УДК 321.7:004.896
DOI https://doi.org/10.32782/2663-6170/2025.43.9

ОСНОВНІ ПРИНЦИПИ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ПРОЦЕСІ ПРИЙНЯТТЯ ПОЛІТИЧНИХ РІШЕНЬ

BASIC PRINCIPLES OF USING ARTIFICIAL INTELLIGENCE IN POLITICAL DECISION-MAKING

Вітушко М.Ю.,
orcid.org/0009-0003-2002-8473
аспірант сектору правових проблем політичних інститутів і процесів
Інституту держави і права імені В.М. Корецького Національної академії наук України

Цифровізація послуг, зокрема використання штучного інтелекту, спрощує доступ та підвищує ефективність державних органів, сприяє зниженню витрат і боротьбі з корупцією, а також зменшує ймовірність людських помилок. Водночас використання ШІ може становити загрозу для захисту та ефективного здійснення основних прав і свобод людини.

Однією з найбільших проблем є прозорість внутрішнього функціонування алгоритмів ШІ, процесів прийняття рішень та використання даних для різних зацікавлених сторін, включаючи розробників, користувачів, політиків та громадськість. Це стосується пояснюваності – здатності системи ШІ пояснювати свої дії, рішення, рекомендації, результати або процес прийняття рішень у доступній формі, оскільки більшість користувачів систем ШІ не знають, як генеруються конкретні результати.

Іншою проблемою є упередженість ШІ – несправедливе або нерівне ставлення до певних груп на основі раси, статі або інших захищених характеристик. Ця упередженість може набувати різних форм, таких як дискримінаційні практики найму, упереджене схвалення кредитів тощо. Така упередженість часто призводить до дискримінації – випадкового або навмисного диференційованого ставлення до особи та/або групи осіб на основі певної характеристики (етнічної, релігійної, расової тощо).

Щоб забезпечити відповідальне використання систем ШІ, низка наукових спільнот, аналітичних центрів та міжнародних організацій запропонували своє бачення основних керівних принципів для ШІ, які включають елементи гуманності.

З метою забезпечення відповідального використання систем ШІ низка наукових спільнот, аналітичних центрів та міжнародних організацій запропонували своє бачення основних керівних принципів ШІ, які включають елементи прав людини, захисту даних та етичних стандартів. У найзагальніших рисах універсальні принципи ШІ включають: принцип прозорості, який передбачає незалежну підзвітність за автоматизовані рішення; обов'язок ідентифікації; обов'язок справедливості; обов'язок оцінки та підзвітності; обов'язки точності, надійності та валідності; обов'язок припинення; обов'язок громадської безпеки та кібербезпеки; заборону унітарного оцінювання.

Ключові слова: штучний інтелект, прийняття рішень, права людини, прозорість, підзвітність, технологія, цифровізація, держава, справедливість, безпека.

The digitisation of services, in particular the use of artificial intelligence, simplifies access and increases the efficiency of public authorities, helps to reduce costs and combat corruption, and reduces the likelihood of human error. At the same time, the use of AI can pose a threat to the protection and effective exercise of fundamental human rights and freedoms.

One of the biggest challenges is the transparency of the internal functioning of AI algorithms, decision-making processes and data use for various stakeholders, including developers, users, policymakers and the public. This refers to explainability – the ability of an AI system to explain its actions, decisions, recommendations, results or decision-making process in an accessible way, as most users of AI systems do not know how specific results are generated.

Another problem is AI bias – unfair or unequal treatment of certain groups based on race, gender or other protected characteristics. This bias can take many forms, such as discriminatory hiring practices, biased credit approvals, and so on. Such bias often leads to discrimination – accidental or intentional differential treatment of a person and/or group of persons on the basis of a particular characteristic (ethnic, religious, racial, etc.).

To ensure the responsible use of AI systems, a number of scientific communities, think tanks and international organisations have proposed their vision of basic guiding principles for AI, which include elements of human rights, data protection and ethical standards. In the most general terms, universal AI principles include: the principle of transparency, which provides for independent accountability for automated decisions; identification obligation; fairness obligation; assessment and accountability obligation; accuracy, reliability, and validity obligations; termination obligation; public safety and cybersecurity obligation; prohibition on unitary scoring.

Key words: artificial intelligence, decision-making, human rights, transparency, accountability, technology, digitalisation, state, justice, security.

Постановка проблеми. Одним із найважливіших досягнень, що визначатимуть майбутнє, є цифрова трансформація повсякденного життя, зокрема цифровізація державних і приватних послуг. Доступ до Інтернету, технології, цифрові навички та цифрові послуги дедалі більше стають необхідними умовами для повсякденного життя та доступу до державних послуг. Цифровий світ

ХХІІІ ст., зокрема штучний інтелект (ШІ), став реальністю, яка супроводжується низкою викликів.

