять її ефективнішою, актуальним завданням стає вироблення етичних принципів, які б дозволили збільшити користь та зменшити ризики застосування таких систем. Це може стати одним із напрямів подальшої розробки тематики етичного та соціального виміру ААПР.

Література

1. Бех Ю.В. Філософія управління соціальними системами : монографія. Київ : Вид-во НПУ імені М.П. Драгоманова, 2012. 623 с.

2. Карпенко Ю.В. Етичні принципи застосування штучного інтелекту в публічному управлінні. *Вісн. НАДУ. Серія «Державне управління»*. 2019. № 4. Вип. 95. С. 93–97. DOI: https://doi.org/10.36030/2310-2837-4(95)-2019-93-97.

3. Слободяник В., Додонова В. Філософія штучного інтелекту – філософія майбутнього. *Збірник наукових праць ГО* . 2020. Вип. 4. С. 12–14. DOI: https://doi.org/10.36074/25.12.2020.v4.03.

4. Харарі Ю.Н. 21 урок для 21 століття. Київ : Book Chef, 2018. 416 с.

5. Харарі Ю.Н. Homo Deus: за лаштунками майбутнього. Київ : Book Chef, 2018. 512 с.

6. Algorithm. *Merriam-Webster* : вебсайт. URL: https://www.merriam-webster.com/dictionary/algorithm (дата звернення: 10.05.2021).

7. Chesterman S. Privacy and our digital selves. Simone Chesterman : вебсайт. URL: https://simonchesterman.com /blog/2017/09/02/our-digital-selves (дата звернення: 10.05.2021).

8. Deleuze G. Postscript on the Societies of Control. *October*. 1992. No. 59. P. 3–7.

9. Jernigan C., Mistree B. Gaydar: Facebook friendships expose sexual orientation. *First Monday*. 2009. Vol. 14, No. 10. DOI: https://doi.org/10.5210/fm.v14i10.2611. URL: https://firstmonday.org/article/view/2611/2302 (дата звернення: 10.05.2021).

10. Karanicolas M. Travel guide to the digital world: Surveillance and international standard. London : Global Partners Digital, 2014. 98 p.

11. Lubin G. The incredible story of how Target exposed a teen girl's pregnancy. *Insider* : вебсайт. 2012. URL: https://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2 (дата звернення: 10.05.2021).

12. Mumford L. Technics and the nature of man. *Technology and Culture*. 1966. Vol. 7, No. 1. P. 303–317. DOI: https://doi.org/10.2307/3101930.

13. Newman N. Racial and Economic Profiling in Google Ads: A Preliminary Investigation. *Huffpost* : вебсайт. 2017. URL: https://www.huffpost.com/entry/racial-and-economic-profi_b_970451 (дата звернення: 10.05.2021).

14. O'Neil C. Weapons of math destruction: How Big Data increases inequality and threatens democracy. New York : Crown, 2016. 272 p.

15. Opinions concerning Accelerating the Construction of Credit Supervision, Warning and Punishment Mechanisms for Persons Subject to Enforcement for Trust-Breaking. URL: https://chinacopyrightandmedia.wordpress.com/2016/09/25/opinions-concerning-accelerating-the-construction-of-credit-supervision-warning-and-punishment-mechanisms-for-persons-subject-to-enforcement-for-trust-breaking/ (дата звернення: 10.05.2021).

16. Pasquale F. The black box society: The secret algorithms that control money and Information. Oxford : Harvard University Press, 2016. 320 p.

17. Sweeney L. Discrimination in online ad delivery. *SSRN*. 2013. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2208240 (дата звернення: 10.05.2021).

18. Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. New York : PublicAffairs, 2019. 704 p.

Анотація

Ланюк Є. Ю. **Проблеми застосування автоматичних алгоритмів прийняття рішень: етичний та соціальний аналіз.** – Стаття.