Системи ШІ можуть варіюватися від систем, що базуються на правилах, до більш складних моделей навчання, таких як машинне навчання (МН) та глибоке навчання (ГН), де системи вдосконалюються та адаптуються на основі даних, на яких вони навчаються. Їх алгоритми призначені для навчання на прикладах, виявлення закономірностей, групування об'єктів із подібними атрибутами та вдосконалення з часом завдяки досвіду. Якщо МН можна використовувати для структурованих даних з метою прогнозного аналізу, то ГН виконує складний аналіз величезних обсягів даних.

Переваги цифровізації послуг, спрямованої на спрощення та підвищення ефективності доступу, включають підвищення ефективності роботи органів публічної влади, зниження витрат та зменшення ймовірності людських помилок. Більш оптимізовані та гармонізовані адміністративні процеси допомагають уникнути зайвої бюрократії, сприяють сталому розвитку та захисту навколишнього середовища, а також зменшують корупцію. Однак такі системи можуть становити загрозу для захисту та ефективного здійснення основних прав і свобод людини.

В Європі штучний інтелект використовується в сфері державних послуг, освіти та соціальних програм. Наприклад, Данія експериментувала з використанням штучного інтелекту для найму шкільного персоналу, а Італія використовувала його для визначення права на соціальну допомогу [1]. Однак ці системи зіткнулися з проблемами, такими як помилки, що призвели до несправедливого скорочення виплат або необґрунтованого розподілу робочих місць [2].

Аналіз останніх досліджень і публікацій. Проблематика використання штучного інтелекту у суспільно-політичних процесах поки що залишається малодослідженою через новизну технології. У світі дослідженням політико-соціальних аспектів ШІ займаються Дж. Баллок, М. Даббер, А. Корінек, Ф. Паскуале, Дж. Хіммельрейх, Ю-Че Чень та ін. Серед вітчизняних дослідників слід згадати О. Стойко та Н. Хому.

Вирішення невирішених раніше частин загальної проблеми. Позитивні і негативні наслідки використання штучного інтелекту для прийняття політичних рішень у вітчизняній політології не досліджувалися.

Формування цілей статті (постановка завдання). Метою дослідження є визначення принципів використання ШІ, які дозволять мінімізувати негативні наслідки його впровадження в процес прийняття рішень.

Виклад основного матеріалу дослідження. Позитивний вплив від використання ШІ у процесі прийняття політичних рішень у вигляді підвищення ефективності, стимулювання інновацій, автоматизації процесів та розв'язання складних проблем може бути знівельований негативними наслідками, зумовленими браком знань про штучний інтелект. Останнє передбачає глибше розуміння та усвідомлення того, як використовуються особисті дані, як навчаються моделі, як ці системи генерують конкретні результати, а також розуміння ризиків упередженості та необхідності критично оцінювати результати.

Однією з найбільших проблем є зрозумілість внутрішнього функціонування алгоритмів ШІ, процесів прийняття рішень і використання даних для різних зацікавлених сторін, включаючи розробників, користувачів, політиків і громадськість. Йдеться про пояснюваність – здатність системи ШІ доступно пояснювати свої дії, рішення, рекомендації, результати або спосіб прийняття рішень, оскільки більшість користувачів систем ШІ не знають, як генеруються конкретні результати. Відсутність інтерпретованості в моделях ШІ означає, що користувачі мають приймати результати системи прийняття рішень, не знаючи про складнощі, які призвели до цього рішення. Різні моделі ШІ можуть базуватися на різних математичних моделях та вчитися на різних базах даних, що неминуче призводить до розбіжностей у результатах. З огляду на це, пояснюваний ШІ (ПШІ) є новою галуззю, яка зосереджується на розробці технік, що роблять моделі ШІ більш інтерпретованими [3]. Оскільки моделі ШІ часто функціонують як «чорні скриньки», внутрішній механізм роботи яких невідомий більшості користувачів, то це викликає ряд етичних занепокоєнь. ПШІ має на меті розвінчати міф про «чорні ящики» і перетворити їх на більш зрозумілу форму, щоб підвищити прозорість моделі та зміцнити довіру кінцевих користувачів до результатів, виданих ШІ. Такі додаткові гарантії необхідні для широкого впровадження цих моделей, особливо в сферах з високим рівнем ризику, зокрема і в публічному управлінні.

Ще однією проблемою є упередженість ШІ – несправедливе або нерівне ставлення до певних груп на основі раси, статі або інших захищених характеристик. Ця упередженість може проявлятися у багатьох формах, таких як дискримінаційні практики найму, упереджене схвалення кредитів або навіть смертельні рішення, прийняті автономними транспортними засобами. Така упередженість часто призводить до дискримінації – випадкового або умисного відмінного «ставлення до особи та/або групи осіб за ознакою етнічного чи соціального походження, громадянства, національності, раси, релігії та вірувань, віку, статі, сексуальної орієнтації, ґендерної ідентичності, інвалідності або за іншими ознаками, що призводить до обмеження у визнанні,

реалізації або користуванні правами і свободами або до надання привілеїв у будь-якій формі, крім випадків, коли такі обмеження або привілеї мають правомірну об'єктивно обґрунтовану мету, способи досягнення якої є належними, необхідними та пропорційними». [4, с. 11-12]. Вона може бути наслідком: 1) інституційних чи індивідуальних шкідливих упереджень, які були перенесені в процеси упродовж життєвого циклу ШІ або представлені в даних системи ШІ; 2) технічних обмежень в апаратному забезпеченні чи комп'ютерних програмах; 3) невідповідність системи ШІ контексту використання [4, с. 11-12].

Так система виявлення шахрайства на основі ШІ, розроблена податковою службою Нідерландів, позначила тисячі сімей з низьким доходом як потенційних шахраїв, які зловживають соціальною допомогою [5]. Система була зорієнтована на сім'ї з подвійним громадянством або етнічним походженням. Це призвело до несправедливих звинувачень і необґрунтованих вимог повернення допомоги на утримання дітей. Цей інцидент висвітлив не тільки недоліки в дизайні ШІ, зокрема упередженість цих алгоритмів, але й необхідність підзвітності органів публічної влади при використанні систем ШІ. В Австрії впровадження чат-ботів для надання допомоги службам зайнятості в Австрії призвело до посилення гендерних стереотипів, що ще раз підкреслило небезпеку вибіркового дотримання алгоритмічних рекомендацій у процесі прийняття рішень у державному секторі [6].

Щоб забезпечити відповідальне використання систем ШІ низка наукових спільнот, аналітичних центрів та міжнародних організацій запропонували своє бачення універсальних керівних принципів для ШІ, які включають елементи прав людини, захисту даних та етичних стандартів [7]. У найбільш загальному вигляді до базових принципів ШІ можна віднести:

1. Принцип прозорості, який передбачає забезпечення незалежної підзвітності за автоматизовані рішення, з основним акцентом на праві особи знати підстави для прийняття негативного рішення. На практиці особа може не мати можливості інтерпретувати підстави для прийняття конкретного рішення, але це не скасовує необхідності забезпечення можливості такого пояснення. Прозорість є одним з найфундаментальніших принципів у всіх демократичних системах, такий як проведення відкритих засідань, встановлення положень щодо інформації та забезпечення права на доступ до документів [8]. Так, Лісабонський договір (Договір про реформування ЄС, 2007) року містить кілька статей, які закликають до застосування принципів прозорості, вимагають доступу до будь-якої інформації та комунікації, пов'язаної з процесами або персональними даними, та забезпечують легкість комунікації та доступну, зрозумілу мову. Цей принцип стосується, зокрема, надання суб'єктам даних інформації щодо особи контролера та цілей, що лежать в основі обробки. Крім того, він охоплює поширення додаткової інформації для гарантування справедливої та прозорої обробки щодо залучених фізичних осіб, а також їхнього права на отримання підтвердження та повідомлення щодо персональних даних, що стосуються їх і підлягають обробці. У контексті алгоритмічних процесів прийняття рішень, у преамбулі Загального регламенту захисту даних (2016) зазначено, що: «… будь-яка інформація та повідомлення, що стосуються обробки… персональних даних, повинні бути легкодоступними та зрозумілими, а також викладені чіткою та простою мовою…». Крім того, вимагається, щоб «… суб'єктам даних надавалася інформація про особу контролера та цілі обробки, а також додаткова інформація для забезпечення справедливої та прозорої обробки…» [9]. Більше того, прозорість вважається важливою складовою доробку Співтовариств (acquis communautaire) для відстеження інформації [10].

2. Обов'язок ідентифікації, що має за мету усунення асиметрії ідентифікації, яка виникає у взаємодії між людьми та системами ШІ. Система ШІ зазвичай знає про людину дуже багато, тоді як людина може навіть не знати, хто є оператором системи ШІ. Обов'язок ідентифікації закладає основу відповідальності ШІ, яка полягає у чіткому визначенні ідентичності системи ШІ та відповідальної установи.

3. Обов'язок справедливості визнає, що всі автоматизовані системи приймають рішення, які відображають упередженість і дискримінацію, але такі рішення не повинні бути нормативно несправедливими. Тобто для оцінки системи ШІ недостатньо лише об'єктивних результатів, а також слід враховувати нормативні наслідки, включаючи ті, що існували раніше або можуть бути посилені системою ШІ.