У статті досліджуються етичні та соціальні проблеми, що постають із застосування автоматичних алгоритмів прийняття рішень (ААПР), які визначено як комп'ютерні програми, що перетворюють вхідні дані (оцифровану інформацію про індивідів) на вихідні (рішення) на основі своїх запрограмованих інструкцій. Такі алгоритми щораз ширше застосовують у багатьох сферах людської діяльності, зокрема банківській, страховій, маркетинговій та ін. Хоча їхнє використання має очевидні переваги, вони водночас порушують низку етичних, філософських та соціальних проблем, окремі з яких розглядаються у цій статті. Ґрунтуючись на ідеях Льюїса Мамфорда, вказано, що суспільство в інформаційну епоху можна розглядати як абстракцію його ключової технології – комп'ютера. У руслі підходу Жиля Дельоза суспільство, яке вибудовується довкола алгоритмів, що перетворюють дані, визначено як «суспільство контролю». Таке суспільство утверджується як нова історична форма влади, котра приходить на зміну «дисциплінарній» владі, концепцію якої обґрунтував Мішель Фуко. Аналізуються особливості формування такого суспільства у Китаї та на Заході. У Китаї його метою є винагорода і покарання «корисних» та «некорисних» для держави громадян, тоді як на Заході – підштовхування людей до дій, які приносять прибуток технологічним компаніям та їхнім клієнтам. Розглянуто кілька контроверсійних прикладів застосування ААПР і подано їхню етичну та соціальну інтерпретацію. Зокрема, обґрунтовано такі проблемні аспекти, які постають із їхнього використання: односторонній та непрозорий характер прийнятих рішень і труднощі їх оскарження, втручання у приватність, байдужість до соціальних контекстів, редукціонізм, стигматизація, формування петель позитивного зворотного зв'язку, а також упереджене ставлення до меншин. Аналіз пов'язаних з діджиталізацією у нашій країні етичних, філософських та соціальних проблем стає актуальним завданням у вітчизняних гуманітарних та суспільних науках. Важливим, зокрема, є обґрунтування етичних принципів, на які повинен спиратись

процес діджиталізації, що є перспективним напрямом подальшої розробки проблематики цієї статті.

Ключові слова: автоматичні алгоритми прийняття рішень, етика, інформаційна епоха, суспільство контролю, дисциплінарна влада, приватність, редукціонізм, стигматизація, упередження.

Summary

Laniuk Ye. Yu. **Problems of application of automatic decision making algorithms: ethical and social analysis.** – Article.

The article examines the ethical and social problems arising from the use of automatic decision-making algorithms. These tools are defined as computer programs, which convert data inputs (digitized information about individuals) into outputs (decisions) based on their pre-programmed instructions. These instruments have become increasingly widespread in a wide range of spheres, including banking, insurance, marketing, employment, etc. Despite their utilization has obvious advantages, they simultaneously raise several ethical, philosophical, and social problems, some of which are discussed in this article. Based on the ideas of Lewis Mumford, it is indicated that society in the information age can be viewed as an abstraction of its key technology – the computer. In light of the approach of Gilles Deleuze, the society, which is formed around switches converting data, is defined as a «society of control». In such a society, a new and unprecedented historical type of power is exercised, which comes to replace the «disciplinary» power substantiated by Michel Foucault. The features of the «society of control» are different in China and in the West. In China, its task is rewarding and punishing citizens who are considered «useful» or «useless» by the state. In the West, it meant to «nudge» people into actions that bring profit for digital corporations and their clients. A number of controversial cases of the application of algorithms are discussed in this article. In particular, the following problematic aspects of their utilization are highlighted: their decisions are one-sided and opaque, they interfere with privacy, are blind to social contexts, entail reductionism and stigmatization, create pernicious feedback loops, which tend to become «self-fulfilling prophecies», and can be biased to minorities. The analysis of ethical, philosophical, and social problems, which are associated with the «digitalization» reform in Ukraine, is an urgent task for our country's social and humanitarian studies. In particular, it is important to substantiate the ethical principles, on which the process of «digitalization» should be based, which is a promising direction for a future research in this field.

Key words: automatic decision-making algorithms, ethics, information age, society of control, disciplinary power, privacy, reductionism, stigmatization, prejudice.