4. Обов'язок оцінки та підзвітності передбачає проведення оцінки системи ШІ до та під час її впровадження. Результатом проведення такої оцінки є визначення доцільності створення системи ШІ. Якщо в ході оцінки було виявлено істотні ризики, то така система не може використовуватися.

5. Обов'язки щодо точності, надійності та валідності, а також використання якісних даних визначають основні відповідальності, пов'язані з результатами автоматизованих рішень.

6. Обов'язок припинення передбачає, що системи повинні залишатися під контролем людини. Якщо це більше неможливо, система повинна припинити своє функціонування.

7. Зобов'язання щодо громадської безпеки та кібербезпеки визнає, що системи ШІ контролюють пристрої у фізичному світі, тому організації повинні як оцінювати ризики, так і вживати відповідних запобіжних заходів. Користувачі повинні враховувати, що навіть добре розроблені системи можуть стати мішенню для небезпечних зовнішніх втручань.

8. Заборона на унітарну оцінку (присвоєння урядом фізичній особі єдиного багатоцільового номера) та таємне профілювання. Унітарна оцінка відображає не тільки унітарний профіль, але й заздалегідь визначений результат у багатьох сферах людської діяльності. Існує певний ризик, що унітарні оцінки проникнуть і у приватний сектор, тому у законодавстві про захист даних різних країн використання універсальних ідентифікаторів, що дозволяють створювати профілі фізичних осіб, жорстко регулюються або ж заборонені.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямку. Отже, впровадження систем штучного інтелекту у процеси прийняття політичних рішень зумовлене як логікою розвитку сучасних інформаційних технологій та їх дедалі глибшим проникненням у всі сфери життя, так і потенційними позитивними наслідками: спрощенням доступу до органів публічної влади та підвищенням ефективності їх роботи, зменшення видатків на їх функціонування та мінімізація людських помилок. Водночас запровадження ШІ породжує нові виклики у сфері етики та прав людини, які можна вирішити дотриманням універсальних правил, зокрема щодо прозорості, підзвітності, справедливості, ідентифікації, надійності та безпеки. Але дане питання потребує подальших досліджень теоретичних засад використання ШІ та аналізу практики його застосування у політичній сфері суспільного життя.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. How the Italian Social Security and Welfare Administration (INPS) Used Artificial Intelligence to Streamline Services. URL: https://joinup.ec.europa.eu/ collection/public-sector-tech-watch/how-italian-social-security-and-welfare-adminis- tration-inps-used-artificial-intelligence-streamline

2. Holm J.R., Lorenz E. The impact of artificial intelligence on skills at work in Denmark. *New Technology, Work and Employment.* 2022. Vol. 37. №1. P. 79–101. https://doi. org/10.1111/ntwe.12215

3. Ali S., Abuhmed T., El-Sappagh S., Muhammad K., Alonso-Moral J.M., Confalonieri R., Guidotti R., Del Ser J., Díaz-Rodríguez N., Herrera F. Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion.* 2023. Vol. 99: 101805. https://doi.org/10.1016/j.inffus.2023.101805

4. Словник термінів у сфері штучного інтелекту / упорядники: Чумаченко Д., Мішкін Д., Андрієнко О., Краковецький О., Турута О., Дубно О., Хрущова Д., Кобрін А., Авдєєва Т., Кравець І., Герасимяк В., Шабанов О., Бистрицька А. Київ: Міністерство цифрової трансформації України, 2024.

5. Cath C., Jansen F. Dutch Comfort: The limits of AI governance through municipal registers. URL: https://doi.org/10.48550/arXiv.2109.02944

6. Alon-Barkat S., Busuioc M. Human-AI interactions in public sector decision making: «Automation bias» and «selective adherence» to algorithmic advice. *Journal of Public Administration Research and Theory.* 2023. Vol. 33. №1. P. 153–169.

7. Universal Guidelines for AI. URL: https://www.caidp.org/universal-guidelines-for-ai/

8. Kossow N., Windwehr S., Jenkins M. Algorithmic Transparency and Accountability. URL: https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf

9. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

10. Williams R., Cloete R., Cobbe J., Cottrill C., Edwards P., Markovic M., Pang, W. From transparency to accountability of intelligent systems: Moving beyond aspirations. *Data & Policy*. 2022. № 4. URL: https://doi.org/10.1017/dap.2021.37

Дата першого надходження рукопису до видання: 28.11.2025
Дата прийнятого до друку рукопису після рецензування: 11.12.2025
Дата публікації: 31.12.2025